OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

WORKING PAPER 228

Serial Correlation in Contingency Tables

Helmut Elsinger

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.

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

Publisher and editor *Oesterreichische Nationalbank*
Otto-Wagner-Platz 3, 1090 Vienna, Austria
PO Box 61, 1011 Vienna, Austria
www.oenb.at
oenb.info@oenb.at
Phone (+43-1) 40420-6666
Fax (+43-1) 40420-046698

Coordinating editor Martin Summer

Design *Communications and Publications Division*

DVR 0031577

ISSN 2310-5321 (Print)
ISSN 2310-533X (Online)

Serial Correlation in Contingency Tables

Helmut Elsinger*

Abstract

Pearson's chi-squared test for independence in two-way contingency tables is developed under the assumption of multinomial sampling. In this paper I consider the case where draws are not independent but exhibit serial dependence. I derive the asymptotic distribution and show that adjusting Pearson's statistic is simple and works reasonably well irrespective whether the processes are Markov chains or m-dependent. Moreover, I propose a test for independence that has a simple limiting distribution if at least one of the two processes is a Markov chain. For three-way tables I investigate the Cochrane-Mantel-Haenszel (CMH) statistic and show that there exists a closely related procedure that has power against a larger class of alternatives. This new statistic might be used to test whether a Markov chain is simple against the alternative of being a Markov chain of higher order. Monte Carlo experiments are used to illustrate the small sample properties.

Keywords: Goodness of Fit, Independence Tests, Cochrane-Mantel-Haenszel Test, Markov chain

JEL-Classification: C12, C14, 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.

Non-Technical Summary

Contingency tables display the relative or absolute frequencies of two categorical variables. A classical application in economics would be an early warning indicator. Two possible values of the indicator (warning/no warning) are grouped by the actual values (crisis occurred/no crisis occurred). After observing both variables for $T = 100$ periods the table could look like this:

observed	crisis	no crisis	sum
warning	7	13	20
no warning	13	67	80
sum	20	80	100

To assess the quality of the indicator we compare the performance to a pure random guess. In this particular example we would expect that the table for 100 draws should look like this

expected	crisis	no crisis	sum
warning	4	16	20
no warning	16	64	80
sum	20	80	100

The null hypothesis that the indicator and the actual value are independent of each other is commonly tested using Pearson's chi-squared statistic by comparing the observed values to the expected ones. This test is asymptotically chi-squared distributed under the assumption of multinomial sampling. In the example the Pearson statistic equals 3.52. The p-value is 6.1% and hence we would reject at a level of 10%.

In the above example we reckon that both the indicator and the actual event at a particular point in time might depend on the outcomes in preceding periods. Even if both series are independent from each other, the serial correlation of the individual series would induce a spurious relationship. The assumption of multinomial sampling is not justified in such a situation. An autocorrelation of 0.5 would change the p-value of the test statistic to 14.6% and we would not reject at a level of 10%.

It is a well known but persistently ignored fact that Pearson's test is not chi-squared distributed if the variables exhibit serial correlation. In this paper I show that it is quite simple to adjust the test statistic in a way such that serial or spatial correlation is accounted for. The proposed method works for a fairly large class of Markov chains and m-dependent processes. I extend the approach to higher dimensional tables. Finally, I discuss a procedure to test for conditional independence. Simulations illustrate that all these methods work quite well.

1 Introduction

Serial correlation in contingency table data is a widespread phenomenon (for examples and references see Pesaran and Timmermann [2009]). The dependencies between observations make the assumption of a multinomial sampling scheme invalid. It follows that the Pearson chi-squared test for independence in two-way contingency tables with serial correlation is asymptotically not χ^2 distributed.

The larger part of the literature on serial correlation in contingency tables dates back to the 1980s. In a series of papers the effect of stratification and clustering on the asymptotic distribution of Pearson statistics is investigated (Holt et al. [1980], Rao and Scott [1979], Rao and Scott [1981], Rao and Scott [1984], Rao and Scott [1987]). It is shown that the Pearson statistic is asymptotically distributed as a weighted sum of independent χ_1^2 variables with weights related to design effects. The derived asymptotic distribution differs considerably from the null distribution under serial independence.

A second strand of literature deals with the case of Markov chains. Tavaré [1983] and Tavaré and Altham [1983] show that for reversible Markov chains the weights in the asymptotic distribution are simple functions of the eigenvalues of the transition matrices. Porteous [1987] extends the results to multi-way tables. The most recent contribution is Pesaran and Timmermann [2009]. They choose a novel route and propose new tests based on canonical correlations. They illustrate via Monte Carlo simulations that these tests perform very well under serial correlation.

In this paper I assume that a central limit theorem holds for the observed frequencies. Stationary and ergodic Markov chains and m-dependent processes satisfy this assumption. The asymptotic distribution of the standard Pearson chi-squared statistics for goodness of fit and independence in two-way tables is given by a weighted sum of independent χ_1^2 variables. Simple examples are used to illustrate the difference between the asymptotic distribution with serial correlation and that without. For Markov chains I generalize the result of Tavaré [1983]. The sum of the weights in the asymptotic distribution is a simple function of the eigenvalues of the transition matrices. This result is useful insofar as the exact critical values of weighted sums of independent χ_1^2 variables are difficult to obtain but can be approximated quite well by matching the first moment which corresponds to the sum of the weights. I propose a test for independence based on appropriately filtered processes which works if at least one of the two processes is a Markov chain.

For conditional independence in multi-way tables I demonstrate that the classical Cochrane-Mantel-Haenszel statistic (CMH) tests for a rather weak implication of conditional independence and hence has no power against a class of relevant alternatives. I show that a simple adjustment solves the problem and illustrate via Monte Carlo simulations that this test is as powerful as CMH against standard alternatives. Using an argument from Kullback et al. [1962] I show that this adjusted version of CMH works well when we test whether a process is a simple Markov chain against the alternative of being a Markov chain of higher order.

The paper is structured as follows. In the next section I introduce the relevant notation and discuss the consequences of serial correlation on the asymptotic distribution of standard Pearson chi-squared statistic for goodness of fit. Section 3 continues with independence tests for two-way tables under serial correlation. Section 4 deals with three-way tables. Tests for mutual independence, joint independence of two series from the third, and conditional independence are investigated. In section 5 I present the results of Monte Carlo simulations. Section 6 concludes.

2 Goodness of Fit

Let $\{U_t\}_{t\in\mathbb{Z}}$ be a stationary series of categorical variables taking values in some finite set of states m_u. The variable is represented by the column vector $U_t = (U_{1,t}, \ldots, U_{m_u,t})'$. $U_{i,t} = 1$ if at time t the ith category occurs and 0 otherwise. The unconditional probability of $U_{i,t} = 1$ is denoted by p_i. The $m_u \times 1$ vector of unconditional state probabilities is given by $\mathbf{p} = (p_1, \cdots, p_{m_u})'$. The l-step transition probability $\mathbb{P}(U_{j,t+l} = 1|U_{i,t} = 1)$ is denoted by $q_{i,j}^{(l)}$. These probabilities are summarized in the matrix $\mathbf{Q}^{(l)}$. The joint probabilities, $p_{i,j}^{(l)} = \mathbb{P}(U_{i,t} = 1, U_{j,t+l} = 1)$, are summarized in the matrix

$\mathbf{P}^{(l)}$ concordantly to $\mathbf{Q}^{(l)}$. $\mathbf{P}^{(0)}$ is a diagonal matrix with $\mathbf{p}$ on its diagonal, denoted by $\mathbf{D}_p$. $\mathbf{Q}^{(0)}$ is the identity matrix. Observe that $\mathbf{P}^{(-l)} = \mathbf{P}^{(l)\prime}$, $\mathbf{D}_p\mathbf{Q}^{(l)} = \mathbf{P}^{(l)}, \mathbf{D}_p\mathbf{Q}^{(-l)} = \mathbf{Q}^{(l)\prime}\mathbf{D}_p, \mathbf{p}'\mathbf{Q}^{(l)} = \mathbf{p}'$, and $\mathbf{Q}^{(l)}\mathbf{1}_{m_u} = \mathbf{1}_{m_u}$ for all $l \in \mathbb{Z}$. $\mathbf{1}_{m_u}$ denotes the $m_u \times 1$ vector of ones.[1] A process is reversible if $\mathbf{P}^{(l)} = \mathbf{P}^{(-l)} = \mathbf{P}^{(l)\prime}$ for all $l \in \mathbb{Z}$. In this case $\mathbf{Q}^{(l)} = \mathbf{Q}^{(-l)}$ and $\mathbf{Q}^{(l)} = \mathbf{D}_p^{-1}\mathbf{Q}^{(l)\prime}\mathbf{D}_p$. Processes with two states are always reversible.

The vector $\mathbf{n}(T) = \sum_{t=1}^{T} U_t$ counts the number of occurrences of the different categories between $t = 1$ and T. The relative frequencies $(1/T)\mathbf{n}(T)$ are denoted by $\hat{\mathbf{p}}$. All results in this paper are based on the following assumption which is satisfied if $\{U_t\}$ is a stationary and ergodic Markov chain or an m-dependent process.

Assumption 1. *$\{U_t\}_{t\in\mathbb{Z}}$ with $\mathbf{p} > \mathbf{0}$ is such that a central limit theorem holds, i.e. $\sqrt{T}\,(\hat{\mathbf{p}} - \mathbf{p}) \sim^a \mathcal{N}\,(\mathbf{0}, \mathbf{\Sigma})$ with*

$$\mathbf{\Sigma} = \lim_{T\to\infty} \mathbf{D}_p \sum_{l=-T+1}^{T-1} \frac{T-|l|}{T} \left(\mathbf{Q}^{(l)} - \mathbf{1}\mathbf{p}'\right) > \mathbf{0}.$$

We assume that the above series is absolutely convergent. The rank of $\mathbf{\Sigma}$ is denoted by $d \leq m_u - 1$.

Consider testing the hypothesis H_0: $\mathbf{p}$ specified. The Pearson χ^2 statistic is given by

$$P(T) = \sum_{i=1}^{m_u} \frac{(n_i(T) - Tp_i)^2}{Tp_i} = T\,(\hat{\mathbf{p}} - \mathbf{p})'\,\mathbf{D}_p^{-1}\,(\hat{\mathbf{p}} - \mathbf{p})\,. \qquad (1)$$

The asymptotic distribution of $P(T)$ under H_0 can be represented as a weighted sum of independent χ^2_1 variables, i.e. $\sum_{i=1}^{d} \rho_i Z_i^2$. $\{\rho_i\}$ are the nonzero eigenvalues of $\mathbf{D}_p^{-1}\mathbf{\Sigma}$.

[1] In the sequel I will omit subscripts whenever the dimension is clear from the context.

Example 1. *Let* $\{U_t\}$ *be a simple two-state Markov chain with transition matrix*

$$\mathbf{Q} = \begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix} \text{ for } a \in (0,1).$$

$\mathbf{p} = (1/2, 1/2)'$ *and*

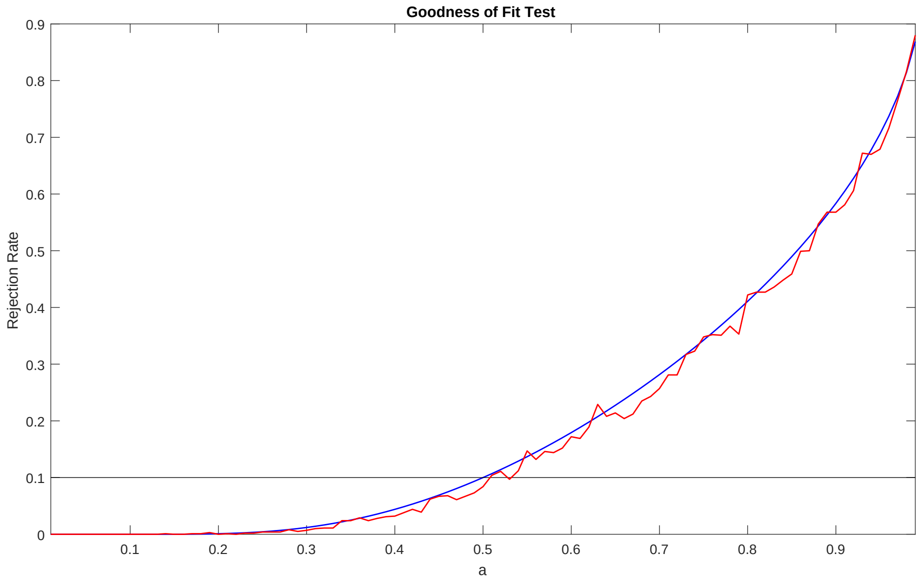
$$\boldsymbol{\Sigma} = \frac{a}{4(1-a)} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The eigenvalues of $\mathbf{D}_p^{-1}\boldsymbol{\Sigma}$ *are given by* $\{0, a/(1-a)\}$*. Therefore,* $P(T) \sim^a \frac{a}{1-a}\chi_1^2$*. Figure 1 illustrates the rejection rates if we use the incorrect* χ_1^2 *distribution. The level differs considerably from the nominal level of* 10%*, i.e. a critical value of* 2.71*, unless* $a = 0.5$*. Let* $U_t^* = \sum_{i=1}^{m_u} iU_{i,t}$*. The correlation of* U_{t-1}^* *and* U_t^* *equals* $2a-1$*. Even moderate levels of autocorrelation lead to sizable deviations from the* 10% *level. If* $a \geq 1/2$ *then* $\{U_t\}$ *is positively dependent in the sense of Gleser and Moore [1985]. A similar example for an m-dependent process can be found in Appendix A.*

The critical values of a linear combination of independent χ_1^2 variables have to be determined via simulations. The literature (e.g. Holt et al. [1980] and Yuan and Bentler [2010]) shows that two simple approximations work reasonably well. The first approximation is by matching the mean. Let $\bar{\rho}$ equal the average nonzero eigenvalue of $\mathbf{D}_p^{-1}\boldsymbol{\Sigma}$, i.e. $\bar{\rho} = \sum_{i=1}^{d} \rho_i/d$. The distribution of $P(T)$ is approximated by $\bar{\rho}\chi_d^2$. Alternatively, we may match the first two moments via $a\chi_b^2$. This implies

$$a = \frac{\sum_{i=1}^{d} \rho_i^2}{\sum_{i=1}^{d} \rho_i} \text{ and } b = \frac{\left(\sum_{i=1}^{d} \rho_i\right)^2}{\sum_{i=1}^{d} \rho_i^2}.$$

Figure 1: The rejection rate at a nominal level of 10% for the Markov chain given in Example 1. The blue line is the theoretical rejection rate. The red line is based on 1000 simulations for $a = 0.01$ to $a = 0.99$ in steps of 0.01. The number of observations in each simulation equals 200.



Remark 1. *Using that* $\mathbf{Q}^{(-l)} = \mathbf{D}_p^{-1}\mathbf{Q}^{(l)\prime}\mathbf{D}_p$ *and that the trace is invariant under cyclic permutations we get*

$$\sum_{i=1}^{d} \rho_i = tr(\mathbf{D}_p^{-1}\boldsymbol{\Sigma}) = 2\sum_{l=0}^{\infty}(tr(\mathbf{Q}^{(l)}) - 1) - (m_u - 1). \tag{2}$$

This is true irrespective whether the process is reversible or not. Observe that only the diagonal elements of $\mathbf{Q}^{(l)}$ *are needed, i.e. the probabilities that* $U_{i,t+l} = 1$ *conditional on* $U_{i,t} = 1$ *for all* $l > 0$ *and* i.

If $\{\mathbf{U}_t\}$ is a simple Markov chain the variance of the asymptotic distribution may be rewritten as

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathbf{D}_p \sum_{l=1}^{\infty} (\mathbf{Q} - \mathbf{1p}')^l + \sum_{l=1}^{\infty} (\mathbf{Q}' - \mathbf{p1}')^l \mathbf{D}_p + (\mathbf{D}_p - \mathbf{pp}') \\ &= \mathbf{D}_p\mathbf{Z} + \mathbf{Z}'\mathbf{D}_p - \mathbf{D}_p - \mathbf{pp}' \end{aligned}$$

with $\mathbf{Z} = (\mathbf{I} - \mathbf{Q} + \mathbf{1p}')^{-1}$. Tavaré and Altham [1983] show that for reversible Markov chains the eigenvalues of $\mathbf{D}_p^{-1}\boldsymbol{\Sigma}$ are simple functions of the nonunit eigenvalues of the transition matrix $\mathbf{Q}$. If we drop the reversibility assumption, we get that the sum of the eigenvalues of $\mathbf{D}_p^{-1}\boldsymbol{\Sigma}$ is a simple function of the eigenvalues of the transition matrix.

Theorem 1. *If* $\{\mathbf{U}_t\}$ *is a simple Markov chain and fulfills Assumption 1, the asymptotic distribution of* $P(T)$ *under* H_0 *equals* $\sum_{i=1}^{d} \rho_i Z_i^2$ *where* $\{\rho_i\}$ *are the eigenvalues of* $\mathbf{D}_p^{-1}\boldsymbol{\Sigma}$ *and* $\{Z_i\}$ *are independent* $\mathcal{N}(0,1)$. *It holds that*

$$\sum_{i=1}^{d} \rho_i = \sum_{i=1}^{m_u - 1} \frac{1 + \lambda_i}{1 - \lambda_i} \tag{3}$$

with $\{\lambda_i\}$ *being the nonunit eigenvalues of* $\mathbf{Q}$. *If* $\{\mathbf{U}_t\}$ *is additionally reversible,* $\{\frac{1+\lambda_i}{1-\lambda_i}\}$ *are the eigenvalues of* $\mathbf{D}_p^{-1}\boldsymbol{\Sigma}$.

Proof. The first part of the theorem is well established. It remains to be shown that

$$\sum_{i=1}^{d} \rho_i = \sum_{i=1}^{m_u-1} \frac{1+\lambda_i}{1-\lambda_i}.$$

For simple Markov chains $\mathbf{Q}^{(l)} = \mathbf{Q}^l$ for $l \geq 0$. From Equation (2) we know that

$$\sum_{i=1}^{d} \rho_i = 2\sum_{l=0}^{\infty}(tr(\mathbf{Q}^l) - 1) - (m_u - 1).$$

Now, $tr(\mathbf{Q}^l) = \sum_{i=1}^{m_u} \lambda_i^l$. $\mathbf{1}$ is a right eigenvector of $\mathbf{Q}$ with an eigenvalue of 1. Assumption 1 implies that the infinite sum on the right hand side converges. Hence, 1 is the unique largest eigenvalue in absolute value. W.l.o.g. we may order the eigenvalues such that $\lambda_{m_u} = 1$. This implies

$$\begin{aligned}\textstyle\sum_{i=1}^{d} \rho_i &= 2\textstyle\sum_{l=0}^{\infty}(\sum_{i=1}^{m_u} \lambda_i^l - 1) - (m_u - 1)\\ &= 2\textstyle\sum_{l=0}^{\infty} \sum_{i=1}^{m_u-1} \lambda_i^l - (m_u - 1) = \sum_{i=1}^{m_u-1}(2\sum_{l=0}^{\infty} \lambda_i^l - 1)\\ &= \textstyle\sum_{i=1}^{m_u-1}\left(\frac{2}{1-\lambda_i} - 1\right) = \sum_{i=1}^{m_u-1} \frac{1+\lambda_i}{1-\lambda_i}.\end{aligned}$$

The claim for reversible Markov chains is proved in Tavaré and Altham [1983]. □

If we approximate the distribution of $P(T)$ by matching the first moment only, it suffices to determine the eigenvalues of $\mathbf{Q}$ and the rank of Σ irrespective whether the process is reversible or not.

For Markov chains there is yet another alternative by adjusting the Pearson test. If we left multiply $\sqrt{T}(\hat{\mathbf{p}} - \mathbf{p})$ by $(\mathbf{I} - \mathbf{Q}')$, the covariance matrix of the asymptotic distribution equals $\mathbf{D}_p - \mathbf{Q}'\mathbf{D}_p\mathbf{Q}$ which is singular. To define a quadratic form, we need the following definition.

Definition 1. *An $n \times m$ matrix $\mathbf{G}$ is said to be a generalized inverse (g inverse) of an $m \times n$ matrix $\mathbf{A}$ if $\mathbf{AGA} = \mathbf{A}$. We denote $\mathbf{G}$ by $\mathbf{A}^-$. A g inverse exists for any matrix. If additionally $\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-$ then we call $\mathbf{A}^-$ a reflexive g inverse and denote it by $\mathbf{A}^g$. If both $\mathbf{A}\mathbf{A}^g$ and $\mathbf{A}^g\mathbf{A}$ are*

Hermitian, we call $\mathbf{A}^g$ a Moore-Penrose inverse and denote it by $\mathbf{A}^+$. The Moore-Penrose inverse exists and is unique.

The quadratic form

$$P_a(T) = T\,(\hat{\mathbf{p}} - \mathbf{p})'\,(\mathbf{I} - \mathbf{Q})(\mathbf{D}_p - \mathbf{Q}'\mathbf{D}_p\mathbf{Q})^{-}(\mathbf{I} - \mathbf{Q}')\,(\hat{\mathbf{p}} - \mathbf{p}) \tag{4}$$

is asymptotically χ^2_d distributed for any g inverse of $(\mathbf{D}_p - \mathbf{Q}'\mathbf{D}_p\mathbf{Q})$ by the results of Ogasawara and Takahashi [1951]. Note that in general the quadratic form is not invariant with respect to the chosen g inverse but the asymptotic distribution is the same.

If $\mathbf{Q}$ is not known a priori but can be estimated consistently by $\hat{\mathbf{Q}}$ than

$$\hat{P}_a(T) = T\,(\hat{\mathbf{p}} - \mathbf{p})'\,(\mathbf{I} - \hat{\mathbf{Q}})(\mathbf{D}_p - \hat{\mathbf{Q}}'\mathbf{D}_p\hat{\mathbf{Q}})^{-}(\mathbf{I} - \hat{\mathbf{Q}}')\,(\hat{\mathbf{p}} - \mathbf{p}) \tag{5}$$

is asymptotically equivalent to $P_a(T)$.

3 Two-way tables

Suppose that $U_t = X_t \otimes Y_t$ where $\otimes$ denotes the Kronecker product. Both series of category variables take values in some finite set of states m_x and m_y with $m_x m_y = m_u$. The unconditional probability that $X_{i,t} = 1$ and $Y_{j,t} = 1$ is denoted by $p_{i,j}$. The $m_u \times 1$ vector of unconditional state probabilities is given by $\mathbf{p} = (p_{1,1}, \cdots, p_{1,m_y}, p_{2,1}, \cdots, p_{m_x,m_y})'$. $q^{(l)}_{i,j}$ is the l-step transition probability $\mathbb{P}(X_{u,t+l} = 1 \text{ and } Y_{v,t+l} = 1 | X_{a,t} = 1 \text{ and } Y_{b,t} = 1)$ with $i = (a-1)m_y + b$ and $j = (u-1)m_y + v$.

The marginal probabilities of $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ are given by

$$\mathbf{p}_x = (\sum_{i=1}^{m_y} p_{1,i}, \cdots, \sum_{i=1}^{m_y} p_{m_x,i})' = (\mathbf{I}_{m_x} \otimes \mathbf{1}'_{m_y})\mathbf{p}$$

and $\mathbf{p}_y = (\mathbf{1}'_{m_x} \otimes \mathbf{I}_{m_y})\mathbf{p}$. $\mathbf{n}_x(T)$, $\hat{\mathbf{p}}_x$, $\mathbf{n}_y(T)$, and $\hat{\mathbf{p}}_y$ are defined analogously.

$\mathbf{P}^{(l)}_x = (\mathbf{I}_{m_x} \otimes \mathbf{1}'_{m_y})\mathbf{P}^{(l)}(\mathbf{I}_{m_x} \otimes \mathbf{1}_{m_y})$ and $\mathbf{P}^{(l)}_y = (\mathbf{1}'_{m_x} \otimes \mathbf{I}_{m_y})\mathbf{P}^{(l)}(\mathbf{1}_{m_x} \otimes \mathbf{I}_{m_y})$ are the (marginal) joint probabilities of the pairs (X_{t+l}, X_t) and (Y_{t+l}, Y_t). The

l-step transition probabilities for $\{X_t\}$ and $\{Y_t\}$ are given by $\mathbf{Q}_x^{(l)} = \mathbf{D}_{p_x}^{-1}\mathbf{P}_x^{(l)}$ and $\mathbf{Q}_y^{(l)} = \mathbf{D}_{p_y}^{-1}\mathbf{P}_y^{(l)}$.

Consider the null hypothesis that the series $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ are independent of each other. In this case $\mathbf{P}^{(l)} = \mathbf{P}_x^{(l)} \otimes \mathbf{P}_y^{(l)}$ for all $l \in \mathbb{Z}$ and

$$\sqrt{T}\,(\hat{\mathbf{p}} - \mathbf{p}) = \sqrt{T}\,(\hat{\mathbf{p}} - \mathbf{p}_x \otimes \mathbf{p}_y) \sim^a \mathcal{N}\,(\mathbf{0}, \boldsymbol{\Sigma})$$

with

$$\boldsymbol{\Sigma} = (\mathbf{D}_{p_x} \otimes \mathbf{D}_{p_y}) \sum_{l=-\infty}^{\infty} \left(\mathbf{Q}_x^{(l)} \otimes \mathbf{Q}_y^{(l)} - (\mathbf{1}_{m_x}\mathbf{p}_x') \otimes (\mathbf{1}_{m_y}\mathbf{p}_y')\right).$$

If we condition on the observed marginal proportions $\hat{\mathbf{p}}_x$ and $\hat{\mathbf{p}}_y$, Tavaré [1983] showed that

$$\sqrt{T}\,(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y) = \sqrt{T}\mathbf{B}'\,(\hat{\mathbf{p}} - \mathbf{p}_x \otimes \mathbf{p}_y) + O_p(T^{-1/2})$$

with $\mathbf{B} = (\mathbf{I}_{m_x} - \mathbf{1}_{m_x}\mathbf{p}_x') \otimes (\mathbf{I}_{m_y} - \mathbf{1}_{m_y}\mathbf{p}_y')$.

Remark 2. *To be precise, Tavaré [1983] showed that* $\sqrt{T}\,(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y) = \sqrt{T}(\mathbf{B}' - \mathbf{p}\mathbf{1}')\,(\hat{\mathbf{p}} - \mathbf{p}_x \otimes \mathbf{p}_y) + O_p(T^{-1/2})$. *It is easy to see that* $\mathbf{1}'\,(\hat{\mathbf{p}} - \mathbf{p}) \equiv \mathbf{0}$ *and we may use* $\mathbf{B}$ *as specified. Holt et al. [1980] derive the same result using the delta method. Let* $\mathbf{h}(\hat{\mathbf{p}}) = \hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y$. *The Jacobian matrix of* $\mathbf{h}(\mathbf{p})$ *equals* $(\mathbf{B}' - \mathbf{p}\mathbf{1}')$.

Theorem 2. *Under Assumption 1 and the null hypothesis that* $\{X_t\}_{t\in\mathbb{Z}}$ *and* $\{Y_t\}_{t\in\mathbb{Z}}$ *are independent of each other,* $\sqrt{T}\,(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y)$ *is asymptotically distributed* $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ *with*

$$\boldsymbol{\Omega} = (\mathbf{D}_{p_x} \otimes \mathbf{D}_{p_y}) \sum_{l=-\infty}^{\infty} \left(\mathbf{Q}_x^{(l)} - \mathbf{1}_{m_x}\mathbf{p}_x'\right) \otimes \left(\mathbf{Q}_y^{(l)} - \mathbf{1}_{m_y}\mathbf{p}_y'\right).$$

If at least one of the two series is i.i.d. then $\boldsymbol{\Omega} = (\mathbf{D}_{p_x} - \mathbf{p}_x\mathbf{p}_x') \otimes \left(\mathbf{D}_{p_y} - \mathbf{p}_y\mathbf{p}_y'\right)$ *as in the case of multinomial sampling.*

Proof. Using $\mathbf{B}'\mathbf{D}_p = \mathbf{D}_p\mathbf{B}$ we get

$$\mathbf{B}\left(\mathbf{Q}_x^{(l)} \otimes \mathbf{Q}_y^{(l)} - (\mathbf{1}_{m_x}\mathbf{p}_x') \otimes (\mathbf{1}_{m_y}\mathbf{p}_y')\right)\mathbf{B} = \left(\mathbf{Q}_x^{(l)} - \mathbf{1}_{m_x}\mathbf{p}_x'\right) \otimes \left(\mathbf{Q}_y^{(l)} - \mathbf{1}_{m_y}\mathbf{p}_y'\right)$$

for all $l \in \mathbb{Z}$. To verify the second claim assume w.l.o.g. that $\{X_t\}$ is i.i.d. In this case $\mathbf{Q}_x^{(l)} = \mathbf{1}_{m_x}\mathbf{p}_x' \quad \forall l \neq 0$. Plugging this in yields the desired result. $\square$

The asymptotic distribution of Pearson's chi-squared test

$$\begin{aligned} X^2 &= T\left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y\right)' \left(\mathbf{D}_{\hat{p}_x}^{-1} \otimes \mathbf{D}_{\hat{p}_y}^{-1}\right) \left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y\right) \\ &= T\left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y\right)' \left(\mathbf{D}_{p_x}^{-1} \otimes \mathbf{D}_{p_y}^{-1}\right) \left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y\right) + O_p(T^{-1/2}). \end{aligned}$$

is equal to the distribution of a weighted sum of χ_1^2 variables.

$$X^2 \sim^a \sum_{i=1}^{d} \rho_i Z_i^2$$

with $d = rank(\mathbf{\Omega})$ and $\{\rho_i\}$ being the eigenvalues of

$$\mathbf{\Theta} = \mathbf{D}_p^{-1}\mathbf{\Omega} = \sum_{l=-\infty}^{\infty} \left(\mathbf{Q}_x^{(l)} - \mathbf{1}_{m_x}\mathbf{p}_x'\right) \otimes \left(\mathbf{Q}_y^{(l)} - \mathbf{1}_{m_y}\mathbf{p}_y'\right).$$

Tavaré [1983] proved that for reversible Markov chains the eigenvalues $\{\rho_i\}$ of $\mathbf{\Theta}$ are given by $\{(1 + \lambda_{x,i}\lambda_{y,j})/(1 - \lambda_{x,i}\lambda_{y,j})\}$ where $\{\lambda_{x,i}\}$ and $\{\lambda_{y,j}\}$ are the nonunit eigenvalues of $\mathbf{Q}_x$ and $\mathbf{Q}_y$ respectively.

Theorem 3. *Under Assumption 1 and the null hypothesis that $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ are independent Markov chains, the asymptotic distribution of Pearson's chi-squared test equals that of $\sum_{i=1}^{d} \rho_i Z_i^2$ where $\{\rho_i\}$ are the eigenvalues of $\mathbf{\Theta}$ and $\{Z_i\}$ are independent $\mathcal{N}(0,1)$ variables. d is the rank of $\mathbf{\Omega}$. It holds that*

$$\sum_{i=1}^{d} \rho_i = \sum_{i=1}^{m_x-1} \sum_{j=1}^{m_y-1} \frac{1 + \lambda_{x,i}\lambda_{y,j}}{1 - \lambda_{x,i}\lambda_{y,j}}. \tag{6}$$

If both processes are additionally reversible, $\{\frac{1+\lambda_{x,i}\lambda_{y,j}}{1-\lambda_{x,i}\lambda_{y,j}}\}$ are the eigenvalues *of* $\mathbf{D}_p^{-1}\boldsymbol{\Omega}$.

A proof of this theorem can be established analogously to the proof of Theorem 1. The significance of this result stems again from the fact that for a mean approximation we only need to know the eigenvalues of $\mathbf{Q}_x$ and $\mathbf{Q}_y$ and the rank of $\boldsymbol{\Theta}$. The rather restrictive reversibility assumption can thus be dropped.

The Pearson test can be adjusted analogously to the discussion in the previous section. If $\{X_t\}$ and $\{Y_t\}$ are independent Markov chains then we get

$$\sqrt{T}\left(\mathbf{I}-\mathbf{Q}_x'\otimes\mathbf{Q}_y'\right)(\hat{\mathbf{p}}-\mathbf{p}_x\otimes\mathbf{p}_y)\sim^a\mathcal{N}\left(\mathbf{0},\boldsymbol{\Omega}_b\right)$$

with $\boldsymbol{\Omega}_b=\mathbf{D}_{p_x}\otimes\mathbf{D}_{p_y}-\mathbf{Q}_x'\mathbf{D}_{p_x}\mathbf{Q}_x\otimes\mathbf{Q}_y'\mathbf{D}_{p_y}\mathbf{Q}_y$. If we condition on the observed marginals $\hat{\mathbf{p}}_x$ and $\hat{\mathbf{p}}_y$ we get

$$\sqrt{T}\left(\mathbf{I}-\mathbf{Q}_x'\otimes\mathbf{Q}_y'\right)(\hat{\mathbf{p}}-\hat{\mathbf{p}}_x\otimes\hat{\mathbf{p}}_y)\sim^a\mathcal{N}\left(\mathbf{0},\boldsymbol{\Omega}_a\right)$$

with

$$\begin{aligned}\boldsymbol{\Omega}_a &= (\mathbf{D}_{p_x}-\mathbf{p}_x\mathbf{p}_x')\otimes(\mathbf{D}_{p_y}-\mathbf{p}_y\mathbf{p}_y')\\ &- \mathbf{Q}_x'(\mathbf{D}_{p_x}-\mathbf{p}_x\mathbf{p}_x')\mathbf{Q}_x\otimes\mathbf{Q}_y'(\mathbf{D}_{p_y}-\mathbf{p}_y\mathbf{p}_y')\mathbf{Q}_y.\end{aligned}$$

The quadratic form

$$X_a^2(T)=T\,(\hat{\mathbf{p}}-\hat{\mathbf{p}}_x\otimes\hat{\mathbf{p}}_y)'\,(\mathbf{I}-\mathbf{Q}_x\otimes\mathbf{Q}_y)\boldsymbol{\Omega}_a^-(\mathbf{I}-\mathbf{Q}_x'\otimes\mathbf{Q}_y')\,(\hat{\mathbf{p}}-\hat{\mathbf{p}}_x\otimes\hat{\mathbf{p}}_y)$$

is asymptotically χ_d^2 distributed where d is the rank of $\boldsymbol{\Omega}_a$. Observe that in the derivation $\mathbf{Q}_x$ and $\mathbf{Q}_y$ are known a priori. But again, if $\mathbf{Q}_x$ and $\mathbf{Q}_y$ can be estimated consistently by $\hat{\mathbf{Q}}_x$ and $\hat{\mathbf{Q}}_y$ the asymptotic distribution does not change, i.e.

$$\hat{X}_a^2(T)=T\,(\hat{\mathbf{p}}-\hat{\mathbf{p}}_x\otimes\hat{\mathbf{p}}_y)'\,(\mathbf{I}-\hat{\mathbf{Q}})\hat{\boldsymbol{\Omega}}_a^-(\mathbf{I}-\hat{\mathbf{Q}}')\,(\hat{\mathbf{p}}-\hat{\mathbf{p}}_x\otimes\hat{\mathbf{p}}_y)$$

with

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_a &= (\mathbf{D}_{\hat{p}_x}-\hat{\mathbf{p}}_x\hat{\mathbf{p}}_x')\otimes(\mathbf{D}_{\hat{p}_y}-\hat{\mathbf{p}}_y\hat{\mathbf{p}}_y')\\ &- \hat{\mathbf{Q}}_x'(\mathbf{D}_{\hat{p}_x}-\hat{\mathbf{p}}_x\hat{\mathbf{p}}_x')\hat{\mathbf{Q}}_x\otimes\hat{\mathbf{Q}}_y'(\mathbf{D}_{\hat{p}_y}-\hat{\mathbf{p}}_y\hat{\mathbf{p}}_y')\hat{\mathbf{Q}}_y\end{aligned}$$

is also asymptotically χ_d^2 distributed.

Finally, consider $\epsilon_t^x = X_t - \mathbf{Q}_x^{(1)\prime} X_{t-1}$ and $\epsilon_t^y = Y_t - \mathbf{Q}_y^{(1)\prime} Y_{t-1}$ and let

$$H_b(T) = \frac{1}{\sqrt{T-1}} \sum_{t=2}^{T} \epsilon_t^x \otimes \epsilon_t^y.$$

The expected value of $H_b(T)$ is $\mathbf{0}$. Under H_0 the variance is given by

$$\mathbb{V}(H_b(T)) = \sum_{l=-T+2}^{T-2} \frac{T-1-|l|}{T-1} \left(\mathbf{A}_x^{(l)} - \mathbf{Q}_x^{(1)\prime}\mathbf{A}_x^{(l+1)}\right) \otimes \left(\mathbf{A}_y^{(l)} - \mathbf{Q}_y^{(1)\prime}\mathbf{A}_y^{(l+1)}\right)$$

with

$$\mathbf{A}_x^{(l)} = \mathbf{P}_x^{(l)} - \mathbf{P}_x^{(l-1)}\mathbf{Q}_x^{(1)} \text{ and } \mathbf{A}_y^{(l)} = \mathbf{P}_y^{(l)} - \mathbf{P}_y^{(l-1)}\mathbf{Q}_y^{(1)}.$$

Theorem 4. *If Assumption 1 holds, if $\{X_t\}$ and $\{Y_t\}$ are independent of each other, and if at least one of the two series is a Markov chain then $H_b(T) \sim^a \mathcal{N}(\mathbf{0}, (\mathbf{D}_{p_x} - \mathbf{Q}_x^{(1)\prime}\mathbf{D}_{p_x}\mathbf{Q}_x^{(1)}) \otimes (\mathbf{D}_{p_y} - \mathbf{Q}_y^{(1)\prime}\mathbf{D}_{p_y}\mathbf{Q}_y^{(1)\prime}))$. The quadratic form $X_b^2(T) = H_b(T)' \left\{(\mathbf{D}_{p_x} - \mathbf{Q}_x^{(1)\prime}\mathbf{D}_{p_x}\mathbf{Q}_x^{(1)}) \otimes (\mathbf{D}_{p_y} - \mathbf{Q}_y^{(1)\prime}\mathbf{D}_{p_y}\mathbf{Q}_y^{(1)\prime})\right\}^- H_b(T)$ is asymptotically χ_d^2 distributed where d is the rank of the covariance matrix.*

Proof. Assumption 1 implies that $\mathbb{V}(H_b(T))$ converges for $T \to \infty$. W.l.o.g. let $\{X_t\}$ be a Markov chain. $\mathbf{A}_x^{(l)} = \mathbf{D}_{p_x}\mathbf{Q}_x^l - \mathbf{D}_{p_x}\mathbf{Q}_x^{l-1}\mathbf{Q}_x = \mathbf{0}$ for $l > 0$. If $l < 0$ then $\mathbf{A}_x^{(l)} - \mathbf{Q}_x^{(1)\prime}\mathbf{A}_x^{(l+1)} = \mathbf{P}_x^{(-l)\prime} - \mathbf{P}_x^{(-l+1)\prime}\mathbf{Q}_x - \mathbf{Q}_x'\mathbf{P}_x^{(-l-1)\prime} + \mathbf{Q}_x'\mathbf{P}_x^{(-l)\prime}\mathbf{Q}_x$. By setting $k = -l$ we get $\mathbf{Q}_x^{\prime k}\mathbf{D}_{p_x} - \mathbf{Q}_x^{\prime k+1}\mathbf{D}_{p_x}\mathbf{Q}_x - \mathbf{Q}_x'\mathbf{Q}_x^{\prime k-1}\mathbf{D}_{p_x} + \mathbf{Q}_x'\mathbf{Q}_x^{\prime k}\mathbf{D}_{p_x}\mathbf{Q}_x = \mathbf{0}$. Plugging this in yields $\mathbb{V}(H_b(T)) = \mathbf{A}_x^{(0)} \otimes \left(\mathbf{A}_y^{(0)} - \mathbf{Q}_y^{(1)\prime}\mathbf{A}_y^{(1)}\right)$. Now, $\mathbf{A}_y^{(1)} = \mathbf{P}_y^{(1)} - \mathbf{P}_y^{(0)}\mathbf{Q}_y^{(1)} = \mathbf{D}_{p_y}\mathbf{Q}_y^{(1)} - \mathbf{D}_{p_y}\mathbf{Q}_y^{(1)} = \mathbf{0}$ irrespective whether $\{Y_t\}$ is a Markov chain or not. Hence,

$$\mathbb{V}(H_b(T)) = (\mathbf{D}_{p_x} - \mathbf{Q}_x^{(1)\prime}\mathbf{D}_{p_x}\mathbf{Q}_x^{(1)}) \otimes (\mathbf{D}_{p_y} - \mathbf{Q}_y^{(1)\prime}\mathbf{D}_{p_y}\mathbf{Q}_y^{(1)}).$$

□

Now, suppose that the transition probabilities are not known but the estimators $\hat{\mathbf{Q}}_x^{(1)}$ and $\hat{\mathbf{Q}}_y^{(1)}$ satisfy $\hat{\mathbf{Q}}_x^{(1)} = \mathbf{Q}_x^{(1)} + O_p(T^{-1/2})$ and $\hat{\mathbf{Q}}_y^{(1)} = \mathbf{Q}_y^{(1)} +$

$O_p(T^{-1/2})$. Define $\hat{\epsilon}_t^x = X_t - \hat{\mathbf{Q}}_x^{(1)\prime} X_{t-1}$, $\hat{\epsilon}_t^y = Y_t - \hat{\mathbf{Q}}_y^{(1)\prime} Y_{t-1}$, and

$$\hat{H}_b(T) = \frac{1}{\sqrt{T-1}} \sum_{t=2}^{T} \hat{\epsilon}_t^x \otimes \hat{\epsilon}_t^y.$$

Theorem 5. *If the assumptions of Theorem 4 hold and if* $\hat{\mathbf{Q}}_x^{(1)} = \mathbf{Q}_x^{(1)} + O_p(T^{-1/2})$ *and* $\hat{\mathbf{Q}}_y^{(1)} = \mathbf{Q}_y^{(1)} + O_p(T^{-1/2})$ *then*

$$\hat{X}_b^2(T) = \hat{H}_b(T)' \left\{ (\mathbf{D}_{\hat{p}_x} - \hat{\mathbf{Q}}_x^{(1)\prime} \mathbf{D}_{\hat{p}_x} \hat{\mathbf{Q}}_x^{(1)}) \otimes (\mathbf{D}_{\hat{p}_y} - \hat{\mathbf{Q}}_y^{(1)\prime} \mathbf{D}_{\hat{p}_y} \hat{\mathbf{Q}}_y^{(1)\prime}) \right\}^{-} \hat{H}_b(T)$$

is asymptotically equivalent to $X_b^2(T)$.

Proof. The term in curly brackets equals $(\mathbf{D}_{p_x} - \mathbf{Q}_x^{(1)\prime} \mathbf{D}_{p_x} \mathbf{Q}_x^{(1)}) \otimes (\mathbf{D}_{p_y} - \mathbf{Q}_y^{(1)\prime} \mathbf{D}_{p_y} \mathbf{Q}_y^{(1)\prime}) + O_p(T^{-1/2})$. $\hat{H}_b(T)$ may be expressed as

$$\begin{aligned}
\hat{H}_b(T) &= H_b(T) \\
&+ \tfrac{1}{\sqrt{T-1}} \left\{ \mathbf{I}_x \otimes \left(\mathbf{Q}_y^{(1)} - \hat{\mathbf{Q}}_y^{(1)} \right)' \right\} \textstyle\sum_{t=2}^{T} \epsilon_t^x \otimes Y_{t-1} \\
&+ \tfrac{1}{\sqrt{T-1}} \left\{ \left(\mathbf{Q}_x^{(1)} - \hat{\mathbf{Q}}_x^{(1)} \right)' \otimes \mathbf{I}_y \right\} \textstyle\sum_{t=2}^{T} X_{t-1} \otimes \epsilon_t^y \\
&+ \tfrac{1}{\sqrt{T-1}} \left\{ \left(\mathbf{Q}_x^{(1)} - \hat{\mathbf{Q}}_x^{(1)} \right)' \otimes \left(\mathbf{Q}_y^{(1)} - \hat{\mathbf{Q}}_y^{(1)} \right)' \right\} \textstyle\sum_{t=2}^{T} X_{t-1} \otimes Y_{t-1} \\
&= H_b(T) \\
&+ \tfrac{1}{\sqrt{T-1}} O_p(T^{-1/2}) \textstyle\sum_{t=2}^{T} \epsilon_t^x \otimes Y_{t-1} \\
&+ \tfrac{1}{\sqrt{T-1}} O_p(T^{-1/2}) \textstyle\sum_{t=2}^{T} X_{t-1} \otimes \epsilon_t^y \\
&+ O_p(T^{-3/2}) \textstyle\sum_{t=2}^{T} X_{t-1} \otimes Y_{t-1}
\end{aligned}$$

X_t and Y_t are vectors consisting of zeros and ones. Hence, the last summand is $O_p(T^{-1/2})$. Now, the expected value $(1/\sqrt{T-1}) \sum_{t=2}^{T} \epsilon_t^x \otimes Y_{t-1}$ equals zero and the variance equals

$$\sum_{l=-T+2}^{T-2} \frac{T-1-|l|}{T-1} \left(\mathbf{A}_x^{(l)} - \mathbf{Q}_x^{(1)\prime} \mathbf{A}_x^{(l+1)} \right) \otimes \mathbf{P}_y^{(l)}$$

which is finite by Assumption 1. The same is true for $(1/\sqrt{T-1}) \sum_{t=2}^{T} X_{t-1} \otimes \epsilon_t^y$. Hence, $\hat{H}_b(T) = H_b(T) + O_p(T^{-1/2})$ and $\hat{X}_b^2(T) = X_b^2(T) + O_p(T^{-1/2})$. □

Remark 3. *I show in Appendix B that $X_b^2(T)$ is asymptotically equivalent to the trace statistic developed in Theorem 1 in Pesaran and Timmermann [2009]. The main difference is that only one of the two series has to be a Markov chain for the theorem to hold. Chou and Chu [2010] proposed a similar test using stronger assumptions.*

4 Three-way tables

Let $U_t = X_t \otimes Y_t \otimes Z_t$. All three series fulfill Assumption 1. The notation is analogous to the last section. We analyze three different notions of independence: mutual independence, joint independence of two series from the third, and finally, conditional independence.

Mutual independence implies that $\mathbf{P}^{(l)} = \mathbf{P}_x^{(l)} \otimes \mathbf{P}_y^{(l)} \otimes \mathbf{P}_z^{(l)}$ for all $l \in \mathbb{Z}$, in particular $\mathbf{p} = \mathbf{p}_x \otimes \mathbf{p}_y \otimes \mathbf{p}_z$. We know that $\sqrt{T}(\hat{\mathbf{p}} - \mathbf{p}_x \otimes \mathbf{p}_y \otimes \mathbf{p}_z) \sim^a \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ with

$$\mathbf{\Sigma} = (\mathbf{D}_{p_x} \otimes \mathbf{D}_{p_y} \otimes \mathbf{D}_{p_z}) \sum_{l=-\infty}^{\infty} \left(\mathbf{Q}_x^{(l)} \otimes \mathbf{Q}_y^{(l)} \otimes \mathbf{Q}_z^{(l)} - \mathbf{1}_{m_x}\mathbf{p}_x' \otimes \mathbf{1}_{m_y}\mathbf{p}_y' \otimes \mathbf{1}_{m_z}\mathbf{p}_z'\right).$$

Porteous [1987] showed that

$$\sqrt{T}\left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y \otimes \hat{\mathbf{p}}_y\right) = \sqrt{T}\mathbf{B}'\left(\hat{\mathbf{p}} - \mathbf{p}_x \otimes \mathbf{p}_y \otimes \mathbf{p}_z\right) + O_p(T^{-1/2})$$

with $\mathbf{B} = \mathbf{I} - \mathbf{I}_{m_x} \otimes \mathbf{1}_{m_y}\mathbf{p}_y' \otimes \mathbf{1}_{m_z}\mathbf{p}_z' - \mathbf{1}_{m_x}\mathbf{p}_x' \otimes \mathbf{I}_{m_y} \otimes \mathbf{1}_{m_z}\mathbf{p}_z' - \mathbf{1}_{m_x}\mathbf{p}_x' \otimes \mathbf{1}_{m_y}\mathbf{p}_y' \otimes \mathbf{I}_{m_z}$. Porteous [1987] used $\mathbf{B}' + 2 \cdot \mathbf{p}\mathbf{1}'$. But again, as $\mathbf{1}'(\hat{\mathbf{p}} - \mathbf{p}) = \mathbf{0}$ we may drop the last summand. The asymptotic covariance matrix $\mathbf{\Omega}$ of $\sqrt{T}\left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y \otimes \hat{\mathbf{p}}_y\right)$ is rather clumsy.

$$\begin{aligned}
\mathbf{\Omega} = \mathbf{B}'\mathbf{\Sigma}\mathbf{B} &= \mathbf{D}_p\mathbf{B}\textstyle\sum_{l=-\infty}^{\infty}\left(\mathbf{Q}_x^{(l)} \otimes \mathbf{Q}_y^{(l)} \otimes \mathbf{Q}_z^{(l)} - \mathbf{1}\mathbf{p}'\right)\mathbf{B} \\
&= \mathbf{D}\textstyle\sum_{l=-\infty}^{\infty}\left\{\mathbf{Q}_x^{(l)} \otimes \mathbf{Q}_y^{(l)} \otimes \mathbf{Q}_z^{(l)} - \mathbf{1}\mathbf{p}'\right. \\
&- (\mathbf{Q}_x^{(l)} - \mathbf{1}_{m_x}\mathbf{p}_x) \otimes \mathbf{1}_{m_y}\mathbf{p}_y' \otimes \mathbf{1}_{m_z}\mathbf{p}_z' \\
&- \mathbf{1}_{m_x}\mathbf{p}_x' \otimes (\mathbf{Q}_y^{(l)} - \mathbf{1}_{m_x}\mathbf{p}_y) \otimes \mathbf{1}_{m_z}\mathbf{p}_z' \\
&- \left.\mathbf{1}_{m_x}\mathbf{p}_x' \otimes \mathbf{1}_{m_y}\mathbf{p}_y' \otimes (\mathbf{Q}_z^{(l)} - \mathbf{1}_{m_z}\mathbf{p}_z)\right\}.
\end{aligned}$$

Pearson's statistic for testing mutual independence of the three variables is given by

$$X^2_{I(3)} = T\left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y \otimes \hat{\mathbf{p}}_z\right)' \left(\mathbf{D}^{-1}_{\hat{p}_x} \otimes \mathbf{D}^{-1}_{\hat{p}_y} \otimes \mathbf{D}^{-1}_{\hat{p}_z}\right) \left(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y \otimes \hat{\mathbf{p}}_z\right).$$

The asymptotic distribution is equal to the distribution of a weighted sum of χ^2_1 variables. The weights correspond to the eigenvalues of $\mathbf{D}_p^{-1}\boldsymbol{\Omega}$.

If the three series are Markov chains, we get by using the same reasoning as above that

$$\begin{aligned} \textstyle\sum_{i=1}^{d} \rho_i &= \underbrace{\sum_{i=1}^{m_x}\sum_{j=1}^{m_y}\sum_{i=1}^{m_z}}_{i+j+k<m_x+m_y+m_z} \tfrac{1+\lambda_{x,i}\lambda_{y,j}\lambda_{z,k}}{1-\lambda_{x,i}\lambda_{y,j}\lambda_{z,k}} \\ &\quad - \textstyle\sum_{i=1}^{m_x-1} \tfrac{1+\lambda_{x,i}}{1-\lambda_{x,i}} - \sum_{i=1}^{m_y-1} \tfrac{1+\lambda_{y,i}}{1-\lambda_{y,i}} - \sum_{i=1}^{m_z-1} \tfrac{1+\lambda_{z,i}}{1-\lambda_{z,i}} \end{aligned}$$

where $\{\lambda_{x,i}\}$, $\{\lambda_{y,j}\}$, and $\{\lambda_{z,k}\}$ are the ordered eigenvalues of $\mathbf{Q}_x$, $\mathbf{Q}_y$, and $\mathbf{Q}_z$. $\lambda_{x,m_x} = \lambda_{y,m_y} = \lambda_{z,m_z} = 1$. Porteous [1987] showed that if all three series are reversible, the eigenvalues of $\mathbf{D}_p\boldsymbol{\Omega}$ correspond to $\{(1+\lambda_{x,i}\lambda_{y,j}\lambda_{z,k})/(1-\lambda_{x,i}\lambda_{y,j}\lambda_{z,k})\}$.

If two of the three series are i.i.d., we get

$$\begin{aligned} \boldsymbol{\Omega} &= (\mathbf{D}_p - \mathbf{p}\mathbf{p}') - (\mathbf{D}_{p_x} \otimes \mathbf{p}_y\mathbf{p}_y' \otimes \mathbf{p}_z\mathbf{p}_z' - \mathbf{p}\mathbf{p}') \\ &\quad -(\mathbf{p}_x\mathbf{p}_x' \otimes \mathbf{D}_{p_y} \otimes \mathbf{p}_z\mathbf{p}_z' - \mathbf{p}\mathbf{p}') - (\mathbf{p}_x\mathbf{p}_x' \otimes \mathbf{p}_y\mathbf{p}_y' \otimes \mathbf{D}_{p_z} - \mathbf{p}\mathbf{p}'). \end{aligned}$$

In this case $\mathbf{D}_p^{-1}\boldsymbol{\Omega}$ is idempotent and the Pearson test is asymptotically χ^2_d distributed with $d = (m_x m_y m_z - 1) - (m_x - 1) - (m_y - 1) - (m_z - 1)$.

To test whether $\{(Y_t, Z_t)\}$ are jointly independent from $\{X_t\}$ is a simple extension of the two-way tables case. The null hypothesis implies that $\mathbf{P}^{(l)} = \mathbf{P}^{(l)}_x \otimes \mathbf{P}^{(l)}_{y\otimes z}$ for all $l \in \mathbb{Z}$ where the subscript $y \otimes z$ is shorthand for the random variable $Y_t \otimes Z_t$. All the results from section 2 apply. A weaker implication of this null hypothesis is that both $\mathbf{P}^{(l)}_{x\otimes y} = \mathbf{P}^{(l)}_x \otimes \mathbf{P}^{(l)}_y$ and $\mathbf{P}^{(l)}_{x\otimes z} = \mathbf{P}^{(l)}_x \otimes \mathbf{P}^{(l)}_z$ for all $l \in \mathbb{Z}$. I show in Appendix B that the test proposed by Pesaran and Timmermann [2009] tests actually for this weaker implication.

Let us finally assume that Y_t and Z_t are independent conditionally on X_t, i.e. $p_{ijk} = p_{ij+}p_{i+k}/p_{i++}$. As usual the subscript + denotes that we sum across this dimension. Hence, we focus on

$$d^i_{jk} = \frac{1}{T}\left(n_{ijk} - \frac{n_{ij+}n_{i+k}}{n_{i++}}\right) = \hat{p}_{ijk} - \frac{\hat{p}_{ij+}\hat{p}_{i+k}}{\hat{p}_{i++}}.$$

The first observation is that

$$\begin{aligned} d^i_{jk} &= \hat{p}_{ijk} - \tfrac{p_{ij+}p_{i+k}}{p_{i++}} - \tfrac{p_{ij+}(\hat{p}_{i+k}-p_{i+k})}{p_{i++}} - \tfrac{p_{i+k}(\hat{p}_{ij+}-p_{ij+})}{p_{i++}} \\ &+ \tfrac{p_{ij+}p_{i+k}}{p_{i++}}(\hat{p}_{i++} - p_{i++}) + O_p(T^{-1}). \end{aligned}$$

Let $\hat{\mathbf{r}}^i = (\hat{p}_{i11}, \ldots, \hat{p}_{i,m_y,m_z})'/p_{i++}$ and $\mathbf{r}^i = (p_{i11}, \ldots, p_{i,m_y,m_z})'/p_{i++}$. For $\mathbf{d}^i = (d^i_{1,1}, d^i_{1,2} \ldots, d^i_{m_y,m_z})'$ we may write

$$\begin{aligned} \mathbf{d}^i &= p_{i++}\left(\mathbf{I}_{m_y} \otimes \mathbf{I}_{m_z} - \mathbf{r}^i_y\mathbf{1}' \otimes \mathbf{I}_{m_z} - \mathbf{I}_{m_y} \otimes \mathbf{r}^i_z\mathbf{1}'\right)\left(\hat{\mathbf{r}}^{\mathbf{i}} - \mathbf{r}^i\right). \\ &+ (\hat{p}_{i++} - p_{i++})\mathbf{r}^i_y \otimes \mathbf{r}^i_z + O_p(T^{-1}). \end{aligned}$$

$\mathbf{r}^i_y$ and $\mathbf{r}^i_z$ are the conditional marginal probabilities of Y_t and Z_t ($\mathbf{r}^i_y = (\mathbf{I}_{m_y} \otimes \mathbf{1}'_{m_z})\mathbf{r}^i$ and $\mathbf{r}^i_z = (\mathbf{1}'_{m_y} \otimes \mathbf{I}_{m_z})\mathbf{r}^i$). For $\mathbf{d} = (\mathbf{d}^{1\prime}, \mathbf{d}^{2\prime}, \ldots, \mathbf{d}^{m_x\prime})'$ we get $\mathbf{d} = \mathbf{A}'(\hat{\mathbf{p}} - \mathbf{p}) + O_p(T^{-1})$ where $\mathbf{A}$ is the block diagonal matrix consisting of $\mathbf{A}_i = (\mathbf{I}_{m_y} - \mathbf{1}\mathbf{r}^{i\prime}_y) \otimes (\mathbf{I}_{m_z} - \mathbf{1}\mathbf{r}^{i\prime}_z)$.

Without serial correlation the covariance matrix of $\sqrt{T}(\hat{\mathbf{p}} - \mathbf{p})$ is given by $\mathbf{\Sigma} = \mathbf{D}_p - \mathbf{p}\mathbf{p}'$ with

$$\mathbf{p} = \begin{pmatrix} p_{1++}\mathbf{r}^1_y \otimes \mathbf{r}^1_z \\ \vdots \\ p_{m_x++}\mathbf{r}^{m_x}_y \otimes \mathbf{r}^{m_x}_z \end{pmatrix}.$$

It is easy to check that $\mathbf{A}'_i(\mathbf{r}^i_y \otimes \mathbf{r}^i_z) = \mathbf{0}$. The covariance matrix $\mathbf{\Omega}$ of $\sqrt{T}\mathbf{d}$ is therefore the block matrix given by the direct sum $\bigoplus_{i=1}^{m_x} p_{i++}(\mathbf{D}_{r^i_y} - \mathbf{r}^i_y\mathbf{r}^{i\prime}_y) \otimes (\mathbf{D}_{r^i_z} - \mathbf{r}^i_z\mathbf{r}^{i\prime}_z)$. An immediate consequence is that $\mathbf{d}^i$ and $\mathbf{d}^j$ are asymptotically independent for $i \neq j$. Moreover,

$$CMH^*_i = T\mathbf{d}^{i\prime}\left(p_{i++}(\mathbf{D}_{r^i_y} - \mathbf{r}^i_y\mathbf{r}^{i\prime}_y) \otimes (\mathbf{D}_{r^i_z} - \mathbf{r}^i_z\mathbf{r}^{i\prime}_z)\right)^- \mathbf{d}^i \sim^a \chi^2_{(m_y-1)(m_z-1)}.$$

$(1/p_{i++})\mathbf{D}_{r_y^i}^{-1} \otimes \mathbf{D}_{r_z^i}^{-1}$ is a g inverse of $p_{i++}(\mathbf{D}_{r_y^i} - \mathbf{r}_y^i\mathbf{r}_y^{i\prime}) \otimes (\mathbf{D}_{r_z^i} - \mathbf{r}_z^i\mathbf{r}_z^{i\prime})$. Hence,

$$CMH_i^* = \frac{T}{p_{i++}}\mathbf{d}^{i\prime}(\mathbf{D}_{r_y^i}^{-1} \otimes \mathbf{D}_{r_z^i}^{-1})\mathbf{d}^i = T\sum_{j=1}^{m_y}\sum_{k=1}^{m_z}\frac{\left(\hat{p}_{ijk} - \frac{\hat{p}_{ij+}\hat{p}_{i+k}}{\hat{p}_{i++}}\right)^2}{\frac{p_{ij+}p_{i+k}}{p_{i++}}}.$$

The asymptotic independence of $\mathbf{d}^i$ and $\mathbf{d}^j$ implies that $CMH^* = T\mathbf{d}'\mathbf{\Omega}^-\mathbf{d} = \sum_{i=1}^{m_x} CMH_i^*$ is χ_d^2 distributed with $d = m_x(m_y-1)(m_z-1)$. Plugging in the observed frequencies instead of p_{ij+}, p_{i+k}, and p_{i++} in the denominator does not change the result, i.e.

$$T\sum_{i=1}^{m_x}\sum_{j=1}^{m_y}\sum_{k=1}^{m_z}\frac{\left(\hat{p}_{ijk} - \frac{\hat{p}_{ij+}\hat{p}_{i+k}}{\hat{p}_{i++}}\right)^2}{\frac{\hat{p}_{ij+}\hat{p}_{i+k}}{\hat{p}_{i++}}} \sim^a \chi_d^2.$$

The classical Cochran–Mantel–Haenszel test (CMH) focuses on $\mathbf{d}^* = \sum_{i=1}^{m_x}\mathbf{d}^i = (\mathbf{1}'_{m_x} \otimes \mathbf{I}_{m_y} \otimes \mathbf{I}_{m_z})\mathbf{d}$. CMH tests the weaker hypothesis that $p_{+jk} = \sum_{i=1}^{m_x} p_{ij+}p_{i+k}/p_{i++}$. The covariance matrix of $\sqrt{T}\mathbf{d}^*$ is given by

$$\mathbf{\Omega}^* = \sum_{i=1}^{m_x} p_{i++}(\mathbf{D}_{r_y^i} - \mathbf{r}_y^i\mathbf{r}_y^{i\prime}) \otimes (\mathbf{D}_{r_z^i} - \mathbf{r}_z^i\mathbf{r}_z^{i\prime}).$$

Again, using the observed frequencies instead of the true probabilities in $\mathbf{\Omega}^*$ does not change the result and $CMH = T\mathbf{d}^{*\prime}(\hat{\mathbf{\Omega}}^*)^+\mathbf{d}^* \sim^a \chi_d^2$ with $d = (m_y-1)(m_z-1)$. A detailed analysis that the above definition of CMH is equivalent to the classical definition as in Agresti [2002] can be found in Appendix C. It is important to note that the equivalence holds only for the Moore-Penrose inverse. If we use an arbitrary g inverse, $T\mathbf{d}^{*\prime}(\hat{\mathbf{\Omega}}^*)^-\mathbf{d}^*$ is asymptotically χ_d^2 distributed but does not equal the standard definition of CMH. The test for conditional independence proposed in Pesaran and Timmermann [2009] is closely related to CMH. As I show in Appendix B they use

$$\left(\sum_{i=1}^{m_x} p_{i++}(\mathbf{D}_{r_y^i} - \mathbf{r}_y^i\mathbf{r}_y^{i\prime})\right) \otimes \left(\sum_{i=1}^{m_x} p_{i++}(\mathbf{D}_{r_z^i} - \mathbf{r}_z^i\mathbf{r}_z^{i\prime})\right)$$

instead of $\mathbf{\Omega}^*$. The difference is small but not necessarily negligible.[2]

If we allow for serial correlation the situation becomes a lot more complex. The fact that Y_t and Z_t are independent conditional on X_t does not imply that Y_t and Z_t are independent conditional on X_t and some lagged variables.

Example 2. *Let $\{U_t\}$ be a Markov process with transition matrix*

$$\mathbf{Q} = \begin{pmatrix} a & 1-a & 1-a & a & 1-a & a & a & 1-a \\ a & 1-a & 1-a & a & 1-a & a & a & 1-a \\ 1-a & a & a & 1-a & a & 1-a & 1-a & a \\ 1-a & a & a & 1-a & a & 1-a & 1-a & a \\ a & 1-a & 1-a & a & 1-a & a & a & 1-a \\ a & 1-a & 1-a & a & 1-a & a & a & 1-a \\ 1-a & a & a & 1-a & a & 1-a & 1-a & a \\ 1-a & a & a & 1-a & a & 1-a & 1-a & a \end{pmatrix}.$$

$a \in (0,1)$. *If $U_t = X_t \otimes Y_t \otimes Z_t$ then it is straightforward to verify that $\mathbb{P}(Y_t, Z_t|X_t) = \mathbb{P}(Y_t|X_t)\mathbb{P}(Z_t|X_t)$. But if we condition additionally on Y_{t-1} the two series are not independent anymore, i.e. $\mathbb{P}(Y_t, Z_t|X_t, Y_{t-1})$ does not equal $\mathbb{P}(Y_t|X_t, Y_{t-1})\mathbb{P}(Z_t|X_t, Y_{t-1})$.*

So it seems reasonable to put further restrictions on the transition matrices. In the first case assume that only the conditioning variable might be serially correlated. Let $q_{i,j}^{x,l}$ be the probability of $X_{t+l} = j$ conditional on $X_t = i$. For $m_x = 2$ and $l \neq 0$ the transition matrix is given by

$$\mathbf{Q}^{(l)} = \begin{pmatrix} q_{1,1}^{x,l}\mathbf{1}_{m_y m_z}(\mathbf{r}_y^1 \otimes \mathbf{r}_z^1)' & q_{1,2}^{x,l}\mathbf{1}_{m_y m_z}(\mathbf{r}_y^2 \otimes \mathbf{r}_z^2)' \\ q_{2,1}^{x,l}\mathbf{1}_{m_y m_z}(\mathbf{r}_y^1 \otimes \mathbf{r}_z^1)' & q_{1,1}^{x,l}\mathbf{1}_{m_y m_z}(\mathbf{r}_y^2 \otimes \mathbf{r}_z^2)' \end{pmatrix}.$$

It is not surprising that the covariance matrix of $\sqrt{T}\mathbf{d}$ in this case is equal to the case without serial correlation, i.e. $\mathbf{A}'\mathbf{Q}^{(l)} = \mathbf{0}$ for $l \neq 0$. Both, CMH and CMH* have the above specified asymptotic distributions under H_0.

A second application is to test whether a Markov chain is of order 1 against the alternative of a higher order as in Kullback et al. [1962]. Sup-

[2]See Appendix B for details.

pose that $\{X_t\}$ is a simple Markov chain with transition matrix $\mathbf{Q}$. Define $U_t = X_{t-1} \otimes X_{t-2} \otimes X_t$ and let $p_{i,j,k}$ denote $\mathbb{P}(X_{t-1} = i, X_{t-2} = j, X_t = k)$. The particular ordering is chosen such that the first index corresponds to the conditioning variable. Under the null hypothesis X_{t-2} and X_t are independent conditional on X_{t-1}, or

$$\mathbb{P}(X_{t-1} = i, X_{t-2} = j, X_t = k) = p_{i,j,k} = \frac{p_{i,j,+}p_{i,+,k}}{p_{i++}} = p_{i,j,+}q_{i,k}$$

where $q_{i,k}$ denotes $\mathbb{P}(X_t = k | X_{t-1} = i)$. For $m_x = 2$ this equals in vector notation

$$\mathbf{p} = \begin{pmatrix} p_{1++}\mathbf{r}_y^1 \otimes \mathbf{r}_z^1 \\ p_{2++}\mathbf{r}_y^2 \otimes \mathbf{r}_z^2 \end{pmatrix} = \begin{pmatrix} \mathbf{p}_{1,\bullet,+} \otimes \mathbf{q}'_{1,\bullet} \\ \mathbf{p}_{2,\bullet,+} \otimes \mathbf{q}'_{2,\bullet} \end{pmatrix}$$

with $\mathbf{p}_{i,\bullet,+} = (p_{i,1,+}, p_{i,2,+})'$ and $\mathbf{q}_{i,\bullet} = (q_{i,1}, q_{i,2})$. The transition matrix for U_t is given by

$$\mathbf{Q} = \begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} q_{1,1} & q_{1,2} \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ q_{2,1} & q_{2,2} \end{pmatrix} \\ & \\ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} q_{1,1} & q_{1,2} \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ q_{2,1} & q_{2,2} \end{pmatrix} \end{pmatrix}.$$

Now, $\mathbf{A}'\mathbf{DQA} = \mathbf{0}$ and together with the discussion from above we get that CMH and CMH* are asymptotically χ^2 distributed under the null hypothesis.

If the values of X_t correspond to m_x different strata, Assumption 1 is violated. To invoke asymptotic theory we have to take the limit for each stratum separately. Under this assumption the transition matrices are given by

$$\mathbf{Q}^{(l)} = \begin{pmatrix} \mathbf{Q}_y^{(1,l)} \otimes \mathbf{Q}_z^{(1,l)} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_y^{(2,l)} \otimes \mathbf{Q}_z^{(2,l)} \end{pmatrix}.$$

The covariance matrix of $\sqrt{T}\mathbf{d}^i$ is given by

$$\mathbf{\Omega}^i = p_{i++}(\mathbf{D}_{r_y^i} \otimes \mathbf{D}_{r_z^i}) \sum_l (\mathbf{Q}_{r_y^i}^{(i,l)} - \mathbf{1r}_y^{i\prime}) \otimes (\mathbf{Q}_{r_z^i}^{(i,l)} - \mathbf{1r}_z^{i\prime})$$

and the covariance matrix $\mathbf{\Omega}$ of $\sqrt{T}\mathbf{d}$ is the direct sum of the $\mathbf{\Omega}^i$. We may proceed as in the standard $I \times J$ case. Yet, some caution with respect to the asymptotic result is necessary. In particular, it has to be guaranteed that the convergence of the estimators not only within but also across the strata, i.e. for $\hat{p}_{i++}$, is sufficiently fast for all $i = 1, \ldots, m_x$.

5 Simulation Results

I undertake a couple of simulations to illustrate the finite sample properties of the various test statistics.[3] The realizations of U_t are denoted by the column vector $\mathbf{u}_t$. They are summarized in the matrix $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_T)'$. The relative frequencies are calculated by $\hat{\mathbf{p}} = (1/T)\mathbf{U}'\mathbf{1}_T$ and $\mathbf{D}_{\hat{p}} = (1/T)\mathbf{U}'\mathbf{U}$. For $l > 0$ the joint probabilities $\mathbf{P}^{(l)}$ are estimated by $\hat{\mathbf{P}}^{(l)} = \frac{1}{T-l}\sum_{t=1+l}^{T} \mathbf{u}_{t-l}\mathbf{u}_t'$. For notational convenience I denote $\sum_{t=1+l}^{T} \mathbf{u}_{t-l}\mathbf{u}_t'$ by $\mathbf{U}_{-l}'\mathbf{U}$ where it is understood that $\mathbf{U}_{-l}$ includes the first $T-l$ observations and that in $\mathbf{U}$ the first l rows are deleted. As $\mathbf{P}^{(-l)} = \mathbf{P}^{(l)\prime}$ we use $\hat{\mathbf{P}}^{(-l)\prime}$ for $l < 0$ to estimate $\mathbf{P}^{(l)}$. The transition probabilities $\mathbf{Q}^{(l)}$ are estimated by $\hat{\mathbf{Q}}^{(l)} = \mathbf{D}_{\hat{p}}^{-1}\hat{\mathbf{P}}^{(l)}$.[4]

5.1 Goodness of Fit

In this section the classical Pearson test is compared to $\hat{P}_a$ and the Wald test. I use two procedures to estimate the covariance matrix $\mathbf{\Sigma}$ of $\sqrt{T}(\hat{\mathbf{p}} - \mathbf{p})$. The

[3] Appendix D includes additional simulations.

[4] Observe that $\hat{\mathbf{p}}'\hat{\mathbf{Q}}^{(l)} \neq \hat{\mathbf{p}}'$ and $\hat{\mathbf{Q}}^{(l)}\mathbf{1} \neq \mathbf{1}$. If we use $\mathbf{D}_{\hat{p}_{-l}} = (1/(T-l))\mathbf{U}_{-l}'\mathbf{U}_{-l}$ instead of $\mathbf{D}_{\hat{p}}$, we would get $\hat{\mathbf{Q}}^{(l)}\mathbf{1} = \mathbf{1}$. I tried both versions. The differences in the results were negligible so I stick with the first definition.

naive estimator is defined by

$$\begin{aligned}\hat{\mathbf{\Sigma}}_n &= \mathbf{D}_{\hat{p}} \sum_{l=-\tau}^{\tau} \tfrac{T-|l|}{T} \left(\hat{\mathbf{Q}}^{(l)} - \mathbf{1}\hat{\mathbf{p}}'\right) = \sum_{l=-\tau}^{\tau} \tfrac{T-|l|}{T} \left(\hat{\mathbf{P}}^{(l)} - \hat{\mathbf{p}}\hat{\mathbf{p}}'\right) \\ &= \mathbf{D}_{\hat{p}} - \hat{\mathbf{p}}\hat{\mathbf{p}}' + \sum_{l=1}^{\tau} \tfrac{T-|l|}{T} \left(\hat{\mathbf{P}}^{(l)} - \hat{\mathbf{p}}\hat{\mathbf{p}}' + \hat{\mathbf{P}}^{(l)\prime} - \hat{\mathbf{p}}\hat{\mathbf{p}}'\right)\end{aligned}$$

with the number of lags τ specified. This estimator is not consistent unless $\{U_t\}$ is an m-dependent process and $\tau > m$. For Markov chains I use that

$$\mathbf{\Sigma} = \mathbf{D}_p(\mathbf{I} - \mathbf{Q} + \mathbf{1}\mathbf{p}')^{-1} + (\mathbf{I} - \mathbf{Q} + \mathbf{1}\mathbf{p}')^{-1\prime}\mathbf{D}_p - \mathbf{D}_p - \mathbf{p}\mathbf{p}'$$

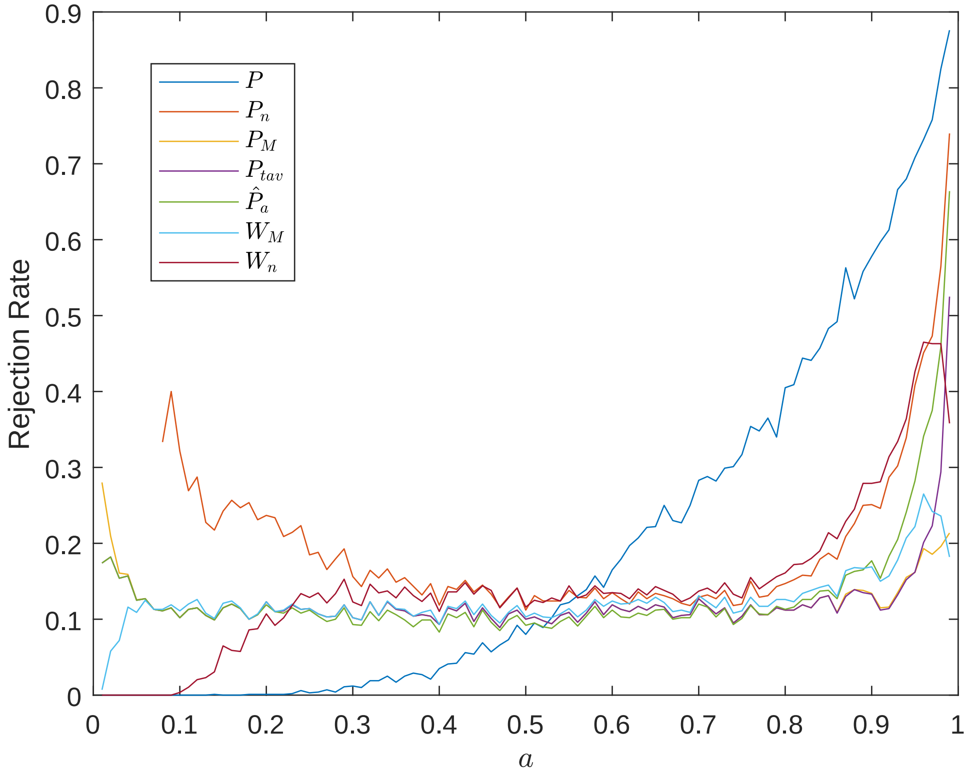
and define

$$\hat{\mathbf{\Sigma}}_M = \mathbf{D}_{\hat{p}}(\mathbf{I} - \hat{\mathbf{Q}} + \mathbf{1}\hat{\mathbf{p}}')^{-1} + (\mathbf{I} - \hat{\mathbf{Q}} + \mathbf{1}\hat{\mathbf{p}}')^{-1\prime}\mathbf{D}_{\hat{p}} - \mathbf{D}_{\hat{p}} - \hat{\mathbf{p}}\hat{\mathbf{p}}'.$$

I approximate the asymptotic distribution of the Pearson test by matching the first moment. Three variants to estimate $\bar{\rho} = (1/d)\sum \rho_i$ are employed. The corrected statistic is denoted by $P_n(T)$ ($P_M(T)$) if the trace of $\mathbf{D}_{\hat{p}}^{-1}\hat{\mathbf{\Sigma}}_n$ ($\mathbf{D}_{\hat{p}}^{-1}\hat{\mathbf{\Sigma}}_M$) is used. The estimated mean eigenvalue is denoted by $\hat{\bar{\rho}}_n$ ($\hat{\bar{\rho}}_M$). If we estimate $\bar{\rho}$ using Theorem 1, we denote the statistic by $P_{tav}(T)$ and the estimated mean eigenvalue by $\hat{\bar{\rho}}_{tav}$. The Wald test based on $\hat{\mathbf{\Sigma}}_n$ ($\hat{\mathbf{\Sigma}}_M$) is denoted by W_n (W_M).

To illustrate the size of the various tests consider the Markov chain given in example 1. I perform 1000 simulations for each a from $a = 0.01$ to $a = 0.99$ in steps of 0.01. The number of observations T in each run is 100. $\tau = 3$ for the naive estimator. The results are displayed in Figure 2. We find that $P_{tav}(T)$, $P_M(T)$, W_M, and $\hat{P}_a$ perform very well. $\hat{\bar{\rho}}_M$ and $\hat{\bar{\rho}}_{tav}$ are close to the true value. The mean absolute percentage error is around 20% for a in the range of 0.15 to 0.85. Not surprisingly, estimators based on the naive approach ($P_n(T)$ and W_n) perform poorer. The estimated mean of the eigenvalues ($\hat{\bar{\rho}}_n$) is consistently negative for $a < 0.18$ (this is equivalent to $\bar{\rho} < 0.22$ and an autocorrelation of $\{U_t^*\}$ of less than -0.64). For a in the range between 0.3 and 0.7 the performance is fine. It deteriorates again beyond $a = 0.7$. Increasing τ would improve the performance slightly yet not solve the general problem. Before calculating the rejection rates I eliminated

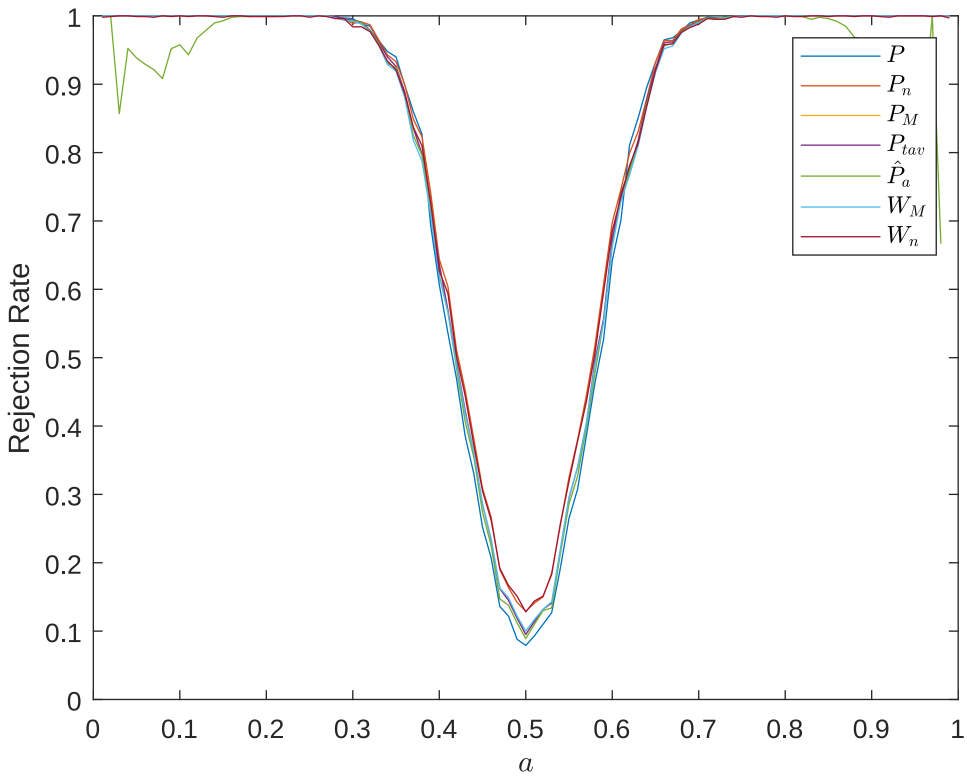
Figure 2: Rejection rates at a nominal level of 10% for the Markov chain given in Example 1. The results are based on 1000 simulations for $a = 0.01$ to $a = 0.99$ in steps of 0.01. The number of observations in each simulation equals 100.



all flawed values (negative values, NaNs, and Inf). For $a < 0.25$ more than 10% of the values of $P_n(T)$ and W_n had to be deleted. Appendix A includes a size simulation for an m-dependent process. In this case the statistics based on the naive estimator perform better.

In terms of power I consider the following example. Let $\{U_t\}$ be an i.i.d. sequence with $\mathbf{p} = (0.5, 0.5)'$. For $a = 0.01$ to $a = 0.99$ in steps of 0.01 we test separately and independently the null hypothesis that $\mathbf{p} = (a, 1 - a)'$. For each a we use 1000 simulations with $T = 100$ observations. Figure 3 illustrates that the alternative tests are as powerful as the standard Pearson test.

Figure 3: Rejection rates at a nominal level of 10% for an i.i.d. process with $\mathbf{p} = (0.5, 0.5)'$. The null hypothesis $\mathbf{p}^0 = (a, 1-a)'$ varies from $a = 0.01$ to $a = 0.99$ in steps of 0.01. The results are based on 1000 simulations. The number of observations in each simulation equals 100.



5.2 Independence of two series

In this section I will analyze the performance of Pearson's classical χ^2 test, the Wald test, $\hat{X}_a^2(T)$, and $\hat{X}_b^2(T)$ and compare the results to the trace statistic in Pesaran and Timmermann [2009]. To control for serial dependence I will use the analogous procedures as above to estimate the covariance matrices. The naive estimator is defined by

$$\begin{aligned}\hat{\mathbf{\Omega}}_n &= (\mathbf{D}_{\hat{p}_x} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}_x') \otimes (\mathbf{D}_{\hat{p}_y} - \otimes\hat{\mathbf{p}}_y\hat{\mathbf{p}}_y') \\ &+ \sum_{l=1}^{\tau} \tfrac{T-|l|}{T} \left(\hat{\mathbf{P}}_x^{(l)} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}_x'\right) \otimes \left(\hat{\mathbf{P}}_y^{(l)} - \hat{\mathbf{p}}_y\hat{\mathbf{p}}_y'\right) \\ &+ \sum_{l=1}^{\tau} \tfrac{T-|l|}{T} \left(\hat{\mathbf{P}}_x^{(l)\prime} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}_x'\right) \otimes \left(\hat{\mathbf{P}}_y^{(l)\prime} - \hat{\mathbf{p}}_y\hat{\mathbf{p}}_y'\right)\end{aligned}$$

with $\tau = 3$. For Markov chains I use that

$$\begin{aligned}\mathbf{\Omega} &= (\mathbf{D}_{p_x} - \mathbf{p}_x\mathbf{p}_x') \otimes (\mathbf{D}_{p_y} - \mathbf{p}_y\mathbf{p}_y') - 2(\mathbf{D}_{p_x} \otimes \mathbf{D}_{p_y}) \\ &+ (\mathbf{D}_{p_x} \otimes \mathbf{D}_{p_y})(\mathbf{I} - (\mathbf{Q}_x - \mathbf{1}_{m_x}\mathbf{p}_x') \otimes (\mathbf{Q}_y - \mathbf{1}_{m_y}\mathbf{p}_y'))^{-1} \\ &+ (\mathbf{I} - (\mathbf{Q}_x' - \mathbf{p}_x\mathbf{1}_{m_x}') \otimes (\mathbf{Q}_y' - \mathbf{p}_y\mathbf{1}_{m_y}'))^{-1}(\mathbf{D}_{p_x} \otimes \mathbf{D}_{p_y}).\end{aligned}$$

Let $\hat{\mathbf{\Omega}}_M$ be the estimator that is obtained by plugging in $\hat{p}_x$, $\hat{p}_y$, $\hat{\mathbf{Q}}_x$, and $\hat{\mathbf{Q}}_y$.

To illustrate size and power of the above tests I use the same process as Pesaran and Timmermann [2009]. Let

$$x_t = \phi x_{t-1} + \nu_t \text{ and } y_t = \phi y_{t-1} + \epsilon_t$$

with $\nu_t = r_{xy}\epsilon_t + \sqrt{1 - r_{xy}^2}\eta_t$. Both, η_t and ϵ_t are i.i.d. standard normal variables. The cross-correlation of the increments r_{xy} is set to 0 for the size and to 0.2 for the power simulations. The simulated data is categorized into $m = m_x = m_y$ equally probable bins.[5] The sample sizes are $T = 20, 50, 100, 500$, and 1000. Table 1 displays the simulation results for $\phi = 0$ and $r_{xy} = 0$ at a nominal level of 5%.[6] By and large the tests have the right size. Yet, there are some notable exceptions. W_n and $\hat{X}_b^2$ have a poor performance for small samples.

[5] In the simulations I use the correct bounds for the bins.

[6] I use the critical values from the asymptotic distribution. The results for estimated small sample critical values can be found in Appendix E.

m	T	X^2	X_n^2	X_M^2	X_{tav}^2	W_M	W_n	$\hat{X}_a^2$	$\hat{X}_b^2$	Trace
2	20	0.056	0.054	0.052	0.054	0.052	0.028	0.054	0.029	0.064
2	50	0.055	0.055	0.056	0.056	0.056	0.055	0.056	0.049	0.052
2	100	0.051	0.048	0.049	0.049	0.049	0.048	0.049	0.046	0.047
2	500	0.058	0.057	0.058	0.058	0.058	0.057	0.058	0.058	0.058
2	1000	0.044	0.045	0.044	0.044	0.044	0.045	0.044	0.046	0.046
3	20	0.050	0.047	0.045	0.049	0.043	0.011	0.044	0.020	0.095
3	50	0.051	0.048	0.048	0.049	0.050	0.040	0.048	0.052	0.072
3	100	0.049	0.048	0.049	0.049	0.049	0.047	0.050	0.049	0.053
3	500	0.046	0.047	0.046	0.046	0.046	0.046	0.046	0.045	0.046
3	1000	0.050	0.050	0.050	0.050	0.050	0.051	0.050	0.047	0.047
4	20	0.040	0.037	0.047	0.043	0.031	0.006	0.040	0.029	0.038
4	50	0.028	0.024	0.024	0.026	0.029	0.027	0.028	0.029	0.071
4	100	0.048	0.047	0.047	0.048	0.049	0.048	0.049	0.048	0.064
4	500	0.040	0.039	0.040	0.040	0.040	0.040	0.040	0.038	0.045
4	1000	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.048	0.049

Table 1: Size of independence tests: level 5%, no serial correlation ($\psi = 0$), no cross-correlation ($r_{xy} = 0$), 1000 simulations

The case with serial correlation ($\phi = 0.8$) but no cross-correlation is summarized in Table 2. The rejection rates of the standard Pearson test differ substantially from 5%. The rejection rates for almost all tests increase with sample size. For very small samples ($T = 20$) $\hat{X}_b^2$ and W_n have a rather poor performance .

If the two series are correlated with $r_{xy} = 0.2$, the power of all tests is on a similar level as the power of the standard Pearson test. The power of the tests is a little bit higher without serial correlation (Table 3) than with serial correlation (Table 4).

5.3 Conditional Independence

In a first simulation I compare CMH with CMH^*. For $m_x = m_y = m_z = 2$ I draw independently $\mathbf{r}_y^1$, $\mathbf{r}_z^1$, $\mathbf{r}_y^2$, $\mathbf{r}_z^2$, and $\mathbf{p}_x$. Set $\mathbf{p} = (p_{x,1}\mathbf{r}_y^1 \otimes \mathbf{r}_z^1, p_{x,2}\mathbf{r}_y^2 \otimes \mathbf{r}_z^2)'$. In this case Y_t and Z_t are independent conditional on X_t. For 100 different values of $\mathbf{p}$ I simulated 1000 series of 100 observations. The results are given

m	T	X^2	X_n^2	X_M^2	X_{tav}^2	W_M	W_n	$\hat{X}_a^2$	$\hat{X}_b^2$	Trace
2	20	0.151	0.059	0.073	0.063	0.044	0.014	0.063	0.025	0.061
2	50	0.202	0.073	0.077	0.077	0.077	0.044	0.077	0.060	0.067
2	100	0.251	0.072	0.089	0.088	0.090	0.069	0.088	0.057	0.061
2	500	0.210	0.053	0.071	0.071	0.071	0.053	0.071	0.052	0.052
2	1000	0.248	0.059	0.080	0.080	0.080	0.059	0.080	0.073	0.074
3	20	0.128	0.059	0.089	0.047	0.028	0.004	0.034	0.021	0.026
3	50	0.247	0.072	0.067	0.066	0.056	0.032	0.052	0.051	0.050
3	100	0.280	0.079	0.086	0.086	0.071	0.061	0.069	0.081	0.088
3	500	0.297	0.072	0.093	0.093	0.071	0.048	0.070	0.060	0.060
3	1000	0.308	0.074	0.089	0.089	0.082	0.058	0.082	0.051	0.051
4	20	0.093	0.047	0.093	0.029	0.014	0.003	0.020	0.011	0.004
4	50	0.194	0.059	0.053	0.048	0.041	0.024	0.038	0.048	0.013
4	100	0.263	0.080	0.073	0.072	0.061	0.060	0.056	0.046	0.016
4	500	0.292	0.081	0.091	0.091	0.057	0.041	0.057	0.048	0.043
4	1000	0.310	0.079	0.096	0.096	0.071	0.055	0.070	0.051	0.051

Table 2: Size of independence tests: nominal level 5%, serial correlation ($\psi = 0.8$), no cross-correlation ($r_{xy} = 0$), 1000 simulations

m	T	X^2	X_n^2	X_M^2	X_{tav}^2	W_M	W_n	$\hat{X}_a^2$	$\hat{X}_b^2$	Trace
2	20	0.089	0.080	0.083	0.082	0.083	0.036	0.082	0.047	0.100
2	50	0.164	0.154	0.160	0.161	0.160	0.153	0.161	0.144	0.156
2	100	0.252	0.245	0.249	0.249	0.249	0.245	0.249	0.235	0.240
2	500	0.823	0.821	0.824	0.824	0.824	0.821	0.824	0.814	0.814
2	1000	0.983	0.983	0.983	0.983	0.983	0.983	0.983	0.983	0.983
3	20	0.073	0.065	0.076	0.072	0.069	0.010	0.073	0.027	0.115
3	50	0.114	0.109	0.112	0.109	0.117	0.107	0.114	0.108	0.130
3	100	0.197	0.196	0.196	0.195	0.198	0.192	0.198	0.189	0.211
3	500	0.819	0.821	0.820	0.820	0.818	0.818	0.817	0.819	0.827
3	1000	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990
4	20	0.062	0.053	0.059	0.057	0.056	0.007	0.060	0.028	0.050
4	50	0.081	0.079	0.082	0.080	0.080	0.079	0.081	0.068	0.136
4	100	0.146	0.147	0.145	0.145	0.146	0.143	0.146	0.145	0.185
4	500	0.786	0.786	0.786	0.786	0.787	0.789	0.788	0.783	0.793
4	1000	0.986	0.986	0.986	0.986	0.986	0.986	0.986	0.987	0.987

Table 3: Power of independence tests: nominal level 5%, no serial correlation ($\psi = 0$), cross-correlation $r_{xy} = 0.2$, 1000 simulations

m	T	X^2	X^2_n	X^2_M	X^2_{tav}	W_M	W_n	$\hat{X}^2_a$	$\hat{X}^2_b$	Trace
2	20	0.178	0.083	0.085	0.076	0.053	0.020	0.075	0.035	0.066
2	50	0.284	0.107	0.131	0.129	0.131	0.066	0.129	0.092	0.097
2	100	0.387	0.170	0.197	0.199	0.197	0.165	0.199	0.129	0.134
2	500	0.722	0.456	0.507	0.507	0.507	0.456	0.507	0.442	0.445
2	1000	0.906	0.744	0.782	0.782	0.782	0.744	0.782	0.717	0.717
3	20	0.153	0.074	0.104	0.060	0.034	0.008	0.042	0.026	0.032
3	50	0.293	0.111	0.107	0.105	0.073	0.040	0.074	0.083	0.072
3	100	0.405	0.157	0.177	0.176	0.112	0.092	0.110	0.133	0.140
3	500	0.747	0.510	0.545	0.545	0.400	0.356	0.401	0.498	0.505
3	1000	0.922	0.779	0.799	0.799	0.670	0.632	0.670	0.846	0.849
4	20	0.102	0.056	0.090	0.038	0.022	0.009	0.030	0.013	0.004
4	50	0.264	0.105	0.085	0.078	0.060	0.039	0.061	0.071	0.018
4	100	0.398	0.150	0.142	0.141	0.080	0.076	0.081	0.116	0.035
4	500	0.761	0.485	0.504	0.504	0.298	0.269	0.296	0.513	0.452
4	1000	0.917	0.782	0.809	0.809	0.538	0.507	0.537	0.853	0.844

Table 4: Power of independence tests: nominal level 5%, serial correlation ($\psi = 0.8$), cross-correlation $r_{xy} = 0.2$, 1000 simulations

in Figure 4. Both tests perform reasonably well.

Now, assume that $\mathbf{p} = (1/4)(a, 1-a, 1-a, a, 1-a, a, a, 1-a)'$. It is easy to verify that in this case p_{ijk} does not equal $p_{ij+}p_{i+k}/p_{i++}$ unless $a = 1/2$. Y_t and Z_t are not independent conditionally on X_t. But $p_{+jk} = p_{1j+}p_{1+k}/p_{1++} + p_{2j+}p_{2+k}/p_{2++}$. Figure 5 displays the rejection rates of both tests in relation to a which ranges from 0 to 0.5 in steps of 0.01. Not surprisingly, CMH has no power against these kind of processes.

Finally, I draw randomly 100 values for $\mathbf{p}$ to illustrate the power of both tests. The rejection rates are plotted against the distance between $\mathbf{p}$ and $\mathbf{p}^0 = (p_{1++}\mathbf{r}^1_y \otimes \mathbf{r}^1_z, p_{2++}\mathbf{r}^2_y \otimes \mathbf{r}^2_z)'$. The performance of CMH^* is considerably better than that of CMH as illustrated in Figure 6.

We may use CMH and CMH^* to test whether a series is Markov of order one against the alternative of a higher order. Suppose that $\{X_t\}$ is a Markov chain of order 2 with two states. Define $\tilde{X}_t = X_{t-1} \otimes X_t$. The transition

Figure 4: Rejection rates of CMH and CMH^* at a nominal level of 10% for a conditionally independent process. The results are based on 1000 simulations. The number of observations in each simulation equals 100.

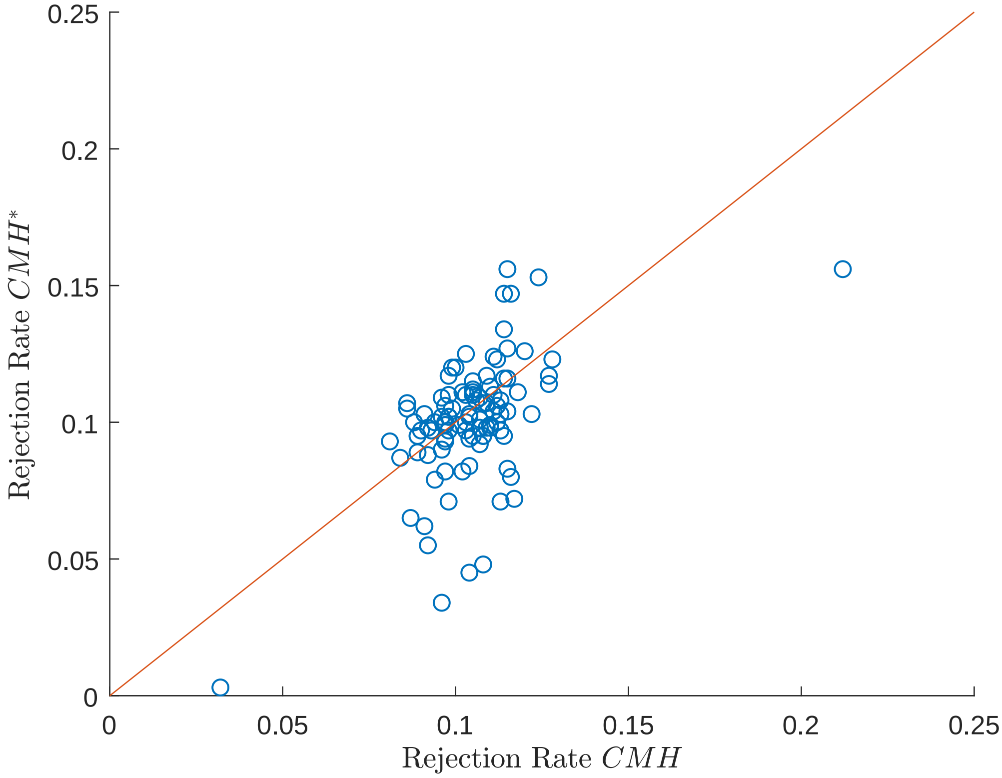


Figure 5: Rejection rates at a nominal level of 10%. $p_{+jk} = \sum_i p_{ij+}p_{i+k}/p_{i++}$ but $p_{ijk} \neq p_{ij+}p_{i+k}/p_{i++}$, i.e. Y_t and Z_t are not independent conditional on X_t. The results are based on 1000 simulations. The number of observations in each simulation equals 100.

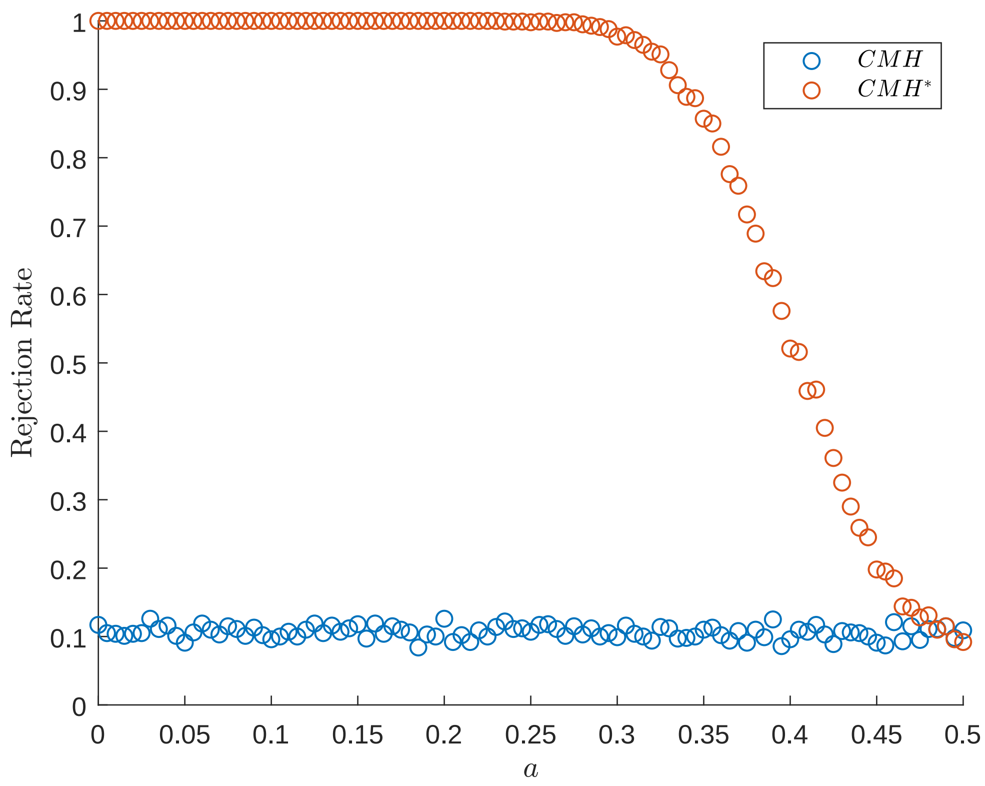
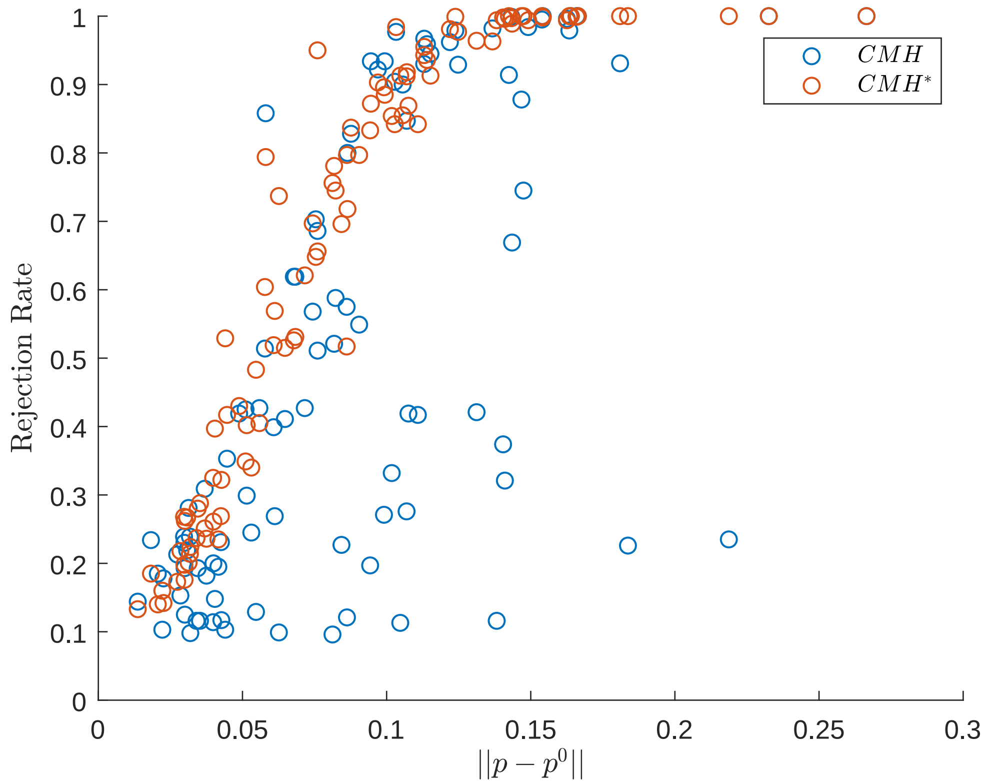


Figure 6: Rejection rates at a nominal level of 10%. One hundred **p** are randomly drawn. The results are based on 1000 simulations for each **p**. The number of observations in each simulation equals 100.



matrix of $\{\tilde{X}_t\}$ is given by

$$\mathbf{Q}_{\tilde{X}} = \begin{pmatrix} a & 1-a & 0 & 0 \\ 0 & 0 & b & 1-b \\ c & 1-c & 0 & 0 \\ 0 & 0 & d & 1-d \end{pmatrix}.$$

If $\{X_t\}$ is a simple Markov chain then $c = a$ and $d = b$. To illustrate the power of CMH and CMH^* I set $b = c = 1-a$ and $d = a$, i.e.

$$\mathbf{Q}_{\tilde{X}} = \begin{pmatrix} a & 1-a & 0 & 0 \\ 0 & 0 & 1-a & a \\ 1-a & a & 0 & 0 \\ 0 & 0 & a & 1-a \end{pmatrix}.$$

In this case $\{X_t\}$ is a simple Markov chain if and only if $a = 0.5$. Figure 7 displays the simulation result for a ranging from 0.1 to 0.9 in steps of 0.01. CMH^* performs very well compared to CMH.

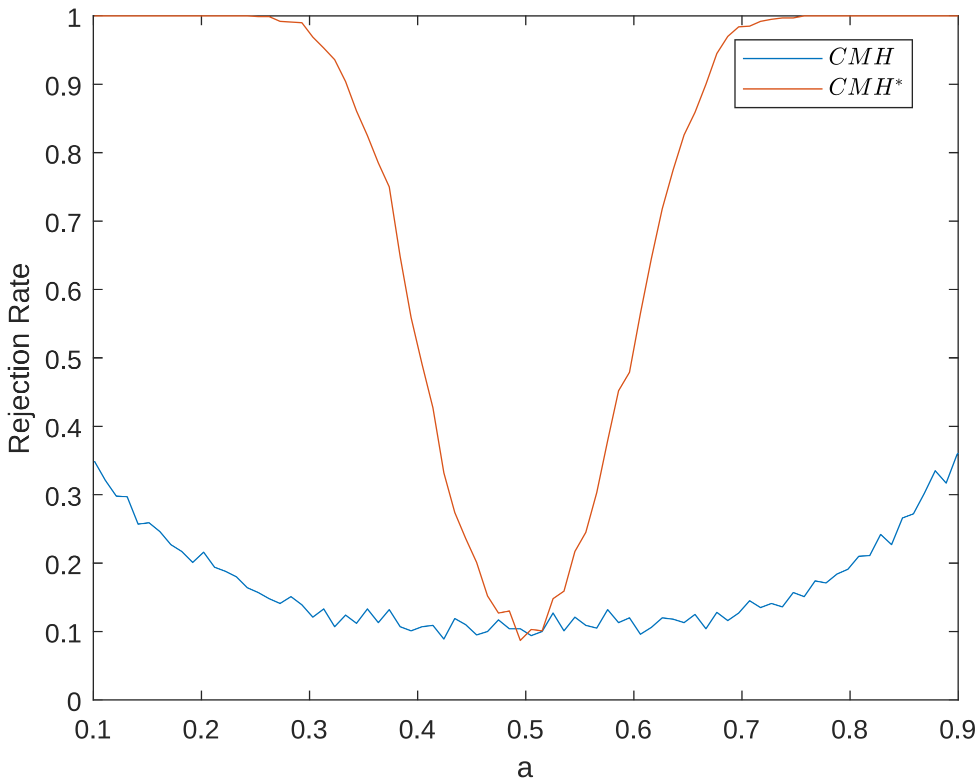
6 Conclusion

Based on the assumption that a central limit theorem holds I analyze the consequences of serial correlation on the distribution of Pearson statistics. Simple adjustments that can easily be implemented suffice to correct for the serial dependence in the data. Simulations illustrate that the loss of power caused by this adjustments in cases without serial correlation is small.

The results in Tavaré [1983] and Porteous [1987] can be generalized to non reversible processes. For independence tests in two-way tables an alternative test using filtered observations is proposed. If one of the two series is a Markov chain, the test has a simple asymptotic distribution and performs well in simulations.

The classical CMH test for conditional independence is discussed in detail. It is shown that CMH tests for a rather weak implication of conditional independence. A minor modification suffices to get a statistic that tests for

Figure 7: Power of CMH and CMH^* for a two-state Markov chain of order two. The rejection rates are based on a nominal level of 10%. The results are based on 1000 simulations for each a. The number of observations in each simulation equals 100.



a larger class of alternatives without losing power in those cases for which CMH is tailor-made. Finally, simulations indicate that the variant of CMH performs very well as a test for Markovity of order one.

References

A. Agresti. *Categorical Data Analysis.* John Wiley & Sons, second edition, 2002.

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

C. Chou and C.-S. J. Chu. Testing independence of two autocorrelated binary time series. *Statistics and Probability Letters*, 80:69–75, 2010.

L. J. Gleser and D. S. Moore. The effect of positive dependence on chi-squared tests for categorical data. *Journal of the Royal Statistical Society. Series B.*, 47(3):459–465, 1985.

D. Holt, A. J. Scott, and P. D. Ewings. Chi-squared tests with survey data. *Journal of the Royal Statistical Society. Series A (General)*, 143(3):303–320, 1980.

S. Kullback, M. Kupperman, and H. H. Ku. Tests for contingency tables and markov chains. *Technometrics*, 4(4):573–608, 1962.

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

M. H. Pesaran and A. Timmermann. Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337, 2009.

B. T. Porteous. The mutual independence hypothesis for categorical data in complex sampling schemes. *Biometrika*, 74(4):857–862, 1987.

C. R. Rao and S. K. Mitra. Generalized inverse of a matrix and its applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pages 601–620, Berkeley, Calif., 1972. University of California Press. URL `https://projecteuclid.org/euclid.bsmsp/1200514113`.

J. N. K. Rao and A. J. Scott. Chisquared tests for analysis of categorical data from complex surveys. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, 1979.

J. N. K. Rao and A. J. Scott. The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, 76(374): 221 – 230, June 1981.

J. N. K. Rao and A. J. Scott. On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics*, 12(1):46–60, Mar. 1984.

J. N. K. Rao and A. J. Scott. On simple adjustments to chi-square tests with sample survey data. *The Annals of Statistics*, 15(1):385–397, 03 1987. doi: 10.1214/aos/1176350273. URL `https://doi.org/10.1214/aos/1176350273`.

R. L. Sandland. Application of methods of testing for independence between two markov chains. *Biometrics*, 32(3):629–636, 1976.

A. R. Solow, W. K. Smith, and C. Recchia. A conditional test of independence of markov chains. *Biometrical Journal*, 37(8):973–977, 1995.

S. Tavaré. Serial dependence in contingency tables. *Journal of the Royal Statistical Society B*, 45(1):100–106, 1983.

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

K.-H. Yuan and P. M. Bentler. Two simple approximations to the distributions of quadratic forms. *The British journal of mathematical and statistical psychology*, 63(2):273–291, 2010.

A m-dependent Process

Let $\{X_t\}$ be a sequence of independent random variables which are distributed uniformly on the unit interval. Define $U_t = 1$ if $X_{t-1} \leq aX_t$ and 2 otherwise. $\{U_t\}$ is a 1-dependent process. For $a \in (0, 1]$ we get $p = (a/2, 1 - a/2)'$,

$$\mathbf{Q}^{(1)} = \begin{pmatrix} \frac{a^2}{3} & \frac{3-a^2}{3} \\ \frac{3a-a^3}{6-3a} & \frac{6-6a+a^3}{6-3a} \end{pmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \frac{a(6 - 9a + 4a^2)}{12} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

The non-zero eigenvalue of $\mathbf{D}_p^{-1}\mathbf{\Sigma}$ equals $(6 - 9a + 4a^2)/(6 - 3a)$. Hence, the asymptotic distribution of $P(T)$ is not χ_1^2 but $\frac{6-9a+4a^2}{6-3a}\chi_1^2$. Figure 8 illustrates the rejection rates at a 10% level for $a \in (0, 1]$ if the critical value from the χ_1^2 distribution is used. Let $U_t^* = \sum_{i=1}^{m_u} iU_{i,t}$. The correlation of U_{t-1}^* and U_t^* equals $(2a^2 - 3a)/(6 - 3a)$. Moderate levels of autocorrelations lead to sizable deviations from the 10% level.

Figure 9 illustrates the performance of the tests specified in subsection 5.1. The naive estimation of $\mathbf{\Sigma}$ and $\bar{\rho}$ with $\tau = 3$ works well in this environment as long as a is larger than 0.2

Figure 8: The rejection rate at a nominal level of 10% for a 1-dependent process. The blue line is the theoretical rejection rate. The red line is based on 1000 simulations for $a = 0.01$ to $a = 0.99$ in steps of 0.01. The number of observations in each simulation equals 200.

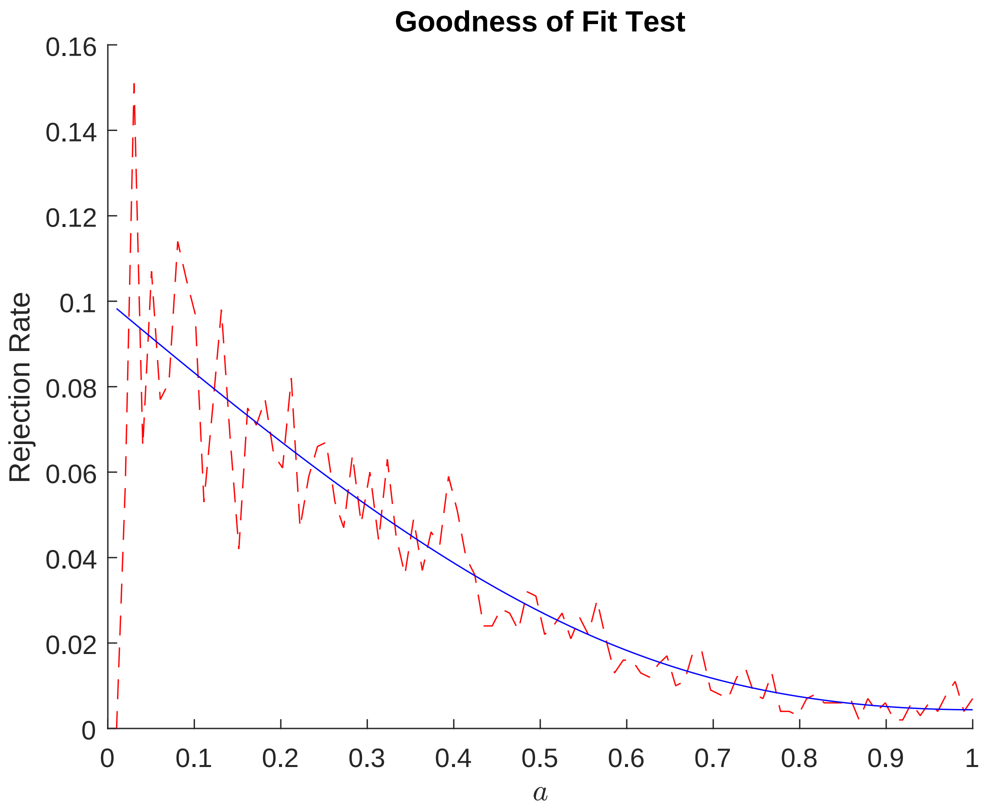
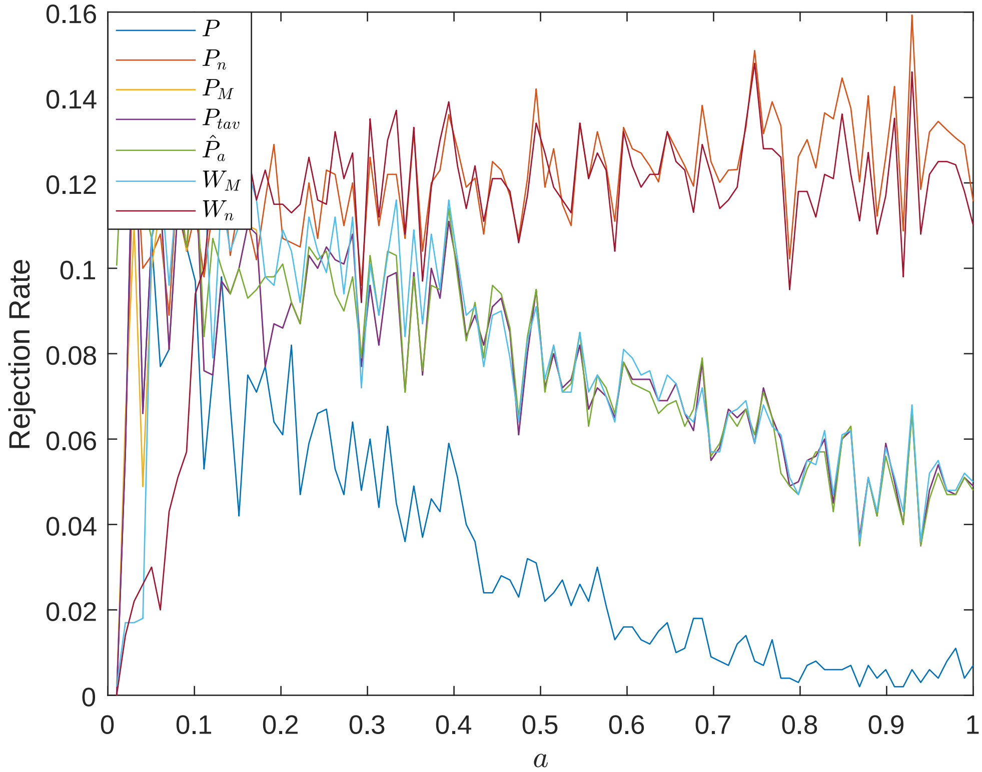


Figure 9: Rejection rates at a nominal level of 10% for a 1-dependent process. The results are based on 1000 simulations for $a = 0.01$ to $a = 0.99$ in steps of 0.01. The number of observations in each simulation equals 200.



B Pesaran and Timmermann's trace statistic

Pesaran and Timmermann [2009] propose new procedures to test for independence of two or more stationary and ergodic Markov chains based on canonical correlations. To be more specific, consider the category variables $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{m_x,t})'$ with $X_{i,t} = 1$ if category i occurs at time t and $X_{i,t} = 0$ otherwise (analogously for $Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{m_x,t})'$). To avoid multicollinearity in the regression framework Pesaran and Timmermann [2009] delete the last entries of X_t and Y_t. I denote these vectors by X_t^0 and Y_t^0, respectively. The observations of X_t are summarized in the matrix $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_T})'$ (analogously for X_t^0, Y_t, and Y_t^0). Pesaran and Timmermann [2009] show that under independence

$$Ttr(\mathbf{S}^{-1}_{y^0y^0,w^0}\mathbf{S}_{y^0x^0,w^0}\mathbf{S}^{-1}_{x^0x^0,w^0}\mathbf{S}_{x^0y^0,w^0}) \sim^a \chi^2_{(m_x-1)(m_y-1)}$$

with $\mathbf{M}_{w^0} = \mathbf{I}_T - \mathbf{W^0}(\mathbf{W^0}'\mathbf{W^0})^{-1}\mathbf{W^0}'$, $\mathbf{S}_{y^0y^0,w^0} = T^{-1}\mathbf{Y^0}'\mathbf{M}_{w^0}\mathbf{Y^0}$, $\mathbf{S}_{x^0x^0,w^0} = T^{-1}\mathbf{X^0}'\mathbf{M}_{w^0}\mathbf{X^0}$, and $\mathbf{S}_{x^0y^0,w^0} = T^{-1}\mathbf{X^0}'\mathbf{M}_{w^0}\mathbf{Y^0} = \mathbf{S}'_{y^0x^0,w^0}$. In the case of no serial correlation $\mathbf{W}^0 = \mathbf{1}_T$. For simple Markov chains $\mathbf{W}^0 = (\mathbf{X}^0_{-1}, \mathbf{Y}^0_{-1}, \mathbf{1}_T)$. For notational convenience I follow Pesaran and Timmermann [2009] and assume that we do have T observations of the lagged variables. $\mathbf{X}^0_{-1}$ denotes the $T \times (m_x - 1)$ matrix of observations on X^0_{t-1}. All simulations in the main text are based on a fixed number of T observations. The statistic has to be adjusted in the obvious way.

To get a clearer understanding of the relation between contingency tables and the above trace statistic I show that we may use X_t and Y_t instead of X_t^0 and Y_t^0 if we substitute generalized inverses for inverse matrices. This would be only a minor contribution. Yet, this restatement of the theorems in Pesaran and Timmermann [2009] allows a more comprehensible representation of the trace statistic.

In the above definitions of $\mathbf{M}_{w^0}$ we used that $\mathbf{W^0}'\mathbf{W^0}$ is invertible. Now, suppose that this is not the case and define $\mathbf{M}^*_w = \mathbf{I}_T - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}'$ for some g inverse of $\mathbf{W}'\mathbf{W}$.

Lemma 1. *Let* $\mathbf{W}^0$ *be an* $m \times k$ *matrix of full column rank and* $\mathbf{A}$ *be a* $k \times n$ *matrix of full row rank. For* $\mathbf{W} = \mathbf{W}^0\mathbf{A}$ *it holds that (a)* $\mathbf{W}(\mathbf{W}'\mathbf{W})^-\mathbf{W}' = \mathbf{W}^0(\mathbf{W}^{0\prime}\mathbf{W}^0)^{-1}\mathbf{W}^{0\prime}$*, (b)* $\mathbf{W}(\mathbf{W}'\mathbf{W})^-\mathbf{W}'\mathbf{W} = \mathbf{W}$*, and (c)* $\mathbf{W}'\mathbf{W}(\mathbf{W}'\mathbf{W})^-\mathbf{W}' = \mathbf{W}'$ *for any g inverse of* $\mathbf{W}'\mathbf{W}$.

Proof. Rao and Mitra [1972] show in Theorem 2.4 that $\mathbf{W}(\mathbf{W}'\mathbf{W})^-\mathbf{W}'$ is invariant for any choice of $(\mathbf{W}'\mathbf{W})^-$, in particular it equals $\mathbf{W}(\mathbf{W}'\mathbf{W})^+\mathbf{W}'$. It is straightforward to check that

$$(\mathbf{W}'\mathbf{W})^+ = (\mathbf{A}'\mathbf{W}^{0\prime}\mathbf{W}^0\mathbf{A})^+ = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}(\mathbf{W}^{0\prime}\mathbf{W}^0)^{-1}(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}.$$

Plugging this in yields $\mathbf{W}(\mathbf{W}'\mathbf{W})^+\mathbf{W}' = \mathbf{W}^0(\mathbf{W}^{0\prime}\mathbf{W}^0)^{-1}\mathbf{W}^{0\prime}$. (b) and (c) are proved in Theorem 2.4 of Rao and Mitra [1972]. □

Lemma 1 implies that $\mathbf{M}_w^*$ is invariant with respect to the chosen g inverse and that $\mathbf{M}_w^* = \mathbf{M}_{w^0}$ provided that $\mathbf{W} = \mathbf{W}^0\mathbf{A}$ holds. Under this assumption $\mathbf{S}_{yy,w}^* = T^{-1}\mathbf{Y}'\mathbf{M}_w^*\mathbf{Y} = T^{-1}\mathbf{Y}'\mathbf{M}_{w^0}\mathbf{Y} = \mathbf{S}_{yy,w^0}$. The same is true for $\mathbf{S}_{xx,w}^*$ and $\mathbf{S}_{xy,w}^*$.

Lemma 2. *Suppose that* $\mathbf{S}_{y^0y^0,w^0}^{-1}$*,* $\mathbf{S}_{x^0x^0,w^0}^{-1}$*, and* $(\mathbf{W}^{0\prime}\mathbf{W}^0)^{-1}$ *exist. If* $\mathbf{1}_T$ *is in the column space of* $\mathbf{W}^\mathbf{0}$ *and* $\mathbf{W} = \mathbf{W}^0\mathbf{A}$ *where* $\mathbf{A}$ *is of full row rank then* $tr(\mathbf{S}_{y^0y^0,w^0}^{-1}\mathbf{S}_{y^0x^0,w^0}\mathbf{S}_{x^0x^0,w^0}^{-1}\mathbf{S}_{x^0y^0,w^0}) = tr((\mathbf{S}_{yy,w}^*)^+\mathbf{S}_{yx,w}^*(\mathbf{S}_{xx,w}^*)^+\mathbf{S}_{xy,w}^*)$.

Proof. $\mathbf{Y}$ may be expressed as $\mathbf{Y} = (\mathbf{Y}^0, \mathbf{1}_T - \mathbf{Y}^0\mathbf{1}_{m_y-1})$. Plugging this in yields $\mathbf{M}_{w^0}(\mathbf{Y}^0, \mathbf{1}_T - \mathbf{Y}^0\mathbf{1}_{m_y-1}) = \mathbf{M}_{w^0}\mathbf{Y}^0(\mathbf{I}_{m_y-1}, -\mathbf{1}_{m_y-1})$. The last equality follows from the fact that $\mathbf{1}_T$ is in the column space of $\mathbf{W}^\mathbf{0}$. Consequently,

$$\mathbf{Y}'\mathbf{M}_w^*\mathbf{Y} = \begin{pmatrix} \mathbf{I}_{m_y-1} \\ -\mathbf{1}_{m_y-1}' \end{pmatrix} \mathbf{Y}^{\mathbf{0}\prime}\mathbf{M}_{w^0}\mathbf{Y}^\mathbf{0} \left(\mathbf{I}_{m_y-1}, -\mathbf{1}_{m_y-1}\right)$$

and

$$(\mathbf{Y}'\mathbf{M}_w^*\mathbf{Y})^+ = \begin{pmatrix} \mathbf{I} \\ -\mathbf{1}' \end{pmatrix} \left(\mathbf{I} + \mathbf{1}\mathbf{1}'\right)^{-1} \left(\mathbf{Y}^{\mathbf{0}\prime}\mathbf{M}_{w^0}\mathbf{Y}^\mathbf{0}\right)^{-1} \left(\mathbf{I} + \mathbf{1}\mathbf{1}'\right)^{-1} \quad (\mathbf{I}, -\mathbf{1})$$

where I dropped the subscript of $\mathbf{I}_{m_y-1}$ and $\mathbf{1}_{m_y-1}$. We get analogous results for $\mathbf{S}^*_{yx,w}$, $(\mathbf{S}^*_{xx,w})^+$, and $\mathbf{S}^*_{xy,w}$. Plugging this in yields

$$(\mathbf{S}^*_{yy,w})^+\mathbf{S}^*_{yx,w}(\mathbf{S}^*_{xx,w})^+\mathbf{S}^*_{xy,w} = \begin{pmatrix} \mathbf{I} \\ -\mathbf{1}' \end{pmatrix} \left(\mathbf{I}+\mathbf{1}\mathbf{1}'\right)^{-1} \mathbf{S}^{-1}_{y^0y^0,w^0}\mathbf{S}_{y^0x^0,w^0} \times \mathbf{S}^{-1}_{x^0x^0,w^0}\mathbf{S}_{x^0y^0,w^0}\left(\mathbf{I},-\mathbf{1}\right).$$

The trace is invariant under cyclic permutations. Hence,

$$tr(\mathbf{S}^{-1}_{y^0y^0,w^0}\mathbf{S}_{y^0x^0,w^0}\mathbf{S}^{-1}_{x^0x^0,w^0}\mathbf{S}_{x^0y^0,w^0}) = tr((\mathbf{S}^*_{yy,w})^+\mathbf{S}^*_{yx,w}(\mathbf{S}^*_{xx,w})^+\mathbf{S}^*_{xy,w})$$

□

Given the last result it is understood that from now on the projection matrix $\mathbf{M}^*_w$ is used instead of $\mathbf{M}_w$ and the asterisk is dropped from $\mathbf{M}^*_w$, $\mathbf{S}^*_{yy,w}$, $\mathbf{S}^*_{yx,w}$, $\mathbf{S}^*_{xx,w}$, and $\mathbf{S}^*_{xy,w}$.

The next step is to verify that the trace does not depend on using the Moore-Penrose inverse, i.e.

$$tr(\mathbf{S}^+_{yy,w}\mathbf{S}_{yx,w}\mathbf{S}^+_{xx,w}\mathbf{S}_{xy,w}) = tr(\mathbf{S}^-_{yy,w}\mathbf{S}_{yx,w}\mathbf{S}^-_{xx,w}\mathbf{S}_{xy,w})$$

for arbitrary g inverses of $\mathbf{S}_{yy,w}$ and $\mathbf{S}_{xx,w}$. The trace is rotation invariant, $tr(A'B) = vec(A)'vec(B)$, and $vec(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})vec(\mathbf{B})$. We get the desired result if

$$vec(\mathbf{S}_{yx,w})'\left(\mathbf{S}^+_{xx,w} \otimes \mathbf{S}^+_{yy,w}\right)vec(\mathbf{S}_{yx,w}) = vec(\mathbf{S}_{yx,w})'\left(\mathbf{S}^-_{xx,w} \otimes \mathbf{S}^-_{yy,w}\right)vec(\mathbf{S}_{yx,w}).$$

Lemma 3. *For $v \in \mathbb{R}^m$ in the column span of the symmetric $m \times m$ matrix Σ the quadratic form $v'\Sigma^- v$ is invariant with respect to the particular g inverse chosen.*

Proof. If v is in the column span then there exists a vector $u \in \mathbb{R}^m$ such that $v = \Sigma u$. Hence, $v'\Sigma^- v = u'\Sigma'\Sigma^-\Sigma u = u'\Sigma\Sigma^-\Sigma u = u'\Sigma u$ for any g inverse. □

The trace is therefore invariant with respect to the chosen g inverse if

$vec(\mathbf{S}_{yx,w})$ is in the column span of $\mathbf{S}_{xx,w} \otimes \mathbf{S}_{yy,w}$. Observe that

$$\mathbf{S}_{yy,w}\mathbf{S}^{+}_{yy,w}\mathbf{S}_{yx,w}\mathbf{S}^{+}_{xx,w}\mathbf{S}_{xx,w} = \mathbf{S}_{yx,w}.$$

This corresponds to

$$(\mathbf{S}_{xx,w} \otimes \mathbf{S}_{yy,w})vec(\mathbf{S}^{+}_{yy,w}\mathbf{S}_{yx,w}\mathbf{S}^{+}_{xx,w}) = vec(\mathbf{S}_{yx,w}).$$

$vec(\mathbf{S}_{yx,w})$ is in the column span of $\mathbf{S}_{xx,w} \otimes \mathbf{S}_{yy,w}$. Lemma 2 can be generalized.

Lemma 4. *Suppose that* $\mathbf{S}^{-1}_{y^0y^0,w^0}$, $\mathbf{S}^{-1}_{x^0x^0,w^0}$, *and* $(\mathbf{W}^{0\prime}\mathbf{W}^0)^{-1}$ *exist. If* $\mathbf{1}_T$ *is in the column span of* $\mathbf{W^0}$ *and* $\mathbf{W} = \mathbf{W}^0\mathbf{A}$ *where* $\mathbf{A}$ *is of full row rank then* $tr(\mathbf{S}^{-1}_{y^0y^0,w^0}\mathbf{S}_{y^0x^0,w^0}\mathbf{S}^{-1}_{x^0x^0,w^0}\mathbf{S}_{x^0y^0,w^0}) = tr(\mathbf{S}^{-}_{yy,w}\mathbf{S}_{yx,w}\mathbf{S}^{-}_{xx}\mathbf{S}_{xy,w})$ *for arbitrary g inverses* $\mathbf{S}^{-}_{yy,w}$ *and* $\mathbf{S}^{-}_{xx,w}$.

Remark 4. *If either* $\mathbf{X}$ *or* $\mathbf{Y}$ *are in the column span of* $\mathbf{W}$*, the trace is identical to zero. This would be the case if for instance* $X_t = Y_{t-1}$.

B.1 Two-way Tables without Serial Correlation

In the case of no serial correlation $\mathbf{W} = \mathbf{W^0} = \mathbf{1_T}$. $\mathbf{1}_T$ is in the space spanned by $\mathbf{W^0}$ and $\mathbf{W}^{0\prime}\mathbf{W}^0 = T$ is invertible. Hence, Lemma 4 may be applied. Pesaran and Timmermann [2009] showed that in this case the trace statistic is equivalent to the standard χ^2 test for two-way tables. Using the new notation this result can be derived quite easily.

Lemma 5. *In the case without serial correlation, i.e.* $\mathbf{W} = \mathbf{W^0} = \mathbf{1_T}$*, we get*

$$Ttr(\mathbf{S}^{-}_{yy,w}\mathbf{S}_{yx,w}\mathbf{S}^{-}_{xx,w}\mathbf{S}_{xy,w}) = Tvec(\hat{\mathbf{P}}_{xy} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_y)'(\hat{\mathbf{D}}^{-1}_y \otimes \hat{\mathbf{D}}^{-1}_x)vec(\hat{\mathbf{P}}_{xy} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_y)$$

where $\hat{\mathbf{P}}_{xy} = (1/T)\mathbf{X}'\mathbf{Y}$. *The trace may be rewritten as Pearson's chi-squared test for independence*

$$T\sum_{i=1}^{m_x}\sum_{j=1}^{m_y}\frac{(\hat{p}_{ij} - \hat{p}_{i+}\hat{p}_{+j})^2}{\hat{p}_{i+}\hat{p}_{+j}}.$$

Proof. $\mathbf{S}_{yy,w} = T^{-1}\mathbf{Y}'(\mathbf{I}-\mathbf{1}_T(\mathbf{1}_T'\mathbf{1}_T)^{-1}\mathbf{1}_T')\mathbf{Y} = \mathbf{D}_{\hat{p}_y} - \hat{\mathbf{p}}_y\hat{\mathbf{p}}_y'$ and $\mathbf{S}_{xx,w} = \mathbf{D}_{\hat{p}_x} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}_x'$. Analogously we get $\mathbf{S}_{yx,w} = T^{-1}\mathbf{Y}'(\mathbf{I}-\mathbf{1}_T(\mathbf{1}_T'\mathbf{1}_T)^{-1}\mathbf{1}_T')\mathbf{X} = \hat{\mathbf{P}}_{yx} - \hat{\mathbf{p}}_y\hat{\mathbf{p}}_x'$. Hence, $vec(\mathbf{S}_{yx,w}) = \hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y$. $\mathbf{D}_{\hat{p}_x}^{-1}$ is a g inverse of $\mathbf{D}_{\hat{p}_x} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}_x'$. Hence, we get

$$\begin{aligned}
& tr(\mathbf{S}_{yy,w}^{-}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-}\mathbf{S}_{xy,w}) \\
= \; & (\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y)'(\mathbf{D}_{\hat{p}_y}^{-1} \otimes \mathbf{D}_{\hat{p}_x}^{-1})(\hat{\mathbf{p}} - \hat{\mathbf{p}}_x \otimes \hat{\mathbf{p}}_y).
\end{aligned}$$

□

B.2 Two-way Tables with Serial Correlation

In this subsection I will show that the dynamically augmented trace statistic is asymptotically equivalent to X_b^2 as defined in the main text if both $\{X_t\}$ and $\{Y_t\}$ are simple, stationary, and ergodic Markov chains. In the dynamically augmented case $\mathbf{W}^0 = (\mathbf{X}_{-1}^0, \mathbf{Y}_{-1}^0, \mathbf{1}_T)$. $\mathbf{W} = (\mathbf{X}_{-1}, \mathbf{Y}_{-1})$ may be written as

$$\begin{aligned}
\mathbf{W} &= (\mathbf{X}_{-1}^0, \mathbf{1}_T - \mathbf{X}_{-1}^0\mathbf{1}_{m_x-1}, \mathbf{Y}_{-1}^0, \mathbf{1}_T - \mathbf{Y}_{-1}^0\mathbf{1}_{m_y-1}) \\
&= \mathbf{W}^0 \begin{pmatrix} \mathbf{I}_{m_x-1} & -\mathbf{1}_{m_x-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{m_y-1} & -\mathbf{1}_{m_y-1} \\ \mathbf{0} & 1 & \mathbf{0} & 1 \end{pmatrix}
\end{aligned}$$

The matrix in brackets is of full row rank and $\mathbf{1}_T$ is in the span of $\mathbf{W}^0$. We may apply Lemma 4 again.

$$\begin{aligned}
\mathbf{S}_{yy,w} &= \mathbf{D}_{\hat{p}_y} - (\hat{\mathbf{P}}_{yx_{-1}}, \hat{\mathbf{P}}_{yy_{-1}}) \begin{pmatrix} \mathbf{D}_{\hat{p}_{x_{-1}}} & \hat{\mathbf{P}}_{x_{-1}y_{-1}} \\ \hat{\mathbf{P}}_{y_{-1}x_{-1}} & \mathbf{D}_{\hat{p}_{y_{-1}}} \end{pmatrix}^{-} \begin{pmatrix} \hat{\mathbf{P}}_{x_{-1}y} \\ \hat{\mathbf{P}}_{y_{-1}y} \end{pmatrix} \\
&= \mathbf{D}_{p_y} - (\mathbf{p}_y\mathbf{p}_x', \mathbf{P}_y') \begin{pmatrix} \mathbf{D}_{p_x} & \mathbf{p}_x\mathbf{p}_y' \\ \mathbf{p}_y\mathbf{p}_x' & \mathbf{D}_{p_y} \end{pmatrix}^{-} \begin{pmatrix} \mathbf{p}_x\mathbf{p}_y' \\ \mathbf{P}_y \end{pmatrix} + O_p(T^{-1/2})
\end{aligned}$$

with $\hat{\mathbf{P}}_{yx_{-1}} = (1/T)\mathbf{Y}'\mathbf{X}_{-1} = \hat{\mathbf{P}}_{x_{-1}y}'$, $\hat{\mathbf{P}}_{yy_{-1}} = (1/T)\mathbf{Y}'\mathbf{Y}_{-1} = \hat{\mathbf{P}}_{y_{-1}y}'$, $\mathbf{D}_{\hat{p}_{x_{-1}}} = (1/T)\mathbf{X}_{-1}'\mathbf{X}_{-1}$, $\mathbf{D}_{\hat{p}_{y_{-1}}} = (1/T)\mathbf{Y}_{-1}'\mathbf{Y}_{-1}$, and $\hat{\mathbf{P}}_{y_{-1}x_{-1}} = (1/T)\mathbf{Y}_{-1}'\mathbf{X}_{-1} = \hat{\mathbf{P}}_{x_{-1}y_{-1}}'$. For stationary and ergodic Markov chains these estimators are consistent.[7] The second equality is then a consequence of the null hypothesis of independence. Using that

$$\begin{pmatrix} \mathbf{D}_{p_x}^{-1} - \frac{3}{4}\mathbf{1}_{m_x}\mathbf{1}'_{m_x} & \frac{1}{4}\mathbf{1}_{m_x}\mathbf{1}'_{m_y} \\ \frac{1}{4}\mathbf{1}_{m_y}\mathbf{1}'_{m_x} & \mathbf{D}_{p_y}^{-1} - \frac{3}{4}\mathbf{1}_{m_y}\mathbf{1}'_{m_y} \end{pmatrix}$$

is a (reflexive) g inverse of

$$\begin{pmatrix} \mathbf{D}_{p_x} & \mathbf{p}_x\mathbf{p}'_y \\ \mathbf{p}_y\mathbf{p}'_x & \mathbf{D}_{p_y} \end{pmatrix}$$

yields

$$\mathbf{S}_{yy,w} = \mathbf{D}_{p_y} - \mathbf{Q}'_y\mathbf{D}_{p_y}\mathbf{Q}_y + O_p(T^{-1/2}).$$

Analogously we get $\mathbf{S}_{xx,w} = \mathbf{D}_{p_x} - \mathbf{Q}'_x\mathbf{D}_{p_x}\mathbf{Q}_x + O_p(T^{-1/2})$. We see that $\mathbf{S}_{xx,w} \otimes \mathbf{S}_{yy,w}$ equals $(\mathbf{D}_{p_x} - \mathbf{Q}'_x\mathbf{D}_{p_x}\mathbf{Q}_x) \otimes (\mathbf{D}_{p_y} - \mathbf{Q}'_y\mathbf{D}_{p_y}\mathbf{Q}_y) + O_p(T^{-1/2})$ and is asymptotically equivalent to the covariance matrix of $H_b(T)$.

$\mathbf{S}_{yx,w}$ is given by $\frac{1}{T}\mathbf{Y}'\mathbf{X} - \frac{1}{T}\mathbf{Y}'\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{X}$. Under H_0

$$\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{X} = \begin{pmatrix} \mathbf{D}_{p_x} & \mathbf{p}_x\mathbf{p}'_y \\ \mathbf{p}_y\mathbf{p}'_x & \mathbf{D}_{p_y} \end{pmatrix}^{-} \begin{pmatrix} \mathbf{P}_{x_{-1}x} \\ \mathbf{p}_y\mathbf{p}'_x \end{pmatrix} + O_p(T^{-1/2}).$$

We use the same g inverse as above to get

$$\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{X} = \begin{pmatrix} \mathbf{Q}_x - \frac{1}{2}\mathbf{1}_{m_x}\mathbf{p}'_x \\ \frac{1}{2}\mathbf{1}_{m_y}\mathbf{p}'_x \end{pmatrix} + O_p(T^{-1/2}) = \mathbf{A}_x + O_p(T^{-1/2}).$$

Analogously we get for $\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{Y}$

$$\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{Y} = \begin{pmatrix} \frac{1}{2}\mathbf{1}_{m_x}\mathbf{p}'_y \\ \mathbf{Q}_y - \frac{1}{2}\mathbf{1}_{m_y}\mathbf{p}'_y \end{pmatrix} + O_p(T^{-1/2}) = \mathbf{A}_y + O_p(T^{-1/2}).$$

[7] See for instance Billingsley [1961].

Plugging this in yields

$$\begin{aligned}
\mathbf{S}_{yx,w} &= \tfrac{1}{T}\mathbf{Y}'\mathbf{X} - \mathbf{Y}'\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-}\left(\tfrac{1}{T}\mathbf{W}'\mathbf{W}\right)\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{X} \\
&= \tfrac{1}{T}\mathbf{Y}'\mathbf{X} - \mathbf{Y}'\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-}\left(\tfrac{1}{T}\mathbf{W}'\mathbf{W}\right)\mathbf{A}_x \\
&- \mathbf{A}_y'\left(\tfrac{1}{T}\mathbf{W}'\mathbf{W}\right)\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{X} \\
&+ \mathbf{A}_y'\left(\tfrac{1}{T}\mathbf{W}'\mathbf{W}\right)\mathbf{A}_x \\
&- \left(\mathbf{Y}'\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-} - \mathbf{A}_y\right)'\left(\tfrac{1}{T}\mathbf{W}'\mathbf{W}\right)\left(\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{X} - \mathbf{A}_x\right).
\end{aligned}$$

The last term is of the order $O_p(T^{-1})$. Now, by Lemma 1

$$\mathbf{W}'\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{X} = \mathbf{W}'\mathbf{X} \quad \text{and} \quad \mathbf{W}'\mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-}\mathbf{W}'\mathbf{Y} = \mathbf{W}'\mathbf{Y}.$$

Hence,

$$\begin{aligned}
\mathbf{S}_{yx,w} &= \tfrac{1}{T}\mathbf{Y}'\mathbf{X} - \left(\tfrac{1}{T}\mathbf{Y}'\mathbf{W}\right)\mathbf{A}_x - \mathbf{A}_y'\left(\tfrac{1}{T}\mathbf{W}'\mathbf{X}\right) \\
&+ \mathbf{A}_y'\left(\tfrac{1}{T}\mathbf{W}'\mathbf{W}\right)\mathbf{A}_x + O_p(T^{-1}).
\end{aligned}$$

Moreover,

$$\mathbf{W}\mathbf{A}_x = \mathbf{X}_{-1}\mathbf{Q}_x \quad \text{and} \quad \mathbf{W}\mathbf{A}_y = \mathbf{Y}_{-1}\mathbf{Q}_y.$$

Plugging this in yields

$$\begin{aligned}
\mathbf{S}_{yx,w} &= \tfrac{1}{T}\mathbf{Y}'\mathbf{X} - \tfrac{1}{T}\mathbf{Y}'\mathbf{X}_{-1}\mathbf{Q}_x - \tfrac{1}{T}\mathbf{Q}_y'\mathbf{Y}_{-1}'\mathbf{X} \\
&+ \tfrac{1}{T}\mathbf{Q}_y'\mathbf{Y}_{-1}'\mathbf{X}_{-1}\mathbf{Q}_x + O_p(T^{-1}) \\
&= \tfrac{1}{T}\left(\mathbf{Y} - \mathbf{Y}_{-1}\mathbf{Q}_y\right)'\left(\mathbf{X} - \mathbf{X}_{-1}\mathbf{Q}_x\right) + O_p(T^{-1}).
\end{aligned}$$

If both processes, $\{X_t\}$ and $\{Y_t\}$, are stationary and ergodic Markov chains then $\sqrt{T}vec(\mathbf{S}_{yx,w}) = H_b(T) + O_p(T^{-1/2})$. Using the results for $\mathbf{S}_{yy,w}$ and $\mathbf{S}_{xx,w}$ we get that $X_b(T)$ is asymptotically equivalent to the trace statistic $Ttr(\mathbf{S}_{yy,w}^{-}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-}\mathbf{S}_{xy,w})$.

B.3 Joint Independence in Three-way Tables

The test for joint independence of $(\{X_t\},\{Z_t\})$ from $\{Y_t\}$ proposed in Pesaran and Timmermann [2009] reads as $Ttr(\mathbf{S}_{yy,w}^{-}\mathbf{S}_{yq,w}\mathbf{S}_{qq}^{-}\mathbf{S}_{qy,w})$ with $\mathbf{Q} = (\mathbf{X}, \mathbf{Z})$.

The statistic is asymptotically $\chi^2_{(m_x+m_z-2)(m_y-1)}$ distributed. Let us ignore any serial correlation for the moment and set $\mathbf{W} = \mathbf{1}_T$. We get $\mathbf{S}^-_{yy,w} = \mathbf{D}^{-1}_{\hat{p}_y}$ and $\mathbf{S}_{yq,w} = (\hat{\mathbf{P}}_{yx}, \hat{\mathbf{P}}_{yz}) - (\hat{\mathbf{p}}_y\hat{\mathbf{p}}'_x, \hat{\mathbf{p}}_y\hat{\mathbf{p}}'_z)$. Furthermore,

$$\mathbf{S}_{qq,w} = \begin{pmatrix} \mathbf{D}_{\hat{p}_x} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_x & \hat{\mathbf{P}}_{xz} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_z \\ \hat{\mathbf{P}}_{zx} - \hat{\mathbf{p}}_z\hat{\mathbf{p}}'_x & \mathbf{D}_{\hat{p}_z} - \hat{\mathbf{p}}_z\hat{\mathbf{p}}'_z \end{pmatrix}.$$

Using the same arguments as above we get that $Ttr(\mathbf{S}^-_{yy,w}\mathbf{S}_{yq,w}\mathbf{S}^-_{qq}\mathbf{S}_{qy,w})$ equals

$$Tvec\begin{pmatrix} \hat{\mathbf{P}}_{xy} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_y \\ \hat{\mathbf{P}}_{zy} - \hat{\mathbf{p}}_z\hat{\mathbf{p}}'_y \end{pmatrix}' \left(\mathbf{D}^{-1}_{\hat{p}_y} \otimes \mathbf{S}^-_{qq,w}\right) vec\begin{pmatrix} \hat{\mathbf{P}}_{xy} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_y \\ \hat{\mathbf{P}}_{zy} - \hat{\mathbf{p}}_z\hat{\mathbf{p}}'_y \end{pmatrix}.$$

The typical elements of the *vec* operator are given by $\hat{p}_{ij+} - \hat{p}_{i++}\hat{p}_{+j+}$ and $\hat{p}_{+jk} - \hat{p}_{+j+}\hat{p}_{++k}$. Joint independence of $(\{X_t\}, \{Z_t\})$ from $\{Y_t\}$ implies that $p_{ijk} = p_{+j+}p_{i+k}$. This in turn implies that $p_{+jk} = p_{+j+}p_{++k}$ and $p_{ij+} = p_{+j+}p_{i++}$ but not vice versa. The trace statistic defined in Pesaran and Timmermann [2009] tests for simultaneous pairwise rather than joint independence.

For illustrative purposes suppose that additionally $p_{i+k} = p_{i++}p_{++k}$. Under this assumption

$$\mathbf{S}_{qq,w} = \begin{pmatrix} \mathbf{D}_{\hat{p}_x} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_x & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\hat{p}_z} - \hat{\mathbf{p}}_z\hat{\mathbf{p}}'_z \end{pmatrix} + O_p(T^{-1/2}).$$

A g inverse is given by

$$\mathbf{S}^-_{qq,w} = \begin{pmatrix} \mathbf{D}^{-1}_{\hat{p}_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}_{\hat{p}_z} \end{pmatrix} + O_p(T^{-1/2}).$$

The trace may be written as

$$Tvec\begin{pmatrix} \hat{\mathbf{P}}_{xy} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_y \\ \hat{\mathbf{P}}_{zy} - \hat{\mathbf{p}}_z\hat{\mathbf{p}}'_y \end{pmatrix}' \left(\mathbf{D}^{-1}_{\hat{p}_y} \otimes \begin{pmatrix} \mathbf{D}^{-1}_{\hat{p}_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1}_{\hat{p}_z} \end{pmatrix}\right) vec\begin{pmatrix} \hat{\mathbf{P}}_{xy} - \hat{\mathbf{p}}_x\hat{\mathbf{p}}'_y \\ \hat{\mathbf{P}}_{zy} - \hat{\mathbf{p}}_z\hat{\mathbf{p}}'_y \end{pmatrix}$$

or

$$T\sum_{i=1}^{m_x}\sum_{j=1}^{m_y}\frac{(\hat{p}_{ij+}-\hat{p}_{i++}\hat{p}_{+j+})^2}{\hat{p}_{i++}\hat{p}_{+j+}}+T\sum_{j=1}^{m_y}\sum_{k=1}^{m_z}\frac{(\hat{p}_{+jk}-\hat{p}_{+j+}\hat{p}_{++k})^2}{\hat{p}_{+j+}\hat{p}_{++k}}.$$

Example 3. *Let $m_x = m_y = m_z = 2$. The joint probabilities are given by $\mathbf{p} = (1/12)(1, 2, 2, 1, 2, 1, 1, 2)'$. It is easy to verify that $\mathbf{p}_{i++} = \mathbf{p}_{+j+} = \mathbf{p}_{++k} = (1/2)(1, 1)'$ and $\mathbf{p}_{ij+} = \mathbf{p}_{i+k} = \mathbf{p}_{+jk} = (1/4)(1, 1, 1, 1)'$. Hence, $\mathbf{P}_{xz} = \mathbf{p}_x\mathbf{p}_z'$, $\mathbf{P}_{xy} = \mathbf{p}_x\mathbf{p}_y'$, and $\mathbf{P}_{yz} = \mathbf{p}_y\mathbf{p}_z'$. But $p_{ijk} \neq p_{+j+}p_{i+k}$. $(\{X_t\}, \{Z_t\})$ are not jointly independent from $\{Y_t\}$. The test has no power against this alternative.*

If we use $\tilde{Q}_t = X_t \otimes Z_t$ instead of $Q_t = (X_t, Z_t)$ then $Ttr(\mathbf{S}_{yy,w}^{-}\mathbf{S}_{y\tilde{q},w}\mathbf{S}_{\tilde{q}\tilde{q}}^{-}\mathbf{S}_{\tilde{q}y,w})$ tests whether $(\{X_t\}, \{Z_t\})$ are jointly independent from $\{Y_t\}$. The statistic is asymptotically $\chi^2_{(m_x m_z-1)(m_y-1)}$ distributed.

B.4 Conditional Independence in Three-way Tables

In Part 1 of Theorem 2 Pesaran and Timmermann [2009] propose a test for conditional independence of $\{X_t\}$ and $\{Y_t\}$ given $\{Z_t\}$. They show that

$$Ttr(\mathbf{S}_{yy,w}^{-}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-}\mathbf{S}_{xy,w}) \sim^a \chi^2_{(m_x-1)(m_y-1)}$$

with $\mathbf{W} = \mathbf{Z}$ in the case of no serial correlation and $\mathbf{W} = (\mathbf{Z}, \mathbf{Y}_{-1}, \mathbf{X}_{-1}, \mathbf{Z}_{-1})$ for Markov chains.[8]

Let us assume no serial correlation. Straightforward calculations along

[8]Pesaran and Timmermann [2009] use $\mathbf{W}^0 = (\mathbf{Z}^0, \mathbf{1}_T)$ and $\mathbf{W}^0 = (\mathbf{Y}_{-1}^0, \mathbf{X}_{-1}^0, \mathbf{Z}^0, \mathbf{Z}_{-1}^0, \mathbf{1}_{T-1})$ respectively. Lemma 4 can be applied.

the above lines yield

$$
\begin{aligned}
& Ttr(\mathbf{S}_{yy,w}^{-}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-}\mathbf{S}_{xy,w}) \\
= \; & Ttr(\mathbf{S}_{xx,w}^{-}\mathbf{S}_{xy,w}\mathbf{S}_{yy,w}^{-}\mathbf{S}_{yx,w}) \\
= \; & Tvec\left(\hat{\mathbf{P}}_{yx} - \hat{\mathbf{P}}_{yz}\mathbf{D}_{\hat{p}_z}^{-1}\hat{\mathbf{P}}_{zx}\right)' \\
\times \; & \left(\left(\mathbf{D}_{\hat{p}_x} - \hat{\mathbf{P}}_{xz}\mathbf{D}_{\hat{p}_z}^{-1}\hat{\mathbf{P}}_{zx}\right)^{-} \otimes \left(\mathbf{D}_{\hat{p}_y} - \hat{\mathbf{P}}_{yz}\mathbf{D}_{\hat{p}_z}^{-1}\hat{\mathbf{P}}_{zy}\right)^{-}\right) \\
\times \; & vec\left(\hat{\mathbf{P}}_{yx} - \hat{\mathbf{P}}_{yz}\mathbf{D}_{\hat{p}_z}^{-1}\hat{\mathbf{P}}_{zx}\right)
\end{aligned}
$$

Let $\tilde{\mathbf{r}}^k = (\hat{p}_{1,1,k} \ldots \hat{p}_{m_x,m_y,k})'/\hat{p}_{++k}$ be the conditional observed fractions and denote the marginal conditional fractions by $\tilde{\mathbf{r}}_x^k$ and $\tilde{\mathbf{r}}_y^k$. A closer look at the matrix $(\mathbf{D}_{\hat{p}_x} - \hat{\mathbf{P}}_{xz}\mathbf{D}_{\hat{p}_z}^{-1}\hat{\mathbf{P}}_{zx})$ reveals that it can be written as

$$
\sum_{k=1}^{m_z} \hat{p}_{++k}\left(\mathbf{D}_{\tilde{\mathbf{r}}_x^k} - \tilde{\mathbf{r}}_x^k\tilde{\mathbf{r}}_x^{k\prime}\right).
$$

Analogously, we get

$$
(\mathbf{D}_{\hat{p}_y} - \hat{\mathbf{P}}_{yz}\mathbf{D}_{\hat{p}_z}\hat{\mathbf{P}}_{zy}) = \sum_{k=1}^{m_z} \hat{p}_{++k}\left(\mathbf{D}_{\tilde{\mathbf{r}}_y^k} - \tilde{\mathbf{r}}_y^k\tilde{\mathbf{r}}_y^{k\prime}\right).
$$

The term in the *vec* operator equals

$$
\begin{aligned}
& vec\left(\hat{\mathbf{P}}_{yx} - \hat{\mathbf{P}}_{yz}\mathbf{D}_{\hat{p}_z}^{-1}\hat{\mathbf{P}}_{zx}\right) \\
= \; & \sum_{k=1}^{m_z}\left\{\begin{pmatrix} \hat{p}_{11k} \\ \hat{p}_{12k} \\ \vdots \\ \hat{p}_{m_x,m_y,k} \end{pmatrix} - \frac{1}{\hat{p}_{++k}}\begin{pmatrix} \hat{p}_{1+k}\hat{p}_{+1k} \\ \hat{p}_{1+k}\hat{p}_{+2k} \\ \vdots \\ \hat{p}_{m_x,+,k}\hat{p}_{+,m_y,k} \end{pmatrix}\right\} \\
= \; & \sum_{k=1}^{m_z} \hat{p}_{++k}\left(\tilde{\mathbf{r}}^k - \tilde{\mathbf{r}}_x^k\tilde{\mathbf{r}}_y^{k\prime}\right) \\
= \; & \sum_{k=1}^{m_z} \mathbf{d}^i
\end{aligned}
$$

with $\mathbf{d}^i$ as defined in the main text.[9] The typical elements of the *vec* operator are given by $\hat{p}_{ij+} - \sum_{k=1}^{m_z} \hat{p}_{i+k}\hat{p}_{+jk}/\hat{p}_{++k}$. But $p_{ij+} = \sum_{k=1}^{m_z} p_{i+k}p_{+jk}/p_{++k}$ is

[9] In the main text I used X_t as conditioning variable for notational convenience. Here I use Z_t as in Pesaran and Timmermann [2009].

only a weak implication of conditional independence compared to the usual $p_{ijk} = p_{i+k}p_{+jk}/p_{++k}$. Without serial correlation the trace statistic equals

$$
\begin{aligned}
& Ttr(\mathbf{S}_{yy,w}^{-}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-}\mathbf{S}_{xy,w}) \\
= \; & T\textstyle\sum_{k=1}^{m_z} \mathbf{d}^{i\prime} \\
\times \; & \left(\textstyle\sum_{k=1}^{m_z} \hat{p}_{++k}\left(\mathbf{D}_{\tilde{\mathbf{r}}_x^k} - \tilde{\mathbf{r}}_x^k\tilde{\mathbf{r}}_x^{k\prime}\right)\right)^{-} \otimes \left(\textstyle\sum_{k=1}^{m_z} \hat{p}_{++k}\left(\mathbf{D}_{\tilde{\mathbf{r}}_y^k} - \tilde{\mathbf{r}}_y^k\tilde{\mathbf{r}}_y^{k\prime}\right)\right)^{-} \\
\times \; & \textstyle\sum_{k=1}^{m_z} \mathbf{d}^{i}.
\end{aligned}
$$

Comparing this to the results from the main text we see that the covariance matrix in the CMH statistic differs from that in the trace statistic by the cross products between the different strata. The covariance matrix in the CMH test as derived in the main text is given by

$$
\begin{aligned}
& \textstyle\sum_{i=1}^{m_x} p_{i++}\left(\mathbf{D}_{\mathbf{r}_y^i} - \mathbf{r}_y^i\mathbf{r}_y^{i\prime}\right) \otimes \left(\mathbf{D}_{\mathbf{r}_z^i} - \mathbf{r}_z^i\mathbf{r}_z^{i\prime}\right) \\
= \; & \mathbb{E}_X\left[\left(\mathbf{D}_{\mathbf{r}_y} - \mathbf{r}_y\mathbf{r}_y'\right) \otimes \left(\mathbf{D}_{\mathbf{r}_z} - \mathbf{r}_z\mathbf{r}_z'\right)\right]
\end{aligned}
$$

where the expectation is taken across the different categories of $\{X_t\}$. If X_t is the conditioning variable, the matrix used in the trace statistic equals

$$
\begin{aligned}
& \left(\textstyle\sum_{i=1}^{m_x} \hat{p}_{i++}\left(\mathbf{D}_{\tilde{\mathbf{r}}_y^i} - \tilde{\mathbf{r}}_y^i\tilde{\mathbf{r}}_y^{i\prime}\right)\right) \otimes \left(\textstyle\sum_{i=1}^{m_x} \hat{p}_{i++}\left(\mathbf{D}_{\tilde{\mathbf{r}}_z^i} - \tilde{\mathbf{r}}_z^i\tilde{\mathbf{r}}_z^{i\prime}\right)\right) \\
= \; & \mathbb{E}_X\left[\mathbf{D}_{\mathbf{r}_y} - \mathbf{r}_y\mathbf{r}_y'\right] \otimes \mathbb{E}_X\left[\mathbf{D}_{\mathbf{r}_z} - \mathbf{r}_z\mathbf{r}_z'\right] + O_p(T^{-1/2}).
\end{aligned}
$$

The difference between these two is the covariance of $\mathbf{D}_{\mathbf{r}_y} - \mathbf{r}_y\mathbf{r}_y'$ with $\mathbf{D}_{\mathbf{r}_z} - \mathbf{r}_z\mathbf{r}_z'$ across the different categories of $\{X_t\}$. The covariance matrix used in the trace statistic will in general not be correct.

Example 4. *Suppose that $\{Y_t\}$ and $\{Z_t\}$ are independent conditional on $\{X_t\}$. The series are not serially correlated and $m_x = m_y = m_z = 2$. Let $p_{1++} = p_{2++} = 0.5$, $\mathbf{r}_y^1 = \mathbf{r}_z^1 = (0.5, 0.5)'$, and $\mathbf{r}_y^2 = \mathbf{r}_z^2 = (1 - 1/n, 1/n)'$. We get*

$$
\begin{aligned}
& \mathbb{E}_X\left[\left(\mathbf{D}_{\mathbf{r}_y} - \mathbf{r}_y\mathbf{r}_y'\right) \otimes \left(\mathbf{D}_{\mathbf{r}_z} - \mathbf{r}_z\mathbf{r}_z'\right)\right] \\
= \; & \tfrac{1}{2}\left(\tfrac{1}{16} + \tfrac{(n-1)^2}{n^4}\right)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}
\end{aligned}
$$

and

$$\begin{aligned}&\mathbb{E}_X\left[\mathbf{D}_{\mathbf{r}_y}-\mathbf{r}_y\mathbf{r}_y'\right]\otimes\mathbb{E}_X\left[\mathbf{D}_{\mathbf{r}_z}-\mathbf{r}_z\mathbf{r}_z'\right]\\ =\;&\tfrac{1}{4}\left(\tfrac{1}{4}+\tfrac{n-1}{n^2}\right)^2\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\otimes\begin{pmatrix}1&-1\\-1&1\end{pmatrix}.\end{aligned}$$

Not surprisingly, both expressions are equal for $n=2$. *For* $n=20$ *the rejection rate of the trace statistic for a notional size of* 10% *is* 17.39%. *For* $n\to\infty$ *the scalar in the first term converges to* $1/32$ *and in the second term to* $1/64$. *The rejection rate converges to* 24.48%.

Alternatively, let $p_{1++}=p_{2++}=0.5$, $\mathbf{r}_y^1=(0.5,0.5)'$, $\mathbf{r}_y^2=(1/n,1-1/n)'$, $\mathbf{r}_z^1=(1-1/n,1/n)'$, *and* $\mathbf{r}_z^2=(0.5,0.5)'$. *Hence,*

$$\begin{aligned}&\mathbb{E}_X\left[\left(\mathbf{D}_{\mathbf{r}_y}-\mathbf{r}_y\mathbf{r}_y'\right)\otimes\left(\mathbf{D}_{\mathbf{r}_z}-\mathbf{r}_z\mathbf{r}_z'\right)\right]\\ =\;&\tfrac{1}{4}\tfrac{(n-1)}{n^2}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\otimes\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\end{aligned}$$

and $\mathbb{E}_X\left[\mathbf{D}_{\mathbf{r}_y}-\mathbf{r}_y\mathbf{r}_y'\right]\otimes\mathbb{E}_X\left[\mathbf{D}_{\mathbf{r}_z}-\mathbf{r}_z\mathbf{r}_z'\right]$ *as above. For* $n=20$ *the rejection rate of the trace statistic is* 2.48% *for a notional size of* 10%. *For* $n\to\infty$ *the correct term as used in the CMH test converges to* 0. *The rejection rate converges also to* 0.

C Cochrane-Mantel-Haenszel Test

In this section it is shown that CMH from the main text equals the Cochrane-Mantel-Haenszel (CMH) test. I follow the presentation in Agresti [2002] with the deviation that I condition on the first variable for notational convenience.

n_{ijk} is the number of realizations with $X_t=i$, $Y_t=j$, and $Z_t=k$. Let $\mathbf{n}_i=(n_{i11},\ldots,n_{i,m_y,m_z})'$ be the counts conditional on X_t being equal to i. Given n_{ij+} and n_{i+k} the nonredundant counts are summarized in

$$\mathbf{n}_i^0=(n_{i11},\ldots,n_{i,1,m_z-1},n_{i21}\ldots,n_{i,m_y-1,m_z-1})'.$$

Let $\mu_i=(n_{i1+}n_{i+1},\ldots,n_{i,m_y,+}n_{i,+,m_z})'/n_{i++}$ and μ_i^0 corresponding to $\mathbf{n}_i^0$.

The null covariance matrix $\mathbf{V}_i$ of $\mathbf{n}_i$ is given by [10]

$$\mathbf{V}_i = \frac{1}{n_{i++}} \left(\mathbf{D}_{n_{i\bullet+}} - \frac{1}{n_{i++}} n_{i\bullet+} n'_{i\bullet+} \right) \otimes \left(\mathbf{D}_{n_{i+\bullet}} - \frac{1}{n_{i++}} n_{i+\bullet} n'_{i+\bullet} \right)$$

where $n_{i\bullet+}$ is the vector $(n_{i,1,+}, \cdots, n_{i,m_y,+})'$ and $n_{i+\bullet} = (n_{i,+,1}, \cdots, n_{i,+,m_z})'$. The covariance matrix of $\mathbf{n}_i^0$ under conditional independence is given by

$$\mathbf{V}_i^0 = \frac{1}{n_{i++}} \left(\mathbf{D}_{n^0_{i\bullet+}} - \frac{1}{n_{i++}} n^0_{i\bullet+} n^{0\prime}_{i\bullet+} \right) \otimes \left(\mathbf{D}_{n^0_{i+\bullet}} - \frac{1}{n_{i++}} n^0_{i+\bullet} n^{0\prime}_{i+\bullet} \right)$$

with $n^0_{i\bullet+} = (n_{i,1,+}, \cdots, n_{i,m_y-1,+})'$ and $n^0_{i+\bullet} = (n_{i,+,1}, \cdots, n_{i,+,m_z-1})'$. The definition of the CMH test in Agresti [2002] is given by

$$CMH^a = \sum_{i=1}^{m_x} \left(\mathbf{n}_i^0 - \mu_i^0 \right)' \left(\sum_{i=1}^{m_x} \mathbf{V}_i^0 \right)^{-1} \sum_{i=1}^{m_x} \left(\mathbf{n}_i^0 - \mu_i^0 \right).$$

So the first question is whether we may use the vector of all counts instead of the nonredundant ones. I will show that CMH^a equals

$$\sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i)' \left(\sum_{i=1}^{m_x} \mathbf{V}_i \right)^{+} \sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i).$$

The first step to establish equality is to reorder the elements of $\mathbf{n}_i$ such that the elements of $\mathbf{n}_i^0$ come first. If we premultiply $\mathbf{n}_i$ and μ_i by the permutation matrix

$$\mathbf{P} = \begin{pmatrix} \mathbf{I}_{m_y} \otimes [\mathbf{I}_{m_z-1}, \mathbf{0}_{m_z-1,1}] \\ \mathbf{I}_{m_y} \otimes [\mathbf{0}_{1,m_z-1}, 1] \end{pmatrix},$$

we get

$$\tilde{\mathbf{n}}_i = \mathbf{P}\mathbf{n}_i = \begin{pmatrix} \mathbf{n}_i^0 \\ \mathbf{n}_i^1 \end{pmatrix} \quad \text{and} \quad \tilde{\mu}_i = \mathbf{P}\mu_i = \begin{pmatrix} \mu_i^0 \\ \mu_i^1 \end{pmatrix}$$

with $\mathbf{n}_i^1 = (n_{i,m_y,1}, \ldots, n_{i,m_y,m_z-1}, n_{i,1,m_z}, \ldots, n_{i,m_y,m_z})'$ and the corresponding

[10] I use Cochrane's version and scale by n_{i++} instead of $n_{i++} - 1$.

ordering for μ_i^1. The covariance matrix of $\tilde{\mathbf{n}}_i$ is given by $\tilde{\mathbf{V}}_i = \mathbf{P}\mathbf{V}_i\mathbf{P}'$. This yields

$$\begin{aligned} & \textstyle\sum_{i=1}^{m_x} (\tilde{\mathbf{n}}_i - \tilde{\mu}_i)' \left(\sum_{i=1}^{m_x} \tilde{\mathbf{V}}_i\right)^+ \sum_{i=1}^{m_x} (\tilde{\mathbf{n}}_i - \tilde{\mu}_i) \\ = & \textstyle\sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i)' \mathbf{P}' (\sum_{i=1}^{m_x} \mathbf{P}\mathbf{V}_i\mathbf{P}')^+ \sum_{i=1}^{m_x} \mathbf{P} (\mathbf{n}_i - \mu_i) \\ = & \textstyle\sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i)' (\sum_{i=1}^{m_x} \mathbf{V}_i)^+ \sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i) . \end{aligned}$$

The ordering of the elements is irrelevant. The next step is to show that there exists a matrix $\mathbf{B}$ such that $\mathbf{n}_i^1 - \mu_i^1 = \mathbf{B}(\mathbf{n}_i^0 - \mu_i^0)$. To see this note that

$$\sum_{j=1}^{m_y} \left(n_{ijk} - \frac{n_{ij+}n_{i+k}}{n_{i++}} \right) = \sum_{k=1}^{m_z} \left(n_{ijk} - \frac{n_{ij+}n_{i+k}}{n_{i++}} \right) = 0.$$

Define

$$\mathbf{B} = \begin{pmatrix} -\mathbf{1}_{1,m_y-1} & \otimes & \mathbf{I}_{m_z-1} \\ -\mathbf{I}_{m_y-1} & \otimes & \mathbf{1}_{1,m_z-1} \\ \mathbf{1}_{1,m_y-1} & \otimes & \mathbf{1}_{1,m_z-1} \end{pmatrix}.$$

It is straightforward to verify that

$$\tilde{\mathbf{n}}_i - \tilde{\mu}_i = \begin{pmatrix} \mathbf{I}_{(m_y-1)(m_z-1)} \\ \mathbf{B} \end{pmatrix} (\mathbf{n}_i^0 - \mu_i^0)$$

and

$$\tilde{\mathbf{V}}_i = \begin{pmatrix} \mathbf{I}_{(m_y-1)(m_z-1)} \\ \mathbf{B} \end{pmatrix} \mathbf{V}_i^0 \left(\mathbf{I}_{(m_y-1)(m_z-1)}, \mathbf{B}'\right) .$$

Using that $(\mathbf{I}_{(m_y-1)(m_z-1)}, \mathbf{B}')$ is of full row rank we get

$$\left(\textstyle\sum_{i=1}^{m_x} \tilde{\mathbf{V}}_i\right)^+ = \begin{pmatrix} \mathbf{I} \\ \mathbf{B} \end{pmatrix} (\mathbf{I} + \mathbf{B}'\mathbf{B})^{-1} (\textstyle\sum_{i=1}^{m_x} \mathbf{V}_i^0)^{-1} (\mathbf{I} + \mathbf{B}'\mathbf{B})^{-1} (\mathbf{I}, \mathbf{B}') .$$

Hence,

$$\begin{aligned}
CMH^a &= \sum_{i=1}^{m_x} (\mathbf{n}_i^0 - \mu_i^0)' \left(\sum_{i=1}^{m_x} \mathbf{V}_i^0\right)^{-1} \sum_{i=1}^{m_x} (\mathbf{n}_i^0 - \mu_i^0) \\
&= \sum_{i=1}^{m_x} (\tilde{\mathbf{n}}_i - \tilde{\mu}_i)' \left(\sum_{i=1}^{m_x} \tilde{\mathbf{V}}_i\right)^{+} \sum_{i=1}^{m_x} (\tilde{\mathbf{n}}_i - \tilde{\mu}_i) \\
&= \sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i)' \left(\sum_{i=1}^{m_x} \mathbf{V}_i\right)^{+} \sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i).
\end{aligned}$$

Now, $T\mathbf{d}^i = (\mathbf{n}_i - \mu_i)$ and

$$\mathbf{V}_i = T\hat{p}_{i++} \left(\mathbf{D}_{\tilde{\mathbf{r}}_y^i} - \tilde{\mathbf{r}}_y^i \tilde{\mathbf{r}}_y^{i\prime}\right) \otimes \left(\mathbf{D}_{\tilde{\mathbf{r}}_z^i} - \tilde{\mathbf{r}}_z^i \tilde{\mathbf{r}}_z^{i\prime}\right)$$

yields that $CMH^a = CMH$, i.e.

$$\begin{aligned}
& T\sum_{i=1}^{m_x} \mathbf{d}^{i\prime} \\
\times & \left(\sum_{i=1}^{m_x} \hat{p}_{i++} \left(\mathbf{D}_{\tilde{\mathbf{r}}_y^i} - \tilde{\mathbf{r}}_y^i \tilde{\mathbf{r}}_y^{i\prime}\right) \otimes \left(\mathbf{D}_{\tilde{\mathbf{r}}_z^i} - \tilde{\mathbf{r}}_z^i \tilde{\mathbf{r}}_z^{i\prime}\right)\right)^{+} \\
\times & \sum_{i=1}^{m_x} \mathbf{d}^i.
\end{aligned}$$

What happens if we use an arbitrary g inverse instead of the Moore-Penrose inverse? The quadratic form is invariant if $\sum_{i=1}^{m_x} (\mathbf{n}_i - \mu_i)$ is in the column space of $\sum_{i=1}^{m_x} \mathbf{V}_i$. But this is in general not the case notwithstanding that $(\mathbf{n}_i - \mu_i)$ is in the column space of $\mathbf{V}_i$. The quadratic form is asymptotically χ^2 distributed for arbitrary g inverses but the equivalence with the standard definition of CMH holds for the Moore-Penrose inverse only.

Remark 5. *If we had used the scaling factor $n_{i++} - 1$, the formula would be*

$$\begin{aligned}
& T\sum_{i=1}^{m_x} \mathbf{d}^{i\prime} \\
\times & \left(\sum_{i=1}^{m_x} \frac{\hat{p}_{i++}^2}{\hat{p}_{i++} - 1/T} \left(\mathbf{D}_{\tilde{\mathbf{r}}_y^i} - \tilde{\mathbf{r}}_y^i \tilde{\mathbf{r}}_y^{i\prime}\right) \otimes \left(\mathbf{D}_{\tilde{\mathbf{r}}_z^i} - \tilde{\mathbf{r}}_z^i \tilde{\mathbf{r}}_z^{i\prime}\right)\right)^{+} \\
\times & \sum_{i=1}^{m_x} \mathbf{d}^i.
\end{aligned}$$

D Further Simulations

In this section I provide some more simulation results.

D.1 Independence Tests

It is important to keep in mind that $\mathbf{p} = \mathbf{p}_x \otimes \mathbf{p}_y$ does not imply that X_t is independent from Y_{t-l} for $l \neq 0$. But as all tests aim only at the difference between $\mathbf{p}$ and $\mathbf{p}_x \otimes \mathbf{p}_y$ none of the tests is able to detect these kind of dependencies.

Example 5. *Suppose that the process $U_t = X_t \otimes Y_t$ is a simple Markov chain with transition probability*

$$\mathbf{Q} = \frac{1}{2} \begin{pmatrix} a & 1-a & a & 1-a \\ 1-a & a & 1-a & a \\ 1-a & a & 1-a & a \\ a & 1-a & a & 1-a \end{pmatrix}.$$

It is straightforward to verify that $\mathbf{p} = (1/4)(1,1,1,1)'$ and $\mathbf{p}_x = \mathbf{p}_y = (1/2)(1,1)'$. Hence, $\mathbf{p} = \mathbf{p}_x \otimes \mathbf{p}_y$. Moreover,

$$\mathbf{Q}_x = \mathbf{Q}_y = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \textit{and} \quad \mathbf{Q} \neq \mathbf{Q}_x \otimes \mathbf{Q}_y.$$

If we condition X_t and Y_t on both X_{t-1} and Y_{t-1} then they are in general (for $a \neq 1/2$) not independent. Figure 10 illustrates that none of the tests has any power to detect this kind of dependence.

To handle such a situation we have to test whether $\mathbf{Q} = \mathbf{Q}_x \otimes \mathbf{Q}_y$. The procedure used in Solow et al. [1995] and Sandland [1976] works well as is illustrated in Figure 11.

The next two examples illustrate that the rejection rates of the standard Pearson test for independence deviates considerably from the nominal size.

Example 6. *To illustrate the size of the various independence tests consider the follwing simple Markov chain*

$$\mathbf{Q}_x = \mathbf{Q}_y = \begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix} \quad \textit{for } a \in (0,1).$$

Figure 10: The rejection rate at a nominal level of 10% for Example 5. The results are based on 1000 simulations for a from 0.1 to 0.9 in steps of 0.01. The number of observations in each simulation equals 100. The power equals the size for all test procedures used in Section 5.

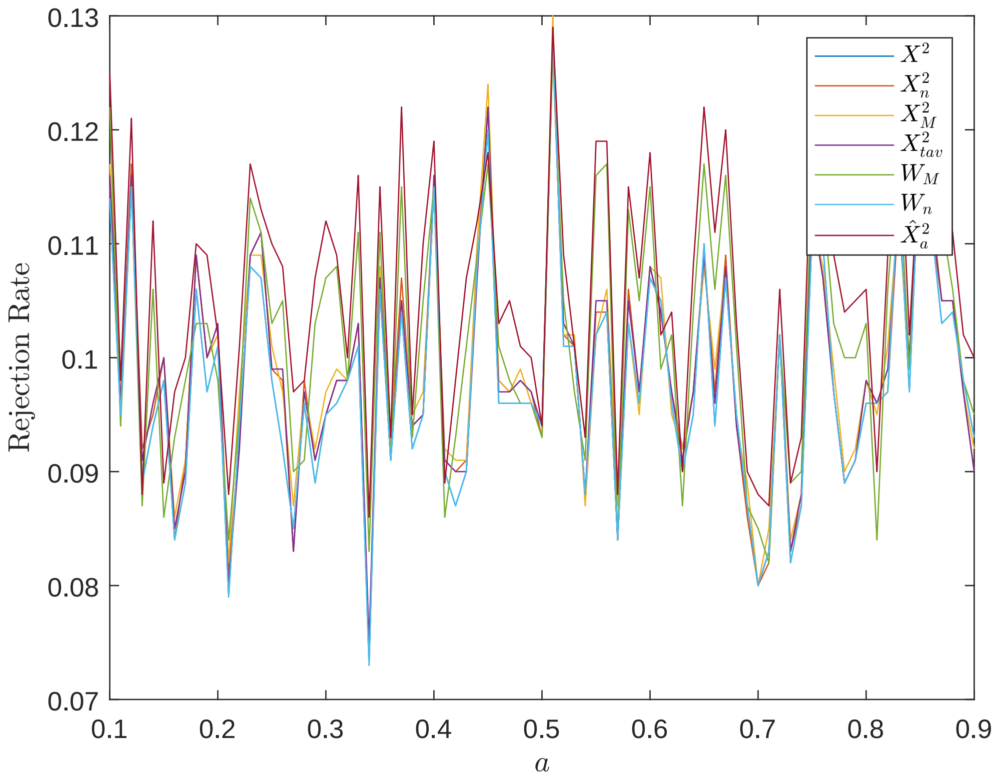


Figure 11: The rejection rate at a nominal level of 10% for Example 5. We test whether $\mathbf{Q} = \mathbf{Q}_x \otimes \mathbf{Q}_y$. The results are based on 1000 simulations for a from 0.1 to 0.9 in steps of 0.01. The number of observations in each simulation equals 100. The rejection rate of Pearson χ^2 test is plotted as a benchmark.

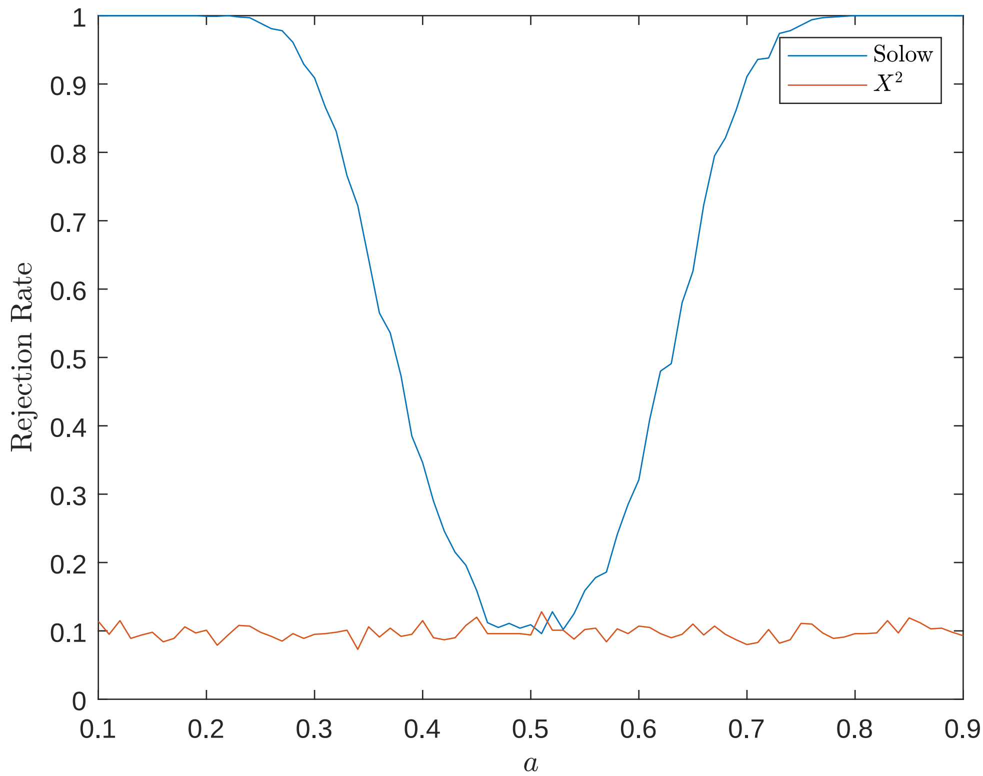
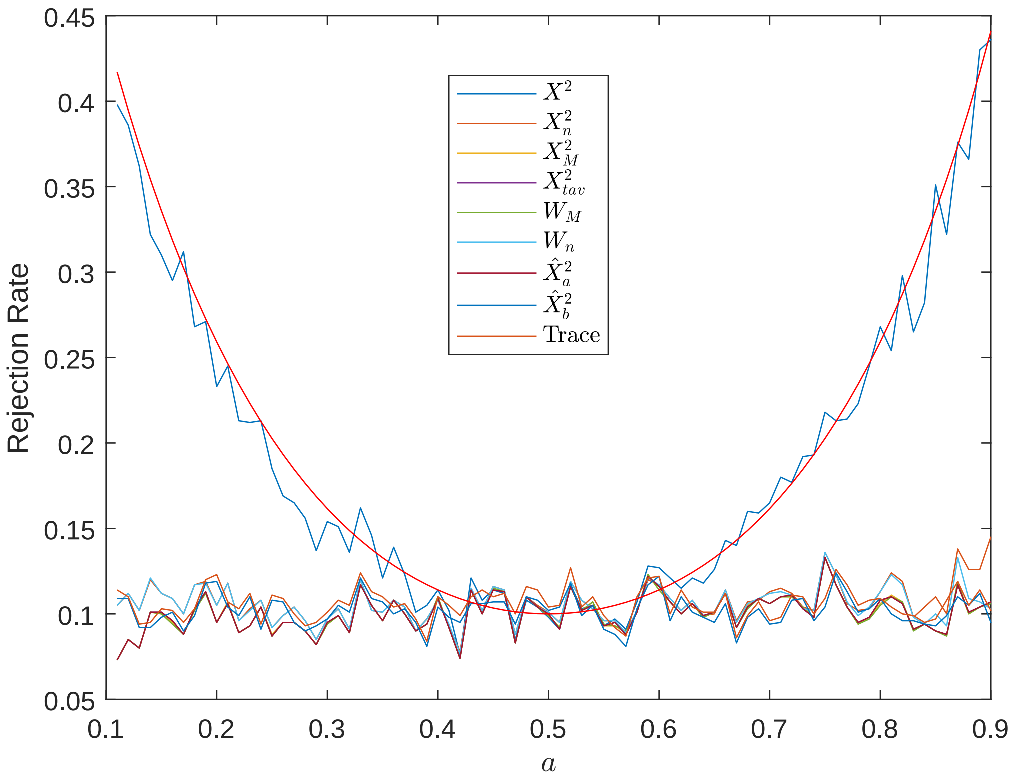


Figure 12: The rejection rate at a nominal level of 10% for the Example 6. The results are based on 1000 simulations for a from 0.1 to 0.9 in steps of 0.01. The number of observations in each simulation equals 100. The smooth red line is the theoretical rejection rate of X^2.



$\{X_t\}$ and $\{Y_t\}$ are independent. $\mathbf{Q}$ equals $\mathbf{Q}_x \otimes \mathbf{Q}_y$ and $\mathbf{p} = \mathbf{p}_x \otimes \mathbf{p}_y$ with $\mathbf{p}_x = \mathbf{p}_y = (0.5, 0.5)'$. The asymptotic distribution of X^2 is equivalent to $(1-2a+2a^2)/(2a(1-a))\chi_1^2$. For the simulations I used $T = 100$ observations. The naive estimator is based on $\tau = 4$ lags. a ranges from 0.1 to 0.9 with increments of 0.01. As already derived in Example 1 the autocorrelation of X_t^* and Y_t^* equals $2a - 1$. All results are based on 1000 simulations. Figure 12 displays the results. The rejection rates of the Pearson test are very high for a close to 0 or 1. All other tests perform reasonably well across the entire range of a.

Example 7. *Suppose that* $\mathbf{Q}_x$ *is as in Example 6 but*

$$\mathbf{Q}_y = \begin{pmatrix} 1-a & a \\ a & 1-a \end{pmatrix} \quad \textit{for } a \in (0,1).$$

$\{X_t\}$ *and* $\{Y_t\}$ *are independent. Again,* $\mathbf{Q}$ *equals* $\mathbf{Q}_x \otimes \mathbf{Q}_y$ *and* $\mathbf{p} = \mathbf{p}_x \otimes \mathbf{p}_y$ *with* $\mathbf{p}_x = \mathbf{p}_y = (0.5, 0.5)'$*. The asymptotic distribution of* X^2 *is equivalent to* $2a(1-a)/(1-2a+2a^2)\chi_1^2$*. In this case the rejection rates of the Pearson test are below the size for* a *close to* 0 *or* 1*. The results are illustrated in Figure 13.*

How should we deal with higher order Markov processes? Pesaran and Timmermann [2009] suggest to include higher lags into the regressor matrix $\mathbf{W}$. This approach is not correct but works mostly very well in simulations. Alternatively, we may transform the process into a simple Markov chain. This correct approach suffers from the problem that the number of states grows exponentially with the inclusion of additional lags.

Example 8. *Suppose that* $\{X_t\}$ *and* $\{Y_t\}$ *are independent Markov processes of order 2. The transition matrices for* $\tilde{X}_t = X_{t-1} \otimes X_t$ *and* $\tilde{Y}_t = Y_{t-1} \otimes Y_t$ *are given by*

$$\mathbf{Q}_{\tilde{x}} = \mathbf{Q}_{\tilde{y}} = \begin{pmatrix} a & 1-a & 0 & 0 \\ 0 & 0 & 1-a & a \\ 1-a & a & 0 & 0 \\ 0 & 0 & a & 1-a \end{pmatrix} \quad \textit{for } a \in (0,1).$$

Both processes are simple Markov chains if and only if $a = 0.5$*. For each* a *from* 0.1 *to* 0.9 *in steps of* 0.01 *I performed* 1000 *simulations with* $T = 100$ *observations. To test whether* $\{X_t\}$ *and* $\{Y_t\}$ *are independent I calculated the trace statistic by including up to* 3 *lags into the regressor matrix* $\mathbf{W}$*. Figure 14 displays the results. For* a *between* 0.3 *and* 0.7 *the rejection rate is very close to the size irrespective how many lags are included. For* $a < 0.3$ *(*$a > 0.7$*) the rejection rate deviates from the level unless we include* 3 *lags although both processes are only of order* 2*. If we perform the trace test on*

Figure 13: The rejection rate at a nominal level of 10% for the Example 7. The results are based on 1000 simulations for a from 0.1 to 0.9 in steps of 0.01. The number of observations in each simulation equals 100. The smooth red line is the theoretical rejection rate of X^2.

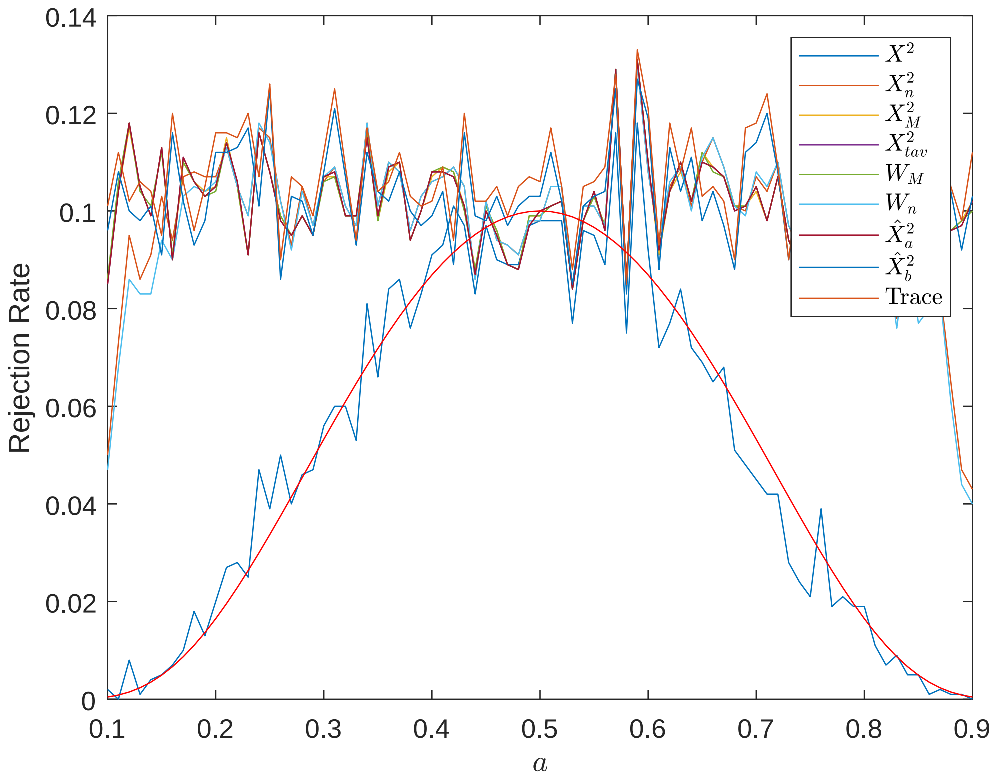
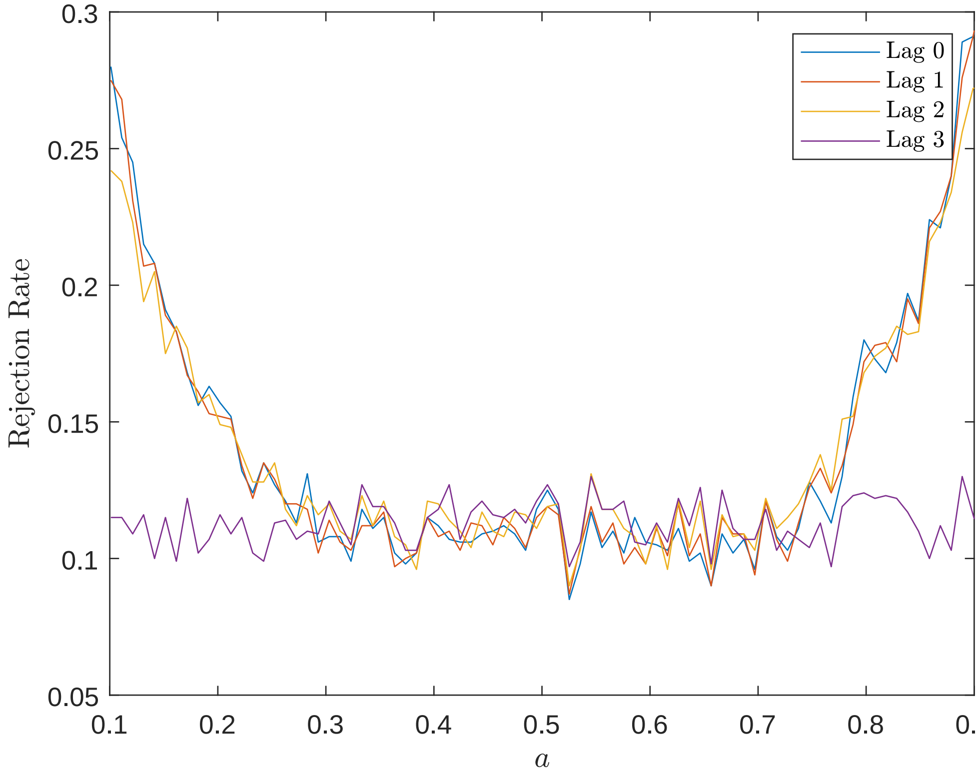


Figure 14: The rejection rate of the trace statistic including different numbers of lags at a nominal level of 10% for Example 8. The results are based on 1000 simulations for a from 0.1 to 0.9 in steps of 0.01. The number of observations in each simulation equals 100.



$\tilde{X}_t$ *and* $\tilde{Y}_t$*, we run into a serious small sample problem. We have only* 100 *observations for* 16 *states. The performance is very weak in particular for* $a < 0.3$ *and* $a > 0.7$*.*

D.2 CMH^* vs CMH: Constant Odds Ratio

Loosely speaking the CMH test aims at situations were the deviation from independence is similar across the different categories of the variable conditioned upon. Suppose that $m_y = m_z = 2$ and that the odds ratio

$$\frac{r^i_{1,1} r^i_{2,2}}{r^i_{1,2} r^i_{2,1}} = c \quad \forall i.$$

Conditional independence implies that $c = 1$. If the odds ratio is constant and different from 1 the CMH statistic should have more power than CMH^*.

Example 9. *Let* $\mathbf{p} = 0.25(a, 1-a, 1-a, a, a, 1-a, 1-a, a)'$. $\mathbf{p}_x = (0.5, 0.5)'$. *The odds ratio equals* $(a/(1-a))^2$ *irrespective whether we condition on* $X_{1,t} = 1$ *or* $X_{2,t} = 1$. *I simulated* 1000 *processes with* $T = 100$ *for each* a *from* 0.01 *to* 0.98 *in steps of* 0.01. *The results are illustrated in Figure 15.*

We observe that the performance of CMH is only marginally better than that of CMH^*. The cost of using CMH^* instead of CMH seems to be low even in this case.

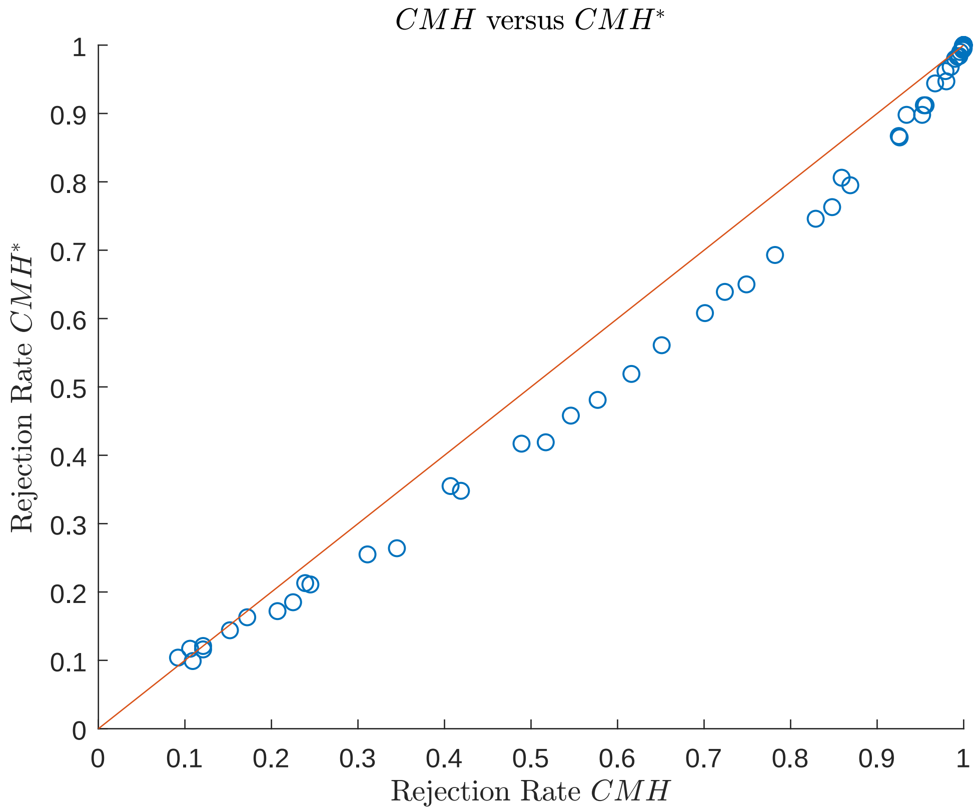
E Small Sample Critical Values

To illustrate size and power of the tests I use the same process as Pesaran and Timmermann [2009]. Let

$$x_t = \phi x_{t-1} + \nu_t \text{ and } y_t = \phi y_{t-1} + \epsilon_t$$

with $\nu_t = r_{xy}\epsilon_t + \sqrt{1 - r^2_{xy}}\eta_t$. Both, η_t and ϵ_t are i.i.d. standard normal variables. The cross-correlation of the increments r_{xy} is set to 0 for the size and to 0.2 for the power simulations. The simulated data is categorized into $m = m_x = m_y$ equally probable bins. In the main text I used the critical values from the asymptotic distribution. W_n, $\hat{X}^2_b$ performed very poorly. Here, I will use estimated critical values for sample sizes of $T = 20, 50, 100$, and 500. The values are summarized in Table 5. Each value is calculated on the basis of 100,000 simulations of discrete i.i.d. random variables with $p_i = 1/m$ for $i = 1, \ldots, m$. The deviations from asymptotic values are sizable for X^2, W_n, $\hat{X}^2_b$, and the trace statistic for $T = 20$. Observe that the

Figure 15: The rejection rate at a nominal level of 10% for Example 9. The results are based on 1000 simulations for a from 0.01 to 0.98 in steps of 0.01. The number of observations in each simulation equals 100.



m	T	X^2	X_n^2	X_M^2	X_{tav}^2	W_M	W_n	$\hat{X}_a^2$	$\hat{X}_b^2$	Trace
2	20	4.105	3.908	3.952	3.925	3.910	2.819	3.925	3.089	4.440
2	50	3.945	3.896	3.922	3.918	3.921	3.883	3.918	3.886	4.073
2	100	3.962	3.921	3.940	3.939	3.939	3.921	3.939	3.898	3.987
2	500	3.867	3.864	3.866	3.866	3.866	3.864	3.866	3.858	3.872
2	∞	3.841	3.841	3.841	3.841	3.841	3.841	3.841	3.841	3.841
3	20	9.228	9.074	9.233	9.141	9.131	6.812	9.163	8.230	11.115
3	50	9.470	9.426	9.479	9.457	9.454	9.367	9.433	9.419	10.256
3	100	9.465	9.451	9.464	9.460	9.464	9.456	9.462	9.477	9.879
3	500	9.532	9.534	9.536	9.535	9.539	9.541	9.538	9.500	9.575
3	∞	9.488	9.488	9.488	9.488	9.488	9.488	9.488	9.488	9.488
4	20	16.250	15.866	16.410	16.088	16.044	12.113	16.025	14.613	15.717
4	50	16.612	16.537	16.589	16.581	16.624	16.350	16.564	16.464	18.704
4	100	16.846	16.839	16.847	16.846	16.854	16.808	16.844	16.784	17.873
4	500	16.890	16.887	16.887	16.887	16.890	16.882	16.886	16.865	17.079
4	∞	16.919	16.919	16.919	16.919	16.919	16.919	16.919	16.919	16.919

Table 5: Critical values for a size of 5% for the various tests. These values are estimated using $100,000$ simulations of discrete i.i.d. random variables with $p_i = 1/m$ for $i = 1, \ldots, m$.

estimated critical values for the trace statistic are quite different from those given in Pesaran and Timmermann [2009].

The performance of the trace statistic is very poor in the case of serial correlation and $m > 2$. The results are displayed in Tables 6 and 7. Tables 8 and 9 summarize the results for the power of the tests.

m	T	X^2	X_n^2	X_M^2	X_{tav}^2	W_M	W_n	$\hat{X}_a^2$	$\hat{X}_b^2$	Trace
2	20	0.051	0.051	0.050	0.053	0.050	0.049	0.053	0.044	0.054
2	50	0.043	0.052	0.048	0.048	0.047	0.053	0.048	0.048	0.047
2	100	0.048	0.046	0.047	0.048	0.047	0.046	0.048	0.045	0.045
2	500	0.054	0.054	0.054	0.054	0.054	0.054	0.054	0.058	0.058
3	20	0.057	0.057	0.053	0.056	0.049	0.050	0.052	0.048	0.046
3	50	0.052	0.049	0.048	0.050	0.050	0.043	0.049	0.052	0.052
3	100	0.049	0.049	0.049	0.049	0.049	0.047	0.050	0.049	0.048
3	500	0.044	0.045	0.044	0.044	0.044	0.044	0.044	0.044	0.045
4	20	0.055	0.055	0.055	0.054	0.046	0.049	0.051	0.059	0.053
4	50	0.030	0.031	0.032	0.031	0.029	0.035	0.031	0.035	0.037
4	100	0.048	0.047	0.048	0.048	0.049	0.048	0.049	0.049	0.047
4	500	0.040	0.039	0.040	0.040	0.040	0.040	0.040	0.039	0.041

Table 6: Size of independence tests: level 5%, no serial correlation ($\psi = 0$), no cross-correlation ($r_{xy=0}$), 1000 simulations.

m	T	X^2	X_n^2	X_M^2	X_{tav}^2	W_M	W_n	$\hat{X}_a^2$	$\hat{X}_b^2$	Trace
2	20	0.146	0.055	0.065	0.058	0.044	0.029	0.058	0.037	0.040
2	50	0.191	0.071	0.076	0.076	0.074	0.042	0.076	0.059	0.059
2	100	0.246	0.068	0.086	0.085	0.087	0.067	0.085	0.055	0.056
2	500	0.210	0.052	0.071	0.071	0.071	0.052	0.071	0.052	0.052
3	20	0.140	0.075	0.098	0.052	0.031	0.028	0.038	0.039	0.019
3	50	0.248	0.073	0.067	0.068	0.056	0.032	0.053	0.053	0.033
3	100	0.282	0.079	0.086	0.087	0.071	0.061	0.070	0.081	0.079
3	500	0.294	0.071	0.090	0.091	0.070	0.046	0.070	0.060	0.059
4	20	0.102	0.063	0.102	0.041	0.021	0.019	0.031	0.025	0.007
4	50	0.205	0.068	0.058	0.052	0.043	0.029	0.038	0.054	0.008
4	100	0.264	0.082	0.074	0.073	0.061	0.063	0.058	0.046	0.009
4	500	0.293	0.081	0.091	0.091	0.057	0.043	0.057	0.050	0.039

Table 7: Size of independence tests: level 5%, serial correlation ($\psi = 0.8$), no cross-correlation ($r_{xy=0}$), 1000 simulations.

m	T	X^2	X^2_n	X^2_M	X^2_{tav}	W_M	W_n	$\hat{X}^2_a$	$\hat{X}^2_b$	Trace
2	20	0.078	0.077	0.076	0.079	0.079	0.077	0.079	0.062	0.080
2	50	0.142	0.152	0.152	0.154	0.152	0.151	0.154	0.142	0.144
2	100	0.244	0.234	0.241	0.240	0.241	0.234	0.240	0.232	0.228
2	500	0.817	0.816	0.818	0.819	0.818	0.816	0.819	0.812	0.812
3	20	0.080	0.085	0.087	0.083	0.081	0.062	0.082	0.053	0.059
3	50	0.114	0.112	0.112	0.111	0.117	0.111	0.116	0.110	0.112
3	100	0.199	0.200	0.198	0.198	0.200	0.194	0.200	0.189	0.190
3	500	0.818	0.819	0.817	0.817	0.817	0.817	0.817	0.818	0.821
4	20	0.073	0.066	0.073	0.075	0.076	0.073	0.079	0.068	0.071
4	50	0.089	0.089	0.089	0.088	0.090	0.089	0.088	0.077	0.078
4	100	0.152	0.155	0.150	0.147	0.147	0.148	0.152	0.150	0.153
4	500	0.787	0.787	0.787	0.787	0.789	0.789	0.790	0.785	0.786

Table 8: Power of independence tests: level 5%, no serial correlation ($\psi = 0$), cross-correlation $r_{xy} = 0.2$, 1000 simulations.

m	T	X^2	X^2_n	X^2_M	X^2_{tav}	W_M	W_n	$\hat{X}^2_a$	$\hat{X}^2_b$	Trace
2	20	0.169	0.081	0.079	0.066	0.052	0.034	0.065	0.048	0.045
2	50	0.275	0.106	0.126	0.123	0.127	0.066	0.123	0.090	0.091
2	100	0.374	0.167	0.194	0.193	0.195	0.163	0.193	0.128	0.127
2	500	0.722	0.455	0.506	0.506	0.506	0.455	0.506	0.441	0.442
3	20	0.163	0.095	0.113	0.065	0.037	0.037	0.049	0.051	0.021
3	50	0.295	0.112	0.107	0.106	0.074	0.042	0.076	0.086	0.055
3	100	0.407	0.161	0.178	0.177	0.113	0.094	0.111	0.133	0.125
3	500	0.746	0.508	0.544	0.544	0.398	0.354	0.397	0.498	0.497
4	20	0.117	0.072	0.101	0.045	0.029	0.031	0.044	0.031	0.006
4	50	0.274	0.113	0.094	0.091	0.064	0.049	0.068	0.078	0.014
4	100	0.399	0.151	0.143	0.142	0.083	0.079	0.083	0.123	0.028
4	500	0.762	0.486	0.507	0.507	0.299	0.272	0.300	0.517	0.443

Table 9: Power of independence tests: level 5%, serial correlation ($\psi = 0.8$), cross-correlation $r_{xy} = 0.2$, 1000 simulations.

Index of Working Papers:

October 13, 2017	Markus Knell	215	Actuarial Deductions for Early Retirement
October 16, 2017	Markus Knell, Helmut Stix	216	Perceptions of Inequality
November 17, 2017	Engelbert J. Dockner, Manuel Mayer, Josef Zechner	217	Sovereign Bond Risk Premiums
December 1, 2017	Stefan Niemann, Paul Pichler	218	Optimal fiscal policy and sovereign debt crises
January 17, 2018	Burkhard Raunig	219	Economic Policy Uncertainty and the Volatility of Sovereign CDS Spreads
February 21, 2018	Andrej Cupak, Pirmin Fessler, Maria Silgoner, Elisabeth Ulbrich	220	Exploring differences in financial literacy across countries: the role of individual characteristics and institutions
May 15, 2018	Peter Lindner, Axel Loeffler, Esther Segalla, Guzel Valitova, Ursula Vogel	221	International monetary policy spillovers through the bank funding channel
May 23, 2018	Christian A. Belabed, Mariya Hake	222	Income inequality and trust in national governments in Central, Eastern and Southeastern Europe
October 16, 2018	Pirmin Fessler, Martin Schürz	223	The functions of wealth: renters, owners and capitalists across Europe and the United States
October 24, 2018	Philipp Poyntner, Thomas Reininger	224	Bail-in and Legacy Assets: Harmonized rules for targeted partial compensation to strengthen the bail-in regime
Dezember 14, 2018	Thomas Breuer, Martin Summer	225	Systematic Systemic Stress Tests
May 20, 2019	Helmut Stix	226	Ownership and purchase intention of crypto-assets – survey results
October 17, 2019	Markus Knell Helmut Stix	227	How Peer Groups Influence Economic Perceptions