OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

WORKING PAPER 165

Independence Tests based on Symbolic Dynamics

Helmut Elsinger

Statement of Purpose

The Working Paper series of the Oesterreichische Nationalbank is designed to disseminate and to provide a platform for discussion of either work of the staff of the OeNB economists or outside contributors on topics which are of special interest to the OeNB. To ensure the high quality of their content, the contributions are subjected to an international refereeing process. The opinions are strictly those of the authors and do in no way commit the OeNB.

Imprint: Responsibility according to Austrian media law: Günther Thonabauer, Communications Division, Oesterreichische Nationalbank.

Published and printed by Oesterreichische Nationalbank, Wien.

The Working Papers are also available on our website (http://www.oenb.at) and they are indexed in RePEc (http://repec.org/).

Editorial

New methods to test whether a time series is i.i.d. are proposed in a recent series of papers (Matilla-García [2007], Matilla-García and Marín [2008], Matilla-García and Marín [2009], and Matilla-García et al. [2010]). The main idea is to map m-histories of a time series onto elements of the symmetric group. The observed frequencies of the different elements are then used to detect dependencies in the original series. The author will demonstrate that the results presented in the above papers are not correct in the suggested generality. Moreover, simulation results indicate that the performance of the original tests are not as good as betoken.

September 15, 2010

Independence Tests based on Symbolic Dynamics

Helmut Elsinger*

New methods to test whether a time series is i.i.d. are proposed in a recent series of papers (Matilla-García [2007], Matilla-García and Marín [2008], Matilla-García and Marín [2009], and Matilla-García et al. [2010]). The main idea is to map m-histories of a time series onto elements of the symmetric group. The observed frequencies of the different elements are then used to detect dependencies in the original series. I will demonstrate that the results presented in the above papers are not correct in the suggested generality. Moreover, simulation results indicate that the performance of the original tests are not as good as betoken.

Keywords: Independence Tests; Symbolic Dynamics; Permutation Entropy
JEL-Classification: C12, C52

*Oesterreichische Nationalbank, Otto-Wagner-Platz 3, A-1011, Wien, Austria. Tel: +43 140420 7212; Email: helmut.elsinger@oenb.at
The views expressed are those of the author and do not necessarily reflect the views of the Oesterreichische Nationalbank. The author would like to thank Markus Knell, Burkhard Raunig, and Martin Summer.

1. Introduction

New methods to test whether a time series is i.i.d. are proposed in a recent series of papers (Matilla-García [2007], Matilla-García and Marín [2008], Matilla-García and Marín [2009], and Matilla-García et al. [2010]). The tests are based on m-tuples of consecutive observations. Depending on the ranking of the consecutive observations they are mapped onto elements of the symmetric group S_m of degree m. The proposed statistics relate the observed frequencies of the elements of S_m to the expected frequencies.

In Matilla-García [2007] independence is tested by using a Pearson χ^2 test. It is claimed that this test is asymptotically $\chi^2_{m!-1}$ distributed.

In Matilla-García and Marín [2008] a G-test based on the likelihood function is defined. The authors claim that the test

- can be applied irrespective whether the data generating distribution is discrete or continuous,
- is invariant to monotonic transformations,
- is asymptotically $\chi^2_{m!-1}$ distributed,
- has finite sample levels that do not differ from the asymptotic level, and
- is consistent under m-dependence[1] of degree $\leq m$.

In Matilla-García et al. [2010] a likelihood ratio test is proposed to test whether two time series are independent of each other.

In Matilla-García and Marín [2009] the authors claim that permutation entropy can be used to detect *the most relevant lag order*.

The paper is organized as follows. In Section 2 I introduce the notation, describe the basic framework, and present the main claims stated in Matilla-García [2007] and Matilla-García and Marín [2008]. In Section 3 I derive the correct asymptotic distribution of the test statistics and prove that an adjusted version of the test statistic is asymptotically $\chi^2_{m!-(m-1)!}$ distributed. The performance of the proposed tests is evaluated in Section 4. Whether permutation entropy can be used to detect the lag order of a time series is the subject-matter of Section 5. The results derived in Matilla-García et al. [2010] are discussed in Section 6. Finally, Section 7 concludes.

2. Notation

Let me start by introducing the notation which is as close as possible to the notation in the above mentioned papers. Let $\{X_t\}_{t\in I}$ be a real valued time series with $I = \{1, \ldots, T\}$. The null hypotheses is that $\{X_t\}$ is an i.i.d. series. Let $\pi_i = (i_1, \ldots, i_m)$ be a permutation of $(0, 1, \ldots, m-1)$. In the context of the above mentioned papers π_i is

[1] A process is m-dependent if $(X_1, X_2, \cdots, X_u)$ and $(X_n, X_{n+1}, \cdots, X_{n+r})$ are independent if $n-u > m$.

called a symbol.[2] Instead of analyzing $\{X_t\}$ directly we focus on m–histories of the time series. So, let $X_m(t) = (X_t, X_{t+1}, \ldots, X_{t+m-1})$ for $t = 1, \cdots, K$ where $K = T - m + 1$.

$X_m(t)$ is of π_i-type if

$$\begin{aligned}&\text{a) } X_{t+i_1} \leq X_{t+i_2} \leq \ldots \leq X_{t+i_m} \text{ and}\\ &\text{b) } i_{s-1} < i_s \text{ if } X_{t+i_{s-1}} = X_{t+i_s}.\end{aligned}$$

The indicator variables $Z_{\pi_i,t}$ are defined by

$$Z_{\pi_i,t} = \begin{cases} 1 & \text{if } X_{t+i_1} \leq X_{t+i_2} \leq \ldots \leq X_{t+i_m} \text{ and} \\ & i_{s-1} < i_s \text{ if } X_{t+i_{s-1}} = X_{t+i_s} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

$Z_{\pi_i,t} = 1$ is equivalent to $X_m(t)$ is of π_i-type.

Under the null hypothesis $Z_{\pi_i,t}$ is a Bernoulli variable with (unconditional) success probability of p_{π_i}. Most of the results in the above mentioned papers rely on the claim that the counting variables

$$Y_{\pi_i,K} = \sum_{t=1}^{K} Z_{\pi_i,t} \tag{2}$$

are binomially distributed $B(K, p_{\pi_i})$ and that as a consequence the vector $Y(K,m) = (Y_{\pi_1,K}, \cdots, Y_{\pi_{m!},K})'$ is multinomially distributed [Matilla-García and Marín, 2008, p. 413]. This is not true as will be shown in the sequel.

In Matilla-García [2007] a Pearson χ^2 test is defined as

$$X^2(m) = \sum_{i=1}^{m!} \frac{(y_{\pi_i,K} - Kp_{\pi_i})^2}{Kp_{\pi_i}} \tag{3}$$

where $y_{\pi_i,K}$ is the realization of $Y_{\pi_i,K}$.

Claim 1. *[Matilla-García, 2007, Theorem, p. 3893] Let $\{X_t\}_{t\in I}$ be an i.i.d. time series from an unknown stochastic process. As long as the $|I|$ observations can be grouped into $m!$ mutually exclusive symbols then the random variable $X^2(m)$ is asymptotically $\chi^2_{m!-1}$ distributed.*

In Section 3 I will show that this claim is not correct. The reason is that $Y(K,m)$ is not multinomially distributed as presumed in Matilla-García [2007].

Under the assumption that $Y(K,m)$ were multinomially distributed the likelihood ratio statistic $G(m)$ would be given by

$$G(m) = -2\left[Kln(K) + \sum_{i=1}^{m!} y_{\pi_i,K} ln\left(\frac{p_{\pi_i}}{y_{\pi_i,K}}\right)\right]. \tag{4}$$

$G(m)$ would be asymptotically $\chi^2_{m!-1}$ distributed and the maximum likelihood estimator

[2] The set of all permutations π_i is the symmetric group S_m of degree m.

for p_{π_i} would be given by

$$\hat{p}_{\pi_i} = \frac{y_{\pi_i,K}}{K}.$$

Observe that

$$G(m) = 2\left[\sum_{i=1}^{m!} y_{\pi_i,K} ln\left(\frac{y_{\pi_i,K}}{Kp_{\pi_i}}\right)\right].$$

Hence, $G(m)$ is a classical G-test.

The permutation entropy $h(m)$ is defined as

$$h(m) = -\sum_{i=1}^{m!} p_{\pi_i} ln(p_{\pi_i})$$

where it is understood that the summation is only across those i for which $p_{\pi_i} > 0$. If we estimate the permutation entropy by plugging in $\hat{p}_{\pi_i}$, we get

$$\hat{h}(m) = -\sum_{i=1}^{m!} \frac{y_{\pi_i,K}}{K} ln\left(\frac{y_{\pi_i,K}}{K}\right).$$

Plugging this into equation (4) yields

$$G(m) = -2K\left(\hat{h}(m) + \sum_{i=1}^{m!} \frac{y_{\pi_i,K}}{K} ln(p_{\pi_i})\right).$$

If X_t is sampled i.i.d. from a continuous distribution all symbols π_i are equally likely, $p_{\pi_i} = 1/m!$ for all i. Under this assumption we get that

$$G^*(m) = 2K\left(ln(m!) - \hat{h}(m)\right) \tag{5}$$

equals $G(m)$. Again, if $Y(K,m)$ were multinomially distributed and $p_{\pi_i} = 1/m!$ for all i, $G^*(m)$ would be asymptotically $\chi^2_{m!-1}$ distributed .[3]

Claim 2. *[Matilla-García and Marín, 2008, Theorem 3.1] Let $\{X_t\}_{t\in I}$ be a real valued time series with $|I| = T$. Denote by $h(m)$ the permutation entropy for a fixed embedding dimension $m > 2$, with $m \in \mathbb{N}$. If the time series $\{X_t\}_{t\in I}$ is i.i.d., then the affine transformation of the permutation entropy*

$$G^*(m) = 2K\left(ln(m!) - \hat{h}(m)\right)$$

is asymptotically $\chi^2_{m!-1}$ distributed.

I will show in Section 3 that this claim is not correct. Neither $G(m)$ nor $G^*(m)$ are

[3] In Matilla-García and Marín [2008] all results are based on $G^*(m)$. It is ignored that $G^*(m) = G(m)$ only under the additional assumption that $p_{\pi_i} = 1/m!$ for all i which is not a consequence of the i.i.d. assumption.

asymptotically $\chi^2_{m!-1}$ distributed.

Finally, in Theorem 3.2 in Matilla-García and Marín [2008] it is claimed that a test based on $G^*(m)$ is consistent for a wide variety of serial dependence, in particular m–dependence.

Claim 3. *[Matilla-García and Marín, 2008, Theorem 3.2] Let $\{X_t\}_{t\in I}$ be a strictly stationary process and $m > 2$ with $m \in \mathbb{N}$. Then $lim_{T\to\infty} Pr(G^*(m) > C) = 1$ under serial dependence of degree $\leq m$ for all $0 < C < \infty$, $C \in \mathbb{R}$.*

I will show that for a large class of alternative hypothesis $Pr(G^*(m) > C)$ and $Pr(G(m) > C)$ do not converge to one for C sufficiently large. Yet, the test may still be consistent.

3. Testing for Independence

To illustrate the results and for simulation purposes I focus on the case $m = 3$. The symbols of S_3 are denoted by $\pi_1 = (0, 1, 2)$, $\pi_2 = (0, 2, 1)$, $\pi_3 = (1, 0, 2)$, $\pi_4 = (2, 0, 1)$, $\pi_5 = (1, 2, 0)$, and $\pi_6 = (2, 1, 0)$.

Two claims in Matilla-García and Marín [2008] can be disproved without any effort. First, it is argued that the test statistic $G^*(m)$ as given in equation (5) can be applied to discrete distributions, too. Now, suppose X_t is either 1 or -1 with equal probability of 1/2 and the draws are independent. For $m = 3$ there are eight different 3-histories possible. Yet, based on condition b) in the definition of $Z_{\pi_i,t}$ the realizations $(-1, -1, -1)$, $(-1, -1, 1)$, $(-1, 1, 1)$, and $(1, 1, 1)$ are mapped onto $\pi_1 = (0, 1, 2)$. The probability of this symbol is therefore 1/2. There is no realization that is mapped onto the symbol $\pi_6 = (2, 1, 0)$. All other symbols occur with a probability of 1/8. Even for moderate sample sizes the test based on $G^*(m)$ will reject the true null hypotheses almost certainly as the implicit assumption that $p_{\pi_i} = 1/m!$ for all i is violated. An answer to this critique might be to use $G(m)$ instead of $G^*(m)$. Yet, I will show in the sequel that this adjustment will not help as neither $G(m)$ nor $G^*(m)$ are asymptotically $\chi^2_{m!-1}$ distributed.

Second, the test based on $G^*(m)$ is not invariant under monotonic transformations. Suppose U_t is i.i.d. standard normal and let $X_t = sign(U_t)$. Again, the test will reject the true null hypotheses almost certainly.

It is evident that $Z_{\pi_i,t}$ and $Z_{\pi_i,t+1}$ are *not* independent for $m > 1$. $Y_{\pi_i,K}$ is therefore not distributed $B(K, p_{\pi_i})$. To see this consider the following example (taken from p.142 in Matilla-García and Marín [2008]) where $\{X_1 = 2, X_2 = 8, X_3 = 6, X_4 = 5\}$. Let $m = 3$ and $\pi_2 = (0, 2, 1) \in S_3$. In this case $Z_{\pi_2,1} = 1$ and $Z_{\pi_2,2}$ is 0. But $Z_{\pi_2,2}$ has to be 0 as $Z_{\pi_2,1} = 1$ implies that $X_2 > X_3$. $X_3(2)$ can only be of the following types: $\pi_3 = (1, 0, 2)$ (the case where $X_3 < X_2 < X_4$), $\pi_5 = (1, 2, 0)$ $(X_3 < X_4 < X_2)$, or $\pi_6 = (2, 1, 0)$ $(X_4 < X_3 < X_2)$. Each symbol π has only three successors out of six possible symbols.[4]

[4]More generally, for any given m each symbol can only have m different successors out of $m!$ symbols.

t	t+1					
	π_1	π_2	π_3	π_4	π_5	π_6
π_1	0.25	0.25	0	0.5	0	0
π_2	0	0	0.25	0	0.25	0.5
π_3	0.25	0.5	0	0.25	0	0
π_4	0	0	0.25	0	0.5	0.25
π_5	0.5	0.25	0	0.25	0	0
π_6	0	0	0.5	0	0.25	0.25

Table 1: Probability that $Z_{\pi_i,t+1} = 1$ conditional on $Z_{\pi_j,t} = 1$ for $m = 3$. $\{X_t\}$ is drawn i.i.d. from some arbitrary continuous distribution.

t	t+2					
	π_1	π_2	π_3	π_4	π_5	π_6
π_1	0.05	0.05	0.15	0.15	0.3	0.3
π_2	0.15	0.15	0.2	0.2	0.15	0.15
π_3	0.05	0.05	0.15	0.15	0.3	0.3
π_4	0.3	0.3	0.15	0.15	0.05	0.05
π_5	0.15	0.15	0.2	0.2	0.15	0.15
π_6	0.3	0.3	0.15	0.15	0.05	0.05

Table 2: Probability that $Z_{\pi_i,t+2} = 1$ conditional on $Z_{\pi_j,t} = 1$ for $m = 3$. $\{X_t\}$ is drawn i.i.d. from some arbitrary continuous distribution.

The conditional distribution of the symbols is independent of the underlying distribution as long as the distribution is continuous and the draws are i.i.d. To see this consider the case where $Z_{\pi_1,t} = 1$ with $\pi_1 = (0, 1, 2)$. The probability that $Z_{\pi_1,t+1} = 1$ conditional on $Z_{\pi_1,t} = 1$ is equal to the probability that observation X_{t+3} is larger than the maximum of the three i.i.d. variables X_t, X_{t+1}, and X_{t+2}. Hence, the probability is 0.25. Table 1 summarizes the probabilities that $Z_{\pi_i,t+1} = 1$ conditional on $Z_{\pi_j,t} = 1$ for all i and j for arbitrary continuous distributions. Note that π_1, π_3, and π_5 have as successors only the symbols π_1, π_2, and π_4. π_2, π_4, and π_6 are succeeded by the symbols π_3, π_5, and π_6. In terms of the counting variable $Y(K, m)$ this implies that $|Y_{\pi_3,K} + Y_{\pi_5,K} - Y_{\pi_2,K} - Y_{\pi_4,K}| \leq 1$ which is at odds with the claim that $Y(K, m)$ is multinomially distributed.

The probability that $Z_{\pi_1,t+2} = 1$ conditional on $Z_{\pi_1,t} = 1$ is just the probability that X_{t+3} is larger than the maximum of the three i.i.d. variables X_t, X_{t+1}, and X_{t+2} and X_{t+4} is larger than X_{t+3}. The probability of this event is $0.05 = (1/4) \cdot (1/5)$. Table 2 summarizes the two step conditional probabilities for all π.

Clearly, $Z_{\pi_i,t+m}$ and $Z_{\pi_i,t}$ are independent of each other.[5] Therefore, we get that $\mathbb{E}[Y_{\pi_i,K}] = K \cdot (1/6)$ for all $\pi_i \in S_3$. For the variance of $Y_{\pi_i,K}$ it holds that

$$\mathbb{V}[Y_{\pi_i,K}] = K\mathbb{V}[Z_{\pi_i,1}] + 2(K-1)Cov(Z_{\pi_i,2}, Z_{\pi_i,1}) + 2(K-2)Cov(Z_{\pi_i,3}, Z_{\pi_i,1})$$

[5] Under the null hypothesis $Z_{\pi_i,t}$ is an $(m-1)-$dependent process.

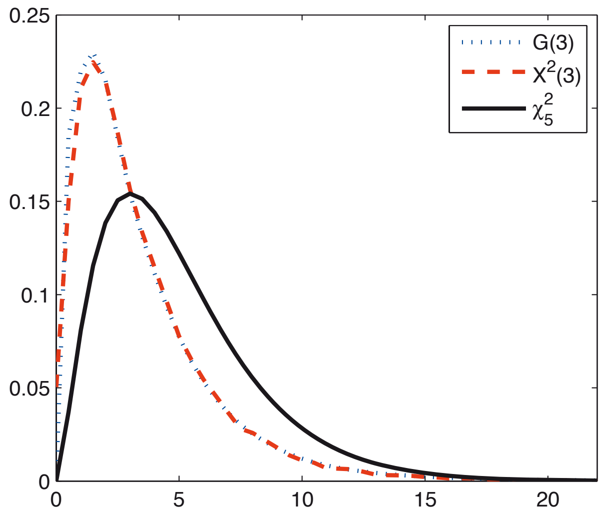


Figure 1: The density functions of $G(3)$, $X^2(3)$, and $\chi^2(5)$. The pdfs of $G(3)$ and $X^2(3)$ are estimated on the basis of 10,000 simulations of 1,000 observations.

Using the conditional probabilities given in Tables 1 and 2 we get

$$\mathbb{V}[Y_{\pi_i,K}] = \begin{cases} p_{\pi_i}(1-p_{\pi_i})\frac{92K+36}{100} = \frac{5}{36}\frac{92K+36}{100} \approx \frac{5}{36}0.92K & \text{for } \pi_1 \text{ and } \pi_6 \\ p_{\pi_i}(1-p_{\pi_i})\frac{56K+48}{100} = \frac{5}{36}\frac{56K+48}{100} \approx \frac{5}{36}0.56K & \text{otherwise.} \end{cases}$$

If $Y_{\pi_i,K}$ were binomially distributed, the variances would equal $Kp_{\pi_i}(1-p_{\pi_i}) = 5K/36$. Thus the variance of a binomial variable with the same expected value is considerably larger than the true variance. This difference does not vanish for $T \to \infty$ and hence $K \to \infty$.

As $Y_{\pi_i,K}$ is not distributed $B(K, p_{\pi_i})$ the vector $Y(K,m)$ is not multinomially distributed either. The proof of Theorem 3.1 in Matilla-García and Marín [2008] is not correct.[6]

Still it could be possible that $G(m)$ or $X^2(m)$ are at least asymptotically $\chi^2_{m!-1}$ distributed. I simulate 100,000 samples each consisting of 1,000 i.i.d. standard normal variates. The estimated distribution functions of $G(3)$ and $X^2(3)$ on the one hand and the pdf of $\chi^2(5)$ on the other hand are quite different as can be seen in Figure 1.[7]

To get a clearer picture of the distribution of $G(m)$ I derive the asymptotic distribution of a suitably normalized $Y(K,m)$ in a more general framework. Let $\{U_1, U_2, \cdots\}$ be a stochastic process taking values in some finite set of states which is assumed for

[6] The simplest remedy to correct the proof would be to use only non-overlapping m–histories. Yet, a major advantage of the proposed test is that it can be used for relatively small samples. This advantage would disappear.

[7] Series consisting of fewer observations (100 and 200) yield the same qualitative results.

notational convenience to consist of the first J positive integers $\{1, \cdots, J\}$. Assume that $\{U_t\}$ is a stationary and m-dependent process, i.e. $(U_t, \cdots, U_{t+l})$ and $(U_{t+r}, \cdots, U_{t+s})$ are independent if $r - l > m$. The (unconditional) probability that a particular state i is reached is denoted by p_i. The vector of unconditional state probabilities is given by $p = (p_1, \cdots, p_J)'$. The transition probabilities from state i at t to state j at $t + l$ for $l > 0$ are denoted by $q_{i,j}^{(l)}$ and summarized in the matrix $Q^{(l)}$. m-dependence implies that $q_{i,j}^{(l)} = p_j$ for $l > m$. Define the indicator variable $Z_{i,t}$ such that $Z_{i,t}$ equals 1 if $U_t = i$ and 0 otherwise. Let $Z_t = (Z_{1,t}, \cdots, Z_{J,t})'$ and denote the $J \times J$ diagonal matrix with p in the diagonal by $diag(p)$.

Theorem 1. *Let $\{U_t\}_{t \in \{1,\dots,T\}}$ be a stationary and m-dependent stochastic process taking values in $\{1, \cdots, J\}$. The indicator variable $Z_{i,t}$ equals 1 if $U_t = i$ and 0 otherwise. Define the counting variable $Y(K)$ by $Y(K) = \sum_{t=1}^{K} Z_t$. Then*

$$\frac{1}{\sqrt{K}} \left(Y(K) - Kp\right) \sim^a MVN(0, \Sigma)$$

where

$$\Sigma = diag(p) - (2m + 1)pp' + diag(p) \sum_{l=1}^{m} Q^{(l)} + \sum_{l=1}^{m} Q^{(l)\prime} diag(p).$$

The proof is given in Appendix A.

Remark 1. *If $\{U_t\}_{t \in I}$ is i.i.d. then $Y(K)$ is multinomially distributed and*

$$\Sigma = diag(p) - pp'.$$

Remark 2. *The variance covariance matrix can be rewritten as*

$$\Sigma = diag(p)W - W' diag(p) - diag(p) - pp'$$

where $W = \sum_{l=1}^{m} \left(Q^{(l)} - \mathbf{1}p'\right) + \mathbf{1}p'$.

Assume for a moment that $\{U_t\}$ is not m-dependent but a simple stationary and ergodic Markov chain. Let $Q = Q^{(1)}$ be the 1-step transition probability. The transition probabilities $Q^{(l)}$ have the property that $Q^{(l)} = Q^l$. The central limit theorem for Markov chains (Billingsley [1961]) states that the variance covariance matrix of $(1/\sqrt{K})Y(K)$ is given by

$$\Sigma = diag(p)W_1 - W_1' diag(p) - diag(p) - pp'$$

where $W_1^{-1} = \mathbf{I} - Q + \mathbf{1}p'$.

The properties of the Pearson χ^2 statistic for Markov chains are discussed in Tavaré and Altham [1983]. In particular, it is shown that the χ^2 statistic is in general not asymptotically χ^2 distributed but has to be adjusted. We will see that the same is true for m-dependence.

In the application under consideration the number of states J equals $m!$ The vector of indicator variables $Z_t = (Z_{\pi_1,t}, \cdots, Z_{\pi_{m!},t})'$ is an $(m - 1)$-dependent process under

H_0. If the original series $\{X_t\}$ is drawn from a continuous distribution then $p_{\pi_i} = 1/m!$ Hence, the variance covariance matrix of the asymptotic distribution of $(1/\sqrt{K})Y(K,m)$ is given by

$$\Sigma = \frac{1}{m!^2}\left(m!\left(\mathbf{I} + \sum_{l=1}^{m-1} Q^{(l)} + \sum_{l=1}^{m-1} Q^{(l)\prime}\right) - (2m-1)\mathbf{1}\mathbf{1}'\right).$$

$\mathbf{I}$ is the $m! \times m!$ identity matrix and $\mathbf{1}$ is an $m! \times 1$ vector of ones.[8]

For the case $m = 2$ the 1-step transition probability matrix is given by

$$Q^{(1)} = \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}$$

and hence,

$$\Sigma_2 = \frac{1}{12}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The transition probability matrices for $m = 3$ are given in Tables 1 and 2. Plugging them in yields

$$\Sigma_3 = \frac{1}{360}\begin{pmatrix} 46 & -23 & -23 & 7 & 7 & -14 \\ -23 & 28 & 10 & -20 & -2 & 7 \\ -23 & 10 & 28 & -2 & -20 & 7 \\ 7 & -20 & -2 & 28 & 10 & -23 \\ 7 & -2 & -20 & 10 & 28 & -23 \\ -14 & 7 & 7 & -23 & -23 & 46 \end{pmatrix}.$$

The rank of Σ_3 equals 4 and not 5 as in the multinomial case. This is a consequence of the mapping onto the elements of S_m and does not depend on the assumption that X_t is drawn from a continuous distribution.

Observe that the $m!$ symbols can be grouped into $(m-1)!$ disjoint sets F_i each of which consists of all symbols that agree on the ordering of $(1, \cdots, m-1)$. Each element of F_i has the same successors and each of the successors has only elements in F_i as predecessors. Denote the set of successors of F_i by $H(F_i)$. The only symbols that are successors of themselves are $(0, 1, \cdots, m-1)$ and $(m-1, m-2, \cdots, 1, 0)$ which are - without loss of generality - labeled π_1 and $\pi_{m!}$. For notational convenience denote the equivalence class of π_1 by F_1 and the equivalence class of $\pi_{m!}$ by $F_{(m-1)!}$. For $m = 3$ we get $F_1 = \{\pi_1, \pi_3, \pi_5\}$ and $F_2 = \{\pi_2, \pi_4, \pi_6\}$. This implies $H(F_1) = \{\pi_1, \pi_2, \pi_4\}$ and $H(F_2) = \{\pi_3, \pi_5, \pi_6\}$.

Each F_i and each $H(F_i)$ has m elements. For each F_i it has to hold that

$$\Big| \sum_{\pi \in F_i} y_{\pi,K} - \sum_{\pi \in H(F_i)} y_{\pi,K} \Big| \leq 1.$$

[8] $\mathbf{1}$ denotes a matrix of ones, $\mathbf{0}$ denotes a matrix of zeros, and $\mathbf{I}$ denotes the identity matrix. Whenever the dimension of these matrices is not obvious, a clarifying subscript will be appended.

As there are $(m-1)!$ (disjoint) sets F_i we get as many constraints.

Define the matrices A and B by

$$a_{i,j} = \begin{cases} 1 & \text{if } \pi_j \in F_i \\ 0 & \text{else} \end{cases} \quad \text{and} \quad b_{i,j} = \begin{cases} 1 & \text{if } \pi_j \in H(F_i) \\ 0 & \text{else} \end{cases}$$

The $(m-1)!$ constraints can be summarized in the matrix $U = A - B$. The mapping onto the symbols implies that $\|UY(K,m)\|_\infty \leq 1$ where $\|\cdot\|_\infty$ is the maximum norm. As a consequence $\mathbb{V}[UY(K,m)] = O(1)$. If we assume that the variance covariance matrix of $Y(K,m)$ is given by $\Omega_{K,m}$ and that $\lim_{K\to\infty}(1/K)\Omega_{K,m} = \Sigma_m$ exists then

$$\mathbf{0} = \lim_{K\to\infty} \mathbb{V}\left[\frac{1}{\sqrt{K}} UY(K,m)\right] = U\Sigma_m U'$$

The rank of Σ_m has to be less than or equal to $m! - rank(U)$.

As in the case of a multinomial distribution there is one additional constraint, namely $\mathbf{1}'Y(K,m) = K$. Given that by construction the first and the last column of U consist of zeros only, the additional constraint is not redundant. We get that $rank(\Sigma_m) \leq m! - 1 - rank(U)$.

Theorem 2. *Assume that the variance covariance matrix $\Omega_{K,m}$ of $Y(K,m)$ has the property that $\lim_{K\to\infty}(1/K)\Omega_{K,m} = \Sigma_m$ exists. The mapping onto the elements of S_m implies that for $m \geq 2$ the rank of Σ_m is less than or equal to $m! - (m-1)!$*

The proof is given in Appendix A.

An immediate consequence of Theorems 1 and 2 is the following corollary.

Corollary 1. *Let $\{X_t\}_{t\in I}$ be independently, identically, and continuously distributed with $K = T - m + 1$. The expected value of $Y_{\pi_i,K}$ equals $K/m!$ for each $\pi_i \in S_m$ and*

$$\frac{1}{\sqrt{K}}\left(Y(K,m) - \frac{K}{m!}\mathbf{1}\right) \sim^a MVN(0, \Sigma_m)$$

where

$$\Sigma_m = \frac{1}{m!^2}\left(m!\left(\mathbf{I} + \sum_{l=1}^{m-1} Q^{(l)} + \sum_{l=1}^{m-1} Q^{(l)\prime}\right) - (2m-1)\mathbf{1}\mathbf{1}'\right).$$

The rank of Σ_m is less than or equal to $m! - (m-1)!$

If $\{X_t\}_{t\in I}$ is i.i.d but the distribution is not continuous then the expected value and the variance covariance matrix have to be adjusted. I illustrate this using a simple example.

Example 1. *Suppose X_t is either 1 or -1 with equal probability of $1/2$ and the draws are independent. For $m = 3$ there are eight different 3-histories possible. Based on condition b) in the definition of $Z_{\pi_i,t}$ the realizations $(-1,-1,-1)$, $(-1,-1,1)$, $(-1,1,1)$, and $(1,1,1)$ are mapped onto $\pi_1 = (0,1,2)$. The probability of this symbol is therefore $1/2$. There is no realization that is mapped onto the symbol $\pi_6 = (2,1,0)$. All other symbols* *occur with a probability of* $1/8$*, i.e.* $p = (1/2, 1/8, 1/8, 1/8, 1/8, 0)'$*. The transition probabilities are given by*

$$Q^{(1)} = \tfrac{1}{8}\begin{pmatrix} 5 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 0 & 4 & 0 \\ 4 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 4 & 0 \\ 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \textit{and} \quad Q^{(2)} = \tfrac{1}{16}\begin{pmatrix} 6 & 1 & 3 & 3 & 3 & 0 \\ 12 & 4 & 0 & 0 & 0 & 0 \\ 4 & 0 & 4 & 4 & 4 & 0 \\ 12 & 4 & 0 & 0 & 0 & 0 \\ 12 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

This yields the following asymptotic variance covariance matrix for $(1/\sqrt{K})Y(K, m)$

$$\Sigma = \frac{1}{64}\begin{pmatrix} 16 & -8 & -8 & 0 & 0 & 0 \\ -8 & 7 & 3 & -3 & 1 & 0 \\ -8 & 3 & 7 & 1 & -3 & 0 \\ 0 & -3 & 1 & 3 & -1 & 0 \\ 0 & 1 & -3 & -1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The rank of Σ *is* 3 *and*

$$\frac{1}{\sqrt{K}}\left(Y(K) - Kp\right) \sim^a MVN(0, \Sigma).$$

Given the asymptotic distribution of $(1/\sqrt{K})(Y(K, m) - Kp)$ under H_0, we are able to determine the asymptotic distribution of $X^2(m)$. The test statistic can be written in matrix notation as

$$X^2(m) = \frac{1}{K}\left(Y(K) - Kp\right)' diag(p)^{-1}\left(Y(K) - Kp\right)$$

where it is assumed that $p_{\pi_i} \neq 0$ for all i. If $\{X_t\}_{t \in I}$ is drawn i.i.d. from a continuous distribution then $p = (1/m!)\mathbf{1}$ and

$$X^2(m) = \frac{1}{K/m!}\left(Y(K) - \frac{K}{m!}\mathbf{1}\right)' \mathbf{I}\left(Y(K) - \frac{K}{m!}\mathbf{1}\right).$$

The first observation is that $X^2(m)$ has an asymptotic distribution that can be represented as $\sum_{i=1}^{rank(\Sigma_m)} \lambda_i N_i^2$, where the N_i are independent $N(0, 1)$ random variables, and the λ_i are the nonzero eigenvalues of $diag(p)^{-1}\Sigma_m$. If the X_t are drawn from a continuous distribution, the nonzero eigenvalues of $m!\Sigma_m$ for $m = 3$ are approximately: 1.7071, 0.8, 0.6, and 0.2929. This implies that the expected value of $X^2(3)$ equals 3.4 and the variance of $X^2(3)$ equals 8.

$X^2(m)$ is asymptotically χ^2 distributed if and only if all nonzero eigenvalues λ_i are equal to one. If all λ_i equal a constant λ then $(1/\lambda)X^2(m)$ is χ^2 distributed. For

$m = 2$ there is only one nonzero eigenvalue, $\lambda_1 = 1/3$. Hence, $3X^2(2)$ is χ^2_1 distributed. Straightforward calculations yield that $X^2(m)$ is not χ^2 distributed for $m = 2, \cdots, 5$. The asymptotic distribution of $X^2(m)$ can be represented in a simple form but to apply the test critical values have to be estimated.

On the other hand for any generalized inverse Σ_m^- of Σ_m the statistic

$$X_a^2(m) = \frac{1}{K} (Y(K) - Kp)' \, \Sigma_m^- \, (Y(K) - Kp)$$

is χ^2 distributed with $rank(\Sigma_m) \leq m! - (m-1)!$ degrees of freedom.[9]

To show that $G(m)$ is not distributed $\chi^2_{m!-1}$ I use a Taylor series expansion. To begin, under H_0

$$\begin{aligned} G(m) &= -2\left[Kln(K) + \textstyle\sum_{i=1}^{m!} y_{\pi_i,K} ln\left(\frac{p_{\pi_i}}{y_{\pi_i,K}}\right)\right] \\ &= 2K \textstyle\sum_{i=1}^{m!} \hat{p}_{\pi_i} ln\left(\frac{\hat{p}_{\pi_i}}{p_{\pi_i}}\right). \end{aligned}$$

The Taylor series expansion of $f(x) = xlog(x/x_0)$ about x_0 is

$$f(x) = (x - x_0) + \frac{1}{2}(x - x_0)^2 \frac{1}{x_0} + R_2(x)$$

In our case this yields for $x_{0i} = p_{\pi_i}$

$$G(m) = 2K \sum_{i=1}^{m!} \left[(\hat{p}_{\pi_i} - p_{\pi_i}) + \frac{1}{2} \frac{(\hat{p}_{\pi_i} - p_{\pi_i})^2}{p_{\pi_i}} + R_2(\hat{p}_{\pi_i}) \right].$$

Under H_0 it holds that $\hat{p}_{\pi_i} = p_{\pi_i} + O_p(K^{-1/2})$. Therefore the remainder $R_2(\hat{p}_{\pi_i})$ is $O_p(K^{-3/2})$ and hence

$$G(m) = K \sum_{i=1}^{m!} \frac{(\hat{p}_{\pi_i} - p_{\pi_i})^2}{p_{\pi_i}} + O_p(K^{-1/2})$$

where we use that the probabilities sum to 1. We get that $G(m)$ and $X^2(m)$ are asymptotically equivalent, i.e.

$$G(m) \approx \sum_{i=1}^{m!} \frac{(y_{i,K} - Kp_{\pi_i})^2}{Kp_{\pi_i}} = X^2(m)$$

As a consequence neither $G(m)$ nor $X^2(m)$ are asymptotically $\chi^2_{m!-1}$ distributed. If these statistics are χ^2 distributed at all than with at most $m! - (m-1)!$ instead of $m! - 1$ degrees of freedom. Theorem 3.1 in Matilla-García and Marín [2008] and the Theorem in Matilla-García [2007] are not correct.

Matilla-García and Marín [2008] state in Theorem 3.2 that the test based on $G^*(m)$ is consistent against m-dependent alternatives. The proof of the theorem rests on the

[9]This result is due to Ogasawara and Takahashi [1951].

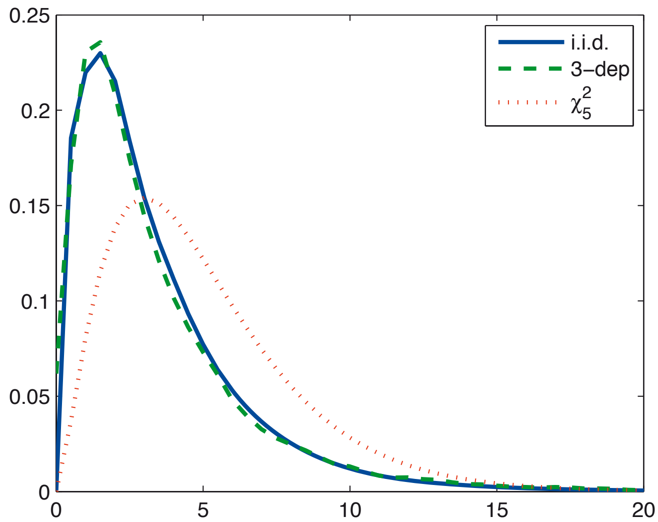


Figure 2: The (estimated) density functions of $G(3)$ for X_t i.i.d. and $X_t = \sqrt{2}(\epsilon_t + \epsilon_{t-3})$ compared to $\chi^2(5)$. The pdfs of $G(3)$ are based on 10,000 simulations of 1,000 observations.

assumption that serial dependence of degree $\leq m$ implies $ln(m!) - h(m) \neq 0$ (see p. 153 in Matilla-García and Marín [2008]). This assumption is not correct. Suppose the data is generated according to $X_t = \sqrt{2}(\epsilon_t + \epsilon_{t-m})$ where ϵ_t are i.i.d. standard normal. X_t is a stationary process with serial dependence of degree m. It is easy to verify that $p_{\pi_i} = \frac{1}{m!}$ for all $\pi_i \in S_m$ and hence $h(m) = ln(m!)$. The power of the test does not converge to 1 but is very low as can be seen in Figure 2. This example is rather special as the dependence structure and the embedding dimension both are equal to m. A test based on a different embedding dimension $m^* > m$ will have power against this particular m-dependent process. As an alternative example consider $X_t = \epsilon_t + \delta_t \epsilon_{t-1}$ where ϵ_t and δ_t are i.i.d. standard normal which are independent of each other. X_t is a $1-$dependent process with $h(m) = ln(m!)$ and $Corr(X_t, X_{t-1}) = 0$ but $Corr(X_t^2, X_{t-1}^2) = 1/7$. Figure 3 illustrates that the test based on $G^*(3)$ has very low power against this alternative, too. In both examples X_t is drawn from continuous distributions. Theorem 3.2 can therefore not be remedied by using $G(m)$ instead of $G^*(m)$.

4. Performance of the Tests

Neither $G(3)$ nor $X^2(3)$ are (asymptotically) χ^2_5 distributed. I estimated critical values under H_0 for significance levels of 10%, 5%, and 1% on the basis of 100,000 simulations for different numbers of observations. They are summarized in Table 3a for $G(3)$ and in Table 3b for $X^2(3)$. The critical values differ only slightly between $G(3)$ and $X^2(3)$ and

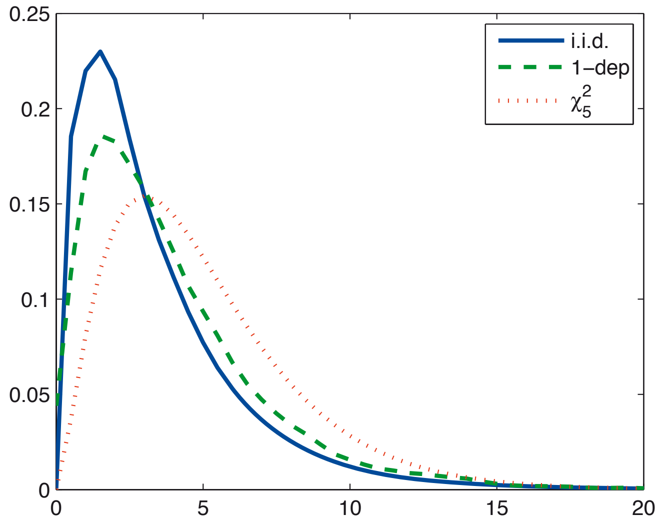


Figure 3: The (estimated) density functions of $G(3)$ for X_t i.i.d. and $X_t = \epsilon_t + \delta_t \epsilon_{t-1}$ compared to $\chi^2(5)$. The pdfs of $G(3)$ are based on 10,000 simulations of 1,000 observations.

vary only moderately with the number of observations. On the other hand , $X^2(m)$ and $G(m)$ have an asymptotic distribution that can be represented as $\sum_{i=1}^{rank(\Sigma)} \lambda_i N_i^2$, where the N_i are independent $N(0,1)$ random variables, and λ_i are the nonzero eigenvalues of $diag(p)^{-1}\Sigma$. By simulating the N_i, multiplying them with the appropriate λ_i, and summing them up the (asymptotic) critical values for $G(3)$ and $X^2(3)$ can be estimated. The values are given in the bottom rows of Tables 3a and 3b. The critical values for the different significance levels may be assumed to equal approximately 7, 8.9, and 13.6. If we use instead the critical values given by a χ^2_5 distribution, simulations show that the implicit p-values are approximately one half of the presumed significance level (4.4%, 2.4%, and 0.6%).

As a benchmark for the performance of $G(3)$ and $X^2(3)$ I use that

$$X_a^2(m) = \frac{1}{K}\left(Y(K) - \frac{K}{m!}\mathbf{1}\right)' \Sigma_m^- \left(Y(K) - \frac{K}{m!}\mathbf{1}\right)$$

is χ^2 distributed with $rank(\Sigma_m)$ degrees of freedom for any generalized inverse Σ_m^- of Σ_m. For simplicity I choose the Moore-Penrose pseudoinverse Σ_3^+ of Σ_3 which is given by

	Mean	Variance	90%	95%	99%
χ^2_5	5,00	10,00	9,24	11,07	15,09
T=100	3,47	8,18	7,10	8,97	13,78
T=500	3,40	7,99	7,00	8,95	13,60
T=1000	3,40	8,00	6,98	8,89	13,63
T=2000	3,39	7,93	6,96	8,88	13,63
T=∞	3.40	8.00	6.98	8.91	13.64

(a) $G(3)$

	Mean	Variance	90%	95%	99%
χ^2_5	5,00	10,00	9,24	11,07	15,09
T=100	3,42	7,77	6,94	8,78	13,43
T=500	3,40	7,91	6,99	8,92	13,49
T=1000	3,40	7,96	6,96	8,87	13,58
T=2000	3,39	7,91	6,95	8,87	13,62
T=∞	3.40	8.00	6.98	8.91	13.64

(b) $X^2(3)$

Table 3: Mean, variance, and (estimated) critical values for $G(3)$ and $X^2(3)$ for different numbers of observations. The first line contains the respective values for the χ^2_5 distribution. The estimates are based on $100,000$ simulations.

$$\Sigma_3^+ = \frac{1}{8}\begin{pmatrix} 68 & 14 & 14 & -34 & -34 & -28 \\ 14 & 49 & 9 & -39 & 1 & -34 \\ 14 & 9 & 49 & 1 & -39 & -34 \\ -34 & -39 & 1 & 49 & 9 & 14 \\ -34 & 1 & -39 & 9 & 49 & 14 \\ -28 & -34 & -34 & 14 & 14 & 68 \end{pmatrix}.$$

To evaluate the performance of the proposed test the same data generating processes (DGPs) are considered as in Matilla-García [2007] and Matilla-García and Marín [2008]. Table 4 lists the DGPs. The labels refer to respective numbers in Matilla-García [2007] and Matilla-García and Marín [2008]. $A-1$ and $A-2$ are the above given examples that illustrate the low power against m-dependent alternatives. For $T = 1,000$ I calculate the rejection rates based on $10,000$ simulations.[10] The results are summarized in Table 5 for a nominal significance level of 10%, in Table 6 for a significance level of 5%, and in Table 7 for a significance level of 1%. The columns labeled $X^2(3)$ and $G(3)$ comprise the rejection rates for H_0 under the (incorrect) assumption that the test are χ^2_5 distributed. $X^2_{no}(3)$ and $G_{no}(3)$ are calculated using non-overlapping $m-$histories. These statistics are χ^2_5 distributed. For the columns labeled $X^2_c(3)$ and $G_c(3)$ the critical values are approximated by 7, 8.9, and 13.6 in line with the simulations from above (Table 3). The test statistic $X^2_a(m)$ is χ^2_4 distributed.

The simulations illustrate that the size of $X^2(3)$ and $G(3)$ is considerably lower than the specified significance level. $X^2_a(3)$ has higher power for almost all DGPs than all

[10]For $T = 100$ and $T = 500$ see Appendix B.

other tests. Still, the test does not detect certain kinds of dependence. For the models $07-8$, $08-3$, $08-6$, $A-1$, and $A-2$ the rejection probability for the false null hypothesis of independence is approximately equal to the significance level. The power against these alternatives is rather low.

07- 1	$X_t \sim i.i.d.N(0,1)$
07- 2	$X_t \sim i.i.d.U[0,1]$
07- 3	$X_t \sim i.i.d.\chi^2_4$
07- 4	$X_t \sim i.i.d.t_4$
07- 5	$X_t = 0.5X_{t-1} + \epsilon_t$
07- 6	$X_t = 0.9X_{t-1} + \epsilon_t$
07- 7	$X_t = 0.8X_{t-1} + 0.15X_{t-2} + \epsilon_t + 0.3\epsilon_{t-1}$
07- 8	$X_t = \epsilon_t + 0.8\epsilon_{t-2}\epsilon_{t-1}$
07- 9	$X_t = 4(1 + 0.5X^2_{t-1})^{0.5}\epsilon_t$
08- 1	$X_t = \epsilon_t + 0.8\epsilon^2_{t-1}$
08- 2	$X_t = \epsilon_t + 0.8\epsilon^2_{t-2}$
08- 3	$X_t = \epsilon_t + 0.8\epsilon^2_{t-3}$
08- 4	$X_t = \epsilon_t + 0.8(\epsilon^2_{t-1} + \epsilon^2_{t-2} + \epsilon^2_{t-3})$
08- 5	$X_t = 0.8\lvert X_{t-1}\rvert^{0.5} + \epsilon_t$
08- 6	$X_t = sign(X_{t-1}) + 0.43\epsilon_t$
08- 7	$X_t = 0.3X_{t-1} + \epsilon_t$
08- 8	$X_t = X_{t-1} + \epsilon_t$
08- 9	$X_t = 0.8\epsilon_{t-1}X_{t-1} + \epsilon_t$
08-10	$X_t = 4X_{t-1}(1 - X_{t-1})$
08-11	$X_t = (1 + 0.8X^2_{t-1})^{0.5}\epsilon_t$
A-1	$X_t = \sqrt{2}(\epsilon_t + \epsilon_{t-m})$
A-2	$X_t = \epsilon_t + \delta_t\epsilon_{t-1}$

Table 4: Different DGPs that are used to evaluate the performance of $G(3)$ and $X^2(3)$. ϵ_t and δ_t were simulated as i.i.d. standard normal variates.

	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	4.35 %	4.31 %	9.90 %	10.06 %	9.69 %	9.76 %	10.21 %
U[0,1]	4.16 %	4.15 %	9.48 %	9.54 %	9.85 %	9.99 %	9.78 %
χ^2_4	4.47 %	4.49 %	10.63 %	10.52 %	10.03 %	10.04 %	10.60 %
t_4	4.52 %	4.52 %	10.46 %	10.50 %	10.04 %	10.12 %	10.36 %
07- 5	99.95 %	99.93 %	80.68 %	80.39 %	99.98 %	99.98 %	99.98 %
07- 6	100.00 %	100.00 %	99.91 %	99.89 %	100.00 %	100.00 %	100.00 %
07- 7	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
07- 8	4.02 %	4.04 %	8.34 %	8.53 %	9.48 %	9.52 %	11.00 %
07- 9	7.77 %	7.63 %	11.18 %	11.46 %	16.46 %	16.33 %	20.40 %
08- 1	88.55 %	88.23 %	54.93 %	54.47 %	93.76 %	93.65 %	91.48 %
08- 2	47.55 %	47.00 %	26.73 %	26.54 %	66.45 %	66.27 %	89.74 %
08- 3	5.02 %	5.09 %	11.00 %	11.15 %	10.82 %	10.83 %	9.76 %
08- 4	99.99 %	99.98 %	90.30 %	89.42 %	100.00 %	100.00 %	100.00 %
08- 5	81.97 %	81.33 %	43.92 %	43.36 %	90.49 %	90.17 %	92.24 %
08- 6	6.19 %	6.11 %	11.11 %	11.14 %	13.64 %	13.47 %	14.73 %
08- 7	81.92 %	81.14 %	41.41 %	41.17 %	91.11 %	90.83 %	93.66 %
08- 8	100.00 %	100.00 %	99.99 %	99.99 %	100.00 %	100.00 %	100.00 %
08- 9	95.80 %	95.80 %	56.07 %	56.39 %	98.58 %	98.58 %	99.28 %
08-10	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	14.49 %	14.27 %	14.12 %	14.31 %	26.92 %	26.77 %	35.43 %
A - 1	5.64 %	5.70 %	25.09 %	25.52 %	10.87 %	10.85 %	7.71 %
A - 2	6.41 %	6.30 %	11.15 %	11.23 %	14.22 %	14.00 %	16.17 %

Table 5: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 1,000$ observations for a nominal level of 10%. $X^2(3)$ and $G(3)$ are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). $X^2_{no}(3)$ and $G_{no}(3)$ are calculated using non-overlapping $m-$histories. For $X^2_c(3)$ and $G_c(3)$ the critical value is approximated by 7. The test statistic $X^2_a(3)$ is χ^2_4 distributed.

	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	2.36 %	2.39 %	4.90 %	4.84 %	4.96 %	4.95 %	4.80 %
U[0,1]	2.16 %	2.18 %	4.78 %	4.81 %	4.72 %	4.78 %	5.21 %
χ^2_4	2.43 %	2.46 %	5.11 %	5.30 %	4.96 %	4.98 %	5.07 %
t_4	2.57 %	2.54 %	5.61 %	5.57 %	4.96 %	5.11 %	4.92 %
07- 5	99.79 %	99.77 %	70.97 %	70.29 %	99.95 %	99.95 %	99.96 %
07- 6	100.00 %	100.00 %	99.80 %	99.76 %	100.00 %	100.00 %	100.00 %
07- 7	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
07- 8	1.88 %	1.89 %	4.26 %	4.39 %	4.61 %	4.58 %	5.82 %
07- 9	4.08 %	3.97 %	5.86 %	5.92 %	8.74 %	8.56 %	12.46 %
08- 1	82.95 %	82.39 %	42.21 %	41.63 %	89.38 %	89.00 %	85.53 %
08- 2	34.50 %	34.09 %	16.71 %	16.62 %	50.35 %	49.94 %	82.76 %
08- 3	2.83 %	2.80 %	5.62 %	5.79 %	5.85 %	5.83 %	4.74 %
08- 4	99.97 %	99.95 %	83.46 %	82.35 %	100.00 %	99.99 %	100.00 %
08- 5	73.29 %	72.03 %	32.02 %	31.30 %	83.52 %	82.75 %	87.10 %
08- 6	3.19 %	3.14 %	5.58 %	5.91 %	6.98 %	6.84 %	8.17 %
08- 7	72.80 %	71.68 %	29.08 %	28.79 %	83.55 %	82.83 %	89.14 %
08- 8	100.00 %	100.00 %	99.98 %	99.98 %	100.00 %	100.00 %	100.00 %
08- 9	91.49 %	91.51 %	42.71 %	43.26 %	96.36 %	96.36 %	98.47 %
08-10	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	8.10 %	7.95 %	7.71 %	7.92 %	15.88 %	15.70 %	24.32 %
A - 1	3.52 %	3.54 %	15.78 %	16.12 %	6.29 %	6.29 %	3.69 %
A - 2	3.50 %	3.44 %	5.74 %	5.93 %	7.22 %	7.06 %	9.02 %

Table 6: Power of various test statistics for the DGPs given in Table 4 based on 10,000 simulations of $T = 1,000$ observations for a nominal level of 5%. $X^2(3)$ and $G(3)$ are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). $X^2_{no}(3)$ and $G_{no}(3)$ are calculated using non-overlapping $m-$histories. For $X^2_c(3)$ and $G_c(3)$ the critical value is approximated by 8.9. The test statistic $X^2_a(3)$ is χ^2_4 distributed.

	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	0.69 %	0.68 %	1.04 %	1.09 %	1.09 %	1.09 %	1.03 %
U[0,1]	0.64 %	0.65 %	0.88 %	0.94 %	0.98 %	1.01 %	1.08 %
χ^2_4	0.65 %	0.69 %	1.15 %	1.16 %	1.04 %	1.11 %	0.99 %
t_4	0.58 %	0.62 %	1.31 %	1.26 %	0.99 %	1.00 %	0.97 %
07- 5	98.58 %	98.39 %	49.26 %	47.97 %	99.20 %	99.13 %	99.86 %
07- 6	100.00 %	100.00 %	98.68 %	98.45 %	100.00 %	100.00 %	100.00 %
07- 7	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
07- 8	0.35 %	0.34 %	0.80 %	0.73 %	0.69 %	0.69 %	1.18 %
07- 9	1.09 %	1.08 %	1.22 %	1.26 %	1.83 %	1.79 %	3.79 %
08- 1	67.75 %	66.56 %	21.26 %	20.65 %	73.47 %	72.67 %	68.47 %
08- 2	15.93 %	15.60 %	5.43 %	5.36 %	21.79 %	21.39 %	63.40 %
08- 3	0.87 %	0.90 %	1.23 %	1.29 %	1.27 %	1.30 %	0.88 %
08- 4	99.77 %	99.74 %	65.24 %	62.62 %	99.89 %	99.88 %	99.98 %
08- 5	53.09 %	50.99 %	14.49 %	13.67 %	60.48 %	58.85 %	72.27 %
08- 6	0.73 %	0.71 %	1.24 %	1.19 %	1.29 %	1.20 %	2.11 %
08- 7	50.34 %	48.63 %	12.32 %	11.96 %	58.31 %	56.74 %	74.94 %
08- 8	100.00 %	100.00 %	99.75 %	99.70 %	100.00 %	100.00 %	100.00 %
08- 9	76.94 %	76.87 %	20.88 %	21.24 %	83.18 %	82.94 %	94.03 %
08-10	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	2.24 %	2.19 %	1.82 %	1.79 %	3.69 %	3.55 %	9.13 %
A - 1	1.16 %	1.21 %	5.85 %	6.00 %	1.77 %	1.80 %	0.70 %
A - 2	0.89 %	0.83 %	1.37 %	1.37 %	1.39 %	1.34 %	2.39 %

Table 7: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 1,000$ observations for a nominal level of 1%. $X^2(3)$ and $G(3)$ are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). $X^2_{no}(3)$ and $G_{no}(3)$ are calculated using non-overlapping $m-$histories. For $X^2_c(3)$ and $G_c(3)$ the critical value is approximated by 13.6. The test statistic $X^2_a(3)$ is χ^2_4 distributed.

5. Detecting the Lag Structure

5.1. A new Method to Detect the Lag Structure

Matilla-García and Marín [2009] propose a test based on permutation entropy to detect *the most relevant lag order*. To discuss the results of Matilla-García and Marín [2009] we need some further notation.

Let $X_{m,d}(t) = (X_t, X_{t+d}, \cdots, X_{t+(m-1)d})$ for $t = 1, \cdots, K(d)$ with $K(d) = T - (m-1)d$. We say $X_{m,d}(t)$ is of $\pi_i = (i_1, \ldots, i_m)$-type if

a) $X_{t+di_1} \leq X_{t+di_2} \leq \ldots \leq X_{t+di_m}$ and
b) $i_{s-1} < i_s$ if $X_{t+di_{s-1}} = X_{t+di_s}$.

The relative frequency of a symbol $\pi \in S_m$ for the lag order d is denoted by

$$\hat{p}(\pi, d) = \frac{\#\{t \in I | X_{m,d}(t) \text{ is of } \pi\text{-type}\}}{T - d(m-1)}.$$

The unconditional probability that $X_{m,d}(t)$ is of the π-type is denoted by $p(\pi, d)$. To guarantee that $p(\pi, d)$ exists for all π and all d it is assumed that the original series is stationary.

Matilla-García and Marín [2009] define the d-permutation entropy for the embedding dimension m as

$$h(m, d) = -\sum_{\pi \in S_m} p(\pi, d) log(p(\pi, d)).$$

where it is understood that the summation is across all π with positive $p(\pi, d)$.

Claim 4. *(Theorem 3.1 in MGM09) Let $\{X_t\}_{t \in I}$ be a stationary time series. Denote by $h(m, d)$ the d-permutation entropy for a fixed embedding dimension $m > 2$ with $m \in \mathbb{N}$. If the most relevant lag order of the time series is d_0 then*

$$h(m, d_0) = min\{h(m, d) | d \in \mathbb{N}\}.$$

First, in the notation of Matilla-García and Marín [2009] $p(\pi, d)$ refers to the relative frequency of the symbols. But given the stated claim my conjecture is that $p(\pi, d)$ actually denotes the unconditional probability of the symbols.

Second, the definition of "most relevant lag order" seems to me rather vague. The authors state the following definition: "In this regard, operationally, we consider the maximum relevant lag order the one (among all possible lags-d) which supplies more dynamic structure (i.e., more information) in order to identify the data generating process. In this fashion, the more relevant lag order that characterize an $AR(p)$ and a $MA(q)$ processes are clearly p and q, respectively."[11] The quintessence of the claim is that the lag order of the original series $\{X_t\}_{t \in I}$ corresponds to the lag order of the symbols as assigned by the d-permutation entropy.

[11] Matilla-García and Marín [2009], p.2

By construction $h(m,d) \in [0, log(m!)]$ where the maximum is attained if each symbol is equally probable and the minimum is attained if the probability of a particular symbol equals 1. The intuition of the claim is that the probabilities of the different symbols have a larger variation at the "most relevant lag order" d_0 than at any other lag d.

The proof of the claim given in the paper is incomplete. As already mentioned d_0 is in my perception not well defined. Second, there is an error in the proof. Let $\mathcal{K}$ be any subset of S_m and $d \in \mathbb{N}$. Then the authors define $\mathcal{K}_d = \{(\pi, d) | \pi \in \mathcal{K}\}$. In the next step the authors state that given d_0 is the most relevant lag order there exists a subset $\tilde{\mathcal{K}}$ such that $p(\tilde{\mathcal{K}}_{d_0}) = p(\tilde{\mathcal{K}}_{d_0} \cup \{(\pi, d_0)\})$ for every $\pi \in S_m \setminus \tilde{\mathcal{K}}$. But this is not a special property of a particular d_0 as such a set tivially exists for each d, namely $\tilde{\mathcal{K}} = S_m$. Finally, Matilla-García and Marín [2009] argue that $p(\tilde{\mathcal{K}}_d) \leq p(\tilde{\mathcal{K}}_{d_0})$ for all $d \in \mathbb{N}$ implies that $h(m, d_0) \leq h(m, d)$ for all $d \in \mathbb{N}$ using the following chain of inequalities[12]

$$\begin{aligned} h(m, d_0) &= -\textstyle\sum_{(\pi, d_0) \in \tilde{\mathcal{K}}_{d_0}} p(\pi, d_0) log(p(\pi, d_0)) \\ &\leq -\textstyle\sum_{(\pi, d) \in \tilde{\mathcal{K}}_d} p(\pi, d) log(p(\pi, d)) \\ &\leq -\textstyle\sum_{\pi \in S_m} p(\pi, d) log(p(\pi, d)) \quad = \quad h(m, d). \end{aligned} \tag{6}$$

Yet, the first inequality is not a consequence of the assumption that $p(\tilde{\mathcal{K}}_d) \leq p(\tilde{\mathcal{K}}_{d_0})$ for all $d \in \mathbb{N}$. To see this consider an example where $m = 3$. Suppose for some d_0 there are only three symbols possible each with equal probability of 1/3. Let $\tilde{\mathcal{K}}$ consist of these three symbols. This subset has the required property that $p(\tilde{\mathcal{K}}_{d_0}) = p(\tilde{\mathcal{K}}_{d_0} \cup \{(\pi, d_0)\})$ for every $\pi \in S_m \setminus \tilde{\mathcal{K}}$. Now, suppose additionally there is a $d \neq d_0$ with the property that all six symbols have a probability of 1/6. Even though $log(3) = h(3, d_0) < h(3, d) = log(6)$ the first inequality in (6) does not hold as both summations are only across the elements of $\tilde{\mathcal{K}}$.

$$-\sum_{(\pi, d_0) \in \tilde{\mathcal{K}}_{d_0}} p(\pi, d_0) log(p(\pi, d_0)) = log(3)$$

and

$$-\sum_{(\pi, d) \in \tilde{\mathcal{K}}_d} p(\pi, d) log(p(\pi, d)) = \frac{1}{2} log(6)$$

which is less than $log(3)$.

The reasoning in the proof is not correct and has to be adjusted.[13] How the minimizer d_0 of $h(m, d)$ is related to the dependence structure of $\{X_t\}_{t \in I}$ remains open. What we do know is that if X_t is a k-dependent time series that is drawn from a continuous distribution then $h(m, k+1) = log(m!) \geq h(m, d)$ for all $d \in \mathbb{N}$.

5.2. Performance of the Method

Still, minimizing d-permutation entropy could be a useful tool to detect serial dependence. Using 1,000 simulations Matilla-García and Marín [2009] try to substantiate

[12] As $p(\tilde{\mathcal{K}}_{d_0}) = 1$ by construction it is true that $p(\tilde{\mathcal{K}}_d) \leq p(\tilde{\mathcal{K}}_{d_0})$ for all $d \in \mathbb{N}$. Again this is not a feature of a particular d_0. For each d^* a set $\mathcal{K}^*$ can be defined such that $p(\mathcal{K}^*_d) \leq p(\mathcal{K}^*_{d*})$ for all $d \in \mathbb{N}$.

[13] An example of such a time series is given in Appendix C.3.

that their procedure detects the correct dynamic structure for a wide variety of different data generating processes (DGPs) by comparing the means of $\hat{h}(m, d)$ for $m = 3$ and $d = 1, \ldots, 5$ where

$$\hat{h}(m, d) = -\sum_{\pi \in S_m} \hat{p}(\pi, d) log(\hat{p}(\pi, d)).$$

The authors are silent about the variance of $\hat{h}(m, d)$ and in particular about the percentages of how often the different lags d are selected. I will complement their results with these two features using $100,000$ simulations instead of $1,000$.

The data generating processes considered are the following:

E- 1 $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$
E- 2 $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$
E- 3 $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$
E- 4 $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$
E- 5 $X_t = |X_{t-1}|^{0.8} + \epsilon_t$
E- 6 $X_t = sign(X_{t-1}) + \epsilon_t$
E- 7 $X_t = 0.8X_{t-1} + \epsilon_t$
E- 8 $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$
E- 9 $X_t = 4X_{t-1}(1 - X_{t-1})$
A-10 $X_t = \epsilon_t$
A-11 $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$
A-12 $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$
A-13 $X_t = 0.9X_{t-2} + \epsilon_t$

where ϵ_t is i.i.d. $N(0, 1)$. The first 9 DGPs are the same as in Matilla-García and Marín [2009]. The i.i.d. series in Model 10 is used as a benchmark. The DGPs 11, 12, and 13 are added to show that the lag selection procedure might be quite misleading even for large sample sizes.

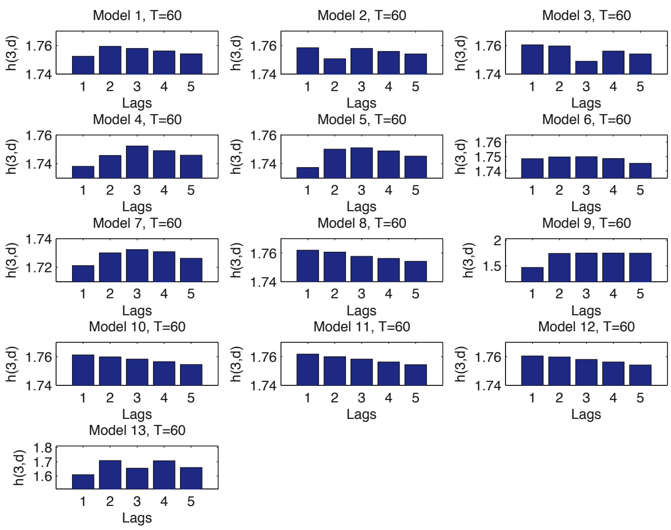


Figure 4: Mean value of $\hat{h}(m, d)$ for $m = 3$ and sample size T=60. $100,000$ simulations.

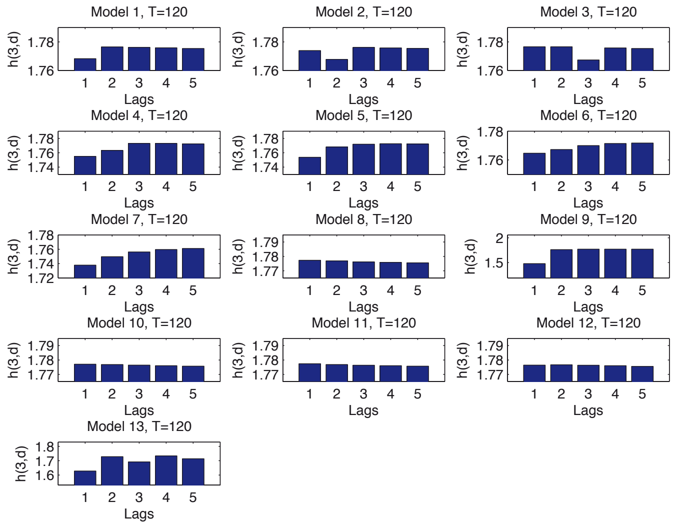


Figure 5: Mean value of $\hat{h}(m, d)$ for $m = 3$ and sample size T=120. $100,000$ simulations.

Figures 4, 5, and 6 and Tables 8 and 9 summarize the simulation results. If we focus on the mean of $\hat{h}(m, d)$ the results in Matilla-García and Marín [2009] are largely confirmed as can be seen in Figures 4, 5, and 6. Table 8 displays the standard deviation of $\hat{h}(m, d)$ which is quite large compared to the differences in the mean. For a sample size of $T = 60$ the "most relevant lag order" in a heuristic sense is found in about $20 - 30\%$ of the cases as can be seen in Table 9. This is at the best only marginally above a random selection of the lag order. The impressive exception is the deterministic model 9 where in virtually all cases a lag of 1 is assigned.[14] For the i.i.d. process 10 the lag order 5 is selected in more than 25% of the simulations. This is driven by the fact that the number of observations, $K(d) = T - (m - 1)d$, decreases with d and hence the variability of $\hat{h}(m, d)$ increases. For the AR(2) process in model 13 a lag of 1 is assigned with high probability. The picture changes only moderatly when $T = 120$. Again, the probability of an incorrect lag order is above 50% for most models.

For a sample size of 500 observations the "most relevant lag order" - again in a heuristic sense - is detected with high probability for models 1, 2, 3, 5, 6, 7, and 9. For the models 8, 11, and 12 all lags are selected with about the same probability. Model 4 is assigned a lag order of 1 even though 3 would probably be the "most relevant". Model 13 is assigned a lag order of 1 instead of 2.

The d-permutation entropy is sensitive to serial dependencies. Yet, the result stated in Matilla-García and Marín [2009] that $h(m, d)$ is minimized at the "most relevant lag order" needs some clarification. Simulations indicate that d-permutation entropy is capable of detecting certain kinds of serial dependence but not others. The reason is that the lag structure of the original series does not necessarily correspond to the lag

[14] For $m = 3$ and $d = 1$ the symbol $\pi = (2, 1, 0)$ is equivalent to $X_{t+2} < X_{t+1} < X_t$. But $X_t > X_{t+1}$ implies that $3/4 < X_t \leq 1$ and $0 \leq X_{t+1} < 3/4$. This in turn implies that $X_{t+2} \geq X_{t+1}$. The symbol $\pi = (2, 1, 0)$ is not possible. This particular DGP is a showcase for the lag selection technique.

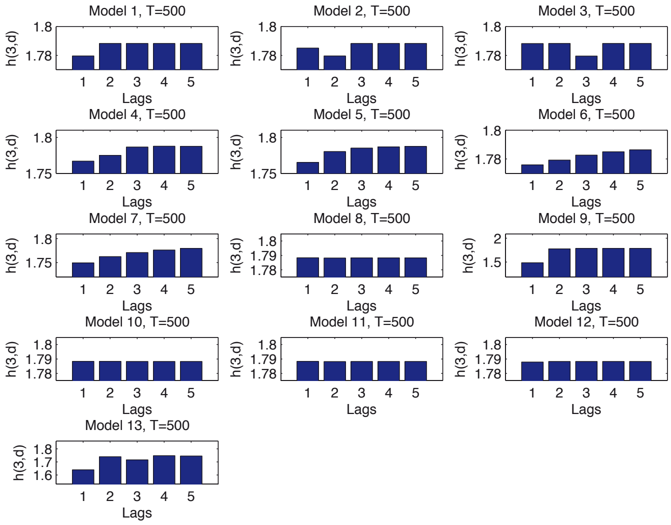


Figure 6: Mean value of $\hat{h}(m,d)$ for $m = 3$ and sample size T=500. $100,000$ simulations.

structure of the symbols.

	1	2	3	4	5
E-1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	0.0331	0.0262	0.0270	0.0280	0.0292
E-2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	0.0260	0.0341	0.0270	0.0283	0.0292
E-3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	0.0266	0.0260	0.0352	0.0280	0.0292
E-4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	0.0381	0.0342	0.0311	0.0329	0.0348
E-5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	0.0373	0.0315	0.0314	0.0323	0.0345
E-6: $X_t = sign(X_{t-1}) + \epsilon_t$	0.0315	0.0310	0.0311	0.0319	0.0337
E-7: $X_t = 0.8X_{t-1} + \epsilon_t$	0.0434	0.0417	0.0422	0.0440	0.0472
E-8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	0.0238	0.0251	0.0271	0.0280	0.0292
E-9: $X_t = 4X_{t-1}(1 - X_{t-1})$	0.0741	0.0441	0.0349	0.0352	0.0366
A-10: $X_t = \epsilon_t$	0.0251	0.0259	0.0268	0.0280	0.0290
A-11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	0.0242	0.0255	0.0268	0.0282	0.0292
A-12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	0.0253	0.0260	0.0269	0.0279	0.0295
A-13: $X_t = 0.9X_{t-2} + \epsilon_t$	0.0943	0.0489	0.1016	0.0542	0.1046

(a) T=60

	1	2	3	4	5
E-1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	0.0193	0.0124	0.0127	0.0128	0.0131
E-2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	0.0136	0.0197	0.0126	0.0129	0.0130
E-3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	0.0130	0.0123	0.0199	0.0128	0.0131
E-4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	0.0234	0.0194	0.0153	0.0150	0.0153
E-5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	0.0228	0.0171	0.0154	0.0149	0.0152
E-6: $X_t = sign(X_{t-1}) + \epsilon_t$	0.0182	0.0171	0.0160	0.0153	0.0150
E-7: $X_t = 0.8X_{t-1} + \epsilon_t$	0.0277	0.0253	0.0237	0.0230	0.0225
E-8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	0.0116	0.0120	0.0127	0.0129	0.0132
E-9: $X_t = 4X_{t-1}(1 - X_{t-1})$	0.0500	0.0244	0.0165	0.0158	0.0160
A-10: $X_t = \epsilon_t$	0.0122	0.0122	0.0125	0.0126	0.0129
A-11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	0.0117	0.0121	0.0125	0.0128	0.0130
A-12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	0.0125	0.0124	0.0126	0.0128	0.0130
A-13: $X_t = 0.9X_{t-2} + \epsilon_t$	0.0689	0.0310	0.0680	0.0318	0.0638

(b) T=120

	1	2	3	4	5
E-1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	0.0079	0.0029	0.0029	0.0029	0.0029
E-2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	0.0044	0.0079	0.0029	0.0029	0.0029
E-3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	0.0031	0.0029	0.0079	0.0029	0.0029
E-4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	0.0101	0.0078	0.0044	0.0034	0.0034
E-5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	0.0099	0.0063	0.0046	0.0039	0.0035
E-6: $X_t = sign(X_{t-1}) + \epsilon_t$	0.0072	0.0064	0.0055	0.0047	0.0040
E-7: $X_t = 0.8X_{t-1} + \epsilon_t$	0.0125	0.0109	0.0096	0.0085	0.0076
E-8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	0.0027	0.0029	0.0029	0.0029	0.0029
E-9: $X_t = 4X_{t-1}(1 - X_{t-1})$	0.0238	0.0095	0.0041	0.0036	0.0035
A-10: $X_t = \epsilon_t$	0.0029	0.0029	0.0028	0.0029	0.0029
A-11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	0.0028	0.0029	0.0029	0.0029	0.0029
A-12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	0.0031	0.0029	0.0029	0.0029	0.0029
A-13: $X_t = 0.9X_{t-2} + \epsilon_t$	0.0341	0.0141	0.0309	0.0140	0.0262

(c) T=500

Table 8: Standard deviation of $\hat{h}(m, d)$ based on 100,000 simulations with $m = 3$. T is the number of observations.

	1	2	3	4	5
E-1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	24.06 %	15.01 %	18.31 %	19.71 %	22.91 %
E-2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	14.83 %	24.13 %	18.20 %	19.94 %	22.91 %
E-3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	14.57 %	13.96 %	29.11 %	19.69 %	22.66 %
E-4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	29.26 %	19.57 %	14.17 %	16.64 %	20.36 %
E-5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	30.97 %	15.80 %	15.55 %	16.68 %	21.00 %
E-6: $X_t = sign(X_{t-1}) + \epsilon_t$	21.04 %	18.31 %	18.43 %	18.82 %	23.40 %
E-7: $X_t = 0.8X_{t-1} + \epsilon_t$	29.17 %	18.22 %	15.36 %	15.72 %	21.54 %
E-8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	15.07 %	15.53 %	21.34 %	22.40 %	25.66 %
E-9: $X_t = 4X_{t-1}(1 - X_{t-1})$	99.92 %	0.05 %	0.00 %	0.01 %	0.02 %
A-10: $X_t = \epsilon_t$	15.83 %	16.26 %	20.37 %	22.07 %	25.47 %
A-11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	15.26 %	16.44 %	20.56 %	22.26 %	25.48 %
A-12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	16.47 %	16.44 %	20.20 %	21.60 %	25.29 %
A-13: $X_t = 0.9X_{t-2} + \epsilon_t$	50.72 %	9.51 %	12.86 %	10.14 %	16.78 %

(a) T=60

	1	2	3	4	5
E-1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	34.62 %	13.79 %	16.54 %	16.62 %	18.43 %
E-2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	16.23 %	33.06 %	16.09 %	16.46 %	18.16 %
E-3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	14.22 %	13.03 %	37.80 %	16.71 %	18.24 %
E-4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	44.48 %	25.48 %	9.63 %	9.48 %	10.93 %
E-5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	49.57 %	16.90 %	11.80 %	10.18 %	11.55 %
E-6: $X_t = sign(X_{t-1}) + \epsilon_t$	30.53 %	22.93 %	17.41 %	14.45 %	14.67 %
E-7: $X_t = 0.8X_{t-1} + \epsilon_t$	45.02 %	21.47 %	13.01 %	10.08 %	10.42 %
E-8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	16.96 %	16.81 %	21.60 %	21.57 %	23.07 %
E-9: $X_t = 4X_{t-1}(1 - X_{t-1})$	100.00 %	0.00 %	0.00 %	0.00 %	0.00 %
A-10: $X_t = \epsilon_t$	17.60 %	17.12 %	21.03 %	21.24 %	23.02 %
A-11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	17.03 %	17.53 %	21.23 %	21.31 %	22.90 %
A-12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	19.15 %	17.44 %	20.46 %	20.46 %	22.49 %
A-13: $X_t = 0.9X_{t-2} + \epsilon_t$	81.66 %	5.63 %	5.53 %	4.13 %	3.06 %

(b) T=120

	1	2	3	4	5
E-1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	74.08 %	5.62 %	6.78 %	6.58 %	6.94 %
E-2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	13.54 %	68.35 %	6.14 %	5.63 %	6.34 %
E-3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	6.53 %	5.63 %	74.62 %	6.50 %	6.72 %
E-4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	73.50 %	24.91 %	0.85 %	0.33 %	0.41 %
E-5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	89.15 %	8.04 %	1.56 %	0.73 %	0.52 %
E-6: $X_t = sign(X_{t-1}) + \epsilon_t$	52.96 %	27.88 %	10.90 %	5.21 %	3.06 %
E-7: $X_t = 0.8X_{t-1} + \epsilon_t$	76.16 %	17.37 %	4.26 %	1.50 %	0.71 %
E-8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	18.28 %	19.31 %	21.34 %	20.06 %	21.02 %
E-9: $X_t = 4X_{t-1}(1 - X_{t-1})$	100.00 %	0.00 %	0.00 %	0.00 %	0.00 %
A-10: $X_t = \epsilon_t$	19.02 %	18.01 %	21.03 %	20.43 %	21.51 %
A-11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	18.44 %	19.47 %	20.79 %	20.14 %	21.16 %
A-12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	23.12 %	17.97 %	19.84 %	19.04 %	20.02 %
A-13: $X_t = 0.9X_{t-2} + \epsilon_t$	99.87 %	0.09 %	0.02 %	0.02 %	0.00 %

(c) T=500

Table 9: Percentages of selected lag as minimizer of $\hat{h}(m, d)$. 100,000 simulations, T observations each

6. Test for Independence between Time Series

In Matilla-García et al. [2010] it is suggested to use permutation entropy to test whether two time series are independent of each other. Let $\{\mathbf{X}_t\}_{t\in I}$ be a two-dimensional time series where $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$. Denote the m-history at time t by

$$\mathbf{X}_m(t) = \begin{pmatrix} X_{1,t}, X_{1,t+1}, \ldots, X_{1,t+m-1} \\ X_{2,t}, X_{2,t+1}, \ldots, X_{2,t+m-1} \end{pmatrix}$$

for $t = 1, \cdots, K$ where $K = T - m + 1$. The null hypothesis is that $\{X_{1,t}\}_{t\in I}$ and $\{X_{2,t}\}_{t\in I}$ are independent of each other. Let $\pi_i = (i_1, \ldots, i_m)$ and $\pi_j = (j_1, \ldots, j_m)$ be two elements of S_m. We say that $\mathbf{X}_m(t)$ is of the (π_i, π_j)-type if

a) $X_{1,t+i_1} \leq X_{1,t+i_2} \leq \ldots \leq X_{1,t+i_m}$,
b) $i_{s-1} < i_s$ if $X_{1,t+i_{s-1}} = X_{1,t+i_s}$,
c) $X_{2,t+j_1} \leq X_{2,t+j_2} \leq \ldots \leq X_{2,t+j_m}$, and
d) $j_{s-1} < j_s$ if $X_{2,t+j_{s-1}} = X_{2,t+j_s}$.

The indicator variable $Z_{i,j,t}$ equals 1 if $\mathbf{X}_m(t)$ is of the (π_i, π_j)-type and 0 otherwise.[15] Matilla-García et al. [2010] claim that $Z_{i,j,t}$ is a Bernoulli variable with some success probability $p_{i,j}$. Yet without any additional assumption on $\{X_{1,t}\}_{t\in I}$ and $\{X_{2,t}\}_{t\in I}$ the success probability need not be time-invariant.[16] Let us assume that $\{X_{1,t}\}_{t\in I}$ and $\{X_{2,t}\}_{t\in I}$ are stationary time series. Jointly with the null hypothesis of independence between the series this implies that the success probability $p_{i,j}$ is time invariant and that $Z_{i,j,t}$ is a Bernoulli variable.

The $m!^2$ counting variables are defined by

$$Y_{i,j,K} = \sum_{t=1}^{K} Z_{i,j,t}.$$

Matilla-García et al. [2010] claim that $Y_{i,j,K}$ is binomially distributed. But even in the case where each of the series $\{X_{1,t}\}$ and $\{X_{2,t}\}$ is i.i.d. the mapping onto the symbols induces dependence between consecutive $Z_{i,j,t}$'s. Hence, $Y_{i,j,K}$ is in general *not* binomially distributed. Accordingly, the vector $Y(K,m) = (Y_{1,1,K}, Y_{1,2,K}, \cdots, Y_{m!,m!,K})'$ is *not* multinomially distributed either. As a consequence

$$\Lambda(m) = -2\left[Kln(K) + \sum_{i=1}^{m!}\sum_{j=1}^{m!} y_{i,j,K} ln\left(\frac{p_{i,j}}{y_{i,j,K}}\right)\right]$$

is not the likelihood ratio statistic and $\Lambda(m)$ is not necessarily asymptotically $\chi^2_{(m!-1)^2}$

[15] To keep the notation simple I use $Z_{i,j,t}$ as an abbreviation for $Z_{\pi_i,\pi_j,t}$.

[16] Matilla-García et al. [2010] state explicitly that they do not need any stationarity assumption on $\{X_{1,t}\}_{t\in I}$ and $\{X_{2,t}\}_{t\in I}$. It is not explained why $p_{i,j}$ should be time invariant.

distributed as claimed in Proposition 1 in Matilla-García et al. [2010]. Simulations[17] indicate that $\Lambda(m)$ is indeed not $\chi^2_{(m!-1)^2}$ distributed even if both $\{X_{1,t}\}$ and $\{X_{2,t}\}$ are i.i.d and independent from each other (see Figure 7a).

Note that $\Lambda(m)$ can be rewritten as

$$\Lambda(m) = 2K \sum_{i=1}^{m!} \sum_{j=1}^{m!} \frac{y_{i,j,K}}{K} ln \left(\frac{y_{i,j,K}/K}{p_{i,j}} \right).$$

Using a Taylor series expansion about $x^0_{i,j} = p_{i,j}$ we get that

$$\Lambda(m) \approx X^2(m) = \sum_{i=1}^{m!} \sum_{j=1}^{m!} \frac{(y_{i,j,K} - Kp_{i,j})^2}{Kp_{i,j}}. \tag{7}$$

The permutation entropy of the two-dimensional process is given by

$$h_{\mathbf{X}}(m) = - \sum_{\pi_i \in S_m} \sum_{\pi_j \in S_m} p_{i,j} ln(p_{i,j}).$$

It is estimated by the relative frequency, i.e.

$$\hat{h}_{\mathbf{X}}(m) = - \sum_{\pi_i \in S_m} \sum_{\pi_j \in S_m} \frac{y_{i,j,K}}{K} ln \left(\frac{y_{i,j,K}}{K} \right) := - \sum_{\pi_i \in S_m} \sum_{\pi_j \in S_m} \hat{p}_{i,j} ln(\hat{p}_{i,j}).$$

Plugging $\hat{h}_{\mathbf{X}}(m)$ into $\Lambda(m)$ yields

$$\Lambda(m) = -2K \left[\hat{h}_{\mathbf{X}}(m) + \sum_{i=1}^{m!} \sum_{j=1}^{m!} \frac{y_{i,j,K}}{K} ln(p_{i,j}) \right].$$

The permutation entropy measures for $\{X_{1,t}\}_{t \in I}$ and $\{X_{2,t}\}_{t \in I}$ are given by $h_{X_1}(m) = -\sum_{\pi_i \in S_m} p_{i+} ln(p_{i+})$ and $h_{X_2}(m) = -\sum_{\pi_i \in S_m} p_{+i} ln(p_{+i})$ where p_{i+} denotes the probability of π_i based on $\{X_{1,t}\}_{t \in I}$ and p_{+i} denotes the probability of π_i based on $\{X_{2,t}\}_{t \in I}$. We estimate p_{i+} and p_{+i} by

$$\hat{p}_{i+} = \frac{y_{i,+,K}}{K} = (1/K) \sum_{j=1}^{m!} y_{i,j,K}$$

and

$$\hat{p}_{+i} = \frac{y_{+,i,K}}{K} = (1/K) \sum_{j=1}^{m!} y_{j,i,K},$$

respectively.

[17] See Figure 7 for some examples.

If $\{X_{1,t}\}$ and $\{X_{2,t}\}$ are independent from each other and $p_{+i} = p_{j+} = 1/m!$ for all i and j then it holds that $\Lambda(m) = 2K(2ln(m!) - \hat{h}_{\mathbf{X}}(m))$. Using that in this case $h_{X_1}(m) = h_{X_2} = ln(m!)$ we may write $\Lambda(m) = 2K(h_{X_1}(m) + h_{X_2}(m) - \hat{h}_{\mathbf{X}}(m))$. Note that $\Lambda(m)$ is a function of the true entropy of the marginals and the estimated entropy of the joint process. On the other hand, Matilla-García et al. [2010] claim in Theorem 1 that

$$\Lambda(m) = 2K(\hat{h}_{X_1}(m) + \hat{h}_{X_2}(m) - \hat{h}_{\mathbf{X}}(m)).$$

This result can not be not correct unless the ture entropy of the marginals equals the estimated entropy.[18] The proof of the theorem is not correct as relative frequencies and probabilities are used interchangeably. In particular it is claimed that $p_{i,j} = \hat{p}_{i+}\hat{p}_{+j}$ [Matilla-García et al., 2010, p. 78]. In general, this is not true.

Let

$$\Lambda^*(m) = 2K(\hat{h}_{X_1}(m) + \hat{h}_{X_2}(m) - \hat{h}_{\mathbf{X}}(m))$$

and observe that

$$\Lambda^*(m) = 2K \sum_{i=1}^{m!} \sum_{j=1}^{m!} \frac{y_{i,j,K}}{K} ln\left(\frac{y_{i,j,K}/K}{\hat{p}_{i+}\hat{p}_{+j}}\right).$$

A Taylor approximation yields

$$\Lambda^*(m) \approx X_c^2(m) = \sum_{i=1}^{m!} \sum_{j=1}^{m!} \frac{\left(y_{i,j,K} - \frac{y_{i,+,K}y_{+,j,K}}{K}\right)^2}{\frac{y_{i,+,K}y_{+,j,K}}{K}}.$$

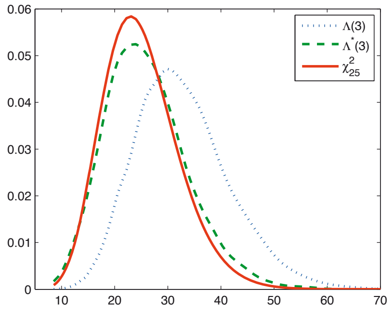
This resembles a Pearson χ^2 test for an $m! \times m!$ contingency table. Under suitable assumptions (e.g. multinomial sampling) $\Lambda^*(m)$ is asymptotically $\chi^2_{(m!-1)^2}$ distributed. But in the application under consideration $Y(K,m)$ is not necessarily multinomially distributed. If we compare the Taylor series expansion of $\Lambda^*(m)$ with that of $\Lambda(m)$ we see that the latter statistic does not condition on the realized marginals $y_{i,+,K}$ and $y_{+,j,K}$.

Figure 7 displays the estimated density functions of $\Lambda(3)$ and $\Lambda^*(3)$ for three different DGPs. The results are based on $10,000$ runs of $T = 1,000$ observations. $\Lambda(3)$ is quite different from the claimed χ^2_{25} distribution. $\Lambda^*(3)$ seems to be well approximated by a χ^2_{25} distribution in two of the three examples.

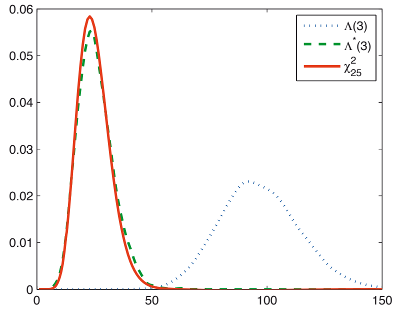
To see that neither $\Lambda(m)$ nor $\Lambda^*(m)$ are asymptotically $\chi^2_{(m!-1)^2}$ distributed we have to take a closer look on the implications of the mapping procedure which induces a fair amount of dependence. As a consequence the rank of the variance covariance matrix is not equal to $(m!-1)^2$.

Observe that the $m!$ symbols in S_m can be grouped into $(m-1)!$ disjoint sets F_i each of which consists of all symbols that agree on the ranking of $(1, \cdots, m-1)$. Each

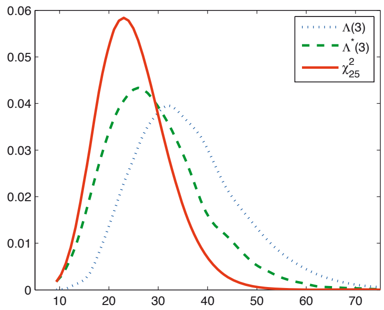
[18] Theorem 1 states that $\hat{\Lambda}(m) = 2K(h_{X_1}(m) + h_{X_2}(m) - h_{\mathbf{X}}(m))$. In their notation $\hat{\Lambda}(m)$ is just the same as $\Lambda(m)$ as can be seen if the first line of the proof on page 78 in Matilla-García et al. [2010] is compared to the representation of Λ in their Proposition 1. Second, the authors denote the estimators of the entropy measure without a hat [Matilla-García et al., 2010, p. 77].



(a) $X_{1,t} = \epsilon_{1,t}$ and $X_{2,t} = \epsilon_{2,t}$ where both $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are i.i.d. $N(0,1)$ and independent from each other.



(b) $X_{1,t} = 0.5X_{1,t-1} + \epsilon_{1,t}$ and $X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t}$ where both $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are i.i.d. $N(0,1)$ and independent from each other. $T = 1,000$.



(c) $X_{1,t} = \sqrt{2}(\epsilon_{1,t} + \epsilon_{1,t-3})$ and $X_{2,t} = \sqrt{2}(\epsilon_{2,t} + \epsilon_{2,t-3})$ where both $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are i.i.d. $N(0,1)$ and independent from each other.

Figure 7: Estimated density functions for $\Lambda(3)$ and $\Lambda^*(3)$ compared to χ^2_{25} based on $10,000$ simulations of $T = 1,000$ observations.

element of F_i has the same successors and each of the successors has only elements in F_i as predecessors. Denote the set of successors of F_i by $H(F_i)$. Each $H(F_i)$ consists of m elements and $H(F_i)$ and $H(F_j)$ are disjoint whenever $F_i \neq F_j$. Each pair (π_i, π_j) belongs to some equivalence class (F_i, F_j) and has m^2 potential successors. The set of successors $H(F_i, F_j)$ of (F_i, F_j) equals the pair $(H(F_i), H(F_j))$. For arbitrary F_i and F_j we get the constraint

$$\mid \sum_{(\pi_a,\pi_b)\in(F_i,F_j)} y_{\pi_a,\pi_b,K} - \sum_{(\pi_a,\pi_b)\in(H(F_i),H(F_j))} y_{\pi_a,\pi_b,K} \mid \leq 1.$$

Let $\pi_1 = (0, 1, \cdots, m-1)$ and $\pi_{m!} = (m-1, m-2, \cdots, 1, 0)$. It is easy to verify that the only pairs of symbols that are potential successors of themselves are (π_1, π_1), $(\pi_1, \pi_{m!})$, $(\pi_{m!}, \pi_1)$, and $(\pi_{m!}, \pi_{m!})$. Without loss of generality denote the equivalence class of π_1 by F_1 and of $\pi_m!$ by $F_{(m-1)!}$.

Define the matrices A and B by

$$a_{i,j} = \begin{cases} 1 & \text{if } \pi_j \in F_i \\ 0 & \text{else} \end{cases} \quad \text{and} \quad b_{i,j} = \begin{cases} 1 & \text{if } \pi_j \in H(F_i) \\ 0 & \text{else} \end{cases}$$

We may summarize these $(m-1)!^2$ constraints in an $(m-1)!^2 \times m!^2$ matrix $U = A \otimes A - B \otimes B$ where $\otimes$ denotes the Kronecker product. The mapping onto the symbols implies that $\|UY(K,m)\|_\infty \leq 1$ where $\|\cdot\|_\infty$ is the maximum norm. As a consequence $\mathbb{V}[UY(K,m)] = O(1)$. If we assume that the variance covariance matrix of $Y(K,m)$ is given by $\Omega_{K,m}$ and that $\lim_{K\to\infty}(1/K)\Omega_{K,m} = \Sigma_m$ exists then

$$\mathbf{0}_{(m-1)!^2,(m-1)!^2} = \lim_{K\to\infty} \mathbb{V}\left[\frac{1}{\sqrt{K}} UY(K,m)\right] = U\Sigma_m U'.$$

The rank of Σ_m has to be less than or equal to $m!^2 - rank(U)$.

The columns of U assigned to (π_1, π_1), $(\pi_1, \pi_{m!})$, $(\pi_{m!}, \pi_1)$ contain zeros only. All other columns contain $+1$ and -1 exactly once. On the one hand this implies that the constraint that $\mathbf{1}'Y(K,m) = K$ is linearly independent of the rows of U, i.e. $rank(\Sigma_m) \leq m!^2 - 1 - rank(U)$. On the other hand, as each column sum is equal to 0, we get that $\mathbf{1}'U = \mathbf{0}$, $rank(U) \leq (m-1)!^2 - 1$.

Any further linear dependence among the rows of U would imply that there is a proper subset $\mathcal{R} \subset \{(F_1, F_1), (F_1, F_2), \cdots, (F_{(m-1)!}, F_{(m-1)!})\}$ such that for any $(F_i, F_j) \in \mathcal{R}$ the set of successors $(H(F_i), H(F_j))$ is in $\mathcal{R}$, too. To see this let $c \neq \mathbf{0}_{(m-1)!^2,1}$ be such that $\mathbf{c}'U = \mathbf{0}_{1,m!^2}$. There exists an i_1 such that $c_{i_1} \neq 0$. We may assume that $c_{i_1} = 1$. U is a matrix of zeros and $rank(U) = 0$ for $m = 2$. Now, let $m > 2$. Each (F_i, F_j) contains at least two pairs (π_a, π_b) with $(\pi_a, \pi_b) \notin (H(F_i), H(F_j))$. Hence, there exists an index j_1 such that $u_{i_1,j_1} = 1$. Column j_1 has to have exactly one other entry that is different from zero in some row i_2, i.e. $u_{i_2,j_1} = -1$. Now, $\mathbf{c}'U = \mathbf{0}_{1,m!2}$ implies that $c_{i_1} = c_{i_2} = 1$. Again, there exist some j_2 such that $u_{i_2,j_2} = 1$ and some i_3 such that $u_{i_3,j_2} = -1$ implying $c_{i_3} = c_{i_1} = 1$. This iteration is terminated as soon as $i_k \in \{i_1, \cdots, i_{k-1}\}$. If $k < (m-1)!$ this implies that there is a proper subset of symbols that is entirely mapped onto itself. In particular, this implies that there is a proper subset $\mathcal{R} \subset \{(F_1, F_1), (F_1, F_2), \cdots, (F_{(m-1)!}, F_{(m-1)!})\}$ such that for any $(F_i, F_j) \in \mathcal{R}$ the set of successors $(H(F_i), H(F_j))$ is in $\mathcal{R}$, too.

Note that each (F_i, F_j) by construction corresponds to some pair $(\pi_i^{m-1}, \pi_j^{m-1}) \in S_{m-1} \times S_{m-1}$. Hence, there has to exist a proper subset $\mathcal{P} \in S_{m-1} \times S_{m-1}$ such that for each $(\pi_i^{m-1}, \pi_j^{m-1}) \in \mathcal{P}$ all successors are again in $\mathcal{P}$. But the only subset of $S_3 \times S_3$ with this particular property is $S_3 \times S_3$ itself. Hence, there exists no proper subset $\mathcal{R} \subset \{(F_1, F_1), (F_1, F_2), \cdots, (F_{(m-1)!}, F_{(m-1)!})\}$ such that for any $(F_i, F_j) \in \mathcal{R}$ the set of successors $(H(F_i), H(F_j))$ is in $\mathcal{R}$. The rank of U equals $(m-1)!^2 - 1$. Therefore there exist exactly $(m-1)!^2$ linearly independent constraints on Σ_m implying that the rank of Σ_m is less than or equal to $m!^2 - (m-1)!^2$.

Theorem 3. *If we assume that the variance covariance matrix of* $Y(K,m)$ *is given by* $\Omega_{K,m}$ *and that* $\lim_{K\to\infty}(1/K)\Omega_{K,m} = \Sigma_m$ *exists then* $rank(\Sigma_m) \le m!^2 - (m-1)!^2$.

$\Lambda(m)$ is asymptotically equivalent to $X^2(m)$ which is a quadratic form in $Y(K,m)$. If $\Lambda(m)$ is asymptotically χ^2 distributed than with at most $m!^2 - (m-1)!^2$ degrees of freedom. In Proposition 1 Matilla-García et al. [2010] derive that $\Lambda(m)$ is asymptotically $\chi^2_{(m!-1)^2}$ distributed. $(m!-1)^2$ is larger than $m!^2 - (m-1)!^2$. Hence, $\Lambda(m)$ can not be $\chi^2_{(m!-1)^2}$ distributed for $m > 4$. But the result is not correct for $m \le 4$ either as will be illustrated in the following example.

Example 2. *Suppose* $\{X_{1,t}\}_{t\in I}$ *and* $\{X_{2,t}\}_{t\in I}$ *are both i.i.d.* $N(0,1)$. *Moreover, assume that both series are independent of each other. Let* $m = 3$. *Applying Theorem 1 we get that*

$$\frac{1}{\sqrt{K}}\left(Y(K,m) - \frac{1}{m!^2}\mathbf{1}\right) \sim^a MVN(0,\Sigma)$$

where

$$\Sigma = \frac{1}{m!^4}\left(m!^2\left(\mathbf{I} + \sum_{l=1}^{m-1}(Q^{(l)}\otimes Q^{(l)}) + \sum_{l=1}^{m-1}(Q^{(l)}\otimes Q^{(l)})'\right) - (2m-1)\mathbf{1}\mathbf{1}'\right).$$

$\mathbf{I}$ *is the* $m!^2 \times m!^2$ *identity matrix,* $\mathbf{1}$ *is an* $m!^2 \times 1$ *vector of ones, and* $Q^{(l)}$ *are the* l*-step transition probabilities induced by the marginals* $\{X_{1,t}\}$ *and* $\{X_{2,t}\}$. $Q^{(1)}$ *and* $Q^{(2)}$ *are given in Tables 1 and 2.* $360^2\Sigma$ *is displayed in Table 10. The rank of this matrix is* $32 = 3!^2 - 2!^2$ *and not* $(3!-1)^2 = 25$.

The test statistic $X^2(m)$ *can be written as*

$$X^2(m) = \frac{1}{K/m!^2}\left(Y(K,m) - \frac{K}{m!^2}\mathbf{1}\right)'\mathbf{I}\left(Y(K,m) - \frac{K}{m!^2}\mathbf{1}\right).$$

Using the results by Ogasawara and Takahashi [1951] we know that $X^2(m)$ *is asymptotically* χ^2 *distributed with* $rank(\Sigma)$ *degrees of freedom if and only if* $m!^2\Sigma\mathbf{I}\Sigma\mathbf{I}\Sigma = \Sigma\mathbf{I}\Sigma$. *Yet, in the example under consideration* $m!^2\Sigma^3 \ne \Sigma^2$. *Hence,* $X^2(3)$ *and* $\Lambda(3)$ *are not asymptotically* χ^2 *distributed.*

Let

$$X_a^2(m) = \frac{1}{K}\left(Y(K,m) - \frac{K}{m!^2}\mathbf{1}\right)'\Sigma^+\left(Y(K,m) - \frac{K}{m!^2}\mathbf{1}\right)$$

where Σ^+ *is the Moore Penrose inverse of* Σ. $X_a^2(m)$ *is asymptotically* χ^2_{32} *distributed. To compare* $X^2(3)$ *and* $\Lambda(3)$ *with* $X_a^2(m)$ *and* χ^2_{32} *I simulated* 10,000 *runs of* $T = 1{,}000$ *observations. Figure 8 displays the (estimated) density functions.* $X_a^2(m)$ *seems to be well approximated by the* χ^2_{32} *distribution.*

Matilla-García et al. [2010] claim in Theorem 1 that not only $\Lambda(m)$ but also $\Lambda^*(m)$ is χ^2 distributed with $(m!-1)^2$ degrees of freedom. The Taylor series approximation

3568	-239	-239	31	31	-392	-239	-185	-446	139	-365	-284	-239	-446	-185	-365	139	-284	31	139	-365	805	-176	-14	31	-365	139	-176	805	-14	-392	-284	-284	-14	-14	148
-239	3154	220	-410	4	31	-446	-392	-212	-302	-167	139	-185	-392	67	-338	-167	-365	-365	-257	112	-68	193	805	139	-257	643	-122	193	-176	-284	-176	-230	40	-176	-14
-239	220	3154	4	-410	31	-185	67	-392	-167	-338	-365	-446	-212	-392	-167	-302	139	139	643	-257	193	-122	-176	-365	112	-257	193	-68	805	-284	-230	-176	-176	40	-14
31	-410	4	3154	220	-239	-365	-338	-167	-392	67	-185	139	-302	-167	-392	-212	-446	-176	-122	193	-257	643	139	805	-68	193	-257	112	-365	-14	40	-176	-176	-230	-284
31	4	-410	220	3154	-239	139	-167	-302	-212	-392	-446	-365	-167	-338	67	-392	-185	805	193	-68	112	-257	-365	-176	193	-122	643	-257	139	-14	-176	40	-230	-176	-284
-392	31	31	-239	-239	3568	-284	-365	139	-446	-185	-239	-284	139	-365	-185	-446	-239	-14	-176	805	-365	139	31	-14	805	-176	139	-365	31	148	-14	-14	-284	-284	-392
-239	-446	-185	-365	139	-284	3154	-392	-392	-257	-257	-176	220	-212	67	112	643	-230	-410	-302	-338	-68	-122	40	4	-167	-167	193	193	-176	31	139	-365	805	-176	-14
-185	-392	67	-338	-167	-365	-392	3262	-365	-230	-338	-257	67	-365	778	-302	310	112	-338	-230	-302	-32	-230	-68	-167	-338	310	-230	112	193	-365	-257	112	-68	193	805
-446	-212	-392	-167	-302	139	-392	-365	3262	-338	-230	-257	-212	472	-365	310	-248	643	-302	-248	-230	-230	-68	-122	-167	310	-338	112	-230	193	139	643	-257	193	-122	-176
139	-302	-167	-392	-212	-446	-257	-230	-338	3262	-365	-392	643	-248	310	-365	472	-212	-122	-68	-230	-230	-248	-302	193	-230	112	-338	310	-167	-176	-122	193	-257	643	139
-365	-167	-338	67	-392	-185	-257	-338	-230	-365	3262	-392	112	310	-302	778	-365	67	-68	-230	-32	-302	-230	-338	193	112	-230	310	-338	-167	805	193	-68	112	-257	-365
-284	139	-365	-185	-446	-239	-176	-257	-257	-392	-392	3154	-230	643	112	67	-212	220	40	-122	-68	-338	-302	-410	-176	193	193	-167	-167	4	-14	-176	805	-365	139	31
-239	-185	-446	139	-365	-284	220	67	-212	643	112	-230	3154	-392	-392	-257	-257	-176	4	-167	-167	193	193	-176	-410	-338	-302	-122	-68	40	31	-365	139	-176	805	-14
-446	-392	-212	-302	-167	139	-212	-365	472	-248	310	643	-392	3262	-365	-230	-338	-257	-167	-338	310	-230	112	193	-302	-230	-248	-68	-230	-122	139	-257	643	-122	193	-176
-185	67	-392	-167	-338	-365	67	778	-365	310	-302	112	-392	-365	3262	-338	-230	-257	-167	310	-338	112	-230	193	-338	-302	-230	-230	-32	-68	-365	112	-257	193	-68	805
-365	-338	-167	-392	67	-185	112	-302	310	-365	778	67	-257	-230	-338	3262	-365	-392	193	-230	112	-338	310	-167	-68	-32	-230	-230	-302	-338	805	-68	193	-257	112	-365
139	-167	-302	-212	-392	-446	643	310	-248	472	-365	-212	-257	-338	-230	-365	3262	-392	193	112	-230	310	-338	-167	-122	-230	-68	-248	-230	-302	-176	193	-122	643	-257	139
-284	-365	139	-446	-185	-239	-230	112	643	-212	67	220	-176	-257	-257	-392	-392	3154	-176	193	193	-167	-167	4	40	-68	-122	-302	-338	-410	-14	805	-176	139	-365	31
31	-365	139	-176	805	-14	-410	-338	-302	-122	-68	40	4	-167	-167	193	193	-176	3154	-392	-392	-257	-257	-176	220	67	-212	643	112	-230	-239	-185	-446	139	-365	-284
139	-257	643	-122	193	-176	-302	-230	-248	-68	-230	-122	-167	-338	310	-230	112	193	-392	3262	-365	-230	-338	-257	-212	-365	472	-248	310	643	-446	-392	-212	-302	-167	139
-365	112	-257	193	-68	805	-338	-302	-230	-230	-32	-68	-167	310	-338	112	-230	193	-392	-365	3262	-338	-230	-257	67	778	-365	310	-302	112	-185	67	-392	-167	-338	-365
805	-68	193	-257	112	-365	-68	-32	-230	-230	-302	-338	193	-230	112	-338	310	-167	-257	-230	-338	3262	-365	-392	112	-302	310	-365	778	67	-365	-338	-167	-392	67	-185
-176	193	-122	643	-257	139	-122	-230	-68	-248	-230	-302	193	112	-230	310	-338	-167	-257	-338	-230	-365	3262	-392	643	310	-248	472	-365	-212	139	-167	-302	-212	-392	-446
-14	805	-176	139	-365	31	40	-68	-122	-302	-338	-410	-176	193	193	-167	-167	4	-176	-257	-257	-392	-392	3154	-230	112	643	-212	67	220	-284	-365	139	-446	-185	-239
31	139	-365	805	-176	-14	4	-167	-167	193	193	-176	-410	-302	-338	-68	-122	40	220	-212	67	112	643	-230	3154	-392	-392	-257	-257	-176	-239	-446	-185	-365	139	-284
-365	-257	112	-68	193	805	-167	-338	310	-230	112	193	-338	-230	-302	-32	-230	-68	67	-365	778	-302	310	112	-392	3262	-365	-230	-338	-257	-185	-392	67	-338	-167	-365
139	643	-257	193	-122	-176	-167	310	-338	112	-230	193	-302	-248	-230	-230	-68	-122	-212	472	-365	310	-248	643	-392	-365	3262	-338	-230	-257	-446	-212	-392	-167	-302	139
-176	-122	193	-257	643	139	193	-230	112	-338	310	-167	-122	-68	-230	-230	-248	-302	643	-248	310	-365	472	-212	-257	-230	-338	3262	-365	-392	139	-302	-167	-392	-212	-446
805	193	-68	112	-257	-365	193	112	-230	310	-338	-167	-68	-230	-32	-302	-230	-338	112	310	-302	778	-365	67	-257	-338	-230	-365	3262	-392	-365	-167	-338	67	-392	-185
-14	-176	805	-365	139	31	-176	193	193	-167	-167	4	40	-122	-68	-338	-302	-410	-230	643	112	67	-212	220	-176	-257	-257	-392	-392	3154	-284	139	-365	-185	-446	-239
-392	-284	-284	-14	-14	148	31	-365	139	-176	805	-14	31	139	-365	805	-176	-14	-239	-446	-185	-365	139	-284	-239	-185	-446	139	-365	-284	3568	-239	-239	31	31	-392
-284	-176	-230	40	-176	-14	139	-257	643	-122	193	-176	-365	-257	112	-68	193	805	-185	-392	67	-338	-167	-365	-446	-392	-212	-302	-167	139	-239	3154	220	-410	4	31
-284	-230	-176	-176	40	-14	-365	112	-257	193	-68	805	139	643	-257	193	-122	-176	-446	-212	-392	-167	-302	139	-185	67	-392	-167	-338	-365	-239	220	3154	4	-410	31
-14	40	-176	-176	-230	-284	805	-68	193	-257	112	-365	-176	-122	193	-257	643	139	139	-302	-167	-392	-212	-446	-365	-338	-167	-392	67	-185	31	-410	4	3154	220	-239
-14	-176	40	-230	-176	-284	-176	193	-122	643	-257	139	805	193	-68	112	-257	-365	-365	-167	-338	67	-392	-185	139	-167	-302	-212	-392	-446	31	4	-410	220	3154	-239
148	-14	-14	-284	-284	-392	-14	805	-176	139	-365	31	-14	-176	805	-365	139	31	-284	139	-365	-185	-446	-239	-284	-365	139	-446	-185	-239	-392	31	31	-239	-239	3568

Table 10: The asymptotic variance covariance matrix Σ for $(1/\sqrt{K})Y(K,m)$ in Example 2. To improve the readability the entries in Σ were multiplied by 360^2.

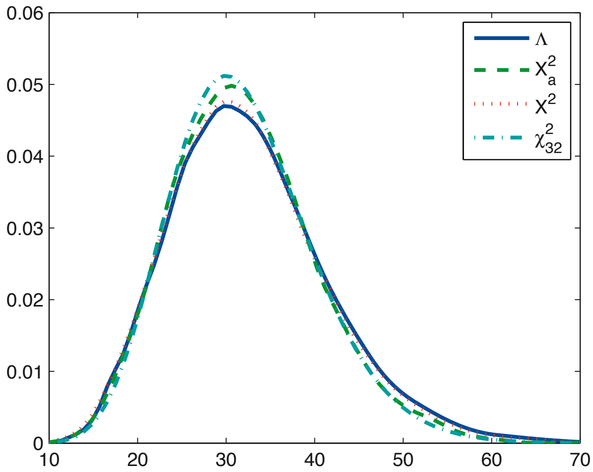


Figure 8: Estimated density functions for $\Lambda(3)$, $X^2(3)$, and $X_a^2(3)$ compared to χ_{32}^2 based on $10,000$ simulations of $T = 1,000$ uncorrelated bivariate i.i.d standard normal variates.

shows that $\Lambda^*(m)$ is asymptotically equivalent to

$$X_c^2(m) = \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{\left(y_{i,j,K} - \frac{y_{i,+,K}y_{+,j,K}}{K}\right)^2}{\frac{y_{i,+,K}y_{+,j,K}}{K}}.$$

$X_c^2(m)$ is conditioned on the marginals $y_{\bullet,+,K} = (y_{1,+,K}, y_{2,+,K}, \ldots, y_{m!,+,K})$ and $y_{+,\bullet,K} = (y_{+,1,K}, y_{+,2,K}, \ldots, y_{+,m!,K})$. The variance covariance matrix of $Y(K,m)$ conditional on the marginals has to comply with an additional set of constraints, namely

$$\sum_{i=1}^{m!} y_{i,j,K} = y_{+,j,K} \text{ for all } j = 1, \ldots, m!$$

and

$$\sum_{j=1}^{m!} y_{i,j,K} = y_{i,+,K} \text{ for all } i = 1, \ldots, m!$$

These $2m!$ constraints may be summarized by the $2m! \times m!^2$ matrix

$$U_1 = \begin{pmatrix} \mathbf{I}_{m!,m!} \\ \mathbf{0}_{m!,m!} \end{pmatrix} \otimes \mathbf{1}_{1,m!} \quad - \quad \mathbf{1}_{1,m!} \otimes \begin{pmatrix} \mathbf{0}_{m!,m!} \\ \mathbf{I}_{m!,m!} \end{pmatrix}.$$

The variance of $U_1 Y(K,m)$ conditional on the marginals is exactly equal to $\mathbf{0}$. Let

$$U_2 = \begin{pmatrix} U \\ U_1 \end{pmatrix}.$$

U_2 contains exactly $2(m-1)!$ redundant constraints. The rank of U_2 is therefore $(m-1)!^2 + 2m! - 2(m-1)!$ and the rank of the asymptotic conditional variance covariance matrix of $(1/\sqrt{K})Y(K,m)$ has to be less than or equal to $m!^2 - (m-1)!^2 - 2(m! - (m-1)!)$ which is less than $(m!-1)^2$. If $X_c^2(m)$ and $\Lambda^*(m)$ are asymptotically χ^2 distributed then with at most $m!^2 - (m-1)!^2 - 2(m! - (m-1)!)$ degrees of freedom.

Example 3. *We continue with the above example. The matrix U_2 is given in Table 11. We know that*

$$\frac{1}{\sqrt{K}}\left(Y(K,m) - \frac{1}{m!^2}\mathbf{1}\right) \sim^a MVN(0,\Sigma)$$

where

$$\Sigma = \frac{1}{m!^4}\left(m!^2\left(\mathbf{I} + \sum_{l=1}^{m-1}(Q^{(l)} \otimes Q^{(l)}) + \sum_{l=1}^{m-1}(Q^{(l)} \otimes Q^{(l)})'\right) - (2m-1)\mathbf{1}\mathbf{1}'\right).$$

Now let $R = (R_1, R_2)' = (1/\sqrt{K})(Y, U_1 Y)'$. R is asymptotically multivariate normally distributed. The variance of R_1 conditional on R_2 is given by

$$\Sigma_{R_1|R_2} = \Sigma - \Sigma U_1' \left(U_1 \Sigma U_1'\right)^{-} U_1 \Sigma$$

where A^- denotes any generalized inverse of A. $\Sigma_{R_1|R_2}$ is displayed in Table 12 (after multiplication by 360^2 for readability). The rank of $\Sigma_{R_1|R_2}$ equals $24 = m!^2 - (m-1)!^2 - 2(m! - (m-1)!)$.

$\Lambda^*(m)$ is asymptotically equivalent to

$$X_c^2(m) = \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{\left(\frac{1}{\sqrt{K}}(y_{i,j,K} - K\hat{p}_{i,+}\hat{p}_{+,j})\right)^2}{\hat{p}_{i,+}\hat{p}_{+,j}}.$$

For the DGPs under consideration the denominator converges sufficiently fast to $p_{i,+}p_{+,j} = 1/m!^2$ such that

$$X_c^2(m) \approx m!^2 \sum_{i=1}^{m!}\sum_{j=1}^{m!} \left(\frac{1}{\sqrt{K}}(y_{i,j,K} - K\hat{p}_{i,+}\hat{p}_{+,j})\right)^2.$$

Under independence $\mathbb{E}[y_{i,j,K} \mid R_2 = r_2] = K\hat{p}_{i,+}\hat{p}_{+,j}$ such that

$$X_c^2(m) \approx m!^2 \sum_{i=1}^{m!}\sum_{j=1}^{m!} \left(R_1 - \mathbb{E}[R_1 \mid R_2]\right)^2.$$

0	1	-1	1	-1	0	1	1	0	1	0	0	-1	0	-1	0	-1	0	1	1	0	1	0	0	-1	0	-1	0	-1	0	0	0	0	0	0	0
0	-1	1	-1	1	0	0	0	1	0	1	1	0	-1	0	-1	0	-1	0	0	1	0	1	1	0	-1	0	-1	0	-1	0	0	0	0	0	0
0	0	0	0	0	0	-1	0	-1	0	-1	0	1	1	0	1	0	0	-1	0	-1	0	-1	0	1	1	0	1	0	0	0	1	-1	1	-1	0
0	0	0	0	0	0	0	-1	0	-1	0	-1	0	0	1	0	1	1	0	-1	0	-1	0	-1	0	0	1	0	1	1	0	-1	1	-1	1	0
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0
0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0
0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0
0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0
0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0
0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0	0	-1

Table 11: The matrix of constraints U on $Y(K,m)$ for $m=3$. The first $(m-1)!^2 = 4$ rows contain the constraints induced by the mapping. The last $2m!$ represent the summations $\sum_{i=1}^{m!} y_{i,j,K} = y_{+,j,K}$ for all $j = 1,\dots,m!$ and $\sum_{j=1}^{m!} y_{i,j,K} = y_{i,+,K}$ for all $i = 1,\dots,m!$ The rank of U is 8.

2648	-469	-469	-499	-499	-712	-469	275	14	299	-205	86	-469	14	275	-205	299	86	-499	299	-205	665	-316	56	-499	-205	299	-316	665	56	-712	86	86	56	56	428
-469	2414	-340	-670	-436	-499	14	-442	-82	128	83	299	275	-442	197	92	83	-205	-205	-607	-58	62	143	665	299	-607	473	8	143	-316	86	-316	-190	380	-16	56
-469	-340	2414	-436	-670	-499	275	197	-442	83	92	-205	14	-82	-442	83	128	299	299	473	-607	143	8	-316	-205	-58	-607	143	62	665	86	-190	-316	-16	380	56
-499	-670	-436	2414	-340	-469	-205	92	83	-442	197	275	299	128	83	-442	-82	14	-316	8	143	-607	473	299	665	62	143	-607	-58	-205	56	380	-16	-316	-190	86
-499	-436	-670	-340	2414	-469	299	83	128	-82	-442	14	-205	83	92	197	-442	275	665	143	62	-58	-607	-205	-316	143	8	473	-607	299	56	-16	380	-190	-316	86
-712	-499	-499	-469	-469	2648	86	-205	299	14	275	-469	86	299	-205	275	14	-469	56	-316	665	-205	299	-499	56	665	-316	299	-205	-499	428	56	56	86	86	-712
-469	14	275	-205	299	86	2414	-442	-442	-607	-607	-316	-340	-82	197	-58	473	-190	-670	128	92	62	8	380	-436	83	83	143	143	-16	-499	299	-205	665	-316	56
275	-442	197	92	83	-205	-442	2702	-745	-310	-598	-607	197	-745	578	-202	230	-58	92	-310	-202	368	-10	62	83	-598	230	-10	152	143	-205	-607	-58	62	143	665
14	-82	-442	83	128	299	-442	-745	2702	-598	-310	-607	-82	272	-745	230	-148	473	128	-148	-310	-10	332	8	83	230	-598	152	-10	143	299	473	-607	143	8	-316
299	128	83	-442	-82	14	-607	-310	-598	2702	-745	-442	473	-148	230	-745	272	-82	8	332	-10	-310	-148	128	143	-10	152	-598	230	83	-316	8	143	-607	473	299
-205	83	92	197	-442	275	-607	-598	-310	-745	2702	-442	-58	230	-202	578	-745	197	62	-10	368	-202	-310	92	143	152	-10	230	-598	83	665	143	62	-58	-607	-205
86	299	-205	275	14	-469	-316	-607	-607	-442	-442	2414	-190	473	-58	197	-82	-340	380	8	62	92	128	-670	-16	143	143	83	83	-436	56	-316	665	-205	299	-499
-469	275	14	299	-205	86	-340	197	-82	473	-58	-190	2414	-442	-442	-607	-607	-316	-436	83	83	143	143	-16	-670	92	128	8	62	380	-499	-205	299	-316	665	56
14	-442	-82	128	83	299	-82	-745	272	-148	230	473	-442	2702	-745	-310	-598	-607	83	-598	230	-10	152	143	128	-310	-148	332	-10	8	299	-607	473	8	143	-316
275	197	-442	83	92	-205	197	578	-745	230	-202	-58	-442	-745	2702	-598	-310	-607	83	230	-598	152	-10	143	92	-202	-310	-10	368	62	-205	-58	-607	143	62	665
-205	92	83	-442	197	275	-58	-202	230	-745	578	197	-607	-310	-598	2702	-745	-442	143	-10	152	-598	230	83	62	368	-10	-310	-202	92	665	62	143	-607	-58	-205
299	83	128	-82	-442	14	473	230	-148	272	-745	-82	-607	-598	-310	-745	2702	-442	143	152	-10	230	-598	83	8	-10	332	-148	-310	128	-316	143	8	473	-607	299
86	-205	299	14	275	-469	-190	-58	473	-82	197	-340	-316	-607	-607	-442	-442	2414	-16	143	143	83	83	-436	380	62	8	128	92	-670	56	665	-316	299	-205	-499
-499	-205	299	-316	665	56	-670	92	128	8	62	380	-436	83	83	143	143	-16	2414	-442	-442	-607	-607	-316	-340	197	-82	473	-58	-190	-469	275	14	299	-205	86
299	-607	473	8	143	-316	128	-310	-148	332	-10	8	83	-598	230	-10	152	143	-442	2702	-745	-310	-598	-607	-82	-745	272	-148	230	473	14	-442	-82	128	83	299
-205	-58	-607	143	62	665	92	-202	-310	-10	368	62	83	230	-598	152	-10	143	-442	-745	2702	-598	-310	-607	197	578	-745	230	-202	-58	275	197	-442	83	92	-205
665	62	143	-607	-58	-205	62	368	-10	-310	-202	92	143	-10	152	-598	230	83	-607	-310	-598	2702	-745	-442	-58	-202	230	-745	578	197	-205	92	83	-442	197	275
-316	143	8	473	-607	299	8	-10	332	-148	-310	128	143	152	-10	230	-598	83	-607	-598	-310	-745	2702	-442	473	230	-148	272	-745	-82	299	83	128	-82	-442	14
56	665	-316	299	-205	-499	380	62	8	128	92	-670	-16	143	143	83	83	-436	-316	-607	-607	-442	-442	2414	-190	-58	473	-82	197	-340	86	-205	299	14	275	-469
-499	299	-205	665	-316	56	-436	83	83	143	143	-16	-670	128	92	62	8	380	-340	-82	197	-58	473	-190	2414	-442	-442	-607	-607	-316	-469	14	275	-205	299	86
-205	-607	-58	62	143	665	83	-598	230	-10	152	143	92	-310	-202	368	-10	62	197	-745	578	-202	230	-58	-442	2702	-745	-310	-598	-607	275	-442	197	92	83	-205
299	473	-607	143	8	-316	83	230	-598	152	-10	143	128	-148	-310	-10	332	8	-82	272	-745	230	-148	473	-442	-745	2702	-598	-310	-607	14	-82	-442	83	128	299
-316	8	143	-607	473	299	143	-10	152	-598	230	83	8	332	-10	-310	-148	128	473	-148	230	-745	272	-82	-607	-310	-598	2702	-745	-442	299	128	83	-442	-82	14
665	143	62	-58	-607	-205	143	152	-10	230	-598	83	62	-10	368	-202	-310	92	-58	230	-202	578	-745	197	-607	-598	-310	-745	2702	-442	-205	83	92	197	-442	275
56	-316	665	-205	299	-499	-16	143	143	83	83	-436	380	8	62	92	128	-670	-190	473	-58	197	-82	-340	-316	-607	-607	-442	-442	2414	86	299	-205	275	14	-469
-712	86	86	56	56	428	-499	-205	299	-316	665	56	-499	299	-205	665	-316	56	-469	14	275	-205	299	86	-469	275	14	299	-205	86	2648	-469	-469	-499	-499	-712
86	-316	-190	380	-16	56	299	-607	473	8	143	-316	-205	-607	-58	62	143	665	275	-442	197	92	83	-205	14	-442	-82	128	83	299	-469	2414	-340	-670	-436	-499
86	-190	-316	-16	380	56	-205	-58	-607	143	62	665	299	473	-607	143	8	-316	14	-82	-442	83	128	299	275	197	-442	83	92	-205	-469	-340	2414	-436	-670	-499
56	380	-16	-316	-190	86	665	62	143	-607	-58	-205	-316	8	143	-607	473	299	299	128	83	-442	-82	14	-205	92	83	-442	197	275	-499	-670	-436	2414	-340	-469
56	-16	380	-190	-316	86	-316	143	8	473	-607	299	665	143	62	-58	-607	-205	-205	83	92	197	-442	275	299	83	128	-82	-442	14	-499	-436	-670	-340	2414	-469
428	56	56	86	86	-712	56	665	-316	299	-205	-499	56	-316	665	-205	299	-499	86	299	-205	275	14	-469	86	-205	299	14	275	-469	-712	-499	-499	-469	-469	2648

Table 12: The asymptotic conditional variance covariance matrix $\Sigma_{R_1|R_2}$ in Example 3. To improve the readability the entries were multiplied by 360^2.

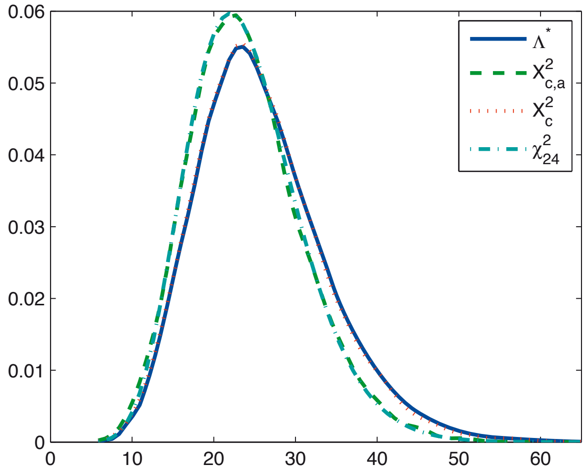


Figure 9: Estimated density functions for $\Lambda^*(3)$, $X^2_c(3)$, and $X^2_{c,a}(3)$ compared to χ^2_{24} based on $10,000$ simulations of $T = 1,000$ uncorrelated bivariate i.i.d standard normal variates.

By the result of Ogasawara and Takahashi [1951] we know that $X^2_c(m)$ is asymptotically χ^2 distributed with $rank(\Sigma_{R_1|R_2})$ degrees of freedom if and only if $m!^2\Sigma^3_{R_1|R_2} = \Sigma^2_{R_1|R_2}$. A simple calculation shows that for $m = 3$ the equality does not hold.

Let

$$X^2_{c,a}(m) = \frac{1}{K}\left(Y(K,m) - \frac{1}{K} y_{\bullet,+,K} \otimes y_{+,\bullet,K}\right)' \Sigma^+_{R_1|R_2}\left(Y(K,m) - \frac{1}{K} y_{\bullet,+,K} \otimes y_{+,\bullet,K}\right)$$

where $\Sigma^+_{R_1|R_2}$ is the Moore Penrose inverse of $\Sigma_{R_1|R_2}$. $X^2_{c,a}(m)$ is asymptotically χ^2_{24} distributed. To compare $X^2_c(3)$ and $\Lambda^*(3)$ with $X^2_{c,a}(m)$ and χ^2_{24} I simulated $10,000$ runs of $T = 1,000$ observations. Figure 9 displays the (estimated) density functions. $X^2_a(m)$ is well approximated by the χ^2_{24} distribution.

If $\{X_{1,t}\}_{t\in I}$ and $\{X_{2,t}\}_{t\in I}$ are both i.i.d. $N(0,1)$ and independent of each other then Figures 8 and 9 illustrate that the adjusted statistics $X^2_a(3)$ and $X^2_{c,a}(3)$ can be approximated by χ^2_{32} and χ^2_{24} distributions reasonably well for a sample size of $T = 1000$ observations. It remains unclear whether this result holds for other DGPs, too.

In a typical application of the test procedures the probabilities of the different symbols are not known and have to be estimated. The unconditional versions of the tests are therefore not applicable. I focus on the conditional test statistics $\Lambda^*(m)$, $X^2_c(m)$ and $X^2_{c,a}(m)$ where the latter can only be calculated if the conditional variance covariance matrix of $Y(K,m)$ is known or can be estimated consistently which is not a trivial task as $Y(K,m)$ is typically observed only once.[19] To estimate the variance covariance matrix

[19]How to estimate the variance of $Y(K,m)$ consistently is discussed in Andrews [1991].

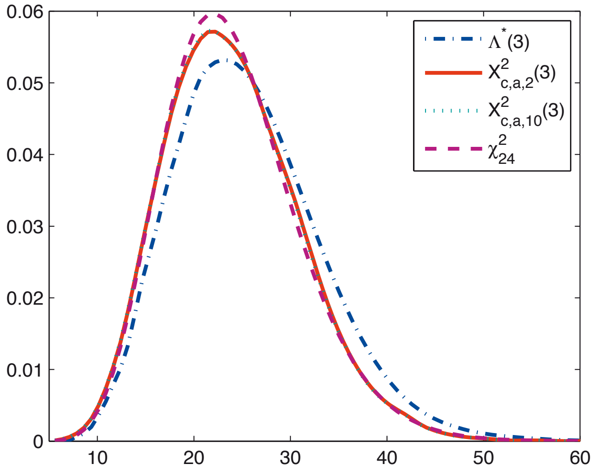


Figure 10: The estimated pdfs of $\Lambda^*(3)$, $X_c^2(3)$, $X_{c,a,2}^2(3)$, and $X_{c,a,10}^2(3)$ compared to χ_{24}^2 for the DGP T-1 based on 10,000 runs of $T = 1,000$ observations.

I assume that $Z_{i,j,t}$ is d-dependent. Under H_0 the asymptotic variance covariance matrix is then given by

$$\begin{aligned}\Sigma &= diag(p^{(1)}) \otimes diag(p^{(2)}) - (2d+1)(p^{(1)}p^{(1)\prime}) \otimes (p^{(2)}p^{(2)\prime}) \\ &+ (diag(p^{(1)}) \otimes diag(p^{(2)})) \textstyle\sum_{l=1}^{d}(Q^{(1,l)} \otimes Q^{(2,l)}) \\ &+ \textstyle\sum_{l=1}^{d}(Q^{(1,l)} \otimes Q^{(2,l)})'(diag(p^{(1)}) \otimes diag(p^{(2)})).\end{aligned}$$

$p^{(1)}$ is the vector of the unconditional probabilities of the different elements of S_m for $\{X_{1,t}\}$. $Q^{(1,l)}$ are the l-step transition probabilities induced by $\{X_{1,t}\}$. $p^{(2)}$ and $Q^{(2,l)}$ are defined analogously. The variance covariance matrix is estimated by substituting relative frequencies for the probabilities. The test statistics based on this estimated Σ are denoted by $X_{c,a,2}^2(3)$ for $d = 2$ and $X_{c,a,10}^2(3)$ for $d = 10$, respectively.

To analyze the performance of the tests consider the following DGPs:

$T-1$ $\quad X_{1,t} = 0.5X_{1,t-1} + \epsilon_{1,t}$ and $X_{2,t} = 0.5X_{2,t-1} + \epsilon_{2,t}$

$T-2$ $\quad X_{1,t} = \sqrt{2}(\epsilon_{1,t} + \epsilon_{1,t-3})$ and $X_{2,t} = \sqrt{2}(\epsilon_{2,t} + \epsilon_{2,t-3})$

Both $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are i.i.d. $N(0,1)$ and independent of each other. Figures 10 and 11 display the estimated density functions for the various test statistics. The quality of the approximation hinges substantially on the estimation of the variance covariance matrix. For model $T-1$ both estimation techniques yield the desired result. In the case of $T-2$ the estimation based on only 2 lags is not precise enough. If we can not be sure that Σ is estimated consistently, the test statistic is of limited value.

Finally, Theorem 2 in Matilla-García et al. [2010] states that if $\{X_{1,t}\}_{t\in I}$ and $\{X_{2,t}\}_{t\in I}$ are jointly stationary and the dependence between these series is of order $\leq m$ then

$$\lim_{T\to\infty} \mathbb{P}(\Lambda^*(m) > C) = 1.$$

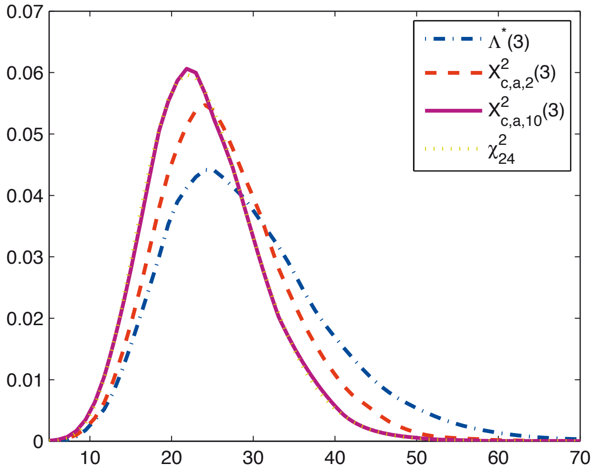


Figure 11: The estimated pdfs of $\Lambda^*(3)$, $X_c^2(3)$, $X_{c,a,2}^2(3)$, and $X_{c,a,10}^2(3)$ compared to χ^2_{24} for the DGP T-2 based on $10,000$ runs of $T = 1,000$ observations.

In the proof the authors show that

$$\hat{h}_{X_1}(m) + \hat{h}_{X_2}(m) - \hat{h}_{\mathbf{X}}(m) = -\sum_{i=1}^{m!}\sum_{j=1}^{m!} \hat{p}_{i,j} ln\left(\frac{\hat{p}_{i+}\hat{p}_{+j}}{\hat{p}_{i,j}}\right) > 0.$$

This is trivially true as even in the case of independence between the series the probability that $\hat{p}_{i,j} = \hat{p}_{i+}\hat{p}_{+j}$ is zero. Matilla-García et al. [2010] conclude that $\hat{h}_{X_1}(m)+\hat{h}_{X_2}(m)-\hat{h}_{\mathbf{X}}(m) > 0$ implies that $\mathbb{P}(\Lambda^*(m) > C)$ tends to one. But this need not be true. If for instance $\hat{h}_{X_1}(m) + \hat{h}_{X_2}(m) - \hat{h}_{\mathbf{X}}(m)$ is $O_p(T^{-1})$ then $\Lambda^*(m)$ is $O_p(1)$ and therefore bounded in probability.

Yet, joint stationarity implies that for $p_{i,j} \neq 0$,

$$\hat{p}_{i,j} ln\left(\frac{\hat{p}_{i+}\hat{p}_{+j}}{\hat{p}_{i,j}}\right) \to p_{i,j} ln\left(\frac{p_{i+}p_{+j}}{p_{i,j}}\right).$$

and hence

$$\hat{h}_{X_1}(m) + \hat{h}_{X_2}(m) - \hat{h}_{\mathbf{X}}(m) \to h_{X_1}(m) + h_{X_2}(m) - h_{\mathbf{X}}(m).$$

If we assume that $\{X_{1,t}\}_{t\in I}$ and $\{X_{2,t}\}_{t\in I}$ are not independent of each other and there exists at least one pair (i,j) such that $p_{i+}p_{+j} \neq p_{i,j}$ then

$$h_{X_1}(m) + h_{X_2}(m) - h_{\mathbf{X}}(m) = c > 0 \quad \text{and} \quad \lim_{T\to\infty} \mathbb{P}(\Lambda^*(m) > C) = 1.$$

7. Conclusions

Testing whether the elements of a time series are i.i.d is a key issue in statistics. In a recent series of papers (Matilla-García [2007], Matilla-García and Marín [2008], Matilla-García and Marín [2009], and Matilla-García et al. [2010]) proposed a new testing procedure. m-tuples of consecutive observations are mapped onto elements of the symmetric group. The observed frequencies of the different elements are then compared to the expected frequencies to detect dependencies in the original series. If the m-tuples overlap then the mapping procedure induces non negligible dependence into the series of symbols. As a consequence the standard G-test and χ^2 test are not necessarily χ^2 distributed and the degrees of freedom differ from the standard case.

Under H_0, the scaled counter of the elements of S_m is asymptotically normally distributed. Due to the mapping procedure the rank of the variance covariance matrix is smaller than in the standard multinomial case, i.e. $m! - (m-1)!$ versus $m! - 1$. The result on the asymptotic distribution of the counter allows on the one hand to represent the G-test and the Pearson χ^2 test by a weighted sum of independent squared standard normal random variables where the weights are known functions of the mean and the variance covariance matrix. On the other hand it is possible to define a quadratic form which is asymptotically χ^2 distributed.

Simulations indicate that all three tests have similar power for almost all DGPs considered. They outperform tests based on non-overlapping m-tuples by a considerable margin. The main drawback of the G-test and the Pearson χ^2 test is that the critical values have to be estimated.

Matilla-García and Marín [2009] suggest to use permutation entropy as a device to detect lag structures. I show that this is only partially appropriate. The proposed procedure has rather low power for small sample sizes. For large sample sizes the lag identified by the procedure need not coincide with the lag order of the original series. The connection between these two lag structures remains an open question.

Finally, it is suggested to use the mapping technique to test whether two time series are independent of each other. Again, the standard G-test and χ^2 test are not necessarily χ^2 distributed. I derive a test statistic that is asymptotically χ^2 distributed if the variance covariance matrix of the counting variable is either known or can be estimated consistently.

References

Donald W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–58, 1991.

Patrick Billingsley. Statistical methods in markov chains. *The Annals of Mathematical Statistics*, 32(1):12–40, 1961.

Kai Lai Chung. *A Course in Probability Theory*. Academic Press, second edition, 2000.

Mariano Matilla-García. A non-parametric test for independence based on symbolic dynamics. *Journal of Economic Dynamics and Control*, 31(12):3889 – 3903, 2007.

Mariano Matilla-García and Manuel Ruiz Marín. A non-parametric independence test using permutation entropy. *Journal of Econometrics*, 144(1):139 – 155, 2008.

Mariano Matilla-García and Manuel Ruiz Marín. Detection of non-linear structure in time series. *Economics Letters*, 105(1):1 – 6, 2009.

Mariano Matilla-García, José Miguel Rodríguez, and Manuel Ruiz Marín. A symbolic test for testing independence between time series. *Journal of Time Series Analysis*, 31(2):76 – 85, 2010.

T. Ogasawara and M. Takahashi. Independence of quadratic forms in normal systems. *J. Sci. Hiroshima University*, 15:1 – 9, 1951.

Simon Tavaré and Patricia M. E. Altham. Serial dependence of observations leading to contingency tables, and corrections to chi-squared statistics. *Biometrika*, 70(1): 139–144, 1983.

A. Proofs

Theorem 1. *Let $\{U_t\}_{t\in\{1,\ldots,T\}}$ be a stationary and m-dependent stochastic process taking values in $\{1,\cdots,J\}$. The indicator variable $Z_{i,t}$ equals 1 if $U_t = i$ and 0 otherwise. Define the counting variable $Y(K)$ by $Y(K) = \sum_{t=1}^{K} Z_t$. Then*

$$\frac{1}{\sqrt{K}}\left(Y(K) - Kp\right) \sim^a MVN(0,\Sigma)$$

where

$$\Sigma = diag(p) - (2m+1)pp' + diag(p)\sum_{l=1}^{m} Q^{(l)} + \sum_{l=1}^{m} Q^{(l)\prime} diag(p).$$

Proof. The proof is an application of Theorem 7.3.1 on page 224 in Chung [2000] which states that if some $\{X_n\}$ is a sequence of m-dependent, uniformly bounded random variables such that

$$\frac{\sqrt{\mathbb{V}[S_n]}}{n^{1/3}} \rightarrow +\infty$$

as $n \rightarrow \infty$. Then

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{V}[S_n]}} \sim^a N(0,1)$$

where $S_n = \sum_{i=1}^{n} X_i$. Using the Cramér - Wold device we have to prove that for any $v \in \mathbb{R}^J$ it holds that $V_t = v'Z_t$ is m-dependent and uniformly bounded and that

$$\frac{\sqrt{\mathbb{V}[S_K]}}{K^{1/3}} \rightarrow +\infty$$

as $K \rightarrow \infty$ where $S_K = v'Y(K) = \sum_{t=1}^{K} V_t$.

Z_t and thereby V_t inherit the stationarity and m-dependence of U_t. Uniform boundedness is not a problem either as $v'Z_t \leq \|v\|_\infty$. The expected value of S_K is given by $Kv'p$. The variance of S_K equals $v'\Omega_K v$ where Ω_K is the variance covariance matrix of $Y(K)$ which has to be determined.

The covariance of $Y_i(K)$ and $Y_j(K)$ can be expressed as

$$\begin{aligned} Cov(Y_i(K), Y_j(K)) &= Cov(\sum_{l=1}^{J} Z_{i,l}, \sum_{t=1}^{J} Z_{j,t}) \\ &= \sum_{l=1}^{J}\sum_{t=1}^{J} \mathbb{E}[(Z_{i,l} - p_i)(Z_{j,t} - p_j)]. \end{aligned}$$

$Z_{i,l}$ and $Z_{j,t}$ are independent whenever $\mid l - t \mid > m$. Observe that for arbitrary l, i, and j

$$\mathbb{E}[Z_{i,t}|Z_{j,t-l}] = \begin{cases} q_{j,i}^{(l)} & \text{for} \quad Z_{j,t-l} = 1 \\ \frac{1}{1-p_j}\sum_{k=1,k\neq j}^{J} p_k q_{k,i}^{(l)} & \text{for} \quad Z_{j,t-l} = 0. \end{cases}$$

Hence

$$p_i = \mathbb{E}[\mathbb{E}[Z_{i,t}|Z_{j,t-l}]] = \sum_{k=1}^{J} p_k q_{k,i}^{(l)}.$$

This implies that $p' = p'Q^{(l)}$ for all l.

To calculate $\mathbb{E}[(Z_{i,l} - p_i)(Z_{j,t} - p_j)]$ consider first the case $l = t$. For $j = i$ it holds that

$$\mathbb{E}[(Z_{i,t} - p_i)(Z_{i,t} - p_i)] = p_i(1 - p_i)$$

and for $j \neq i$ we have

$$\mathbb{E}[(Z_{i,t} - p_i)(Z_{j,t} - p_i)] = -p_i p_j$$

For the second case $l > t$ we get

$$\mathbb{E}[(Z_{i,l} - p_i)(Z_{j,t} - p_j)] = p_j(q_{j,i}^{(l-t)} - p_i)$$

and for third case $l < t$ we derive

$$\mathbb{E}[(Z_{i,l} - p_i)(Z_{j,t} - p_j)] = p_i(q_{i,j}^{(t-l)} - p_j).$$

Hence, the variance-covariance matrix Ω_K of $Y(K)$ is given by $\Omega_K = K\Sigma_K$ with

$$\Sigma_K = diag(p) - pp' + \sum_{l=1}^{m} \frac{K-l}{K}(diag(p)Q^{(l)} - pp') + \sum_{l=1}^{m} \frac{K-l}{K}(Q^{(l)\prime} diag(p) - pp').$$

For $K \to \infty$ we get

$$\Sigma = \lim_{K\to\infty} \Sigma_K = diag(p) - (2m+1)pp' + diag(p)\sum_{l=1}^{m} Q^{(l)} + \sum_{l=1}^{m} Q^{(l)\prime} diag(p).$$

It remains to be shown that

$$\frac{\sqrt{\mathbb{V}[S_K]}}{K^{1/3}} \to +\infty$$

Yet, $S(K) = v'\Omega_K v = O(K)$ implies

$$\frac{\sqrt{\mathbb{V}[S_K]}}{K^{1/3}} = O(K^{1/6}) \to +\infty$$

for $K \to \infty$. Using that $(1/K)\Omega_K \to \Sigma$ yields the desired result that

$$\frac{1}{\sqrt{K}}\,(Y(K) - Kp) \sim^a MVN(0, \Sigma).$$

□

Theorem 2. *Assume that the variance covariance matrix $\Omega_{K,m}$ of $Y(K,m)$ has the property that $\lim_{K\to\infty}(1/K)\Omega_{K,m} = \Sigma_m$ exists. The mapping onto the elements of S_m implies that for $m \geq 2$ the rank of Σ_m is less than or equal to $m! - (m-1)!$*

Proof. For $m = 2$ the fact that $Y_{\pi_1,K} + Y_{\pi_2,K} = K$ implies that $rank(\Sigma_2) \leq 1$.

Assume $m \geq 3$. We have to show that $rank(U) = (m-1)! - 1$. We construct a vector $\mathbf{c} \neq \mathbf{0}_{(\mathbf{m-1})!,\mathbf{1}}$ such that $\mathbf{c}'U = \mathbf{c}'A - \mathbf{c}'B = \mathbf{0}_{1,m!}$. Take an arbitrary index i_1 and set $c_{i_1} = 1$. The row i_1 corresponds to some F_{i_1} which has exactly m elements. Denote the set of all columns that correspond to elements of F_{i_1} by J_1, i.e. $J_1 = \{j \mid a_{i_1,j} = 1\}$. As each symbol is an element of exactly one F_i we can conclude that the j-th coordinate of $\mathbf{c}'A$ equals 1 if $j \in J_1$. This implies that the j-th coordinate of $\mathbf{c}'B$ has to equal 1 if $j \in J_1$. There is at most one element (π_1 or $\pi_{m!}$) that belongs to F_{i_1} and $H(F_{i_1})$ simultaneously. Hence, for $m \geq 3$ there are at least $m - 1 \geq 2$ different columns $j \in J_1$ with $b_{i_1,j} = 0$. Take an arbitrary $j_1 \in J_1$ with $b_{i_1,j} = 0$. As there exists exactly one row i_2 with $b_{i_2,j_1} = 1$, c_{i_2} has to equal 1. The set of all columns that correspond to elements of F_{i_2} is denoted by J_2, i.e. $J_2 = \{j \mid a_{i_2,j} = 1\}$. Iterate this procedure until $i_{k+1} \in I_k = \{i_1, \cdots, i_k\}$. By construction $c_i = 1$ for $i \in I_k$ and $c_i = 0$ otherwise. $\mathbf{c}'A = \mathbf{c}'B$ implies that there exists a subset $\mathcal{R} \subset \{F_1, \cdots, F_{(m-1)!}\}$ such that for any $F_i \in \mathcal{R}$ the set of successors $H(F_i)$ is in $\mathcal{R}$, too. Clearly, for $\mathbf{c} = \mathbf{1}_{(m-1)!,1}$ we find $\mathcal{R} = \{F_1, \cdots, F_{(m-1)!}\}$. At least one constraint is redundant. Any further linear dependence would imply that $\mathcal{R}$ is a *proper* subset of $\{F_1, \cdots, F_{(m-1)!}\}$.

Note that by construction each F_i corresponds to some $\pi^{m-1} \in S_{m-1}$. Hence, there has to exist a *proper* subset $\mathcal{P} \subset S_{m-1}$ such that for each $\pi^{m-1} \in \mathcal{P}$ all successors are again in $\mathcal{P}$. But for $m = 3$ the only subset of S_3 with this particular property is S_3 itself. Hence, by induction there exists no proper subset $\mathcal{R} \subset \{F_1, \cdots, F_{(m-1)!}\}$ such that for any $F_i \in \mathcal{R}$ the set of successors $H(F_i)$ is in $\mathcal{R}$ for all $m \geq 3$. $\mathbf{c}$ has to equal $\mathbf{1}_{(m-1)!,1}$. The rank of U equals $(m-1)! - 1$ and $rank(\Sigma_m) \leq m! - (m-1)!$ $\square$

B. Independence Tests for $T = 100$ and $T = 500$

This section of the Appendix complements the results given in Tables 5, 6, and 7 in Section 3 for smaller sample sizes. Tables 13, 14, and 15 summarize the results for $T = 100$. Tables 16, 17, and 18 summarize the simulations for $T = 500$. The results are extremely similar to those for $T = 1000$. $X_a^2(3)$, $X_c^2(3)$, and $G_c(3)$ have finite sample level that are close to the asymptotic level and have considerably more power than the test based on non overlapping m-tuples.

	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	4.62 %	4.78 %	10.17 %	11.57 %	10.36 %	10.77 %	10.66 %
U[0,1]	4.69 %	4.93 %	9.58 %	10.80 %	9.91 %	10.31 %	10.21 %
χ^2_4	4.39 %	4.52 %	9.67 %	11.19 %	9.78 %	10.33 %	10.68 %
t_4	4.61 %	4.84 %	10.36 %	12.02 %	10.11 %	10.59 %	10.62 %
07- 5	25.01 %	23.42 %	15.55 %	17.61 %	40.04 %	38.63 %	47.18 %
07- 6	71.14 %	69.04 %	31.05 %	32.78 %	83.14 %	81.96 %	87.87 %
07- 7	94.16 %	93.25 %	51.67 %	52.53 %	97.74 %	97.40 %	98.51 %
07- 8	4.37 %	4.47 %	8.27 %	9.60 %	9.42 %	9.68 %	11.08 %
07- 9	5.25 %	5.16 %	10.13 %	11.82 %	11.25 %	11.44 %	12.98 %
08- 1	14.33 %	14.17 %	13.52 %	14.97 %	24.38 %	24.47 %	22.31 %
08- 2	6.52 %	6.82 %	11.31 %	13.17 %	13.77 %	14.26 %	21.81 %
08- 3	5.44 %	5.72 %	10.80 %	12.69 %	11.65 %	12.06 %	10.62 %
08- 4	34.52 %	32.13 %	16.47 %	18.09 %	50.04 %	48.20 %	58.12 %
08- 5	11.74 %	11.04 %	11.93 %	13.85 %	22.33 %	21.65 %	24.39 %
08- 6	4.63 %	4.58 %	9.80 %	11.32 %	10.52 %	10.75 %	12.01 %
08- 7	10.94 %	10.19 %	11.45 %	13.10 %	20.77 %	20.12 %	24.96 %
08- 8	84.40 %	82.69 %	41.58 %	42.32 %	91.98 %	91.31 %	95.39 %
08- 9	14.42 %	13.57 %	13.39 %	15.76 %	26.62 %	25.97 %	33.51 %
08-10	100.00 %	100.00 %	98.17 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	5.77 %	5.55 %	9.97 %	11.58 %	11.76 %	11.90 %	14.87 %
A - 1	5.76 %	5.94 %	24.35 %	26.29 %	11.34 %	11.82 %	7.85 %
A - 2	4.49 %	4.55 %	9.51 %	11.25 %	10.40 %	10.89 %	11.23 %

Table 13: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 100$ observations for a nominal level of 10%. X^2 and G are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). X^2_{no} and G_{no} are calculated using non-overlapping $m-$histories. For X^2_c and G_c the critical value is approximated by 7. The test statistic X^2_a is χ^2_4 distributed.

	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	2.40 %	2.62 %	5.15 %	6.27 %	5.02 %	5.51 %	5.26 %
U[0,1]	2.55 %	2.85 %	4.71 %	5.80 %	5.02 %	5.45 %	5.21 %
χ^2_4	2.25 %	2.43 %	4.71 %	5.93 %	4.73 %	5.15 %	5.26 %
t_4	2.35 %	2.56 %	5.15 %	6.44 %	5.05 %	5.57 %	5.50 %
07- 5	16.08 %	14.78 %	8.68 %	10.06 %	26.12 %	25.49 %	35.00 %
07- 6	60.36 %	57.72 %	20.73 %	22.44 %	72.11 %	71.20 %	81.67 %
07- 7	89.95 %	88.31 %	39.15 %	40.14 %	94.46 %	94.12 %	97.38 %
07- 8	2.23 %	2.36 %	4.00 %	4.74 %	4.66 %	5.08 %	5.76 %
07- 9	2.73 %	2.81 %	5.01 %	6.26 %	5.71 %	5.92 %	6.95 %
08- 1	9.47 %	9.30 %	7.26 %	8.68 %	15.19 %	15.56 %	13.90 %
08- 2	3.28 %	3.58 %	5.81 %	7.32 %	7.07 %	7.86 %	13.05 %
08- 3	2.95 %	3.18 %	5.57 %	6.84 %	5.79 %	6.42 %	5.26 %
08- 4	25.00 %	22.94 %	9.54 %	10.90 %	35.90 %	34.58 %	47.10 %
08- 5	6.77 %	6.42 %	6.15 %	7.40 %	12.53 %	12.38 %	15.42 %
08- 6	2.40 %	2.42 %	4.79 %	5.97 %	5.04 %	5.26 %	6.31 %
08- 7	6.03 %	5.68 %	5.69 %	7.16 %	11.64 %	11.44 %	15.67 %
08- 8	76.49 %	73.96 %	29.49 %	30.19 %	85.07 %	83.96 %	92.08 %
08- 9	8.38 %	7.90 %	7.12 %	8.66 %	15.33 %	15.05 %	23.02 %
08-10	100.00 %	100.00 %	93.87 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	3.06 %	3.00 %	4.89 %	5.71 %	6.14 %	6.21 %	8.44 %
A - 1	3.30 %	3.52 %	15.25 %	17.26 %	6.25 %	6.56 %	3.98 %
A - 2	2.26 %	2.42 %	4.57 %	5.66 %	4.91 %	5.19 %	5.69 %

Table 14: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 100$ observations for a nominal level of 5%. X^2 and G are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). X^2_{no} and G_{no} are calculated using non-overlapping $m-$histories. For X^2_c and G_c the critical value is approximated by 8.9. The test statistic X^2_a is χ^2_4 distributed.

	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	0.57 %	0.63 %	0.96 %	1.57 %	1.02 %	1.17 %	1.15 %
U[0,1]	0.62 %	0.71 %	0.90 %	1.40 %	0.95 %	1.16 %	1.01 %
χ^2_4	0.38 %	0.56 %	0.80 %	1.30 %	0.82 %	1.02 %	1.03 %
t_4	0.53 %	0.62 %	1.11 %	1.61 %	0.97 %	1.08 %	1.08 %
07- 5	5.46 %	5.03 %	2.04 %	3.01 %	8.06 %	7.20 %	16.72 %
07- 6	37.35 %	34.34 %	7.52 %	9.23 %	45.50 %	42.35 %	64.41 %
07- 7	75.67 %	72.78 %	18.15 %	19.91 %	81.58 %	79.17 %	91.69 %
07- 8	0.57 %	0.65 %	0.72 %	1.22 %	0.92 %	1.03 %	1.30 %
07- 9	0.53 %	0.67 %	0.96 %	1.63 %	0.91 %	1.05 %	1.52 %
08- 1	3.63 %	3.62 %	2.04 %	2.70 %	5.12 %	5.08 %	4.67 %
08- 2	0.80 %	0.95 %	1.29 %	1.96 %	1.29 %	1.55 %	3.74 %
08- 3	0.85 %	0.97 %	1.25 %	1.74 %	1.30 %	1.50 %	0.94 %
08- 4	11.62 %	9.85 %	2.41 %	3.36 %	15.43 %	13.51 %	27.81 %
08- 5	2.06 %	1.96 %	1.25 %	1.89 %	3.13 %	3.01 %	5.13 %
08- 6	0.59 %	0.61 %	0.95 %	1.53 %	1.02 %	1.10 %	1.51 %
08- 7	1.69 %	1.47 %	1.19 %	2.12 %	2.66 %	2.40 %	5.38 %
08- 8	57.36 %	52.99 %	12.53 %	13.10 %	64.60 %	61.04 %	82.55 %
08- 9	2.56 %	2.38 %	1.69 %	2.38 %	3.92 %	3.76 %	9.07 %
08-10	100.00 %	100.00 %	73.79 %	96.29 %	100.00 %	100.00 %	100.00 %
08-11	0.69 %	0.73 %	0.91 %	1.35 %	1.22 %	1.18 %	2.31 %
A - 1	1.01 %	1.15 %	5.32 %	6.98 %	1.50 %	1.72 %	0.73 %
A - 2	0.61 %	0.65 %	0.69 %	1.30 %	0.92 %	1.04 %	1.13 %

Table 15: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 100$ observations for a nominal level of 1%. X^2 and G are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). X^2_{no} and G_{no} are calculated using non-overlapping $m-$histories. For X^2_c and G_c the critical value is approximated by 13.6. The test statistic X^2_a is χ^2_4 distributed.

	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	4.44 %	4.54 %	9.46 %	10.10 %	9.75 %	9.88 %	10.08 %
U[0,1]	4.22 %	4.29 %	9.94 %	10.38 %	9.45 %	9.60 %	9.46 %
χ^2_4	4.72 %	4.79 %	9.50 %	9.79 %	9.85 %	9.92 %	10.19 %
t_4	4.30 %	4.32 %	9.75 %	10.06 %	9.88 %	10.03 %	10.09 %
07- 5	92.72 %	92.35 %	48.85 %	48.88 %	97.23 %	97.05 %	98.22 %
07- 6	100.00 %	99.99 %	93.14 %	92.96 %	100.00 %	100.00 %	100.00 %
07- 7	100.00 %	100.00 %	99.74 %	99.74 %	100.00 %	100.00 %	100.00 %
07- 8	4.28 %	4.33 %	8.09 %	8.58 %	9.69 %	9.83 %	10.90 %
07- 9	6.31 %	6.28 %	10.61 %	11.05 %	13.66 %	13.35 %	16.42 %
08- 1	58.00 %	57.38 %	31.96 %	32.31 %	71.18 %	70.74 %	65.00 %
08- 2	22.18 %	22.06 %	17.32 %	17.58 %	37.10 %	37.07 %	62.81 %
08- 3	5.20 %	5.19 %	11.19 %	11.53 %	10.71 %	10.77 %	9.61 %
08- 4	97.05 %	96.64 %	60.99 %	60.14 %	98.92 %	98.78 %	99.48 %
08- 5	47.80 %	46.58 %	25.85 %	25.81 %	64.21 %	63.29 %	68.08 %
08- 6	5.16 %	5.14 %	9.81 %	10.38 %	11.74 %	11.67 %	12.90 %
08- 7	47.74 %	46.59 %	24.38 %	24.72 %	64.80 %	64.21 %	70.92 %
08- 8	100.00 %	100.00 %	97.45 %	97.23 %	100.00 %	100.00 %	100.00 %
08- 9	66.31 %	65.95 %	33.03 %	33.71 %	81.46 %	81.29 %	87.70 %
08-10	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	9.11 %	8.93 %	11.28 %	11.91 %	18.44 %	18.34 %	24.04 %
A - 1	5.98 %	6.10 %	25.12 %	25.85 %	10.91 %	11.07 %	7.72 %
A - 2	5.81 %	5.72 %	10.20 %	10.73 %	12.28 %	12.25 %	13.77 %

Table 16: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 500$ observations for a nominal level of 10%. X^2 and G are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). X^2_{no} and G_{no} are calculated using non-overlapping $m-$histories. For X^2_c and G_c the critical value is approximated by 7. The test statistic X^2_a is χ^2_4 distributed.

	$X^2(3)$	$G(3)$	$X_{no}^2(3)$	$G_{no}(3)$	$X_c^2(3)$	$G_c(3)$	$X_a^2(3)$
N(0,1)	2.29 %	2.32 %	4.72 %	4.92 %	4.96 %	5.02 %	5.24 %
U[0,1]	2.44 %	2.48 %	4.77 %	5.05 %	4.76 %	4.90 %	4.74 %
χ_4^2	2.68 %	2.80 %	4.66 %	4.84 %	5.35 %	5.39 %	4.98 %
t_4	2.33 %	2.40 %	4.75 %	5.03 %	4.79 %	4.84 %	5.06 %
07- 5	87.49 %	86.49 %	36.73 %	36.36 %	93.45 %	93.08 %	96.52 %
07- 6	99.98 %	99.98 %	88.26 %	87.73 %	100.00 %	100.00 %	100.00 %
07- 7	100.00 %	100.00 %	99.32 %	99.25 %	100.00 %	100.00 %	100.00 %
07- 8	2.11 %	2.18 %	4.08 %	4.35 %	4.83 %	4.83 %	5.71 %
07- 9	3.27 %	3.24 %	5.52 %	5.71 %	7.03 %	6.97 %	9.19 %
08- 1	48.55 %	47.83 %	22.00 %	21.80 %	59.85 %	59.35 %	53.24 %
08- 2	14.05 %	13.96 %	9.90 %	10.12 %	23.98 %	23.83 %	50.62 %
08- 3	2.90 %	3.02 %	5.71 %	5.99 %	5.68 %	5.79 %	4.69 %
08- 4	94.22 %	93.56 %	48.79 %	47.42 %	97.40 %	97.10 %	98.97 %
08- 5	36.63 %	34.96 %	16.47 %	16.38 %	50.29 %	48.92 %	55.97 %
08- 6	2.95 %	2.80 %	5.34 %	5.59 %	5.78 %	5.79 %	7.09 %
08- 7	35.47 %	33.79 %	15.03 %	15.33 %	50.10 %	49.10 %	59.61 %
08- 8	100.00 %	100.00 %	94.88 %	94.62 %	100.00 %	100.00 %	100.00 %
08- 9	53.83 %	53.70 %	22.17 %	22.70 %	68.79 %	68.60 %	80.36 %
08-10	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	5.02 %	4.99 %	5.90 %	6.09 %	10.09 %	9.95 %	15.31 %
A - 1	3.71 %	3.76 %	16.24 %	16.33 %	6.63 %	6.69 %	4.13 %
A - 2	3.19 %	3.20 %	5.29 %	5.44 %	6.59 %	6.51 %	7.38 %

Table 17: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 500$ observations for a nominal level of 5%. X^2 and G are based on the (false) assumption that the test are χ_5^2 distributed (columns 1 and 2). X_{no}^2 and G_{no} are calculated using non-overlapping $m-$histories. For X_c^2 and G_c the critical value is approximated by 8.9. The test statistic X_a^2 is χ_4^2 distributed.

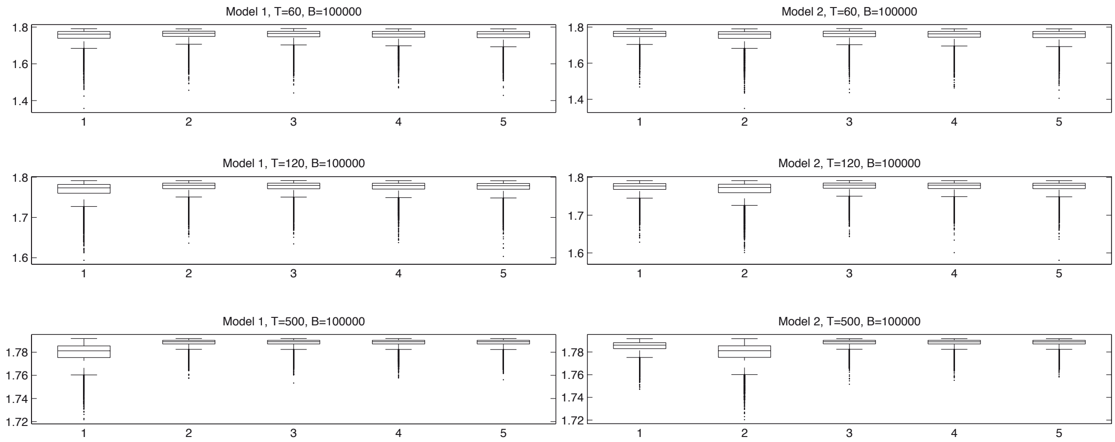
	$X^2(3)$	$G(3)$	$X^2_{no}(3)$	$G_{no}(3)$	$X^2_c(3)$	$G_c(3)$	$X^2_a(3)$
N(0,1)	0.62 %	0.63 %	0.94 %	0.88 %	1.05 %	1.06 %	1.12 %
U[0,1]	0.62 %	0.68 %	0.90 %	1.03 %	1.06 %	1.10 %	0.90 %
χ^2_4	0.66 %	0.66 %	0.88 %	1.03 %	1.05 %	1.04 %	1.10 %
t_4	0.64 %	0.67 %	1.00 %	1.01 %	1.09 %	1.15 %	1.08 %
07- 5	70.37 %	68.41 %	17.65 %	17.15 %	76.85 %	75.33 %	89.06 %
07- 6	99.92 %	99.92 %	73.07 %	71.63 %	99.97 %	99.97 %	99.99 %
07- 7	100.00 %	100.00 %	97.24 %	96.94 %	100.00 %	100.00 %	100.00 %
07- 8	0.54 %	0.54 %	0.79 %	0.88 %	0.90 %	0.91 %	1.27 %
07- 9	0.81 %	0.77 %	1.07 %	1.17 %	1.33 %	1.30 %	2.49 %
08- 1	30.54 %	29.52 %	8.26 %	7.99 %	36.72 %	35.68 %	31.24 %
08- 2	5.18 %	5.10 %	2.87 %	2.84 %	7.62 %	7.58 %	27.38 %
08- 3	0.79 %	0.85 %	1.13 %	1.33 %	1.19 %	1.22 %	0.83 %
08- 4	84.88 %	83.18 %	26.39 %	24.85 %	88.93 %	87.68 %	96.09 %
08- 5	18.51 %	16.99 %	5.63 %	5.32 %	24.32 %	22.86 %	33.19 %
08- 6	0.80 %	0.77 %	1.14 %	1.17 %	1.16 %	1.16 %	1.70 %
08- 7	16.43 %	15.28 %	4.91 %	5.01 %	22.53 %	21.14 %	37.34 %
08- 8	100.00 %	100.00 %	85.57 %	84.51 %	100.00 %	100.00 %	100.00 %
08- 9	30.34 %	29.81 %	8.33 %	8.57 %	38.62 %	38.25 %	60.28 %
08-10	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %
08-11	1.42 %	1.37 %	1.34 %	1.40 %	2.34 %	2.17 %	5.14 %
A - 1	1.39 %	1.44 %	5.69 %	6.02 %	2.05 %	2.13 %	0.88 %
A - 2	0.97 %	0.93 %	1.04 %	1.19 %	1.52 %	1.46 %	1.85 %

Table 18: Power of various test statistics for the DGPs given in Table 4 based on $10,000$ simulations of $T = 500$ observations for a nominal level of 1%. X^2 and G are based on the (false) assumption that the test are χ^2_5 distributed (columns 1 and 2). X^2_{no} and G_{no} are calculated using non-overlapping $m-$histories. For X^2_c and G_c the critical value is approximated by 13.6. The test statistic X^2_a is χ^2_4 distributed.

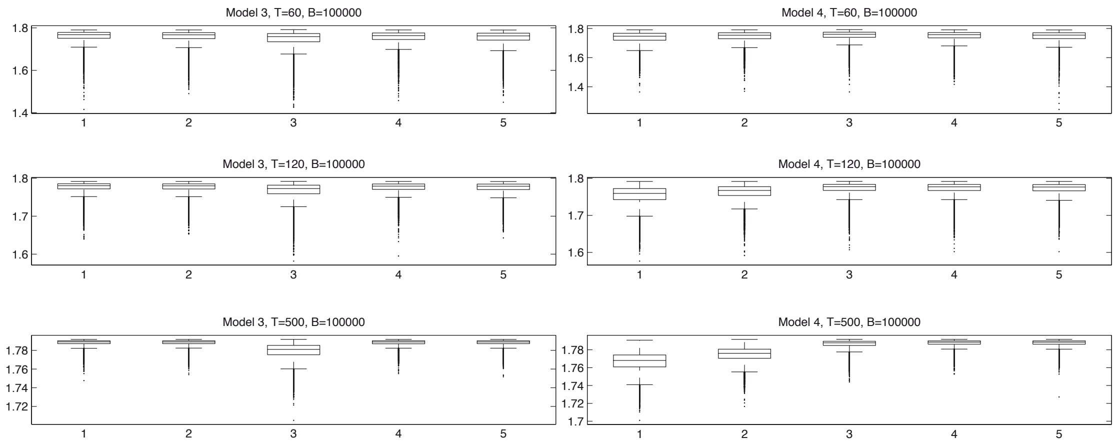
C. More on Matilla-García and Marín [2009]

C.1. Boxplots of $\hat{h}(m, d)$

To get a clearer understanding of the properties of $\hat{h}(m, d)$ the Figures 12, 13, and 14 display boxplots for the various models and numbers of observations.

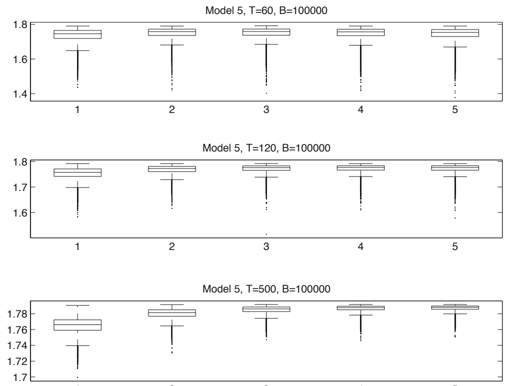


(a) Model 1:$X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$ (b) Model 2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$

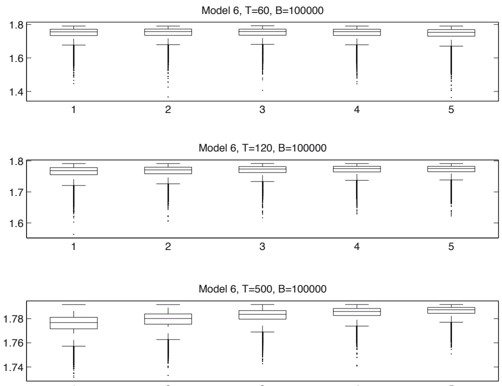


(c) Model 3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$ (d) Model 4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$

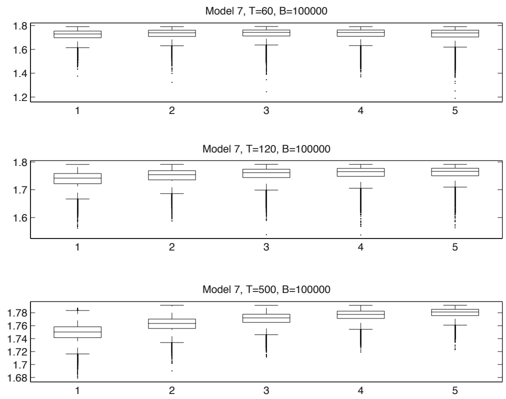
Figure 12: Boxplots for Models 1 to 4. The numbers on the abscissae are the lag orders d. $T = 60, 120, 500$. $100,000$ Simulations.



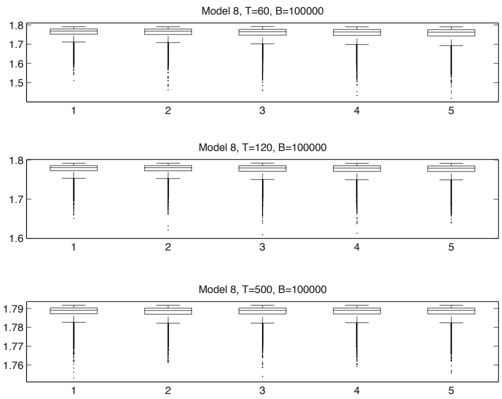
(a) Model 5: $X_t = |X_{t-1}|^{0.8} + \epsilon_t$



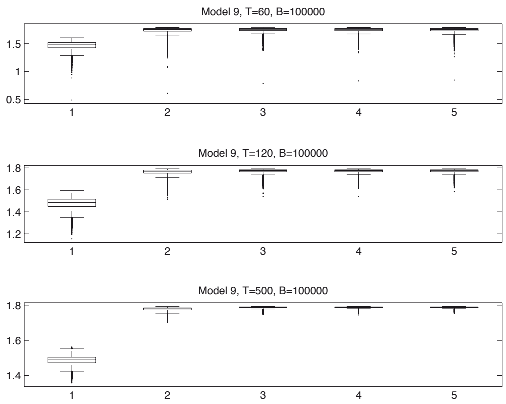
(b) Model 6: $X_t = sign(X_{t-1}) + \epsilon_t$



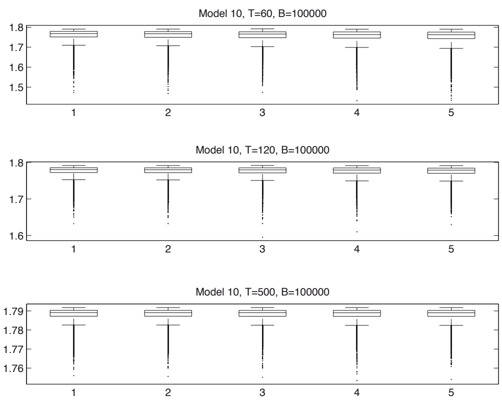
(c) Model 7: $X_t = 0.8X_{t-1} + \epsilon_t$



(d) Model 8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$

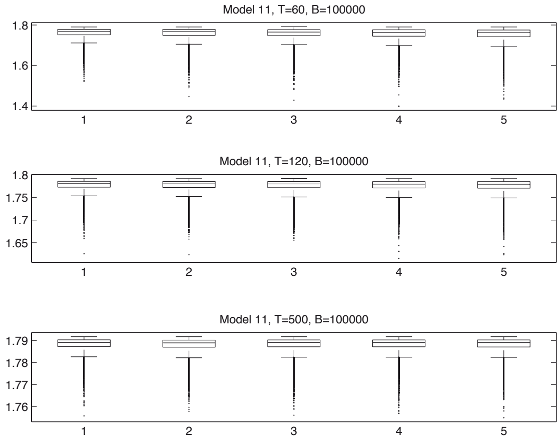


(e) Model 9: $X_t = 4X_{t-1}(1 - X_{t-1})$

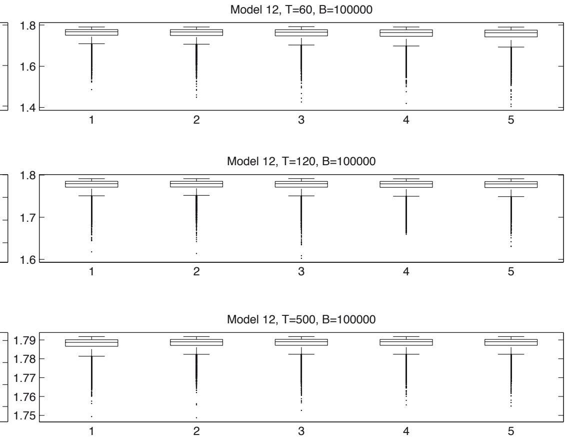


(f) Model 10: $X_t = \epsilon_t$

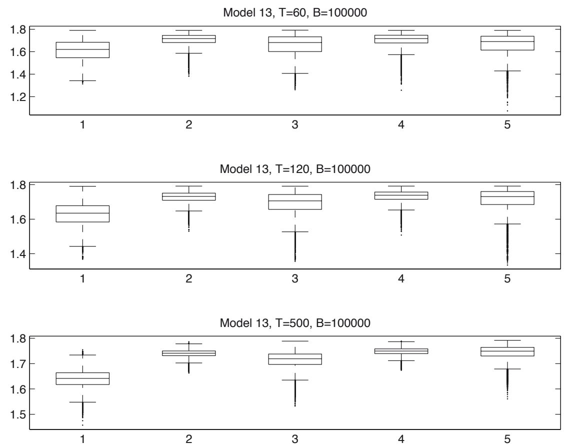
Figure 13: Boxplots for Models 5 to 10. The numbers on the abscissae are the lag orders d. $T = 60, 120, 500$. $100,000$ Simulations.



(a) Model 11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$



(b) Model 12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$



(c) Model 13: $X_t = 0.9X_{t-2} + \epsilon_t$

Figure 14: Boxplots for Models 11 to 13. The numbers on the abscissae are the lag orders d. $T = 60, 120, 500$. $100,000$ Simulations.

C.2. 1,000 Simulations

Matilla-García and Marín [2009] used 1,000 simulations in their paper. For comparative purposes Figures 15, 16, and 17 and Tables 19 and 20 summarize the results based on 1,000 simulations.

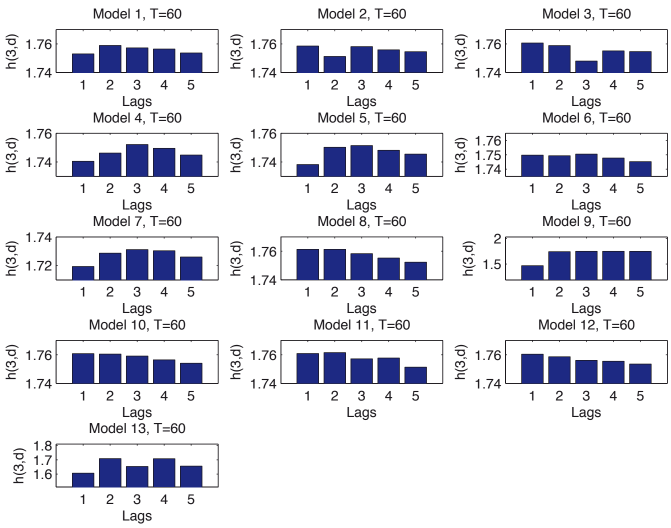


Figure 15: Mean value of $\hat{h}(m,d)$ for $m = 3$ and sample size T=60. 1,000 simulations.

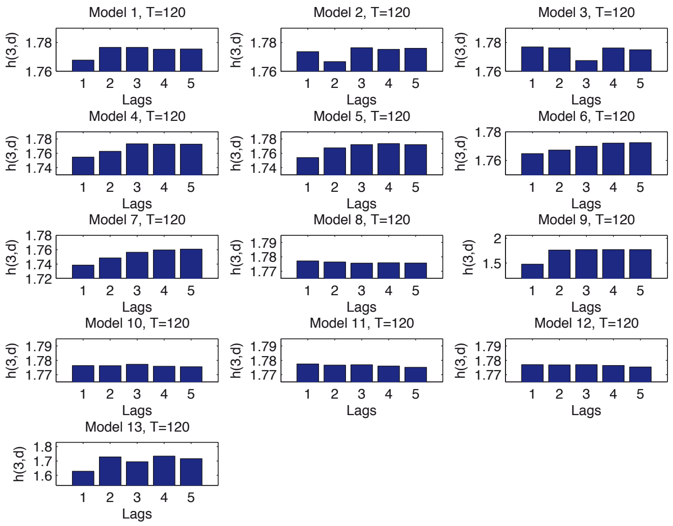


Figure 16: Mean value of $\hat{h}(m, d)$ for $m = 3$ and sample size T=120. $1,000$ simulations.

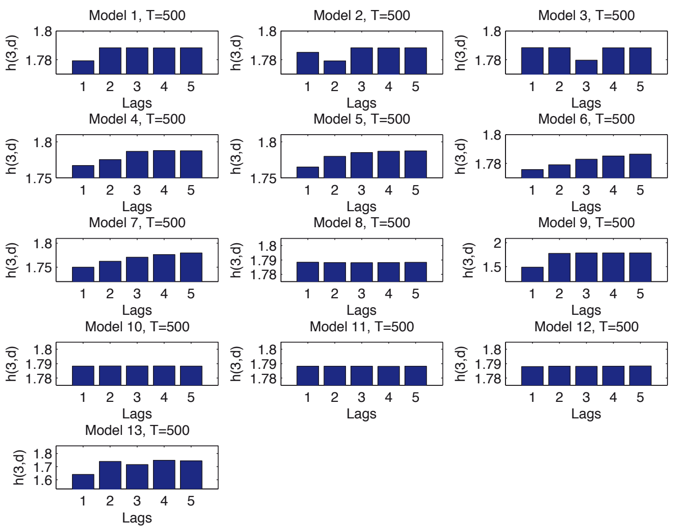


Figure 17: Mean value of $\hat{h}(m, d)$ for $m = 3$ and sample size T=500. $1,000$ simulations.

	1	**2**	**3**	**4**	**5**
Model 1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	0.0321	0.0248	0.0267	0.0266	0.0286
Model 2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	0.0278	0.0335	0.0271	0.0282	0.0300
Model 3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	0.0255	0.0273	0.0351	0.0298	0.0287
Model 4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	0.0372	0.0327	0.0309	0.0333	0.0347
Model 5: $X_t = \|X_{t-1}\|^{0.8} + \epsilon_t$	0.0378	0.0302	0.0301	0.0332	0.0339
Model 6: $X_t = sign(X_{t-1}) + \epsilon_t$	0.0299	0.0308	0.0309	0.0315	0.0330
Model 7: $X_t = 0.8X_{t-1} + \epsilon_t$	0.0464	0.0411	0.0415	0.0445	0.0459
Model 8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	0.0240	0.0255	0.0254	0.0291	0.0324
Model 9: $X_t = 4X_{t-1}(1 - X_{t-1})$	0.0718	0.0470	0.0376	0.0357	0.0349
Model 10: $X_t = \epsilon_t$	0.0245	0.0245	0.0272	0.0261	0.0285
Model 11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	0.0256	0.0247	0.0278	0.0288	0.0318
Model 12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	0.0244	0.0267	0.0283	0.0291	0.0302
Model 13: $X_t = 0.9X_{t-2} + \epsilon_t$	0.0930	0.0472	0.1024	0.0535	0.1043

(a) T=60

	1	**2**	**3**	**4**	**5**
Model 1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	0.0198	0.0122	0.0124	0.0128	0.0125
Model 2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	0.0137	0.0208	0.0121	0.0133	0.0126
Model 3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	0.0122	0.0123	0.0191	0.0127	0.0129
Model 4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	0.0235	0.0197	0.0147	0.0150	0.0146
Model 5: $X_t = \|X_{t-1}\|^{0.8} + \epsilon_t$	0.0227	0.0184	0.0149	0.0147	0.0146
Model 6: $X_t = sign(X_{t-1}) + \epsilon_t$	0.0183	0.0167	0.0156	0.0145	0.0148
Model 7: $X_t = 0.8X_{t-1} + \epsilon_t$	0.0265	0.0265	0.0238	0.0232	0.0230
Model 8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	0.0118	0.0118	0.0131	0.0134	0.0125
Model 9: $X_t = 4X_{t-1}(1 - X_{t-1})$	0.0502	0.0262	0.0168	0.0158	0.0171
Model 10: $X_t = \epsilon_t$	0.0129	0.0120	0.0119	0.0129	0.0138
Model 11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	0.0120	0.0122	0.0115	0.0131	0.0132
Model 12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	0.0123	0.0117	0.0130	0.0125	0.0135
Model 13: $X_t = 0.9X_{t-2} + \epsilon_t$	0.0715	0.0318	0.0675	0.0320	0.0623

(b) T=120

	1	**2**	**3**	**4**	**5**
Model 1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	0.0078	0.0028	0.0030	0.0029	0.0029
Model 2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	0.0044	0.0081	0.0031	0.0029	0.0029
Model 3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	0.0031	0.0027	0.0081	0.0029	0.0030
Model 4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	0.0101	0.0077	0.0046	0.0031	0.0033
Model 5: $X_t = \|X_{t-1}\|^{0.8} + \epsilon_t$	0.0096	0.0066	0.0046	0.0039	0.0036
Model 6: $X_t = sign(X_{t-1}) + \epsilon_t$	0.0072	0.0064	0.0054	0.0045	0.0040
Model 7: $X_t = 0.8X_{t-1} + \epsilon_t$	0.0125	0.0108	0.0096	0.0085	0.0076
Model 8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	0.0026	0.0027	0.0031	0.0030	0.0030
Model 9: $X_t = 4X_{t-1}(1 - X_{t-1})$	0.0238	0.0098	0.0041	0.0035	0.0035
Model 10: $X_t = \epsilon_t$	0.0029	0.0028	0.0029	0.0028	0.0028
Model 11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	0.0029	0.0028	0.0031	0.0031	0.0029
Model 12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	0.0031	0.0029	0.0029	0.0030	0.0028
Model 13: $X_t = 0.9X_{t-2} + \epsilon_t$	0.0349	0.0141	0.0311	0.0139	0.0275

(c) T=500

Table 19: Standard deviation of $\hat{h}(m, d)$ based on 1,000 simulations with $m = 3$. T is the number of observations.

	1	2	3	4	5
Model 1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	22.30 %	16.20 %	19.20 %	19.20 %	23.10 %
Model 2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	14.00 %	23.90 %	20.10 %	20.30 %	21.70 %
Model 3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	13.90 %	13.80 %	31.70 %	19.60 %	21.00 %
Model 4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	27.00 %	21.20 %	13.80 %	15.90 %	22.10 %
Model 5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	30.20 %	15.70 %	14.00 %	18.10 %	22.00 %
Model 6: $X_t = sign(X_{t-1}) + \epsilon_t$	18.60 %	18.90 %	17.90 %	21.00 %	23.60 %
Model 7: $X_t = 0.8X_{t-1} + \epsilon_t$	29.80 %	17.60 %	15.00 %	16.10 %	21.50 %
Model 8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	15.70 %	13.90 %	20.70 %	22.30 %	27.40 %
Model 9: $X_t = 4X_{t-1}(1 - X_{t-1})$	99.80 %	0.10 %	0.00 %	0.10 %	0.00 %
Model 10: $X_t = \epsilon_t$	16.20 %	14.50 %	19.80 %	24.10 %	25.40 %
Model 11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	15.00 %	14.00 %	19.90 %	22.10 %	29.00 %
Model 12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	17.40 %	16.00 %	20.90 %	20.40 %	25.30 %
Model 13: $X_t = 0.9X_{t-2} + \epsilon_t$	52.40 %	8.80 %	11.90 %	9.30 %	17.60 %

(a) T=60

	1	2	3	4	5
Model 1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	36.10 %	11.60 %	17.10 %	17.50 %	17.70 %
Model 2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	16.10 %	36.40 %	15.70 %	15.50 %	16.30 %
Model 3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	14.50 %	13.70 %	37.20 %	14.60 %	20.00 %
Model 4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	43.50 %	24.70 %	10.10 %	11.10 %	10.60 %
Model 5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	50.10 %	16.90 %	11.10 %	10.00 %	11.90 %
Model 6: $X_t = sign(X_{t-1}) + \epsilon_t$	29.90 %	23.60 %	18.20 %	14.10 %	14.20 %
Model 7: $X_t = 0.8X_{t-1} + \epsilon_t$	44.00 %	23.30 %	11.80 %	11.40 %	9.50 %
Model 8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	16.90 %	18.20 %	23.60 %	21.00 %	20.30 %
Model 9: $X_t = 4X_{t-1}(1 - X_{t-1})$	100.00 %	0.00 %	0.00 %	0.00 %	0.00 %
Model 10: $X_t = \epsilon_t$	18.50 %	18.50 %	19.20 %	19.50 %	24.30 %
Model 11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	16.40 %	17.10 %	19.30 %	21.90 %	25.30 %
Model 12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	19.70 %	15.90 %	19.20 %	21.00 %	24.20 %
Model 13: $X_t = 0.9X_{t-2} + \epsilon_t$	82.90 %	6.10 %	4.80 %	3.90 %	2.30 %

(b) T=120

	1	2	3	4	5
Model 1: $X_t = \epsilon_t + 0.8\epsilon_{t-1}^2$	75.60 %	4.90 %	6.20 %	7.30 %	6.00 %
Model 2: $X_t = \epsilon_t + 0.8\epsilon_{t-2}^2$	13.20 %	69.90 %	6.20 %	5.50 %	5.20 %
Model 3: $X_t = \epsilon_t + 0.8\epsilon_{t-3}^2$	7.60 %	5.50 %	72.40 %	6.90 %	7.60 %
Model 4: $X_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$	73.90 %	24.60 %	1.00 %	0.30 %	0.20 %
Model 5: $X_t = \lvert X_{t-1}\rvert^{0.8} + \epsilon_t$	89.70 %	8.10 %	1.40 %	0.30 %	0.50 %
Model 6: $X_t = sign(X_{t-1}) + \epsilon_t$	54.20 %	28.30 %	8.30 %	5.80 %	3.40 %
Model 7: $X_t = 0.8X_{t-1} + \epsilon_t$	74.60 %	19.20 %	3.80 %	1.80 %	0.60 %
Model 8: $X_t = 0.6X_{t-2}\epsilon_{t-1} + \epsilon_t$	18.80 %	19.70 %	22.20 %	19.90 %	19.40 %
Model 9: $X_t = 4X_{t-1}(1 - X_{t-1})$	100.00 %	0.00 %	0.00 %	0.00 %	0.00 %
Model 10: $X_t = \epsilon_t$	18.60 %	18.20 %	19.80 %	22.10 %	21.30 %
Model 11: $X_t = \epsilon_t + 0.8\epsilon_{t-1}\epsilon_{t-2}$	19.50 %	18.10 %	21.70 %	20.40 %	20.30 %
Model 12: $X_t = (1 + 0.5X_{t-1}^2)^{0.5}\epsilon_t$	21.60 %	18.60 %	20.90 %	19.30 %	19.60 %
Model 13: $X_t = 0.9X_{t-2} + \epsilon_t$	99.90 %	0.10 %	0.00 %	0.00 %	0.00 %

(c) T=500

Table 20: Percentage of selected lag as minimizer of $\hat{h}(m, d)$. 1,000 simulations, T observations

C.3. A Rather Special Time Series

Consider the following time series $\{X_t\}_{t \in \mathcal{I}}$:

$$X_t \sim \begin{cases} U[-3,-1] & \text{for} \quad mod(t,3) = 1 \\ U[-1,1] & \text{for} \quad mod(t,3) = 2 \\ U[1,3] & \text{for} \quad mod(t,3) = 0 \end{cases}$$

Let $\pi_1 = (0,1,2)$, $\pi_2 = (0,2,1)$, $\pi_3 = (1,0,2)$, $\pi_4 = (1,2,0)$, $\pi_5 = (2,0,1)$, and $\pi_6 = (2,1,0)$. It is easy to verify that for $d = 1$ and $d = 4$ the sequence of symbols starting at $t = 0$ is given by: $\underbrace{\pi_1, \pi_5, \pi_4}, \underbrace{\pi_1, \pi_5, \pi_4}$, and so on. For $d = 2$ and $d = 5$ the sequence of symbols is $\underbrace{\pi_2, \pi_3, \pi_6}, \underbrace{\pi_2, \pi_3, \pi_6}$, and so on. For $d = 3$ all 6 symbols are possible with equal probability of 1/6.

Let $d_0 = 1$ and $\tilde{\mathcal{K}} = \{\pi_1, \pi_4, \pi_5\}$. Then $\tilde{\mathcal{K}}$ has the property that $p(\tilde{\mathcal{K}}_{d_0}) = p(\tilde{\mathcal{K}}_{d_0} \cup \{(\pi, d_0)\})$ for every $\pi \in S_m \setminus \tilde{\mathcal{K}}$. Moreover, $p(\tilde{\mathcal{K}}_d) \leq p(\tilde{\mathcal{K}}_{d_0})$ for all d and $p(\tilde{\mathcal{K}}_{d_0}) = 1$. We get $h(3,1) = log(3)$, $h(3,2) = log(3)$, $h(3,3) = log(6)$, $h(3,4) = log(3)$, and $h(3,5) = log(3)$. Yet,

$$-\sum_{(\pi,d_0)\in\tilde{\mathcal{K}}_{d_0}} p(\pi, d_0) log(p(\pi, d_0)) \leq -\sum_{(\pi,d)\in\tilde{\mathcal{K}}_d} p(\pi, d) log(p(\pi, d))$$

does not hold. The right hand side of the inequality equals $log(3)$ for $d = 1$ and $d = 4$. It equals 0 for $d = 2$ and $d = 5$. For $d = 3$ the right hand side is equal to $0.5log(6)$.

I am aware of the fact that $\{X_t\}_{t\in\mathcal{I}}$ is not stationary. The point is that the reasoning in the proof is not correct unless stationarity implies that

$$-\sum_{(\pi,d_0)\in\tilde{\mathcal{K}}_{d_0}} p(\pi, d_0) log(p(\pi, d_0)) \leq -\sum_{(\pi,d)\in\tilde{\mathcal{K}}_d} p(\pi, d) log(p(\pi, d))$$

for all d. Yet, this remains to be shown.

C.4. Minor Remarks

In Table 2 in Matilla-García and Marín [2009] critical values at the 5% level for $\hat{h}(m,d)$ are given for $T = 100, 300$ under the assumption that the original series is i.i.d. Given that the number of symbols decreases with d for fixed n and therefore the variance of $\hat{h}(m,d)$ increases I would expect that the critical values decrease monotonically with increasing d. The critical values I get for $T = 60, 100, 120, 300, 500$ using $100,000$ simulations[20] are displayed in Table 21.

[20] Matilla-García and Marín [2009] are silent about the number of simulations they used to determine the critical values.

d	60	100	120	300	500
1	1.7128	1.7457	1.7534	1.7768	1.7828
2	1.7100	1.7441	1.7529	1.7767	1.7827
3	1.7058	1.7433	1.7520	1.7766	1.7827
4	1.7028	1.7418	1.7515	1.7764	1.7826
5	1.6987	1.7409	1.7504	1.7764	1.7826

Table 21: Critical values of $\hat{h}(m, d)$ under the assumption that $\{X_t\}_{t \in \mathcal{I}}$ is an i.i.d. series based on $100,000$ simulations.

Index of Working Papers:

August 31, 2006	John Williamson (comment by Marc Flandreau)	130	A Worldwide System of Reference Rates
September 15, 2006	Sylvia Kaufmann, Peter Kugler	131	Expected Money Growth, Markov Trends and the Instability of Money Demand in the Euro Area
September 18, 2006	Martin Schneider, Markus Leibrecht	132	AQM-06: The Macroeconomic Model of the OeNB
November 6, 2006	Erwin Jericha and Martin Schürz	133	A Deliberative Independent Central Bank
December 22, 2006	Balázs Égert	134	Central Bank Interventions, Communication and Interest Rate Policy in Emerging European Economies
May 8, 2007	Harald Badinger	135	Has the EU's Single Market Programme fostered competition? Testing for a decrease in markup ratios in EU industries
May 10, 2007	Gert Peersman	136	The Relative Importance of Symmetric and Asymmetric Shocks: the Case of United Kingdom and Euro Area
May 14, 2007	Gerhard Fenz and Martin Schneider	137	Transmission of business cycle shocks between unequal neighbours: Germany and Austria
July 5, 2007	Balázs Égert	138	Real Convergence, Price Level Convergence and Inflation Differentials in Europe
January 29, 2008	Michał Brzoza-Brzezina, Jesus Crespo Cuaresma	139	Mr. Wicksell and the global economy: What drives real interest rates?
March 6, 2008	Helmut Stix	140	Euroization: What Factors drive its Persistence? Household Data Evidence for Croatia, Slovenia and Slovakia

April 28, 2008	Kerstin Gerling	141	The Real Consequences of Financial Market Integration when Countries Are Heterogeneous
April 29, 2008	Aleksandra Riedl and Silvia Rocha-Akis	142	Testing the tax competition theory: How elastic are national tax bases in Western Europe?
May 15, 2008	Christian Wagner	143	Risk-Premia, Carry-Trade Dynamics, and Speculative Efficiency of Currency Markets
June 19, 2008	Sylvia Kaufmann	144	Dating and forecasting turning points by Bayesian clustering with dynamic structure: A suggestion with an application to Austrian data.
July 21, 2008	Martin Schneider and Gerhard Fenz	145	Transmission of business cycle shocks between the US and the euro area
September 1, 2008	Markus Knell	146	The Optimal Mix Between Funded and Unfunded Pensions Systems When People Care About Relative Consumption
September 8, 2008	Cecilia García-Peñalosa	147	Inequality and growth: Goal conflict or necessary prerequisite?
September 30, 2008	Fabio Rumler and Maria Teresa Valderrama	148	Comparing the New Keynesian Phillips Curve with Time Series Models to Forecast Inflation
January 30, 2009	Claudia Kwapil, Johann Scharler	149	Expected Monetary Policy and the Dynamics of Bank Lending Rates
February 5, 2009	Thomas Breuer, Martin Jandačka, Klaus Rheinberger, Martin Summer	150	How to find plausible, severe, and useful stress scenarios
February 11, 2009	Martin Schneider, Christian Ragacs	151	Why did we fail to predict GDP during the last cycle? A breakdown of forecast errors for Austria
February 16, 2009	Burkhard Raunig, Martin Scheicher	152	Are Banks Different? Evidence from the CDS Market
March 11, 2009	Markus Knell, Alfred Stiglbauer	153	The Impact of Reference Norms on Inflation Persistence When Wages are Staggered

May 14, 2009	Tarek A. Hassan	154	Country Size, Currency Unions, and International Asset Returns
May 14, 2009	Anton Korinek	155	Systemic Risk: Amplification Effects, Externalities, and Policy Responses
May 29, 2009	Helmut Elsinger	156	Financial Networks, Cross Holdings, and Limited Liability
July 20, 2009	Simona Delle Chiaie	157	The sensitivity of DSGE models' results to data detrending
November 10, 2009	Markus Knell Helmut Stix	158	Trust in Banks? Evidence from normal times and from times of crises
November 27, 2009	Thomas Scheiber Helmut Stix	159	Euroization in Central, Eastern and South-eastern Europe – New Evidence On Its Extent and Some Evidence On Its Causes
January 11, 2010	Jesús Crespo Cuaresma Martin Feldircher	160	Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe
March 29, 2010	Markus Knell	161	Nominal and Real Wage Rigidities. In Theory and in Europe
May 31, 2010	Zeno Enders Philip Jung Gernot J. Müller	162	Has the Euro changed the Business Cycle?
August 25, 2010	Marianna Červená Martin Schneider	163	Short-term forecasting GDP with a DSGE model augmented by monthly indicators
September 8, 2010	Sylvia Kaufmann Johann Scharler	164	Bank-Lending Standards, the Cost Channel and Inflation Dynamics
September 15, 2010	Helmut Elsinger	165	Independence Tests based on Symbolic Dynamics