Testing Dependence Among Serially Correlated Multi-category Variables*

M. Hashem Pesaran
Cambridge University
Allan Timmermann
University of California, San Diego
May 22, 2008

Abstract

The contingency table literature on tests for dependence among discrete multi-category variables is extensive. Standard tests assume, however, that draws are independent and only limited results exist on the effect of serial dependency—a problem that is important in areas such as economics, finance, medical trials and meteorology. This paper proposes new tests of independence based on canonical correlations from dynamically augmented reduced rank regressions. The tests allow for an arbitrary number of categories as well as multi-way tables of arbitrary dimension and are robust in the presence of serial dependencies that take the form of finite-order Markov processes. For three-way or higher order tables we propose new tests of joint and marginal independence. Monte Carlo experiments show that the proposed tests have good finite sample properties. An empirical application to microeconomic survey data on firms' forecasts of changes to their production and prices demonstrates the importance of correcting for serial dependencies in predictability tests.

Keywords: Contingency Tables, Canonical Correlations, Serial Dependence, Markov Chains

*We benefitted from the comments of two anonymous referees, Herman van Dijk and Adrian Pagan and from help with the data analysis from Klaus Abberger and Klaus Wohlrabe of CESifo. Constructive comments and suggestions by the Joint Editor (Stephen Portnoy) and an Associate Editor are much appreciated.

1 Introduction

Categorized data that are serially dependent are commonplace in many areas of scientific research. In psychological or medical trials repeated measurements of the condition of an individual give rise to serial dependencies in contingency table data (Conaway (1989)) as do observations in sociological studies of individuals' decisions over time (e.g. judges' decisions categorized against the type of case and circuit (Sunstein et al. (2006)) or selection of jurors serving on grand juries (Miao and Gastwirth (2004)). In meteorology, forecast accuracy for categorized weather variables is routinely investigated using contingency tables (Katz and Murphy (1997) and Stephenson (2000)), although these variables are often highly serially correlated. Software failure rates also give rise to serial dependencies in categorical data (Ray, Liu and Ravishanker (2006)). Other types of dependencies have been studied in the analysis of clusters of binary data (e.g. in chemical repellency trials (Gerard and Schucany (2007)) and in the analysis of genetic equilibrium in multidimensional contingency tables (Lazzeroni and Lange (1997)). Finally, in economics and finance recession indicators used to track the business cycle and bull and bear market indicators used to characterize stock markets are examples of serially dependent indicator variables (Harding and Pagan (2006) and Lunde and Timmermann (2004)).

There is a literature in statistics that addresses the effect of serial dependencies on the chi-squared tests of independence applied to two-way contingency tables, including the special 2×2 case common to many applications (Tavare and Altham (1983)), and other robust procedures such as sign tests (Gastwirth and Rubin (1971, 1975), Wolf et al. (1967) and Serfling (1968); see also Portnoy and He (2000) for a recent review and further references). This literature has established that the effect of serial persistence on standard tests for dependence between categorized data can be severe. Assuming that the underlying variables of a two-way contingency table are draws from stationary and reversible Markov processes, under the null that the row and column variables in the contingency table are independent, Tavare (1983) shows that the asymptotic distribution of the Pearson test for independence is a mixture of chi-squared variables whose weights depend on the eigenvalues of the transition matrices. Building on this work, Gleser and Moore (1985) demonstrate that under positive dependence between successive observations, the Pearson chi-squared test has a null distribution that is asymptotically larger than that obtained under serial independence. Porteous (1987) extends Tavare's results to multi-way tables and shows that Pearson's test will be valid when all but one of the variables under consideration are serially independent.

Tavare's result can be used to correct the Pearson chi-squared statistic for serial dependence. However, its implementation faces important difficulties, namely the presence of sampling errors in estimates of the eigenvalues of the transition matrices and the requirement that the Markovian processes are reversible. In practice this could be unduly restrictive and, as noted by Porteous (1987), a test of the reversibility assumption might be needed.

To avoid this restrictive assumption and to bypass the need for estimation of the transition matrices, our paper proposes a new dynamically augmented reduced rank regression approach that allows great flexibility and generality in how serial dependence is treated for contingency tables. The approach leads to a test statistic that is consistent under broad conditions and is very easy to implement.

More specifically, in the case of m_y category realizations ($y_{it},\ i = 1, 2, ..., m_y$), and $m_x \leq m_y$ categories for an associated variable ($x_{jt},\ j = 1, 2, ..., m_x$), we cast the relationship between the $m_y - 1$ category realizations, $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{m_y-1,t})'$, and the $m_x - 1$ category variables, $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{m_y-1,t})$, as a regression of $\mathbf{a}'\mathbf{y}_t$ on $\mathbf{b}'\mathbf{x}_t$ where $\mathbf{a}$ and $\mathbf{b}$ are viewed as nuisance parameters. For serially independent outcomes ($\mathbf{y}_t$), we show that a test of independence between $\mathbf{y}_t$ and $\mathbf{x}_t$ can be based on the canonical correlation coefficients between $\mathbf{y}_t$ and $\mathbf{x}_t$. We consider both a maximum canonical correlation test and a trace test based on the average canonical correlation which is identical to the standard Pearson chi-square contingency table test of independence.

In the case of serially dependent outcomes we show that a valid multi-category dependence test can be constructed using canonical correlations of suitably filtered versions of $\mathbf{y}_t$ and $\mathbf{x}_t$ after accounting for the effect of lagged values of $(\mathbf{y}_t', \mathbf{x}_t')'$. This gives rise to trace and maximum canonical correlation tests based on a dynamically augmented reduced rank regression that is very simple to compute. These tests do not rely on making functional form assumptions regarding the specific relationship between $\mathbf{y}_t$ and $\mathbf{x}_t$ and their underlying distributions, nor do they require maintaining assumptions regarding the nature of the time-series dynamics of $\mathbf{y}_t$ and $\mathbf{x}_t$, other than that they are generated by ergodic, finite-order Markov processes.

Small sample properties of the maximum and trace canonical correlation tests are investigated through Monte Carlo experiments. Tests that ignore serial correlation are generally found to be severely oversized and tend to over-reject when the degree of serial correlation in the outcome variable is high. In contrast, the canonical correlation test based on dynamically augmented regressions generally has the right size unless the number of categories is large relative to the sample size.

We finally extend our results to three-way contingency tables with serially dependent outcomes, where two different types of hypotheses are tested, namely tests of the joint independence of all the three variables, and conditional tests of the independence of any two of the variables conditional on the third.

The plan of the paper is as follows. Section 2 describes test statistics for the static case. Section 3 extends the reduced rank regression test to allow for serial dependencies. Section 4 generalizes the results to multi-way contingency tables of third or higher order. Section 5 presents Monte Carlo simulation results, while Section 6 illustrates the proposed tests through an empirical application to microeconomic data on price and production forecasts. Section 7 concludes. Proofs are given in an Appendix with additional material provided in a Supplement.

2 Two-Way Static Case

Consider a discrete time series (ordered set) of T observations on some explanatory or predictive variable, X, that is arranged into m_x categories (states), while observations on the dependent or realized variable, Y, are categorized into m_y states. Denote the X-categories by x_{jt} so that $x_{jt} = 1$ if the j^{th} category occurs at time t and $x_{jt} = 0$ otherwise. Similarly, denote the realized outcomes by y_{it} so $y_{it} = 1$ if category i occurs at time t and $y_{it} = 0$ otherwise.

Convert the categorical observations into quantitative measures by assigning the weights a_i to y_{it} for $i = 1, 2, ..., m_y$ and b_j to x_{jt} for $j = 1, 2, ..., m_x$ and $t = 1, 2, ..., T$ as follows

$$y_t = \sum_{i=1}^{m_y} a_i y_{it}, \text{ and } x_t = \sum_{j=1}^{m_x} b_j x_{jt}.$$

Since the outcome categories are mutually exclusive and $\sum_{i=1}^{m_y} a_i = \sum_{j=1}^{m_x} b_j = 1$, regression of y_t on an intercept and x_t, $y_t = \alpha + \beta x_t + u_t$, can be written as

$$a_{m_y} + \sum_{i=1}^{m_y-1} \left(a_i - a_{m_y}\right) y_{it} = \alpha + \beta b_{m_x} + \beta \left[\sum_{j=1}^{m_x-1} \left(b_j - b_{m_x}\right) x_{jt}\right] + u_t, \quad \text{or}$$

$$\boldsymbol{\theta}' \mathbf{y}_t = c + \boldsymbol{\gamma}' \mathbf{x}_t + u_t, \tag{1}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{m_y-1,t})'$, $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{m_x-1,t})'$, $c = \alpha + \beta b_{m_x} - a_{m_y}$, $\boldsymbol{\theta} = (a_1 - a_{m_y}, a_2 - a_{m_y}, ..., a_{m_y-1} - a_{m_y})'$, and $\boldsymbol{\gamma} = \left[\beta\left(b_1 - b_{m_x}\right), \beta\left(b_2 - b_{m_x}\right), ..., \beta\left(b_{m_x-1} - b_{m_x}\right)\right]'$. A test of independence can now be carried out by testing $\boldsymbol{\gamma} = \mathbf{0}$ in (1), conditional on a given value of $\boldsymbol{\theta}$. The idea is to construct a quantitative measure by reweighting the categorical data (i.e. by forming linear combinations) and applying regression-based tools to the transformed data. When $m_y = m_x = 2$, the test reduces to testing the significance of the slope coefficient in a regression of y_{1t} on x_{1t}. In the more general case the test of $\boldsymbol{\gamma} = \mathbf{0}$ will depend on the "nuisance" parameters, $\boldsymbol{\theta}$.

2.1 Canonical Correlation Tests

We first consider testing the null hypothesis, $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ conditional on a given value of $\boldsymbol{\theta}$ under classical assumptions applied to u_t conditional on $\mathbf{x}_t$ and then examine the properties of the test for other values of $\boldsymbol{\theta}$.

For a given value of $\boldsymbol{\theta} \in \Theta$, where Θ is a compact set, a standard $F-$statistic can be employed to test independence of y_t and x_t :

$$F(\boldsymbol{\theta}) = \left(\frac{T - m_x}{m_x - 1}\right) \frac{\boldsymbol{\theta}'\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\boldsymbol{\theta}}{\boldsymbol{\theta}'\left(\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\right)\boldsymbol{\theta}}, \tag{2}$$

where $\mathbf{S}_{yx} = \mathbf{S}'_{xy} = T^{-1}\mathbf{Y}'\mathbf{M}_\tau\mathbf{X}$, $\mathbf{S}_{yy} = T^{-1}\mathbf{Y}'\mathbf{M}_\tau\mathbf{Y}$, $\mathbf{S}_{xx} = T^{-1}\mathbf{X}'\mathbf{M}_\tau\mathbf{X}$, and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)'$ and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)'$ are the $T \times (m_y - 1)$ and $T \times (m_x - 1)$ observation matrices on the qualitative indicators, respectively. $\mathbf{M}_\tau = \mathbf{I}_T - \boldsymbol{\tau}_T(\boldsymbol{\tau}'_T\boldsymbol{\tau}_T)^{-1}\boldsymbol{\tau}'_T$, where $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of ones. It is not known *a priori* which element of $\boldsymbol{\theta}$ might be non-zero, so we employ the normalization $\boldsymbol{\theta}'\mathbf{S}_{yy}\boldsymbol{\theta} = 1$.

Throughout the paper we make the following assumption:

Assumption 1: The sequence of sample averages $\bar{x}_{jT} = T^{-1}\sum_{t=1}^{T} 1_{\{x_t = j\}}$, $\bar{y}_{iT} = T^{-1}\sum_{t=1}^{T} 1_{\{y_t = i\}}$ satisfy the conditions $\bar{x}_{jT}(1 - \bar{x}_{jT}) \neq 0$ and $\bar{y}_{iT}(1 - \bar{y}_{iT}) \neq 0$ for all $i = 1, 2, ..., m_y$, $j = 1, 2, ..., m_x$, and for all sample sizes, T.

Assumption 1 (uniform representativeness) requires that all categories are represented in a given sample while categories with zero representation must be dropped. It follows from this assumption that $\mathbf{S}_{xx}$ and $\mathbf{S}_{yy}$ are non-singular matrices. To see this, note that due to the multinomial nature of the underlying data the limits of $\mathbf{S}_{xx}$ and $\mathbf{S}_{yy}$ exist for all T. Furthermore, since the events in the m_x or m_y categories are mutually exclusive, $T^{-1}\mathbf{X}'\mathbf{X}$, and $T^{-1}\mathbf{Y}'\mathbf{Y}$ will be diagonal matrices, with their i^{th} diagonal element given by $\bar{x}_{iT} = T^{-1}\sum_{t=1}^{T} x_{it}$ and $\bar{y}_{iT} = T^{-1}\sum_{t=1}^{T} y_{it}$, respectively. For example, the (i, j) element of $\mathbf{S}_{xx}$ is given by $\bar{x}_{iT}(1 - \bar{x}_{iT})$ if $i = j$ and $-\bar{x}_{iT}\bar{x}_{jT}$ if $i \neq j$. To ensure that $\mathbf{S}_{xx}$ and $\mathbf{S}_{yy}$ are non-singular we must have $\bar{x}_{jT}(1 - \bar{x}_{jT}) \neq 0$ and $\bar{y}_{iT}(1 - \bar{y}_{iT}) \neq 0$ for all $i = 1, 2, .., m_y$ and $j = 1, 2, ..., m_x$ which holds by Assumption 1.

2.1.1 Maximum Canonical Correlation Test

A general approach to dealing with the dependence of $F(\boldsymbol{\theta})$ on the nuisance parameters is to base the test on $F_{\max} = Arg\max_{\theta}[F(\boldsymbol{\theta})]$ subject to the normalizing restriction that $\boldsymbol{\theta}'\mathbf{S}_{yy}\boldsymbol{\theta} = 1$. This idea has been used in the literature (e.g. by Davies (1977)) in cases where certain parameters of the statistical model disappear under the null hypothesis, and has been applied in econometrics to the analysis of non-nested models by Pesaran (1981). Notice, however, that in our case the nuisance parameter, $\boldsymbol{\theta}$, does not disappear under the null.

Using (2), the first order condition for optimization of $F(\boldsymbol{\theta})$ is given by

$$\left(\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\right)\hat{\boldsymbol{\theta}} = \hat{\rho}^2\mathbf{S}_{yy}\hat{\boldsymbol{\theta}}, \tag{3}$$

$$\text{where } \hat{\rho}^2 = \frac{F\left(\hat{\boldsymbol{\theta}}\right)\left(\frac{m_x-1}{T-m_x}\right)}{1 + \left(\frac{m_x-1}{T-m_x}\right)F\left(\hat{\boldsymbol{\theta}}\right)}. \tag{4}$$

The value of $\boldsymbol{\theta}$ that maximizes $F(\boldsymbol{\theta})$ is therefore given by the eigenvector associated with the maximum eigenvalue of

$$\mathbf{S} = \mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}. \tag{5}$$

Denoting the non-zero eigenvalues of $\mathbf{S}$ in descending order by $\hat{\rho}_1^2 \geq \hat{\rho}_2^2 \geq ... \geq \hat{\rho}_{m_x-1}^2$, we have (using (4))

$$F_{\max} = \frac{(T - m_x)\hat{\rho}_1^2}{(m_x - 1)\left(1 - \hat{\rho}_1^2\right)}. \tag{6}$$

Note that $\hat{\rho}_i^2$ $i = 1, 2, ..., m_x - 1$ are the squared canonical correlation coefficients between the indicators, $\mathbf{x}_t$, and the realizations, $\mathbf{y}_t$. The concept of canonical correlations was proposed by Hotelling (1935, 1936) and considers the degree of linear dependence between two random vectors. For categorical data this would involve choosing the weights, a_i, $i = 1, 2, ..., m_y - 1$ and b_j, $j = 1, 2, ..., m_x - 1$ such that the simple correlation between $\sum_{i=1}^{m_y-1} a_i y_{it}$, and $x_t = \sum_{j=1}^{m_x-1} b_j x_{jt}$ is maximized, see Anderson (2003, Ch. 12). There are $m_x - 1$ such canonical correlations that are given by the square roots of the ordered non-zero solutions of the determinantal equation (recall that $m_x \leq m_y$) $|\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} - \rho^2\mathbf{S}_{yy}| = 0$. These are the same as the $m_x - 1$ non-zero eigenvalues of the matrix $\mathbf{S}$ defined by (5). The estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_1$, is given by the eigenvector associated with $\hat{\rho}_1^2$, which satisfies

$$\left(\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} - \hat{\rho}_1^2\mathbf{S}_{yy}\right)\hat{\boldsymbol{\theta}}_1 = \mathbf{0}. \tag{7}$$

Since $\hat{\rho}_1^2 < 1$ and $F_{\max}$ is a monotonic function of $\hat{\rho}_1^2$, a test of $\boldsymbol{\gamma} = \mathbf{0}$ in (1) is thus reduced to testing the statistical significance of the largest canonical correlation between $\mathbf{y}_t$ and $\mathbf{x}_t$. The exact joint probability distribution of the canonical correlations, $1 > \hat{\rho}_1^2 > \hat{\rho}_2^2 > ... > \hat{\rho}_{m_x-1}^2$, is provided in Anderson (2003, pp. 543-545) for the case where the distribution of $\mathbf{y}_t$ conditional on $\mathbf{x}_t$ is Gaussian. In the present application where the elements of $\mathbf{y}_t$ (conditional on $\mathbf{x}_t$) can be viewed as independent draws from a multinominal distribution, the *exact* distribution of the canonical correlations will be less tractable but can readily be simulated.

2.1.2 Trace Canonical Correlation Test

The null of independence between x and y implies not only that $\rho_1 = 0$ but that $\rho_1 = \rho_2 = ... = \rho_{m_x - 1} = 0$. An alternative to the maximum canonical correlation test is therefore to base a test of $\boldsymbol{\gamma} = \mathbf{0}$ on an average of the squared canonical correlations which can also be regarded as an average $F-$test defined by $\bar{F} = (m_x - 1)^{-1} \sum_{i=1}^{m_x - 1} \hat{\rho}_i^2$. This test can also be derived in the context of the reduced rank regression

$$\mathbf{y}_t = \mathbf{a} + \mathbf{\Pi} \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{8}$$

where in our application the null hypothesis of interest is $rank\left(\mathbf{\Pi}\right) = 0$.

Suppose that (Y, X) are draws from serially independent processes with m_y and m_x states, respectively, and that Assumption 1 holds for any finite T and as $T \rightarrow \infty$. Under the null hypothesis that $rank\left(\mathbf{\Pi}\right) = 0$, as $T \rightarrow \infty$,

$$T \times Trace(\mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy}) \overset{a}{\sim} \chi^2_{(m_y - 1)(m_x - 1)}. \tag{9}$$

This is a special case of a more general result that is provided in Section 3 and is proved in the Appendix. When X and Y are serially independent, it is common to arrange the data in an $m_x \times m_y$ contingency table and use a Pearson Chi-square test for independence (see, e.g., Agresti (2007)). Haberman (1981) and Goodman (1985, 1986) explore the relation between this type of test, log-linear models and the canonical correlation test. It can be shown that the Pearson test is identical to the trace test based on canonical correlations (see the supplement for details).

3 Markov Dependence

We next turn to the two-way case with serial dependence. A significant advantage of the maximum canonical correlation and reduced rank regression framework is, as we shall see, that it allows a natural extension of the test to dynamic contexts which does not seem possible within the standard contingency table set up. Before setting out our methods we first describe the Tavare (1983) procedure for dealing with serial dependency in the data generating process.

3.1 The Tavare Test

For two-way $m_x \times m_y$ contingency tables generated by stationary and reversible Markov processes, X and Y, with transition matrices $\mathbf{P}_x$ and $\mathbf{P}_y$, Tavare (1983) shows that, under the null that the row and column variables in the contingency table are independent, the asymptotic distribution of the Pearson test for independence is a mixture of chi-squared variables whose weights depend on the eigenvalues of the associated transition matrices

$$\sum_{i=1}^{m_y-1} \sum_{j=1}^{m_x-1} \left(\frac{1+\lambda_{jx}\lambda_{iy}}{1-\lambda_{jx}\lambda_{iy}} \right) Z_{ij}^2, \quad Z_{ij} \sim iid\ N(0,1), \tag{10}$$

where λ_{jx} and λ_{iy} are the (non-unit) eigenvalues of $\mathbf{P}_x$ and $\mathbf{P}_y$, respectively. Although this procedure can handle dynamic dependencies in the underlying data, its practical implementation is restricted. First, even under the null, the distribution of the critical values depends on the unknown eigenvalues, $\lambda_{jx}, \lambda_{iy}$, which are nuisance parameters that in practice have to be estimated. Second, the requirement that the Markov process be reversible is very restrictive when $m_x > 2$ or $m_y > 2$ and excludes many situations of practical interest. These difficulties become even more serious when higher dimensional tables are considered (Porteous (1987)).

3.2 Dynamically Augmented Reduced Rank Regression

We next propose a test that allows for serial dependencies of arbitrary (but finite) order, does not require reversibility and does not depend on estimates of the transition matrices. To this end consider again the regression model (1) and assume that the errors, u_t, could be serially dependent. To simplify the exposition suppose that u_t follows a stationary first order autoregressive process

$$u_t = \varphi u_{t-1} + \varepsilon_t, \ |\varphi| < 1, \tag{11}$$

where ε_t are serially independent. For this error specification, using (1) we have

$$\boldsymbol{\theta}'\mathbf{y}_t = c\,(1-\varphi) + \boldsymbol{\gamma}'\mathbf{x}_t - \varphi\boldsymbol{\gamma}'\mathbf{x}_{t-1} + \varphi\boldsymbol{\theta}'\mathbf{y}_{t-1} + \varepsilon_t.$$

As in the previous section, a consistent test of $\boldsymbol{\gamma} = \mathbf{0}$ can be carried out using the maximum or the average of the canonical correlation coefficients of $\mathbf{Y}$ and $\mathbf{X}$ after filtering both sets of variables for the effects of $\mathbf{y}_{t-1}$ and $\mathbf{x}_{t-1}$. More specifically, we compute the eigenvalues of $\mathbf{S}_w = \mathbf{S}_{yy,w}^{-1}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-1}\mathbf{S}_{xy,w}$, where

$$\begin{aligned} \mathbf{S}_{yy,w} &= T^{-1}\mathbf{Y}'\mathbf{M}_w\mathbf{Y}, \mathbf{S}_{xx,w} = T^{-1}\mathbf{X}'\mathbf{M}_w\mathbf{X}, \mathbf{S}_{xy,w} = T^{-1}\mathbf{X}'\mathbf{M}_w\mathbf{Y}, \\ \mathbf{M}_w &= \mathbf{I}_T - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}', \text{ with } \mathbf{W} = (\boldsymbol{\tau}, \mathbf{X}_{-1}, \mathbf{Y}_{-1}), \end{aligned} \tag{12}$$

and $\mathbf{X}_{-1}$ and $\mathbf{Y}_{-1}$ are $T \times (m_x - 1)$ and $T \times (m_y - 1)$ observation matrices on $\mathbf{x}_{t-1}$ and $\mathbf{y}_{t-1}$, respectively.

It can be shown that the trace test based on $\mathbf{S}_w$ is the same as testing $\boldsymbol{\Pi} = \mathbf{0}$ in the dynamically augmented reduced rank regression

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Pi}' + \mathbf{W}\mathbf{B} + \mathbf{E}, \tag{13}$$

where $\mathbf{E}$ is a $T \times (m_y - 1)$ matrix of serially uncorrelated errors. We next establish a formal result that relates the properties of ergodic Markov chains to the indicator variables used in the above reduced rank regressions.

Proposition 1: Consider a sequence $\{Y_t;\ t = 0, 1, 2, ...\}$ taking values in $\{1, 2, ..., m\}$, and suppose that Y_t follows a Markov chain with transition probability matrix, $\mathbf{P} = (p_{ij})$. Let $\mathbf{y}_t^0 = (y_{1t}, y_{2t}, ..., y_{mt})'$ be an $m \times 1$ vector where $y_{it} = 1$ if $Y_t = i$ and 0 otherwise. Then $\boldsymbol{\varepsilon}_t^0 = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{mt})'$, defined by $\boldsymbol{\varepsilon}_t^0 = \mathbf{y}_t^0 - \mathbf{P}'\mathbf{y}_{t-1}^0$ is a martingale difference process with respect to the information set $\mathcal{I}_{y,t-1} = (\mathbf{y}_{t-1}^0, \mathbf{y}_{t-2}^0, ...)$.

This result readily extends to higher order Markov chains by appropriately defining the state space. See the Appendix for a proof.

The asymptotic distribution of the trace statistic for testing the independence of two Markov chains can now be stated in the following theorem (proved in the Appendix).

Theorem 1: Suppose that Assumption 1 applies to two independent Markov chains $\{Y_t, X_t; t = 0, 1, 2, ...\}$ with ergodic transition matrices $\mathbf{P}_{y^0}$ and $\mathbf{P}_{x^0}$ such that $Rank(\breve{\mathbf{P}}_{y^0} - \mathbf{P}'_{y^0}\breve{\mathbf{P}}_{y^0}\mathbf{P}_{y^0}) = m_y - 1$ and $Rank(\breve{\mathbf{P}}_{x^0} - \mathbf{P}'_{x^0}\breve{\mathbf{P}}_{x^0}\mathbf{P}_{x^0}) = m_x - 1$, where $\breve{\mathbf{P}}_{y^0} = Diag(p_{1y}, p_{2y}, ..., p_{m_y,y})$, and $\breve{\mathbf{P}}_{x^0} = Diag(p_{1x}, p_{2x}, ..., p_{m_x,x})$. Then as $T \to \infty$,

$$T \times Trace(\mathbf{S}_{yy,w}^{-1}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-1}\mathbf{S}_{xy,w}) \overset{a}{\sim} \chi^2_{(m_y-1)(m_x-1)}, \tag{14}$$

where $\mathbf{S}_{yy}$, $\mathbf{S}_{yx,w}$, and $\mathbf{S}_{xx,w}$ are defined by (12).

Again this theorem extends readily to higher order Markov chains. The rank conditions stated in Theorem 1 ensure that the residuals from the dynamically augmented regressions have non-singular unconditional variance matrices. It is easily established that the rank condition is always met for two- and three-state ergodic Markov chains, and it seems reasonable to expect that it is satisfied more generally.

3.3 Ordered Alternatives

Our tests are based on canonical correlations and so do not use information on any potential ordering of the variables. In cases where the variables are ordered, however, one would expect to get a more powerful test by accounting for ordering information. To see how important this is in practice, and as a natural benchmark, we also consider a test that accounts for ordering information but only applies to the static case.

Building on the work of Olsson (1979) and Ronning and Kukuk (1996), a simple score test is obtained for testing the dependence of y and x when these variables are observed as ordinal measures. The y-categories are specified in terms of the thresholds, $a_0 < a_1 < ... < a_{m_y-1} < a_{m_y}$, while the thresholds of the x-categories are $b_0 < b_1 < ... < b_{m_x-1} < b_{m_x}$. In both cases $a_0 = b_0 = -\infty$, and $a_{m_y} = b_{m_x} = +\infty$. The relative frequencies for the joint occurrence of y and x in their i^{th} and j^{th} categories are denoted by $\hat{\pi}_{ij} = n_{ij}/T$, where n_{ij} is the frequency. Assuming joint normality of the underlying latent variables and random draws, the null hypothesis that x and y are independent can be tested using the score statistic computed as follows:

$$S_\rho = \frac{T\left[\sum_{i=1}^{m_y}\sum_{j=1}^{m_x}\left(\frac{\hat{\pi}_{ij}}{\hat{\pi}_{i.}\hat{\pi}_{.j}}\right)\left[\phi(\hat{a}_i)-\phi(\hat{a}_{i-1})\right]\left[\phi(\hat{b}_j)-\phi(\hat{b}_{j-1})\right]\right]^2}{D}, \qquad (15)$$

where $\hat{\pi}_{i.} = \sum_j \hat{\pi}_{ij}$, $\hat{\pi}_{.j} = \sum_i \hat{\pi}_{ij}$, $\phi(.)$ is the density of a normal random variable, Φ is the associated c.d.f., $\hat{a}_i = \Phi^{-1}(\hat{\pi}_{1.} + \hat{\pi}_{2.} + ... + \hat{\pi}_{i.})$, $i = 1, 2, ..., m_y$, $\hat{b}_j = \Phi^{-1}(\hat{\pi}_{.1} + \hat{\pi}_{.2} + ... + \hat{\pi}_{.j})$, $j = 1, 2, ..., m_x$, with $\Phi(\hat{a}_0) = \Phi(\hat{b}_0) = 0$ and $\Phi(\hat{a}_{m_y}) = \Phi(\hat{b}_{m_x}) = 1$ and the term in the denominator of (15) is given by

$$D = \sum_{i=1}^{m_y}\sum_{j=1}^{m_x}\hat{\pi}_{ij}\frac{\left[\phi(\hat{a}_i)-\phi(\hat{a}_{i-1})\right]^2\left[\phi(\hat{b}_j)-\phi(\hat{b}_{j-1})\right]^2}{\hat{\pi}_{i.}^2\hat{\pi}_{.j}^2} - \sum_{i=1}^{m_y}\sum_{j=1}^{m_x}\left(\frac{\hat{\pi}_{ij}}{\hat{\pi}_{i.}\hat{\pi}_{.j}}\right)\left[\hat{a}_i\phi(\hat{a}_i)-\hat{a}_{i-1}\phi(\hat{a}_{i-1})\right]\left[\hat{b}_j\phi(\hat{b}_j)-\hat{b}_{j-1}\phi(\hat{b}_{j-1})\right]. \quad (16)$$

This test makes use of distributional assumptions to set the threshold values and so can be expected to be more powerful when these assumptions are correct or provide a good approximation. On the other hand, the test assumes that the draws are serially independent and it is complicated to derive an easily computable test of this type in the context of general Markov processes.

4 Serially Correlated Multi-Way Tables

The dynamically augmented reduced rank tests developed for the serially correlated two-way tables can be readily generalized to three-way or higher dimensional tables. However, to simplify the exposition we focus on a three-way table where we allow the underlying innovations to be serially correlated. Consider an $m_y \times m_x \times m_z$ contingency table that cross-classifies T possibly serially correlated observations. Denote the t^{th} observation on the i^{th} category of the first variable, Y, by y_{it}, on the j^{th} category of the second variable, X, by x_{jt}, and on the k^{th} category of the third variable, Z, by z_{kt}. As in the two-way case we suppose that these observations are categorical with y_{it} taking the value of unity if category i $(1 \le i \le m_y)$ of variable Y occurs at time t and zero otherwise, and similarly for x_{jt} and z_{kt}.

With more than two levels there are a variety of associations that could be tested. We test for two types of independence, namely (i) joint independence, and (ii) marginal independence of two of the variables conditional on the third. As in the two-way set up, let $y_t = \sum_{i=1}^{m_y} a_i y_{it}$, $x_t = \sum_{j=1}^{m_x} b_j x_{jt}$, and $z_t = \sum_{k=1}^{m_z} c_k z_{kt}$ and consider the following linear regression

$$y_t = \alpha + \beta x_t + \delta z_t + u_t. \tag{17}$$

Joint independence can be formulated as $\beta = \delta = 0$, and the conditional independence hypothesis as $\beta = 0$ (or $\delta = 0$) allowing for the possibility that $\delta \neq 0$ (or $\beta \neq 0$). Clearly, other conditional independence hypotheses can be considered via regressions of x_t (or z_t) on y_t and z_t (or y_t and x_t).

To deal with the nuisance parameters (a_i, b_j, and c_k), as before we write (17) as

$$\boldsymbol{\theta}'\mathbf{y}_t = c + \boldsymbol{\gamma}'\mathbf{x}_t + \boldsymbol{\lambda}'\mathbf{z}_t + u_t, \tag{18}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{m_y-1,t})'$, $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{m_x-1,t})'$, and $\mathbf{z}_t = (z_{1t}, z_{2t}, ..., z_{m_z-1,t})'$. Joint independence can now be tested through the joint hypothesis on $\boldsymbol{\gamma} = \mathbf{0}$ and $\boldsymbol{\lambda} = \mathbf{0}$, taking account of the dependence of the test on the nuisance parameters $\boldsymbol{\theta}$ in the same manner as above. Similarly, the conditional independence hypotheses can be tested through $\boldsymbol{\gamma} = \mathbf{0}$ or $\boldsymbol{\lambda} = \mathbf{0}$, separately. For the validity of these tests Assumption 1 needs to be extended to cover the observations on the third variable, namely that $\bar{z}_{kT}(1 - \bar{z}_{kT}) \neq 0$, for all $k = 1, ..., m_z$, and all T, where $\bar{z}_{kT} = T^{-1}\sum_{t=1}^{T} 1_{\{z_t = k\}}$.

In the case of serially uncorrelated observations, the maximum eigenvalue or the trace tests can be applied to the canonical correlations of $\mathbf{Y}$ and $\mathbf{Q} = (\mathbf{X}, \mathbf{Z})$ where $\mathbf{Y}$ and $\mathbf{X}$ are as defined above and $\mathbf{Z}$ is the $T \times (m_z - 1)$ matrix of observations on the third variable, namely $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T)$. To test the conditional independence of $\mathbf{Y}$ and $\mathbf{X}$, the maximum and average canonical correlations of $\hat{\mathbf{Y}}_w$ and $\hat{\mathbf{X}}_w$ can be used where $\hat{\mathbf{Y}}_w = \mathbf{M}_w\mathbf{Y}$, $\hat{\mathbf{X}}_w = \mathbf{M}_w\mathbf{X}$, $\mathbf{M}_w = \mathbf{I}_T - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$, with $\mathbf{W} = (\boldsymbol{\tau}, \mathbf{Z})$.

The results for the case of serially correlated observations developed for the two-way classifications also readily extend to this more general setting. For example, the dynamically augmented version of the test of the joint independence hypothesis is defined in terms of the maximum and average canonical correlations of $\hat{\mathbf{Y}}_w = \mathbf{M}_w\mathbf{Y}$ and $\hat{\mathbf{Q}}_w = \mathbf{M}_w\mathbf{Q}$ where $\mathbf{W} = (\boldsymbol{\tau}, \mathbf{Y}_{-1}, \mathbf{X}_{-1}, \mathbf{Z}_{-1})$ for a first order Markov process. Similarly, for the conditional independence test in the presence of first order Markov dependence we need to compute canonical correlations of $\hat{\mathbf{Y}}_w$ and $\hat{\mathbf{X}}_w$ where now $\mathbf{W} = (\boldsymbol{\tau}, \mathbf{Y}_{-1}, \mathbf{X}_{-1}, \mathbf{Z}, \mathbf{Z}_{-1})$. We state the result below:

Theorem 2: Suppose that Assumption 1 applies to the categorized observations $(y_t, x_t, z_t,\ t = 0, 1, 2, ...)$, and assume that $\{Y_t, X_t, Z_t; t = 0, 1, 2, ...\}$ are three independent Markov chains with ergodic transition matrices $\mathbf{P}_{a^0}$ for $a = y, x, z$, respectively, satisfying the rank conditions

$$Rank(\breve{\mathbf{P}}_{a^0} - \mathbf{P}'_{a^0}\breve{\mathbf{P}}_{a^0}\mathbf{P}_{a^0}) = m_a - 1, \text{ for } a = y, x, z,$$

with $\breve{\mathbf{P}}_{a^0} = Diag(p_{1a}, p_{2a}, ..., p_{m_a,a})$, for $a = y, x, z$. Then as $T \to \infty$,

(a) a conditional test of independence of Y and X given Z takes the form

$$T \times Trace(\mathbf{S}_{yy,w}^{-1}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-1}\mathbf{S}_{xy,w}) \overset{a}{\sim} \chi^2_{(m_y-1)(m_x-1)}, \tag{19}$$

where $\mathbf{S}_{yy,w}$ $\mathbf{S}_{xx,w}$, and $\mathbf{S}_{xy,w}$ are defined as before by (12) and $\mathbf{W} = (\boldsymbol{\tau}, \mathbf{Y}_{-1}, \mathbf{X}_{-1}, \mathbf{Z}, \mathbf{Z}_{-1})$.

(b) a test for joint independence of Y, X and Z takes the form

$$T \times Trace(\mathbf{S}_{yy,w}^{-1}\mathbf{S}_{yq,w}\mathbf{S}_{qq,w}^{-1}\mathbf{S}_{qy,w}) \overset{a}{\sim} \chi^2_{(m_x+m_z-2)(m_y-1)}, \tag{20}$$

where

$$\begin{aligned} \mathbf{S}_{yq,w} &= T^{-1}\mathbf{Y}'\mathbf{M}_w\mathbf{Q},\ \mathbf{S}_{qq,w} = T^{-1}\mathbf{Q}'\mathbf{M}_w\mathbf{Q},\ \text{with } \mathbf{Q} = (\mathbf{X}, \mathbf{Z}) \\ \mathbf{M}_w &= \mathbf{I}_T - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}', \mathbf{W} = (\boldsymbol{\tau}, \mathbf{X}_{-1}, \mathbf{Y}_{-1}, \mathbf{Z}_{-1}). \end{aligned}$$

A proof of this theorem can be established along the same lines advanced for Theorem 1 in the appendix.

5 Monte Carlo Simulations

To understand the finite sample properties of the tests considered so far, we next undertake some Monte Carlo experiments for the two-way and three-way tables.

5.1 Two-Way Tables

To capture serial dependence, Y_t was simulated from a first-order autoregressive process with parameter, $\varphi = 0$ or $\varphi = 0.8$, and Gaussian increments. To allow for different degrees of dependence between Y_t and X_t, we considered three values for the cross-correlation of their increments, $r_{yx} = 0.0$, 0.2 and 0.8. Finally, the simulated data were categorized into $m_x = m_y \equiv m$ equally probable bins. Two thousand replications were carried out for each experiment. We report results for sample sizes of $T = 20, 50, 100, 500$ and 1000. More extensive simulation results that account for the effect of higher order dynamics or heteroskedasticity are reported in the supplemental material.

In these simulations we first consider the Tavare and the dynamically augmented reduced rank regression tests when $m = 2$. When $m > 2$, we focus on the trace statistics (9), (14) and the corresponding maximum canonical correlation tests. In each case we assume a critical level of five percent, using chi-squared critical values for the trace test, and simulated critical values for the maximum canonical correlation test (available in the supplemental material.) For the dynamically augmented reduced rank regression (13) that includes lags of $\mathbf{X}_t$ and $\mathbf{Y}_t$, we consider up to four lags, in each case selected using the Akaike Information Criterion.

5.1.1 Comparison with the Tavare Test in the 2x2 Case

We first compared the results of our tests for the 2×2 contingency table which is the most frequently encountered case in practice. Assuming a first-order Markov chain, we use maximum likelihood estimation to obtain estimates of the eigenvalues, $\hat{\lambda}_{ix}$ and $\hat{\lambda}_{iy}$, needed for the computation of the critical values of the mixture χ^2 distribution given by (10). The results, available in the supplemental material, showed that the Tavare test handles serial persistence better than the standard Pearson test, but also that the Tavare test tends to be over-sized even in large samples. This could reflect the difficulty of getting precise estimates of the eigenvalues of the transition probability matrices when they are close to unity. In contrast, the dynamically augmented reduced rank regression tests control the size of the test well even for values of φ close to unity and their power is comparable with that of the Tavare test even when the latter over-rejects.

5.1.2 Size and Power of the Reduced Rank Regression Tests

We next turn to the general case with more than two categories. Table 1 reports the size of the test statistics under a zero correlation between Y_t and X_t ($r_{yx} = 0$), while varying the degree of serial dependence in Y_t, as measured by φ. In the absence of any serial correlation ($\varphi = 0$, in Panel A), the static and dynamic canonical correlation tests generally have the right size as does the test under ordered alternatives.

Turning to the case with serially correlated outcomes ($\varphi = 0.8$, in Panel B), size distortions become very serious for the static canonical correlation test and tend to grow with the sample size. At this level of persistence, rejection rates around 20-30% are common for the static test. The test under ordered alternatives displays a similar behavior with overrejection rates that grow as the sample size rises. In contrast, the dynamically augmented tests that allow for serially correlated outcomes generally have the right size except for being mildly oversized when T is very small relative to m, e.g. $T = 20$ and $m = 4$. They also appear to converge to the right limits in the largest sample sizes and hence properly adjust for serial dependencies.

Figures 1 and 2 plot power curves as a function of the cross-sectional correlation between innovations to X and Y under no persistence ($\varphi = 0$, in Figure 1) and persistence ($\varphi = 0.8$, in Figure 2). The figures assume $m = 3$ and $T = 100$. In the absence of serial persistence, the static and dynamic canonical correlation tests have nearly identical power, whereas the test under ordered alternatives provides power gains as is to be expected since the setup replicates the conditions under which this test was derived. Figure 2 again shows that the static canonical correlation test and the test under ordered alternatives are severely oversized whereas the dynamically augmented test is much better behaved in the presence of serial correlation.

5.2 Three-Way Tables

To investigate the performance of the tests for multi-way tables, we present simulation results for the three-way case with variables X, Y and Z generated as follows. Let $Z \sim MN(m)$ follow an IID multinomial process with m categories $Z = 1, 2, .., m$ each of which has probability $1/m$ so $E(Z) = (m+1)/2$, $Var(Z) = (m^2 - 1)/12$.

Values of Y and X are generated according to the equations

$$\begin{aligned} Y_t &= \theta_y Z_t + u_{yt}, \text{ and } X_t = \theta_x Z_t + u_{xt}, \quad \text{where} \\ \theta_y &= \left(\sqrt{\frac{12}{m^2-1}}\right)\left(\sqrt{\frac{R_y^2}{1-R_y^2}}\right), \text{ and } \theta_x = \left(\sqrt{\frac{12}{m^2-1}}\right)\left(\sqrt{\frac{R_x^2}{1-R_x^2}}\right). \end{aligned}$$

Dynamics is introduced through u_{xt} and u_{yt} :

$$u_{yt} = \phi_y u_{y,t-1} + \sqrt{1-\phi_y^2}\varepsilon_{yt}, \text{ and } u_{xt} = \phi_x u_{x,t-1} + \sqrt{1-\phi_x^2}\varepsilon_{x_t},$$

where $\varepsilon_{yt} = \alpha\varepsilon_{x_t} + \sqrt{1-\alpha^2}v_t$, so that ε_{yt} and ε_{xt} can be correlated. We generate ε_{xt} and v_t from $iidN(0,1)$ draws. Using these series and the m discrete values of Z_t we can then generate Y_t and X_t and categorize these into $m = m_y = m_x = m_z$ equiprobable cells.

This setup has a number of convenient features. In particular, $Corr(Y_t, X_t) = R_y R_x + \alpha\sqrt{1-R_y^2}\sqrt{1-R_x^2}$, where $R_y = Corr(Y_t, Z_t)$, and $R_x = Corr(X_t, Z_t)$. When $\alpha = 0$, any correlation between Y and X come from Z, and is controlled by the parameters R_y and R_x. When $\alpha \neq 0$, correlation between Y and X can come either from Z or from the innovation terms ε_{yt} and ε_{xt}, the dependence on the latter source being controlled by α.

All simulations assume that $\phi_x = \phi_y = \phi$ and $R_x = R_y$. The conditional tests marginalize y and x with respect to the effect of z and are thus conducted on the residuals $\hat{u}_{yt|z} = y_t - \hat{\boldsymbol{\beta}}_{yz}'\tilde{\mathbf{z}}_t$ and $\hat{u}_{xt|z} = x_t - \hat{\boldsymbol{\beta}}_{xz}'\tilde{\mathbf{z}}_t$, where $\tilde{\mathbf{z}}_t = (1, z_{1t}, ..., z_{m-1t})'$. All tests are then based on the $m-1$ column vectors $\hat{u}_{yt|z}$ and $\hat{u}_{xt|z}$.

Size and power of the conditional tests are summarized in part I of Table 2. When data is generated under the null, i.e. $R_y = R_x = 0$ and $\alpha = 0$, the static and augmented tests have approximately the correct size in the absence of serial correlation ($\phi = 0$). Introducing serial correlation by increasing ϕ to 0.8, the static tests systematically over-reject with rejection rates of 20-30%. In contrast, the dynamically augmented test continues to perform well and there is no commensurate increase in the rejection rates for this test.

Setting $\alpha = 0.1$ and $R_y = 0.2$, the conditional test should reject the null of independence between X and Y, and this is what we find, with increasing power as the sample expands. The higher rejection rates associated with the static test in the presence of serial correlation reflects the tendency of this test to over-reject under this scenario. The power of the dynamically augmented test is reduced somewhat when the serial persistence rises. This is a result of having fewer independent observations available when conducting this test.

Results for the joint independence tests are summarized in Panel II of Table 2. When data is generated under the null, once again the static tests overreject in the presence of serial correlation ($\phi = 0.8$), whereas the dynamically augmented tests continue to have the right size. When $\alpha = 0.1$ and $R_y = 0.2$, the joint test should reject since Y is now dependent on X through Z. Again this is what we find as the power rises from around 30% when $T = 100$, to close to 100% when T =1,000.

6 Empirical Application

We finally present an application to microeconomic data on price and production forecasts and actual outcomes for German manufacturing firms collected by the Institute for Economic Research (Ifo). Each month firms are asked to forecast whether their prices and production will decline, stay the same or increase. The following month it is recorded which of these three categories prices or production actually fell into. Since it is much easier for firms to predict outcomes for these three categories than it is to produce a precise number for the magnitude of the change, this data is naturally categorical. More details on the data are provided by Becker and Wohlrabe (2008).

Using a sample of observations from 1995-2004 (T =120 months), we obtained a sample of 521 and 448 firms with complete records of answers to the price and production surveys, respectively. For illustrative purposes we present individual test statistics and p-values for 10 anonymous firms in Table 3. We also present rejection rates and average test statistics for the full set of firms.

The theoretical analysis and Monte Carlo simulations suggest that the canonical correlation tests behave quite differently in the presence of serial correlation in the data. It is therefore important to first test if the underlying data is serially correlated. Using our setup, a test for first-order serial correlation can be carried out by applying either equation (6) or (9) to current and lagged values of the dependent variable, namely by computing the canonical correlations between $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, ..., \mathbf{y}_T)$ and $\mathbf{Y}_{-1} = (\mathbf{y}_0, \mathbf{y}_1, ..., \mathbf{y}_{T-1})$. Tests for higher-order serial dependencies follow as simple extensions of this.

We would expect the data to be serially correlated both because production and inflation are known to be persistent processes and also because survey participants are asked to predict these variables for the next quarter, thus creating an overlap in the data which is recorded at the monthly frequency. Table 3 shows that this is what we find as there is evidence of significant serial dependence for the vast majority of firms.

As a consequence, the static test of independence which ignores serial correlation in the data rejects much more frequently than the tests that account for serial correlation, in the case of the dynamically augmented test by selecting the number of lags by means of the AIC. For example, for firm four's data on prices, the static trace test is 16.6 (p-value of 0.002) and the test under ordered alternatives is 15.7 (0.000), while the dynamically augmented test is 3.1 (0.54). These differences can be attributed to the strong serial correlation in this firm's prices, as shown in the last two columns. In the aggregate, the static tests find evidence of predictability in the price data for 89% of the firms and for 72% of the firms in the production data. The dynamically augmented tests find evidence of predictability for 63% and 43% of the firms for prices and production, respectively−25-30% lower than for the static tests. In view of the strong evidence of serial correlation in the data, this suggests that the static tests clearly overstate the extent to which firms can predict changes in their prices or production levels.

7 Conclusion

This paper proposed new canonical correlation test statistics that can be used for robust inference concerning the relationship between multicategory variables in the presence of serial dependencies. The dynamically augmented reduced rank regression approach developed here is extremely easy to implement and only requires computing a multivariate regression of a set of categorized variables on an intercept and another set of explanatory variables. The need for such tests arises in a variety of applications in areas such as meteorology, psychology, business cycle research, market timing analysis and in the analysis of survey data. One additional advantage of the proposed tests lies in the fact that they can be applied irrespective of whether Y or X or both represent categorical or continuous measurements.

Our Monte Carlo simulations and empirical application demonstrate that standard test statistics that are based on a multinomial setup with draws that are assumed to be independent over time can be severely over-sized in the presence of serial dependencies in the underlying data. In contrast, the dynamically augmented maximum and trace canonical correlation statistics control size well and appear to have good power properties. It is our hope that applied researchers will use the proposed test statistics in the analysis of serially dependent multicategory data.

Appendix

Proof of Proposition 1: Suppose that y_{it} is the realization of a stationary, first-order, m-state Markov process at time t, which takes the value of unity if the i^{th} state is realized and is zero otherwise. The states are mutually exclusive and exhaustive and so $\sum_{i=1}^{m} y_{it} = 1$. Denote the $m \times m$ transition matrix of this process by $\mathbf{P} = (p_{ij})$ such that $p_{ij} = \Pr(y_{jt} = 1 \,|y_{i,t-1} = 1)$. Define $\mathbf{y}_t^0 = (y_{1t}, y_{2t}, ...y_{mt})' = (\mathbf{y}_t', y_{mt})'$, and set

$$\boldsymbol{\varepsilon}_t^0 = \mathbf{y}_t^0 - \mathbf{P}'\mathbf{y}_{t-1}^0,$$

where $\boldsymbol{\varepsilon}_t^0 = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{mt})'$. Denote the i^{th} column of the $m \times m$ identity matrix by $\mathbf{e}_i$, and note that $\mathbf{y}_t^0$ and $\mathbf{y}_{t-1}^0$ can only take the vector values $\mathbf{e}_i$ for $i = 1, 2, ..., m$. Accordingly, $\boldsymbol{\varepsilon}_t^0$ can take the m^2 values $\mathbf{e}_j - \mathbf{P}'\mathbf{e}_i$ for i and $j = 1, 2, ..., m$, with

$E\left|\varepsilon_{jt}\right| < 1 + \max_i(p_{ij})$. Also,

$$E(\boldsymbol{\varepsilon}_t^0 \left| y_{k,t-1} = 1 \right.) = E(\boldsymbol{\varepsilon}_t^0 \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.) = E(\mathbf{y}_t^0 \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.) - \mathbf{P}'\mathbf{e}_k.$$

But $E(\mathbf{y}_t^0 \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.) = (E(y_{1t} \left| y_{k,t-1} = 1 \right.), ..., E(y_{mt} \left| y_{k,t-1} = 1 \right.))'$, and

$$E(y_{jt} \left| y_{k,t-1} = 1 \right.) = 1 \times \Pr(y_{jt} = 1 \left| y_{k,t-1} = 1 \right.) + 0 \times \Pr(y_{jt} = 0 \left| y_{k,t-1} = 1 \right.) = p_{kj}.$$

Hence

$$E(\mathbf{y}_t^0 \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.) = (p_{k1}, p_{k2}, ..., p_{km})' = \mathbf{P}'\mathbf{e}_k, \quad \text{and so}$$

$$E(\boldsymbol{\varepsilon}_t^0 \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.) = \mathbf{0}, \text{ for all } k, \text{ or}$$

$$E(\boldsymbol{\varepsilon}_t^0 \left| \mathbf{y}_{t-1}^0 \right.) = \mathbf{0}.$$

Following the same line of reasoning but taking expectations with respect to $\mathcal{I}_{t-1} = (\mathbf{y}_{t-1}^0, \mathbf{y}_{t-2}^0, ...)$ it follows that $E(\boldsymbol{\varepsilon}_t^0 \left| \mathcal{I}_{t-1} \right.) = 0$, and $\boldsymbol{\varepsilon}_t^0$ is a martingale difference process with respect to $\mathcal{I}_{t-1}$. Hence, $E(\boldsymbol{\varepsilon}_t^0) = 0$. Also,

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t^0 \boldsymbol{\varepsilon}_t^{0\prime} \left| \mathcal{I}_{t-1} \right.) &= E\left(\mathbf{y}_t^0 \mathbf{y}_t^{0\prime} \left| \mathcal{I}_{t-1} \right.\right) - E\left(\mathbf{y}_t^0 \left| \mathcal{I}_{t-1} \right.\right) \mathbf{y}_{t-1}^{0\prime} \mathbf{P} \\ &\quad - \mathbf{P}' \mathbf{y}_{t-1}^0 E\left(\mathbf{y}_t^{0\prime} \left| \mathcal{I}_{t-1} \right.\right) + \mathbf{P}' \mathbf{y}_{t-1}^0 \mathbf{y}_{t-1}^{0\prime} \mathbf{P}, \end{aligned}$$

and given the Markov property

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t^0 \boldsymbol{\varepsilon}_t^{0\prime} \left| \mathcal{I}_{t-1} \right.) &= E(\boldsymbol{\varepsilon}_t^0 \boldsymbol{\varepsilon}_t^{0\prime} \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.), \text{ for } k = 1, 2, ..., m. \\ &= E\left(\mathbf{y}_t^0 \mathbf{y}_t^{0\prime} \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.\right) - E\left(\mathbf{y}_t^0 \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.\right) \mathbf{e}_k' \mathbf{P} \\ &\quad - \mathbf{P}' \mathbf{e}_k E\left(\mathbf{y}_t^{0\prime} \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.\right) + \mathbf{P}' \mathbf{e}_k \mathbf{e}_k' \mathbf{P}, \end{aligned}$$

But $E\left(\mathbf{y}_t^0 \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.\right) = \mathbf{P}'\mathbf{e}_k = \mathbf{p}_k = (p_{k1}, p_{k2}, ..., p_{km})'$, and $E\left(\mathbf{y}_t^0 \mathbf{y}_t^{0\prime} \left| \mathbf{y}_{t-1}^0 = \mathbf{e}_k \right.\right) = \breve{\mathbf{P}}_k$, where $\breve{\mathbf{P}}_k$ is an $m \times m$ diagonal matrix with p_{ki} on its i^{th} diagonal element. Using these results

$$E(\boldsymbol{\varepsilon}_t^0\boldsymbol{\varepsilon}_t^{0\prime}\left|\mathcal{I}_{t-1}\right.) = E(\boldsymbol{\varepsilon}_t^0\boldsymbol{\varepsilon}_t^{0\prime}\left|\mathbf{y}_{t-1}^0 = \mathbf{e}_k\right.) = \breve{\mathbf{P}}_k - \mathbf{p}_k\mathbf{p}_k', \ k = 1, 2, .., m,$$

which establishes that $E(\boldsymbol{\varepsilon}_t^0\boldsymbol{\varepsilon}_t^{0\prime}\left|\mathcal{I}_{t-1}\right.)$ is conditionally heteroskedastic and takes the m different values $\boldsymbol{\Sigma}_{\varepsilon\varepsilon,k}^0 = \breve{\mathbf{P}}_k - \mathbf{p}_k\mathbf{p}_k'$, $k = 1, 2, ..., m$, depending on the particular state realized in the previous period (step). But in the case of ergodic Markov chains the unconditional variance matrix of $\boldsymbol{\varepsilon}_t^0$ is time-invariant and is given by

$$V(\boldsymbol{\varepsilon}_t^0) = \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^0 = E(\boldsymbol{\varepsilon}_t^0\boldsymbol{\varepsilon}_t^{0\prime}) = \sum_{k=1}^{m} p_k \left(\breve{\mathbf{P}}_k - \mathbf{p}_k\mathbf{p}_k'\right),$$

where $p_k = \Pr(y_{kt} = 1)$, $k = 1, 2, ..., m$ is the equilibrium probability distribution associated with $\mathbf{P}$, defined by $(\mathbf{I}_m - \mathbf{P}')\mathbf{p} = \mathbf{0}$, where $\mathbf{p} = (p_1, p_2, ..., p_m)'$, with $\sum_{i=1}^m p_i = 1$. $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^0$ can be equivalently written more compactly as $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^0 = \breve{\mathbf{P}} - \mathbf{P}'\breve{\mathbf{P}}\mathbf{P}$, where $\breve{\mathbf{P}}$ is a diagonal $m \times m$ matrix with p_i as its i^{th} diagonal element. Since $\mathbf{P}\boldsymbol{\tau}_m = \boldsymbol{\tau}_m$, where $\boldsymbol{\tau}_m$ is an $m \times 1$ unit vector, then $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^0\boldsymbol{\tau}_m = \mathbf{0}$, and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^0$ can have at most rank $m-1$. In fact it is easily seen that $\boldsymbol{\tau}_m'\boldsymbol{\varepsilon}_t^0 = \mathbf{0}$.

Consider now the first $m-1$ equations of $\boldsymbol{\varepsilon}_t^0 = \mathbf{y}_t^0 - \mathbf{P}'\mathbf{y}_{t-1}^0$, and using $\sum_{i=1}^m y_{it} = 1$ note that we also have

$$\mathbf{y}_t = \mathbf{a} + \boldsymbol{\Psi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{21}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{m-1,t})'$, $a_i = p_{mi}$ and $\boldsymbol{\Psi} = (\psi_{ij})$ with $\psi_{ij} = p_{ji} - p_{mi}$, for i and $j = 1, 2, ..., m-1$. Since $\boldsymbol{\varepsilon}_t^0 = (\boldsymbol{\varepsilon}_t', \varepsilon_{mt})'$, it also follows that $\boldsymbol{\varepsilon}_t$ in the above equation is a martingale difference process with respect to $\mathcal{I}_{t-1}$. Clearly, $\mathbf{a}$ and $\boldsymbol{\Psi}$ are uniquely determined in terms of the parameters of the transition matrix of the underlying Markov process.

Extension of the above results to higher order Markov chains is straightforward because higher order Markov chains can be reduced to first order Markov chains by appropriately defining the state space, see Cox and Miller (1965, pp. 132-133).

Proof of Theorem 1: Under the null hypothesis $\mathbf{y}_t$ and $\mathbf{x}_t$ follow independent, ergodic Markov chains. From Proposition 1 we have

$$\mathbf{y}_t = \mathbf{a}_y + \Psi_y \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{22}$$

$$\mathbf{x}_t = \mathbf{a}_x + \Psi_x \mathbf{x}_{t-1} + \mathbf{u}_t, \tag{23}$$

where $\boldsymbol{\varepsilon}_t$ and $\mathbf{u}_t$ are independent martingale difference processes with respect to the information set $\mathcal{I}_{t-1} = (\mathbf{y}_{t-1}, \mathbf{x}_{t-1}, \mathbf{y}_{t-2}, \mathbf{x}_{t-2}, ...)$. For the sample observations $t = 0, 1, 2, ..., T$, the above relations can be written as

$$\mathbf{Y} = \mathbf{W}\mathbf{B}_y + \mathbf{E}, \text{ and } \mathbf{X} = \mathbf{W}\mathbf{B}_x + \mathbf{U}, \tag{24}$$

where $\mathbf{Y}' = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T)$, $\mathbf{X}' = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)$, are $T \times (m_y - 1)$ and $T \times (m_x - 1)$, observation matrices on y_t and x_t, $\mathbf{E}' = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, ..., \boldsymbol{\varepsilon}_T)$, and $\mathbf{U}' = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_T)$, are the associated error matrices, $\mathbf{W} = (\boldsymbol{\tau}, \mathbf{Y}_{-1}, \mathbf{X}_{-1})$, and $\mathbf{Y}_{-1}$ and $\mathbf{X}_{-1}$ are $T \times (m_y - 1)$ and $T \times (m_x - 1)$ observation matrices on $\mathbf{x}_{t-1}$ and $\mathbf{y}_{t-1}$, respectively. $\mathbf{B}_y$ and $\mathbf{B}_x$ are fixed parameter matrices that, as in Proposition 1, are uniquely determined from the transition probabilities of the underlying Markov processes.

Consider the matrix associated with the canonical correlation test statistics

$$\mathbf{S} = \mathbf{S}_{yy,w}^{-1} \mathbf{S}_{yx,w} \mathbf{S}_{xx,w}^{-1} \mathbf{S}_{xy,w}$$

Under the null hypothesis, using (24), we can write,

$$\begin{aligned} \mathbf{S}_{yy,w} &= T^{-1}\mathbf{E}'\mathbf{M}_w\mathbf{E} = \left(\frac{\mathbf{E}'\mathbf{E}}{T}\right) - \left(\frac{\mathbf{E}'\mathbf{W}}{T}\right)\left(\frac{\mathbf{W}'\mathbf{W}}{T}\right)^{-1}\left(\frac{\mathbf{W}'\mathbf{E}}{T}\right), \\ \mathbf{S}_{xx,w} &= T^{-1}\mathbf{U}'\mathbf{M}_w\mathbf{U} = \left(\frac{\mathbf{U}'\mathbf{U}}{T}\right) - \left(\frac{\mathbf{U}'\mathbf{W}}{T}\right)\left(\frac{\mathbf{W}'\mathbf{W}}{T}\right)^{-1}\left(\frac{\mathbf{W}'\mathbf{U}}{T}\right). \end{aligned}$$

Note that

$$T^{-1}\mathbf{E}'\mathbf{E} = T^{-1}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t', \; T^{-1}\mathbf{E}'\mathbf{W} = T^{-1}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t\mathbf{w}_t',$$

where $\mathbf{w}_t = (1, y_{t-1}, x_{t-1})'$. From Proposition 1, $\boldsymbol{\varepsilon}_t$ is a martingale difference process and all moments of y_{t-r} and x_{t-s}, for all r and s exist. Then by standard central limit theorems for martingale difference processes (for example, given in proposition 7.9 in Hamilton (1994, p.194)), it readily follows that

$$p\lim_{T\to\infty} T^{-1}\mathbf{E}'\mathbf{E} = \mathbf{\Sigma}_{\varepsilon\varepsilon}, \; p\lim_{T\to\infty} T^{-1}\mathbf{E}'\mathbf{W} = \mathbf{0}, \; p\lim_{T\to\infty} T^{-1}\mathbf{W}'\mathbf{W} = \mathbf{\Sigma}_{ww},$$

where $\mathbf{\Sigma}_{\varepsilon\varepsilon} = Var(\boldsymbol{\varepsilon}_t)$ and

$$\mathbf{\Sigma}_{ww} = \begin{pmatrix} 1 & \mathbf{p}_y' & \mathbf{p}_x' \\ \mathbf{p}_y & \check{\mathbf{P}}_y & \mathbf{0} \\ \mathbf{p}_x & \mathbf{0} & \check{\mathbf{P}}_x \end{pmatrix},$$

with $\mathbf{p}_y' = (p_{y1}, p_{y2}, ..., p_{y,m_y-1})$, $\mathbf{p}_x' = (p_{x1}, p_{x2}, ..., p_{x,m_x-1})$, and $\check{\mathbf{P}}_y$ and $\check{\mathbf{P}}_x$ being diagonal matrices with p_{yi} and p_{xj} as their i^{th}, $i = 1, 2, ..., m_y - 1$ and j^{th}, $j = 1, 2, ..., m_x - 1$ diagonal elements, respectively. Note that since the underlying Markov chains are assumed to be ergodic then $p_{yi} = \Pr(y_{it} = 1) > 0$ and $p_{xj} = \Pr(x_{jt} = 1) > 0$, for $i = 1, 2, .., m_y - 1$, and $j = 1, 2, ..., m_x - 1$. Hence $\mathbf{\Sigma}_{ww}$ will be a non-singular matrix. Also $\mathbf{\Sigma}_{\varepsilon\varepsilon}$, which is given by the first $m_y - 1$ rows and columns of $\check{\mathbf{P}}_{y^0} - \mathbf{P}_{y^0}'\check{\mathbf{P}}_{y^0}\mathbf{P}_{y^0}$, is non-singular by assumption. Similarly,

$$p\lim_{T\to\infty} T^{-1}\mathbf{U}'\mathbf{U} = \mathbf{\Sigma}_{\mathbf{uu}}, \text{ and } p\lim_{T\to\infty} T^{-1}\mathbf{U}'\mathbf{W} = \mathbf{0},$$

where $\Sigma_{uu} = Var(\mathbf{u}_t)$, given by the first $m_x - 1$ rows and columns of $\check{\mathbf{P}}_{x^0} - \mathbf{P}_{x^0}'\check{\mathbf{P}}_{x^0}\mathbf{P}_{x^0}$, is a non-singular matrix by assumption. Consider now the cross terms

$$\sqrt{T}\mathbf{S}_{yx,w} = T^{-1/2}\mathbf{E}'\mathbf{M}_w\mathbf{U} = \left(\frac{\mathbf{E}'\mathbf{U}}{\sqrt{T}}\right) - T^{-1/2}\left(\frac{\mathbf{E}'\mathbf{W}}{\sqrt{T}}\right)\left(\frac{\mathbf{W}'\mathbf{W}}{T}\right)^{-1}\left(\frac{\mathbf{W}'\mathbf{U}}{\sqrt{T}}\right).$$

Again, due to the martingale difference properties of $\boldsymbol{\varepsilon}_t$ and $\mathbf{u}_t$ it readily follows that $T^{-1/2}\mathbf{E}'\mathbf{W}$ and $T^{-1/2}\mathbf{U}'\mathbf{W}$ are both $O_p(1)$, and hence

$$\sqrt{T}\mathbf{S}_{yx,w} = \left(\frac{\mathbf{E}'\mathbf{U}}{\sqrt{T}}\right) + O_p(T^{-1/2}).$$

Using these results the trace statistic can be written as

$$\begin{aligned} \mathfrak{S}_T &= T \times Trace\left(\mathbf{S}_{yy,w}^{-1/2}\mathbf{S}_{yx,w}\mathbf{S}_{xx,w}^{-1/2}\mathbf{S}_{xx,w}^{-1/2}\mathbf{S}_{xy,w}\mathbf{S}_{yy,w}^{-1/2}\right) \\ &= Trace\left(\frac{\tilde{\mathbf{E}}'\tilde{\mathbf{U}}}{\sqrt{T}}\frac{\tilde{\mathbf{U}}'\tilde{\mathbf{E}}}{\sqrt{T}}\right) + O_p(T^{-1/2}), \quad \text{where} \end{aligned}$$

$$\tilde{\mathbf{E}}' = \Sigma_{\varepsilon\varepsilon}^{-1/2}\mathbf{E}' \text{ and } \tilde{\mathbf{U}}' = \Sigma_{\text{uu}}^{-1/2}\mathbf{U}'.$$

Denote $\mathbf{G}_T = T^{-1/2}\tilde{\mathbf{E}}'\tilde{\mathbf{U}}$ and note that

$$\mathfrak{S}_T = \sum_{i=1}^{m_{y-1}}\sum_{j=1}^{m_x-1} g_{ij,T}^2 + O_p(T^{-1/2}).$$

where g_{ij} is the $(i,j)^{th}$ element of $\mathbf{G}_T$. Also

$$g_{ij,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{\varepsilon}_{it}\tilde{u}_{jt},$$

where $\tilde{\varepsilon}_{it}$ and $\tilde{u}_{jt}$ are independent martingale difference processes with zero means and unit variances. Hence for each i and j, $g_{ij,T} \to_d N(0,1)$ as $T \to \infty$. It is also easily verified that $g_{ij,T}$ and $g_{rs,T}$ for $i \neq r$ or $j \neq s$ are independently distributed as $T \to \infty$. Therefore, $g_{ij,T}^2$ are independent χ^2 variates with one degree of freedom each, which establishes that for finite m_y and m_x and as $T \to \infty$ we have

$$\mathfrak{S}_T \ \to_d \chi^2_{(m_y-1)(m_x-1)}.$$

The extension of the proof to higher order Markov processes is again straightforward and can be carried out by re-defining $\mathbf{W}$ to include higher order lagged values of the observation matrices on y_t and x_t. Note also that the results from Section 2 (static process) follow as a special case of the proofs here, when $\mathbf{P}_x = \mathbf{P}_y = \mathbf{0}$.

Finally, the results also establish that the distribution of the maximum canonical correlation statistic is asymptotically equivalent to the distribution of the maximum eigenvalue of $T^{-1}\tilde{\mathbf{E}}'\tilde{\mathbf{U}}\tilde{\mathbf{U}}'\tilde{\mathbf{E}}$ (or $T^{-1}\tilde{\mathbf{U}}'\tilde{\mathbf{E}}\tilde{\mathbf{E}}'\tilde{\mathbf{U}}$) which is free of nuisance parameters and whose critical values can be computed by stochastic simulations. Since $T^{-1}\tilde{\mathbf{E}}'\tilde{\mathbf{U}}\tilde{\mathbf{U}}'\tilde{\mathbf{E}}$ and $T^{-1}\tilde{\mathbf{U}}'\tilde{\mathbf{E}}\tilde{\mathbf{E}}'\tilde{\mathbf{U}}$ have the same maximum eigenvalue, one could use $T^{-1}\tilde{\mathbf{E}}'\tilde{\mathbf{U}}\tilde{\mathbf{U}}'\tilde{\mathbf{E}}$ if $m_x < m_y$ or else use $T^{-1}\tilde{\mathbf{E}}'\tilde{\mathbf{U}}\tilde{\mathbf{U}}'\tilde{\mathbf{E}}$ if $m_y < m_x$. However, in general the critical values of the Max eigenvalue test will depend on both $m_x - 1$ and $m_y - 1$.

References

[1] Agresti, A., 2007, An Introduction to Categorical Data Analysis: Second Edition. Wiley: New Jersey.

[2] Anderson, T.W., 2003, An Introduction to Multivariate Statistical Analysis, Third Edition, Wiley Series in Probability and Statistics. Wiley: New York.

[3] Becker, S. and K. Wohlrabe, 2008, Micro Data at the Ifo Institute for Economic Research - The Ifo Business Survey, Usage and Access. Journal of Applied Social Science Studies, 128(2), 307-319.

[4] Conaway, M.R., 1989, Analysis of Repeated Categorical Measurements with Conditional Likelihood Methods. Journal of the American Statistical Association 84, 405, 53-62.

[5] Cox, D.R. and H.D. Miller, 1965, The Theory of Stochastic Processes. Chapman and Hall. London.

[6] Davies, R.B., 1977, Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative, Biometrika, 64, 247-254.

[7] Gastwirth, J.L., and H. Rubin, 1971, Effect of Dependence on the Level of Some One-Sample Tests. Journal of the American Statistical Association 66, 816-820.

[8] Gastwirth, J.L., and H. Rubin, 1975, The Behavior of Robust Estimators on Dependent Data. Annals of Statistics 3, 1070-1100.

[9] Gerard, P.D. and W. R. Schucany, 2007, An Enhanced Sign Test for Dependent Binary Data with Small Numbers of Clusters. Computational Statistics and Data Analysis 51, 4622-4632.

[10] Gleser, L.J. and D.S. Moore, 1985, The Effect of Positive Dependence on Chi-squared Tests for Categorical Data. Journal of Royal Statistical Society B, 47, 3, 459-465.

[11] Goodman, L.A., 1985, The Analysis of Cross-Classified Data Having Ordered and/or Unordered Categories: Association Models, Correlation Models, and Asymmetry Models for Contingency Tables With or Without Missing Entries. Annals of Statistics 13, 10-69.

[12] Goodman, L.A., 1986, Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach in the Analysis of Contingency Tables. International Statistical Review 54, 243-270.

[13] Haberman, S.J., 1981, Tests for Independence in Two-Way Contingency Tables Based on Canonical Correlation and on Linear-By-Linear Interaction. Annals of Statistics 9, 1178-1186.

[14] Hamilton, J.D., 1994, Time Series Analysis. Princeton University Press: Princeton.

[15] Harding, D. and A.R. Pagan, 2006, Synchronization of Cycles. Journal of Econometrics 132, 59-79.

[16] Hotelling, H., 1935, The Most Predictable Criterion. Journal of Educational Psychology, 26, 139-142.

[17] Hotelling, H., 1936, Relations Between Two Sets of Variates. Biometrika, 28, 321-377.

[18] Katz, R.W., and A.H. Murphy, 1997, Economic Value of Weather and Climate Forecasts, (eds.), Cambridge University Press, Cambridge.

[19] Lazzeroni, L.C. and K. Lange, 1997, Markov Chains for Monte Carlo Tests of Genetic Equilibrium in Multidimensional Contingency Tables. Annals of Statistics 25 (1), 138-168.

[20] Lunde, A. and A. Timmermann, 2004, Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets. Journal of Business and Economic Statistics 22, 253-273.

[21] Miao,W., and J.L. Gastwirth, 2004. The Effect of Dependence on Confidence Intervals for a Population Proportion. American Statistician 58, 124–130.

[22] Olsson, U., 1979, Maximum Likelihood Estimation of the Polychloric Correlation Coefficient. Psychometrika 44, 443-460.

[23] Pesaran, M.H., 1981, Pitfalls of Testing Non-nested Hypotheses by the Lagrange Multiplier Method, Journal of Econometrics 17, 323-331.

[24] Porteous, B. T., 1987, The Mutual Independence Hypothesis for Categorical Data in Complex Sampling Schemes. Biometrika 74, 857-862.

[25] Portnoy, S. and X. He, 2000, A Robust Journey in the New Millennium. Journal of American Statistical Association 95, 1331-1335.

[26] Ray, B.K., Z. Liu and N. Ravishanker, 2006, Dynamic Reliability Models for Software Using Time-Dependent Covariates. Technometrics 48(1), 1-10.

[27] Ronning, G. and M. Kukuk, 1996, Efficient Estimation of Ordered Probit Models. Journal of American Statistical Association 91, 1120-1129.

[28] Serfling, R.J., 1968, The Wilcoxon Two-Sample Statistics on Strongly Mixing Processes. Annals of Mathematical Statistics 39, 1202-1209.

[29] Stephenson, D. B., 2000, Use of the "Odds Ratio" for Diagnosing Forecast Skill, Weather and Forecasting 15, 221-232.

[30] Sunstein, C.R., D. Schkade, L.M. Ellman and A. Sawicki, 2006, Are Judges Political? An Empirical Analysis of the Federal Judiciary. Brookings Institution Press: Washington, D.C.

[31] Tavare, S., 1983, Serial Dependence in Contingency Tables. Journal of Royal Statistical Society B, 45(1), 100-106.

[32] Tavare, S. and P. M. E. Altham, 1983, Serial Dependence of Observations Leading to Contingency Tables, and Corrections to Chi-squared Statistics. Biometrika 70(1), 139-144.

[33] Wolf, S.S., J.L. Gastwirth, and H. Rubin, 1967, The Effect of Autoregressive Dependence on a Nonparametric Test. IEEE Transaction on Information Theory, IT-13, 311-13.

Table 1. Size of the Reduced Rank Regression Tests in Two-Way Tables ($r_{xy} = 0$)

A. No Serial Correlation ($\varphi = 0$)						
m	**Sample Size**	**Trace Canonical Correlation**		**Maximum Canonical Correlation**		**Ordered Alternatives**
		Static	**Dyn. Augm.**	**Static**	**Dyn. Augm.**	
2	20	0.043	0.058	0.043	0.058	0.034
2	50	0.030	0.044	0.030	0.044	0.027
2	100	0.059	0.049	0.059	0.049	0.027
2	500	0.052	0.048	0.052	0.048	0.036
2	1000	0.045	0.050	0.045	0.050	0.049
3	20	0.020	0.066	0.022	0.061	0.044
3	50	0.042	0.052	0.040	0.055	0.048
3	100	0.050	0.057	0.047	0.057	0.045
3	500	0.056	0.056	0.058	0.059	0.049
3	1000	0.056	0.054	0.054	0.053	0.053
4	20	0.008	0.091	0.002	0.042	0.034
4	50	0.032	0.062	0.030	0.056	0.044
4	100	0.040	0.059	0.040	0.056	0.048
4	500	0.046	0.049	0.046	0.050	0.050
4	1000	0.050	0.052	0.051	0.053	0.052
B. **Serial Correlation ($\varphi = 0.8$)**						
m	**Sample Size**	**Trace Canonical Correlation**		**Maximum Canonical Correlation**		**Ordered Alternatives**
		Static	**Dyn. Augm.**	**Static**	**Dyn. Augm.**	
2	20	0.143	0.070	0.143	0.070	0.151
2	50	0.185	0.061	0.185	0.061	0.165
2	100	0.232	0.053	0.232	0.053	0.168
2	500	0.236	0.048	0.236	0.048	0.204
2	1000	0.223	0.051	0.223	0.051	0.230
3	20	0.088	0.082	0.086	0.078	0.200
3	50	0.204	0.057	0.204	0.060	0.260
3	100	0.257	0.058	0.258	0.062	0.278
3	500	0.288	0.052	0.294	0.051	0.297
3	1000	0.287	0.051	0.285	0.054	0.288
4	20	0.031	0.117	0.007	0.057	0.203
4	50	0.166	0.070	0.163	0.061	0.268
4	100	0.234	0.057	0.238	0.056	0.282
4	500	0.294	0.052	0.292	0.054	0.310
4	1000	0.305	0.050	0.305	0.051	0.322

Table 2. Size and Power of Conditional and Joint Tests in Three-Way Tables

I. Conditional Tests

	A. Size: No Serial Corr. ($\varphi = 0,\ \alpha = 0,\ R_y = 0$)					**C. Power: No Serial Corr.** ($\varphi = 0,\ \alpha = 0.1,\ R_y = 0.2$)			
m	Sample Size	Trace		Maximum		Trace		Maximum	
		Static	Dyn.	Static	Dyn.	Static	Dyn.	Static	Dyn.
2	100	0.045	0.047	0.045	0.047	0.095	0.098	0.095	0.098
2	500	0.046	0.046	0.046	0.046	0.281	0.284	0.281	0.284
2	1000	0.056	0.056	0.056	0.056	0.503	0.504	0.503	0.504
3	100	0.049	0.052	0.047	0.053	0.075	0.084	0.073	0.084
3	500	0.058	0.059	0.061	0.067	0.253	0.254	0.257	0.264
3	1000	0.049	0.051	0.050	0.050	0.492	0.496	0.495	0.499
4	100	0.039	0.052	0.042	0.053	0.073	0.089	0.072	0.081
4	500	0.047	0.048	0.047	0.052	0.205	0.211	0.212	0.215
4	1000	0.049	0.051	0.058	0.059	0.427	0.429	0.444	0.446
	B. Size: Serial Corr. ($\varphi = 0.8,\ \alpha = 0,\ R_y = 0$)					**D. Power: Serial Corr.** ($\varphi = 0.8,\ \alpha = 0.1,\ R_y = 0.2$)			
2	100	0.242	0.064	0.242	0.064	0.250	0.076	0.250	0.076
2	500	0.227	0.050	0.227	0.050	0.409	0.148	0.409	0.148
2	1000	0.225	0.057	0.225	0.057	0.503	0.234	0.503	0.234
3	100	0.264	0.068	0.271	0.068	0.273	0.092	0.271	0.087
3	500	0.304	0.049	0.301	0.048	0.426	0.158	0.425	0.168
3	1000	0.288	0.058	0.284	0.058	0.550	0.236	0.553	0.246
4	100	0.271	0.076	0.261	0.066	0.272	0.098	0.271	0.091
4	500	0.312	0.061	0.301	0.071	0.431	0.143	0.428	0.153
4	1000	0.306	0.052	0.307	0.059	0.554	0.260	0.564	0.269

II. Joint Tests

	A. Size: No Serial Corr. ($\varphi = 0,\ \alpha = 0,\ R_y = 0$)					**C. Power: No Serial Corr.** ($\varphi = 0,\ \alpha = 0.1,\ R_y = 0.2$)			
m	Sample Size	Trace		Maximum		Trace		Maximum	
		Static	Dyn.	Static	Dyn.	Static	Dyn.	Static	Dyn.
2	100	0.037	0.044	0.040	0.045	0.300	0.301	0.300	0.301
2	500	0.047	0.050	0.048	0.051	0.950	0.952	0.950	0.952
2	1000	0.051	0.048	0.052	0.050	1.000	1.000	1.000	1.000
3	100	0.042	0.055	0.042	0.051	0.194	0.219	0.189	0.208
3	500	0.048	0.050	0.053	0.051	0.902	0.908	0.922	0.921
3	1000	0.043	0.047	0.047	0.050	1.000	0.999	1.000	1.000
4	100	0.035	0.065	0.033	0.055	0.127	0.191	0.128	0.175
4	500	0.047	0.047	0.048	0.051	0.826	0.833	0.872	0.872
4	1000	0.050	0.053	0.049	0.052	0.995	0.994	0.998	0.998
	B. Size: Serial Corr. ($\varphi = 0.8,\ \alpha = 0,\ R_y = 0$)					**D. Power: Serial Corr.** ($\varphi = 0.8,\ \alpha = 0.1,\ R_y = 0.2$)			
2	100	0.197	0.063	0.197	0.063	0.449	0.423	0.449	0.423
2	500	0.187	0.057	0.187	0.057	0.956	0.986	0.956	0.986
2	1000	0.187	0.057	0.187	0.057	1.000	1.000	1.000	1.000
3	100	0.191	0.059	0.199	0.063	0.393	0.376	0.394	0.380
3	500	0.241	0.056	0.237	0.049	0.930	0.990	0.935	0.994
3	1000	0.225	0.055	0.229	0.053	0.999	1.000	1.000	1.000
4	100	0.189	0.073	0.181	0.072	0.311	0.330	0.298	0.306
4	500	0.224	0.059	0.226	0.061	0.876	0.984	0.899	0.993
4	1000	0.222	0.048	0.234	0.052	0.994	1.000	0.998	1.000

Table 3. Empirical results for Microeconomic Survey Data: Price and Production forecasts

	Trace Canonical Corr.		Maximum Canonical Corr.		Ordered	Serial Corr.	
Firm No.	Static	Dyn.Augm.	Static	Dyn.Augm.	Alternative	Actual	Forecast
A. Prices							
1	45.30	10.35	41.46	9.88	36.60	51.33	56.01
	(0.000)	(0.035)	(0.000)	(0.028)	(0.000)	(0.000)	(0.000)
2	48.91	32.44	29.55	21.09	20.42	58.4	36.26
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
3	45.05	16.06	41.36	15.1	34.73	34.75	30.7
	(0.000)	(0.003)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
4	16.65	3.14	16.35	2.97	15.76	32.25	36.88
	(0.002)	(0.535)	(0.001)	(0.499)	(0.000)	(0.000)	(0.000)
5	45.27	26.02	35.62	20.71	33.82	63.69	49.17
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
6	11.68	7.68	11.67	7.68	7.23	13	17.07
	(0.02)	(0.104)	(0.012)	(0.074)	(0.007)	(0.011)	(0.002)
7	4.59	6.14	4.46	6	1.6	15.99	0.7
	(0.332)	(0.189)	(0.286)	(0.152)	(0.206)	(0.003)	(0.951)
8	29.81	15.46	21.44	15.35	9.68	34.87	18.11
	(0.000)	(0.004)	(0.000)	(0.002)	(0.002)	(0.000)	(0.001)
9	56.6	7.79	56.1	7.74	22.32	86.81	136.31
	(0.000)	(0.100)	(0.000)	(0.072)	(0.000)	(0.000)	(0.000)
10	49.11	18.79	46.16	18.73	44.17	45.51	53.53
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Average (522 Firms)	38.03	17.82	30.35	15.16	20.10	34.00	42.99
% Rejections (522 Firms)	89.0	63.4	88.6	64.2	86.2	77.6	93.1
B. Production							
1	11.88	2.70	10.09	2.13	6.32	19.22	8.09
	(0.018)	(0.609)	(0.025)	(0.658)	(0.012)	(0.001)	(0.088)
2	27.39	21.00	24.41	20.01	9.67	11.58	6.84
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.021)	(0.145)
3	34.61	31.44	25.03	31.03	14.81	19.31	26.03
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
4	8.94	5.09	6.5	4.94	3.02	13.81	55.89
	(0.063)	(0.278)	(0.123)	(0.237)	(0.082)	(0.008)	(0.000)
5	17.11	5.30	16.39	5.18	10.17	25.93	84.51
	(0.002)	(0.258)	(0.001)	(0.214)	(0.001)	(0.000)	(0.000)
6	11.39	7.55	10.65	7.13	2.89	17.91	5.45
	(0.023)	(0.11)	(0.019)	(0.095)	(0.089)	(0.001)	(0.244)
7	28.97	11.32	25.82	10.08	22.61	24.97	33.41
	(0.000)	(0.023)	(0.000)	(0.025)	(0.000)	(0.000)	(0.000)
8	4.35	3.45	4.09	3.38	2.44	3.11	27.11
	(0.361)	(0.486)	(0.33)	(0.432)	(0.118)	(0.54)	(0.000)
9	27.23	16.23	18.94	11.17	16.93	4.94	40.84
	(0.000)	(0.003)	(0.000)	(0.015)	(0.000)	(0.294)	(0.000)
10	16.06	8.59	15.55	7.46	10.35	14.16	22.59
	(0.003)	(0.072)	(0.002)	(0.082)	(0.001)	(0.007)	(0.000)
Average (448 Firms)	21.39	11.42	18.16	9.87	11.80	26.32	32.67
% Rejections (448 Firms)	71.9	43.8	71.1	43.8	71.9	78.7	84.6

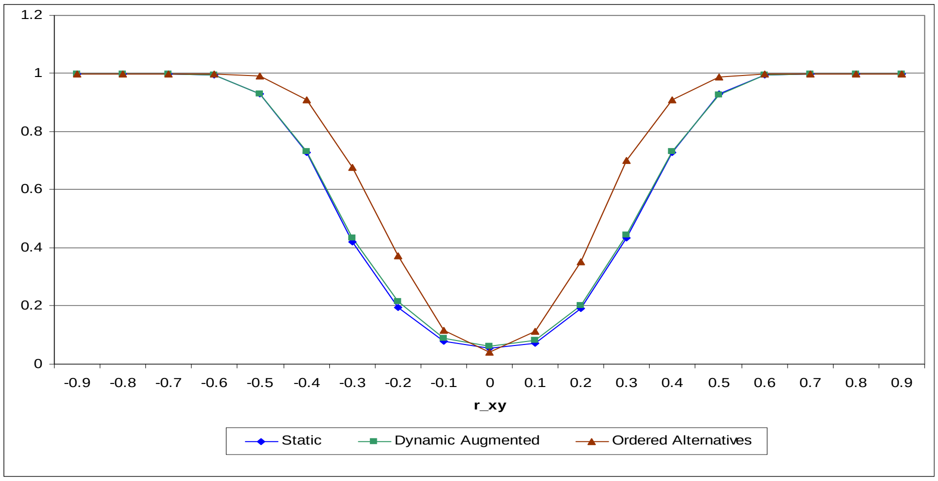

Figure 1: Power Curves for Trace Canonical Correlation and ordered alternative tests under serial independence (ϕ=0)

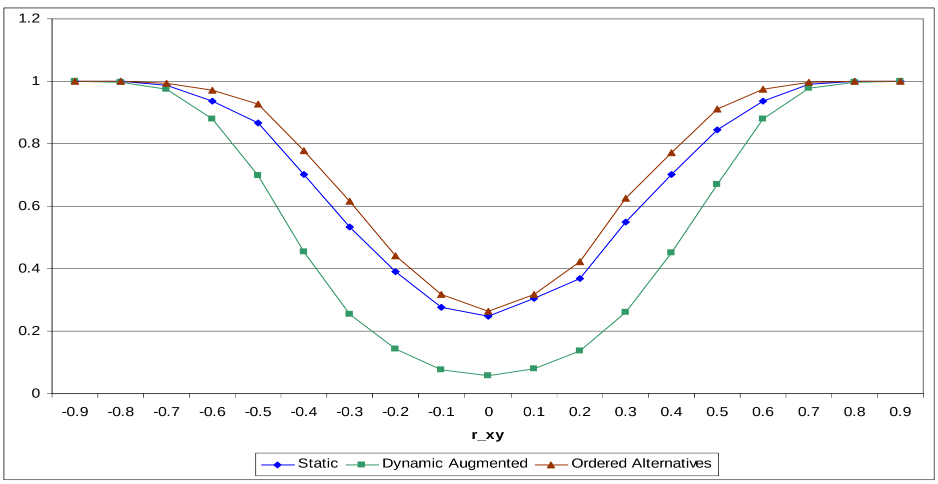

Figure 2: Power Curves for Trace Canonical Correlation and ordered alternative tests under serial dependence (ϕ=0.8)