

Testing Dependence Among Serially Correlated Multi-category Variables*

M. Hashem Pesaran
Cambridge University

Allan Timmermann
University of California, San Diego

July 3, 2006

Abstract

The contingency table literature on tests for dependence among discrete multi-category variables is extensive. Existing tests assume, however, that draws are independent, and there are no tests that account for serial dependencies—a problem that is particularly important in economics and finance. This paper proposes a new test of independence based on the maximum canonical correlation between pairs of discrete variables. We also propose a trace canonical correlation test using dynamically augmented reduced rank regressions or an iterated weighting method in order to account for serial dependence. Such tests are useful, for example, when testing for predictability of one sequence of discrete random variables by means of another sequence of discrete random variables as in tests of market timing skills or business cycle analysis. The proposed tests allow for an arbitrary number of categories, are robust in the presence of serial dependencies and are simple to implement using multivariate regression methods. Monte Carlo experiments show that the proposed tests have good finite sample properties. An empirical application to survey data on forecasts of GDP growth demonstrates the importance of correcting for serial dependencies in predictability tests.

JEL Classifications: C12, C22, C42, C52

Keywords: Contingency Tables, Canonical Correlations, Serial Dependence, Tests of Predictability

1 Introduction

Time-series of discrete data that fall in multiple categories are frequently encountered in economics and finance. For example, in macroeconomic surveys participants are often asked to predict the most likely range for variables such as GDP growth, inflation or interest rates. Similarly, financial analysts and brokerages often categorize stocks into buy, strong buy, hold, sell and strong sell recommendations. The NBER recession/expansion indicator that tracks the evolution of the US business cycle is an example of a binary variable as is the bull and bear market indicator frequently used to characterize conditions in stock markets (see Pagan and Sossounov (2003) and Lunde and Timmermann (2004)).

Besides the discreteness of the data, a key feature that characterizes these examples is the serial dependence or persistence in the underlying variables. This could be due to the construction

*We benefitted from the comments of Herman van Dijk and Adrian Pagan and from participants at the Econometric Institute's 50th anniversary conference in Rotterdam, June 9-10, 2006.

of the data—a single isolated quarter with negative GDP growth is usually not viewed as a recession, nor is the emergence of a short period with negative stock returns sufficient to constitute a bear market—or may reflect the serial dependence properties of the underlying data generating process as in Hamilton’s (1989) regime switching model for GDP growth.

While an extensive literature on contingency tables has developed tests for cross-sectional dependence between discrete random variables, such tests are generally valid only in the case of serially independent outcomes. For example, Fisher’s Chi-square test assumes independence across draws. This assumption is highly unlikely to hold for many economic and financial variables. In such cases, as emphasized recently by Pagan (2005), it becomes important to allow for serial dependence in statistical tests.¹

This paper develops new tests for dependencies between discrete random variables that cover dynamics of general form. An area where tests of such dependencies is of particular interest is in the evaluation of predictive performance when interest lies in testing whether one sequence of discrete random variables (‘outcomes’, $\{\mathbf{y}_t\}$) is predicted by another sequence of discrete random variables (‘forecasts’, $\{\mathbf{x}_t\}$) as in the literature on market timing—see Henriksson and Merton (1981), Cumby and Modest (1987) and Pesaran and Timmermann (1992). While our analysis is ideally suited for this type of application, it is more generally applicable to address the issue of dependencies among discrete random variables that are subject to time-series dynamics.

More specifically, in the case of m_y category realizations (y_{it} , $i = 1, 2, \dots, m_y$), and $m_x \leq m_y$ categories for an associated variable (x_{jt} , $j = 1, 2, \dots, m_x$), we cast the relationship between the $m_y - 1$ category realizations, $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{m_y-1,t})'$, and the $m_x - 1$ category variables, $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{m_x-1,t})$, 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. In the case of 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 . This gives rise to a new maximum canonical correlation test. This test has a non-standard distribution so we compute the finite sample critical values for different numbers of categories and sample sizes by stochastic simulations. We also propose a trace test based on the average canonical correlation and show that this is identical to the standard Fisher 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 $\mathbf{z}_t = (\mathbf{y}'_t, \mathbf{x}'_t)'$. This gives rise to a trace test based on a dynamically augmented reduced rank regression that is very simple to compute. We also propose an iterative procedure that estimates standard errors of the canonical correlation statistic in a way that accounts for serial dependence and heteroskedasticity. A common feature of the proposed tests is that they 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 .

Small sample properties of the proposed maximum and trace canonical correlation tests are investigated through Monte Carlo experiments. Standard 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 forecast categories, m_x , is large relative to the sample size, T . The use of the proposed tests is illustrated by an application to output forecasts using data from the Survey of Professional Forecasters.

The plan of the paper is as follows. Section 2 discusses the frequently encountered setup with

¹Harding and Pagan (2004) and Pagan (2005) propose a test of business cycle synchronization that allows for serial dependencies in the outcomes of pairs of binary variables using the generalized method of moments. These papers do not consider the general multivariate case, however.

binary random variables and contrasts the case with and without serial dependence. Section 3 generalizes the results to the case with multiple categories but assumes a static setting. This is generalized in Section 4 to cover multiple categories and serial dependencies. Section 5 presents Monte Carlo simulation results, while Section 6 reports the empirical application and Section 7 concludes. An appendix provides some of the proofs.

2 Binary Variables

We first establish some familiar results for the 2×2 case involving a pair of binary variables. This case has been extensively discussed in the statistical literature and, as we shall see, a number of existing tests arise as special cases in this setup. To this end, let $I(A)$ be an indicator function that takes the value of unity if $A > 0$ and zero otherwise, and suppose we are interested in testing whether one binary variable, $x_t = I(X_t)$ is related to another binary variable, $y_t = I(Y_t)$ using a sample of observations $(y_1, x_1), (y_2, x_2), \dots, (y_T, x_T)$. We first cover the case without serial dependence and then extend this to allow for such dependencies.

2.1 No Serial Dependence

In the absence of serial dependence in outcomes, the non-parametric sign test proposed by Pesaran and Timmermann (1992) can be used. Let \hat{P} be the so-called hit-rate, i.e. the proportion of cases where Y_t and X_t fall in the same category (have the same sign), while \hat{P}_* is the hit rate expected under the null of independence between x_t and y_t . The PT test statistic is given by

$$PT = \frac{\hat{P} - \hat{P}_*}{\left[\hat{V}(\hat{P}) - \hat{V}(\hat{P}_*) \right]^{\frac{1}{2}}}, \quad (1)$$

where

$$\hat{P} = T^{-1} \sum_{t=1}^T I(Y_t X_t), \quad \hat{P}_* = \bar{y}\bar{x} + (1 - \bar{y})(1 - \bar{x}), \quad (2)$$

$$\hat{V}(\hat{P}) = T^{-1} \hat{P}_*(1 - \hat{P}_*), \quad (3)$$

$$\begin{aligned} \hat{V}(\hat{P}_*) &= T^{-1}(2\bar{y} - 1)^2 \bar{x}(1 - \bar{x}) + T^{-1}(2\bar{x} - 1)^2 \bar{y}(1 - \bar{y}) \\ &\quad + 4T^{-2} \bar{y}\bar{x}(1 - \bar{y})(1 - \bar{x}), \end{aligned} \quad (4)$$

and $\bar{y} = T^{-1} \sum_{t=1}^T y_t$, $\bar{x} = T^{-1} \sum_{t=1}^T x_t$.² Under the null hypothesis that y_t and x_t are distributed independently (namely x_t has no power in predicting y_t), PT is asymptotically distributed as a standard normal, $PT \overset{d}{\sim} N(0, 1)$.

2.1.1 A Regression Approach

As a step towards allowing for serial dependence in the outcomes, we next show how this test can be cast in a regression context. It turns out that the PT statistic can be well approximated by the t-ratio of the coefficient of $x_t = I(X_t)$ in the Ordinary Least Squares (OLS) regression of $y_t = I(Y_t)$ on x_t and an intercept:

$$y_t = \alpha + \beta x_t + u_t, \quad (5)$$

²The PT statistic is undefined when \bar{y} or \bar{x} take the extreme values of zero or unity.

where $E(u_t | x_t, x_{t-1}, \dots) = 0$. We deal with the case where u_t could be serially correlated and/or heteroskedastic below.

The t -ratio of the OLS estimator of β in the above regression is given by

$$t_\beta = \frac{r\sqrt{T-2}}{\sqrt{1-r^2}}, \quad (6)$$

where r is the simple correlation coefficient between y_t and x_t . To establish the relationship between t_β and the PT statistic, note that

$$\begin{aligned} I(Y_t X_t) &= I(Y_t)I(X_t) + [1 - I(Y_t)][1 - I(X_t)] \\ &= 2y_t x_t - y_t - x_t + 1, \end{aligned}$$

and hence

$$\hat{P} = T^{-1} \sum_{t=1}^T I(Y_t X_t) = 2T^{-1} \sum_{t=1}^T y_t x_t - \bar{y} - \bar{x} + 1.$$

Using (2) we have

$$\hat{P} - \hat{P}_* = 2 \left(T^{-1} \sum_{t=1}^T y_t x_t - \bar{y}\bar{x} \right) = 2T^{-1} \sum_{t=1}^T (y_t - \bar{y})(x_t - \bar{x}) = 2S_{yx}.$$

Also, after some algebra, it is easily seen that

$$\hat{V}(\hat{P}) - \hat{V}(\hat{P}_*) = 4T^{-1}\bar{y}(1-\bar{y})\bar{x}(1-\bar{x}) - 4T^{-2}\bar{y}(1-\bar{y})\bar{x}(1-\bar{x}).$$

Ignoring the second term which is of order T^{-2} , and noting that $x_t^2 = x_t$ and $y_t^2 = y_t$, we have

$$\begin{aligned} S_{xx} &= T^{-1} \sum_{t=1}^T (x_t - \bar{x})^2 = \bar{x}(1-\bar{x}), \quad S_{xy} = S_{yx} = T^{-1} \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}), \\ S_{yy} &= T^{-1} \sum_{t=1}^T (y_t - \bar{y})^2 = \bar{y}(1-\bar{y}). \end{aligned}$$

It follows that (up to order T^{-1})

$$PT = \frac{\hat{P} - \hat{P}_*}{\left\{ \hat{V}(\hat{P}) - \hat{V}(\hat{P}_*) \right\}^{\frac{1}{2}}} \approx \frac{\sqrt{T}S_{yx}}{\sqrt{S_{yy}S_{xx}}} = \sqrt{T} r. \quad (7)$$

This in turn establishes that the student- t test of $\beta = 0$ in (6), and the PT test defined by (1), will be asymptotically equivalent. The two test statistics are also likely to be numerically very close in most applications.

2.2 Serial Dependence

As noted in the introduction one of the main shortcomings of the existing tests of predictability for variables that fall in multiple categories, including the PT test, is the assumption that under the null hypothesis ($\beta = 0$), the outcomes, y_t , are serially independent. This assumption is clearly restrictive and unlikely to hold for many economic and financial time series. For example, the presence of regimes whose dynamics is determined by a Markov process as in Hamilton (1989)

might give rise to persistence in output growth. Serial correlation in such variables is likely to generate serial dependence in the qualitative outcomes and could cause distortions in the size of the PT test, typically in the form of over-rejection.

In the context of the regression based test (6), the serial dependence in outcomes under the null hypothesis translates into serial dependence in the errors, u_t . Due to the discrete nature of the $y_t = I(Y_t)$ series, the pattern of serial dependence in y_t could differ from that of Y_t and additionally y_t could be conditionally heteroskedastic even if Y_t is not and vice versa.

In testing $\beta = 0$ in (5), serial dependence in the errors, u_t , can be dealt with either parametrically or by using Bartlett weights as proposed by Newey and West (1987) in constructing a test statistic

$$\tilde{t}_\beta = \frac{\hat{\beta}}{\sqrt{\hat{V}_{NW}(\hat{\beta})}}, \quad (8)$$

where $\hat{\beta}$ is the OLS estimator of β , and $\hat{V}_{NW}(\hat{\beta})$ is the (2, 2) element of

$$\hat{V}_{NW}(\hat{\phi}) = \frac{1}{(T-2)\bar{x}^2(1-\bar{x})^2} \begin{pmatrix} \bar{x} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \hat{\mathbf{F}}_h \begin{pmatrix} \bar{x} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \quad (9)$$

$\hat{\phi} = (\hat{\alpha}, \hat{\beta})'$, h is the length of the lag window,

$$\hat{\mathbf{F}}_h = \hat{\mathbf{\Omega}}_0 + \sum_{j=1}^h \left(1 - \frac{j}{h+1}\right) (\hat{\mathbf{\Omega}}_j + \hat{\mathbf{\Omega}}_j'),$$

and

$$\hat{\mathbf{\Omega}}_j = T^{-1} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j} \begin{pmatrix} 1 & x_{t-j} \\ x_t & x_t x_{t-j} \end{pmatrix}.$$

Clearly, other estimates of the variance of $\hat{\beta}$, based on different estimates of the spectral density of $\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta}x_t$ at zero frequency could be used. We shall return to these additional choices in the Monte Carlo section.

3 Multiple Categories: Static Case

While the binary case is used extensively in empirical work, a number of applications require generalizing the setup to allow for an arbitrary (but countably finite) number of categories. For example, survey data often has multiple categories such as ‘down’, ‘unchanged’ and ‘up’.³ To facilitate analysis of such cases, consider the general case where a time series of T observations on some explanatory or predictive variable, x , is arranged into m_x categories (states) while observations on the dependent or realized variable, y , are categorized into m_y groups.⁴ Without loss of generality we assume that $m_x \leq m_y$. Denote the x -categories by x_{jt} so that $x_{jt} = 1$ if the j^{th} category occurs at time t and zero otherwise. Similarly, denote the realized outcomes by y_{it} so $y_{it} = 1$ if category i occurs at time t and zero otherwise.

³Another example arises in the analysis of contagion where positive as well as negative discrete jumps in market returns or spreads could be of interest (see, for example, Favero and Giavazzi (2002) and Pesaran and Pick (2006)).

⁴That we consider pairs of variables, x, y , is not restrictive. For example, a new variable, x , could be formed from the product space of a set of underlying variables $X_{m_1} \times X_{m_2} \dots \times X_{m_g}$.

Convert the categorical observations into quantitative measures by assigning the weights a_i to y_{it} for $i = 1, 2, \dots, m_y$ and b_j to x_{jt} for $j = 1, 2, \dots, m_x$ and $t = 1, 2, \dots, T$ as follows⁵

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

Since the outcome categories are mutually exclusive, the regression of y_t on an intercept and x_t can be written as

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

or, more compactly,

$$\boldsymbol{\theta}' \mathbf{y}_t = c + \boldsymbol{\gamma}' \mathbf{x}_t + u_t, \quad (10)$$

where $\mathbf{y}_t = (\mathbf{y}_{1t}, \mathbf{y}_{2t}, \dots, \mathbf{y}_{m_y-1,t})'$, $\mathbf{x}_t = (\mathbf{x}_{1t}, \mathbf{x}_{2t}, \dots, \mathbf{x}_{m_x-1,t})'$, $c = \alpha + \beta b_{m_x} - a_{m_y}$ and

$$\boldsymbol{\theta} = \begin{pmatrix} a_1 - a_{m_y} \\ a_2 - a_{m_y} \\ \vdots \\ a_{m_y-1} - a_{m_y} \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \beta (b_1 - b_{m_x}) \\ \beta (b_2 - b_{m_x}) \\ \vdots \\ \beta (b_{m_x-1} - b_{m_x}) \end{pmatrix}.$$

A test of predictability can now be carried out by testing $\boldsymbol{\gamma} = \mathbf{0}$ in (10), conditional on a given value of $\boldsymbol{\theta}$. Under this setup the testing problem reduces to the static regression test of the previous section when $m_y = m_x = 2$. In the more general case the test of $\boldsymbol{\gamma} = \mathbf{0}$ will depend on the “nuisance” parameters, $\boldsymbol{\theta}$.

Throughout our analysis we assume that the following conditions hold:

Assumption 1: The number of categories, m_x and m_y is finite and remains fixed as $T \rightarrow \infty$.

Assumption 2: 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 $i = 1, \dots, m_y$, $j = 1, \dots, m_x$, and for all sample sizes, T , including as $T \rightarrow \infty$.

Assumption 3: $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\gamma} \in \Gamma$ where Θ and Γ are compact sets.

These assumptions are standard from the literature on contingency tables. The assumption that the number of states is finite and fixed is almost always satisfied in empirical applications. The second assumption (uniform representativeness) requires that all categories must be represented in a given sample while categories with zero representation must be dropped. Since this result should hold in the limit as $T \rightarrow \infty$, it follows from this assumption that none of the states can be absorbing. The third assumption is more of a technical requirement and it is hard to think of realistic cases where it is not satisfied.

Under assumptions 1-3 it is easy to establish consistency of the test statistics discussed below. Because the data is discrete, we do not need the underlying variables, Y_t and X_t to satisfy particular moment conditions. The key is assumption 2 that must hold uniformly for each T and as $T \rightarrow \infty$.

In what follows 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 . We then examine the properties of the test for other values of $\boldsymbol{\theta}$. The case of serially correlated and/or heteroskedastic errors will be dealt with subsequently.

⁵We are assuming that these measurement weights remain fixed over time.

3.1 $F(\boldsymbol{\theta})$ -statistic

For a given value of $\boldsymbol{\theta}$, a standard F -statistic can be employed to test independence of y_t and x_t :

$$F(\boldsymbol{\theta}) = \left(\frac{T - m_x}{m_x - 1} \right) \frac{\boldsymbol{\theta}' \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \boldsymbol{\theta}}{\boldsymbol{\theta}' (\mathbf{S}_{yy} - \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy}) \boldsymbol{\theta}}, \quad (11)$$

where

$$\mathbf{S}_{yx} = \mathbf{S}'_{xy} = T^{-1} \mathbf{Y}' \mathbf{M}_\tau \mathbf{X}, \quad \mathbf{S}_{yy} = T^{-1} \mathbf{Y}' \mathbf{M}_\tau \mathbf{Y}, \quad \mathbf{S}_{xx} = T^{-1} \mathbf{X}' \mathbf{M}_\tau \mathbf{X}.$$

$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)'$ and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)'$, are the $T \times (m_y - 1)$ and $T \times (m_x - 1)$ observation matrices on the qualitative indicators, respectively, and $\mathbf{M}_\tau = \mathbf{I}_T - \boldsymbol{\tau}(\boldsymbol{\tau}'\boldsymbol{\tau})^{-1}\boldsymbol{\tau}'$, where $\boldsymbol{\tau} = (1, 1, \dots, 1)'$. Since it is not known *a priori* which element of $\boldsymbol{\theta}$ might be non-zero, we employ the normalizing restriction $\boldsymbol{\theta}' \mathbf{S}_{yy} \boldsymbol{\theta} = 1$. This requires that at least one element of $\boldsymbol{\theta}$ is non-zero.

It follows from Assumption 2 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 (dropping, for simplicity, the T subscript on these sample averages),

$$\mathbf{S}_{xx} = \begin{pmatrix} \bar{x}_1(1 - \bar{x}_1) & -\bar{x}_1\bar{x}_2 & \dots & -\bar{x}_1\bar{x}_{m_x-2} & -\bar{x}_1\bar{x}_{m_x-1} \\ -\bar{x}_2\bar{x}_1 & \bar{x}_2(1 - \bar{x}_2) & \dots & & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\bar{x}_{m_x-2}\bar{x}_1 & -\bar{x}_{m_x-2}\bar{x}_2 & \dots & \bar{x}_{m_x-2}(1 - \bar{x}_{m_x-2}) & -\bar{x}_{m_x-2}\bar{x}_{m_x-1} \\ -\bar{x}_{m_x-1}\bar{x}_1 & -\bar{x}_{m_x-1}\bar{x}_2 & \dots & -\bar{x}_{m_x-1}\bar{x}_{m_x-2} & \bar{x}_{m_x-1}(1 - \bar{x}_{m_x-1}) \end{pmatrix}. \quad (12)$$

Similarly, the (i, j) element of \mathbf{S}_{xy} is given by $T^{-1} \sum_{t=1}^T x_{it}y_{jt} - \bar{x}_i\bar{y}_j$. For these matrices to be non-singular it is necessary that $\bar{x}_j \neq 0$ and $\bar{y}_i \neq 0$ for all i, j . Since $\bar{x}_i = 1$ would necessarily imply that $\bar{x}_j = 0$ for $j \neq i$, to ensure that \mathbf{S}_{xx} and \mathbf{S}_{yy} are non-singular we must have $\bar{x}_j(1 - \bar{x}_j) \neq 0$ and $\bar{y}_i(1 - \bar{y}_i) \neq 0$ for all $i = 1, \dots, m_y$ and $j = 1, \dots, m_x$. This is guaranteed to hold by Assumption 2.

We denote the probability limits of \bar{x}_i , \bar{y}_i , \mathbf{S}_{xx} , \mathbf{S}_{yy} , and \mathbf{S}_{xy} by μ_{ix} , μ_{iy} , $\boldsymbol{\Sigma}_{xx}$, $\boldsymbol{\Sigma}_{yy}$, and $\boldsymbol{\Sigma}_{xy}$, respectively:

$$\begin{aligned} p \lim_{T \rightarrow \infty} \bar{x}_j &= \mu_{jx} \in (0, 1), \quad p \lim_{T \rightarrow \infty} \bar{y}_i = \mu_{iy} \in (0, 1), \\ p \lim_{T \rightarrow \infty} \mathbf{S}_{xx} &= \boldsymbol{\Sigma}_{xx}, \quad p \lim_{T \rightarrow \infty} \mathbf{S}_{yy} = \boldsymbol{\Sigma}_{yy}, \quad p \lim_{T \rightarrow \infty} \mathbf{S}_{xy} = \boldsymbol{\Sigma}_{xy}. \end{aligned}$$

Clearly, $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{yy}$ will be non-singular so long as $\mu_{ix}(1 - \mu_{ix}) \neq 0$ and $\mu_{iy}(1 - \mu_{iy}) \neq 0$, which again is implied by Assumption 2.

3.2 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

$$F_{\max} = \text{Arg max}_{\boldsymbol{\theta}} [F(\boldsymbol{\theta})]$$

subject to the normalizing restriction that $\boldsymbol{\theta}'\mathbf{S}_{yy}\boldsymbol{\theta} = 1$. This idea has been used in the statistical 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), and to testing non-linear effects in dynamic models by Andrews and Ploberger (1994) and Hansen (1996). However, our application of Davies's main idea differs from these applications since the nuisance parameter, $\boldsymbol{\theta}$, does not disappear under the null.

Using (11), the first order condition for optimization of $F(\boldsymbol{\theta})$ is given by

$$(\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy})\hat{\boldsymbol{\theta}} = \hat{\rho}^2\mathbf{S}_{yy}\hat{\boldsymbol{\theta}}, \quad (13)$$

where

$$\hat{\rho}^2 = \frac{F(\hat{\boldsymbol{\theta}}) \left(\frac{m_x - 1}{T - m_x} \right)}{1 + \left(\frac{m_x - 1}{T - m_x} \right) F(\hat{\boldsymbol{\theta}})}. \quad (14)$$

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

$$\mathbf{S} = \mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}. \quad (15)$$

Denoting the non-zero eigenvalues of \mathbf{S} in descending order by $\hat{\rho}_1^2 \geq \hat{\rho}_2^2 \geq \dots \geq \hat{\rho}_{m_x-1}^2$, we have (using (14))

$$F_{\max} = \frac{(T - m_x)\hat{\rho}_1^2}{(m_x - 1)(1 - \hat{\rho}_1^2)}, \quad (16)$$

which is an obvious generalization of (6) and reduces to t_β^2 in the case of $m_x = 2$.

Note that $\hat{\rho}_i^2$ $i = 1, 2, \dots, m_x - 1$ 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. In the context of the categorical data this would involve choosing the weights, a_i , $i = 1, 2, \dots, m_y - 1$ and b_j , $j = 1, 2, \dots, m_x - 1$ 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, for example, Anderson (2003, Ch. 12). There are $m_x - 1$ such canonical correlations, 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 (15). The estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_1$, is given by the eigenvector associated with $\hat{\rho}_1^2$, which satisfies

$$(\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} - \hat{\rho}_1^2\mathbf{S}_{yy})\hat{\boldsymbol{\theta}}_1 = \mathbf{0}. \quad (17)$$

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 (10) 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, $1 > \hat{\rho}_1^2 > \hat{\rho}_2^2 > \dots > \hat{\rho}_{m_x-1}^2$, 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 multinomial distribution, the *exact* distribution of the canonical correlations will be less tractable but can readily be simulated as we show below.

3.3 Reduced Rank Regression Approach: Trace Canonical Correlation Test

The null of independence between x and y implies not only that $\rho_1 = 0$ but that $\rho_1 = \rho_2 = \dots = \rho_{m_x-1} = 0$. An alternative to the maximum canonical correlation test is therefore to base a test of $\boldsymbol{\gamma} = \mathbf{0}$ on an average concept of $F(\boldsymbol{\theta})$ given by⁶

$$\bar{F} = \frac{(T - m_x)}{m_x - 1} \sum_{i=1}^{m_x-1} \frac{\hat{\rho}_i^2}{1 - \hat{\rho}_i^2} \approx \frac{(T - m_x)}{m_x - 1} Tr(\mathbf{S}).$$

This test can also be derived in the context of the reduced rank regression

$$\mathbf{y}_t = \mathbf{a} + \mathbf{\Pi} \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad (18)$$

where in our application the null hypothesis of interest is $\mathbf{\Pi} = \mathbf{0}$, or $rank(\mathbf{\Pi}) = 0$.⁷ Under Assumptions 1 and 2, and assuming that under the null hypothesis the outcomes or $\boldsymbol{\varepsilon}_t$ are serially independent an asymptotic test of $\mathbf{\Pi} = \mathbf{0}$ is given by

$$(T - m_x) \sum_{i=1}^{m_x-1} \hat{\rho}_i^2 \overset{a}{\sim} \chi_{(m_x-1)^2}^2. \quad (19)$$

$\sum_{i=1}^{m_x-1} \hat{\rho}_i^2$ can also be computed by $Tr(\mathbf{S})$ and for this reason is often called the trace test.

As noted earlier in the case of Gaussian errors the exact sample distribution of the squared canonical correlations is provided in the literature. In the case of multi-category observations under the null hypothesis the errors will follow a multinomial distribution and the form of the exact distribution of the squared canonical correlations will be rather complicated to write down but as noted above can be easily simulated. However, since under Assumptions 1 and 2 all moments of $\boldsymbol{\varepsilon}_t$ exist, and \mathbf{S}_{xx} and \mathbf{S}_{yy} tend to finite non-stochastic positive definite matrices, then as $T \rightarrow \infty$, $(T - m_x)Tr(\mathbf{S})$ will follow the same asymptotic distribution as in the case with Gaussian errors established in the literature.

3.4 Relationship to Fisher's Chi-squared Test

In the special case where the realizations of X and Y are serially independent, the standard approach to testing independence of categorized variables is to arrange the outcomes in the form of a contingency table and then compute an appropriate test statistic from the individual cell frequencies. We next show that there is an exact relationship between the trace statistic and Fisher's contingency table χ^2 -test of independence. To this end we first introduce some new notations.

When testing the independence of y_{it} , and x_{jt} for $i = 1, 2, \dots, m_y$, $j = 1, \dots, m_x$ the appropriate

⁶ Average type test statistics have been proposed by Andrews and Ploberger (1994) when testing in the presence of nuisance parameters that disappear under the null hypothesis.

⁷ A detailed account of reduced rank regression techniques is provided in Anderson (2003, section 12.7), a topic which has re-emerged into prominence as a result of the developments in the cointegration literature. See, for example, Johansen (1995). Johansen's setting with integrated variables does not, of course, apply here.

contingency table is given by

Y, X	1	2	\dots	m_x	
1	n_{11}	n_{12}	\dots	n_{1m_x}	$n_{1.}$
2	n_{21}	n_{22}	\dots	n_{2m_x}	$n_{2.}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
m_y	$n_{m_y 1}$	$n_{m_y 2}$	\dots	$n_{m_y m_x}$	$n_{m_y.}$
	$n_{.1}$	$n_{.2}$	\dots	$n_{.m_x}$	n

Here n_{ij} is the frequency of the joint occurrence of y_{it} and x_{jt} , namely $n_{ij} = \sum_{t=1}^T y_{it}x_{jt}$, and

$$n_{i.} = \sum_{j=1}^{m_x} n_{ij} = \sum_{t=1}^T y_{it}, \quad n_{.j} = \sum_{i=1}^{m_y} n_{ij} = \sum_{t=1}^T x_{jt},$$

$$n = \sum_{i=1}^{m_y} n_{i.} = \sum_{j=1}^{m_x} n_{.j} = T.$$

The familiar Fisher Chi-square test of independence for data arranged in a contingency table is given by

$$\chi^2 = T \left(\sum_{i=1}^{m_y} \sum_{j=1}^{m_x} \frac{n_{ij}^2}{n_{i.}n_{.j}} - 1 \right). \quad (20)$$

The appendix contains a proof of the following proposition which shows that the trace test based on the reduced rank regression (18) is in fact identical to the familiar contingency table χ^2 -test:

Proposition 1 *The Fisher Chi-square test for independence of data arranged in an $m_x \times m_y$ contingency table ($m_x \leq m_y$) is identical to a trace test based on the canonical correlations*

$$\sum_{i=1}^{m_y} \sum_{j=1}^{m_x} \frac{n_{ij}^2}{n_{i.}n_{.j}} - 1 = \sum_{i=1}^{m_x-1} \hat{\rho}_i^2,$$

where $\hat{\rho}_i$ is the sample estimate of the i^{th} canonical correlation between $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{m_y-1})$ and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{m_x-1})$, $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$ and $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$. Furthermore,

$$\sum_{i=1}^{m_x-1} \hat{\rho}_i^2 = \text{Tr} \left[(\mathbf{Y}'\mathbf{M}_\tau\mathbf{Y})^{-1} (\mathbf{Y}'\mathbf{M}_\tau\mathbf{X}) (\mathbf{X}'\mathbf{M}_\tau\mathbf{X})^{-1} (\mathbf{X}'\mathbf{M}_\tau\mathbf{Y}) \right],$$

where $\mathbf{M}_\tau = \mathbf{I}_T - T^{-1}\boldsymbol{\tau}_T\boldsymbol{\tau}'_T$, $\boldsymbol{\tau}_T = (1, 1, \dots, 1)'$.

The values or ‘labels’ assigned to the categories for the X and Y variables may have a specific meaning in some applications but are oftentimes arbitrary—think of the convention of labelling recessions as unity and expansions as zeros. It would be unfortunate if such labels had an effect on the outcome of the proposed tests. However, as we next show, a convenient and reassuring property of the tests is that the labelling is irrelevant to the test statistic:⁸

⁸The connection between correlation coefficients and contingency table statistics was first derived by Hirschfeld (1935) and further elaborated by Lancaster (1958) in their attempts to solve the optimal score problem described in the proof of the proposition.

Proposition 2 *The maximum canonical correlation and trace canonical correlation tests are invariant to the values taken by the m_y categories of Y and the values taken by the m_x categories of X .*

This result implies that the trace and the maximum canonical correlation statistics are both invariant to the measurement of the categorical variables. Another way to establish this result which takes advantage of our framework is to notice that the test statistics remain unchanged if \mathbf{y}_t and \mathbf{x}_t are pre-multiplied by non-singular matrices of order $(m_y - 1) \times (m_y - 1)$ and $(m_x - 1) \times (m_x - 1)$, respectively. Denote these matrices by \mathbf{T}_y and \mathbf{T}_x , and let

$$\tilde{\mathbf{y}}_t = \mathbf{T}_y \mathbf{y}_t \text{ and } \tilde{\mathbf{x}}_t = \mathbf{T}_x \mathbf{x}_t,$$

and note that the canonical correlations of $\tilde{\mathbf{y}}_t$ and $\tilde{\mathbf{x}}_t$ are given by the non-zero roots of

$$|\mathbf{S}_{\tilde{y}\tilde{x}} \mathbf{S}_{\tilde{x}\tilde{x}}^{-1} \mathbf{S}_{\tilde{x}\tilde{y}} - \tilde{\rho}^2 \mathbf{S}_{\tilde{y}\tilde{y}}| = 0,$$

where

$$\begin{aligned} \mathbf{S}_{\tilde{y}\tilde{x}} &= \mathbf{S}'_{\tilde{x}\tilde{y}} = T^{-1} \mathbf{T}_y \mathbf{Y}' \mathbf{M}_\tau \mathbf{X} \mathbf{T}'_x, \\ \mathbf{S}_{\tilde{y}\tilde{y}} &= T^{-1} \mathbf{T}_y \mathbf{Y}' \mathbf{M}_\tau \mathbf{Y} \mathbf{T}'_y, \\ \mathbf{S}_{\tilde{x}\tilde{x}} &= T^{-1} \mathbf{T}_x \mathbf{X}' \mathbf{M}_\tau \mathbf{X} \mathbf{T}'_x. \end{aligned}$$

Since \mathbf{T}_y and \mathbf{T}_x are non-singular, we also have

$$\begin{aligned} |\mathbf{S}_{\tilde{y}\tilde{x}} \mathbf{S}_{\tilde{x}\tilde{x}}^{-1} \mathbf{S}_{\tilde{x}\tilde{y}} - \tilde{\rho}^2 \mathbf{S}_{\tilde{y}\tilde{y}}| &= |\mathbf{T}_y (\mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} - \tilde{\rho}^2 \mathbf{S}_{yy}) \mathbf{T}'_y| \\ &= |\mathbf{T}_y|^2 |\mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} - \tilde{\rho}^2 \mathbf{S}_{yy}|. \end{aligned}$$

Finally, since $|\mathbf{T}_y| \neq 0$ this establishes that $\tilde{\rho}_i^2 = \hat{\rho}_i^2$ for $i = 1, 2, \dots, m_x - 1$.

3.5 Simulation of Critical Values

The idea of basing a test of dependence on $(T - m_x) \hat{\rho}_1^2$, where $\hat{\rho}_1^2$ is the largest of the squared canonical correlation has not, to our knowledge, previously been considered. Critical values are therefore unknown for this test. Notably, the ordering of squared canonical correlations induces a non-standard distribution and means that $(T - m_x) \hat{\rho}_1^2$ will not follow a standard chi-squared test even in large samples. Furthermore, even for the trace test where ranking is not an issue, the chi-squared distribution is only achieved asymptotically, so the critical values will differ in small samples.

To compute the critical values of the maximum F (canonical correlation) statistic, $F_{\max} = (T - m_x) \hat{\rho}_1^2$, and the average F (trace canonical correlation) statistic, $\bar{F} = (T - m_x) \sum_{i=1}^{m_x-1} \hat{\rho}_i^2$, we undertook the following simulation experiment. Letting $m_x = m_y = m$, we generated random samples from the multinomial distribution with $2(m-1)$ categories for $m = 2, 3$ or 4 . We carried out 100,000 replications and considered sample sizes of $T = 20, 50, 100, 500$ and $1,000$. Table 1 reports 90% and 95% critical values for the maximum canonical correlation test (panel A) and the trace canonical correlation test (Panel B). The critical values are of course identical for the two tests when $m = 2$.

When $m = 2$, the 95% critical values are below their asymptotic value of 3.84 for the smallest samples of $T = 20$ or 50 but then rise to a value above this when $T = 100$. Hence there is no monotonic relationship between sample size and the 95% critical value when $m = 2$. In contrast, the critical values seem to rise monotonically with the sample size when $m = 3$ or 4 . Notice that a test that ignores the ranking information and hence wrongly assumes that $(T - m) \hat{\rho}_1^2 \sim \chi_1^2$ would be grossly oversized as the critical values are much greater than implied by this distribution.

4 Multiple Categories in the Presence of Error Dependence and Heteroskedasticity

We finally turn to the general case with multiple categories and serial dependence in the outcome variable. 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. Serial dependence in u_t can be allowed for either parametrically or by using a heteroskedasticity and autocorrelation consistent (HAC) procedure.

4.1 Dynamically Augmented Reduced Rank Regression

To allow for possible serial dependencies in the outcomes we consider the regression model (10) and assume that the errors, u_t , could be serially correlated. Suppose that u_t follows a stationary first order autoregressive process

$$u_t = \varphi u_{t-1} + \varepsilon_t, \quad |\varphi| < 1, \quad (21)$$

where ε_t are serially independent. For this error specification, using (10) 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 now 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

$$\mathbf{S}_w = \mathbf{S}_{yy,w}^{-1} \mathbf{S}_{yx,w} \mathbf{S}_{xx,w}^{-1} \mathbf{S}_{xy,w}, \quad (22)$$

where

$$\mathbf{S}_{yy,w} = T^{-1} \mathbf{Y}' \mathbf{M}_w \mathbf{Y}, \quad \mathbf{S}_{xx,w} = T^{-1} \mathbf{X}' \mathbf{M}_w \mathbf{X}, \quad \text{and} \quad \mathbf{S}_{xy,w} = T^{-1} \mathbf{X}' \mathbf{M}_w \mathbf{Y},$$

$$\mathbf{M}_w = \mathbf{I}_T - \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}', \quad \mathbf{W} = (\boldsymbol{\tau}, \mathbf{X}_{-1}, \mathbf{Y}_{-1}),$$

\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 is now easy to show 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

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\Pi}' + \mathbf{W} \mathbf{B} + \mathbf{E}, \quad (23)$$

where \mathbf{E} is a $T \times (m_y - 1)$ matrix of serially uncorrelated errors. Higher order error dynamics can be accommodated by including further lags of y and x as columns of \mathbf{W} . Under $\boldsymbol{\Pi} = \mathbf{0}$, and for T sufficiently large, using results established in Anderson (2003, Section 12.4), we have

$$(T - m_x) \text{Tr}(\mathbf{S}_w) \stackrel{a}{\sim} \chi_{(m_x - 1)^2}^2. \quad (24)$$

4.2 Iterated Method

An alternative to dynamically augmenting the reduced rank regression is to adjust the moment matrices used in calculating the variance matrix of $\hat{\boldsymbol{\gamma}}$ to account for heteroskedasticity and

autocorrelation in the errors in (10). The F -statistic corresponding to (11) in this case is given by

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

where

$$\mathbf{H}_{xx}(\boldsymbol{\theta}) = \lim_{T \rightarrow \infty} E \left[\frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T (\mathbf{x}_t - \bar{\mathbf{x}}) (\mathbf{x}_s - \bar{\mathbf{x}})' u_t(\boldsymbol{\theta}) u_s(\boldsymbol{\theta}) \right],$$

$\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{m_x-1})'$, $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{m_y-1})'$, and under $\boldsymbol{\gamma} = \mathbf{0}$

$$u_t(\boldsymbol{\theta}) = \boldsymbol{\theta}'(\mathbf{y}_t - \bar{\mathbf{y}}).$$

Hence

$$\mathbf{H}_{xx}(\boldsymbol{\theta}) = \lim_{T \rightarrow \infty} E \left[\frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T \boldsymbol{\theta}'(\mathbf{y}_t - \bar{\mathbf{y}}) (\mathbf{x}_t - \bar{\mathbf{x}}) (\mathbf{x}_s - \bar{\mathbf{x}})' (\mathbf{y}_s - \bar{\mathbf{y}})' \boldsymbol{\theta} \right],$$

can be viewed as the long run variance of $T^{-1/2} \sum_{t=1}^T \mathbf{d}_t(\boldsymbol{\theta})$, where $\mathbf{d}_t(\boldsymbol{\theta}) = \boldsymbol{\theta}'(\mathbf{y}_t - \bar{\mathbf{y}}) (\mathbf{x}_t - \bar{\mathbf{x}})$. Since elements of \mathbf{x}_t and \mathbf{y}_t are bounded, $\mathbf{H}_{xx}(\boldsymbol{\theta})$ exists under general assumptions concerning the serial dependence and heteroskedasticity of the error terms, as set out in Newey and West (1987).

Unlike the serially independent case, the first order conditions for maximization of $LM(\boldsymbol{\theta})$ cannot get reduced to solving an eigenvalue problem. An asymptotically equivalent alternative (under $\boldsymbol{\gamma} = \mathbf{0}$) is to use a first-stage consistent estimate of $\mathbf{H}_{xx}(\boldsymbol{\theta})$ that abstracts from the serial dependence of the errors. Such an estimator of $\boldsymbol{\theta}$ is given by (17), and the first-stage estimate of $\mathbf{H}_{xx}(\boldsymbol{\theta})$ can be obtained by (using a Bartlett window)

$$\hat{\mathbf{H}}_{xx,h}(\hat{\boldsymbol{\theta}}_1) = \hat{\boldsymbol{\Gamma}}_0 + \sum_{j=1}^h \left(1 - \frac{j}{h+1} \right) (\hat{\boldsymbol{\Gamma}}_j + \hat{\boldsymbol{\Gamma}}_j'), \quad (25)$$

$$\hat{\boldsymbol{\Gamma}}_j = T^{-1} \sum_{t=j+1}^T \mathbf{d}_t(\hat{\boldsymbol{\theta}}_1) \mathbf{d}_{t-j}'(\hat{\boldsymbol{\theta}}_1),$$

$$\mathbf{d}_t(\hat{\boldsymbol{\theta}}_1) = \hat{\boldsymbol{\theta}}_1'(\mathbf{y}_t - \bar{\mathbf{y}}) (\mathbf{x}_t - \bar{\mathbf{x}}).$$

Using this estimator, one can solve the following eigenvalue problem

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

where $\tilde{\rho}_1^2$ is the largest value of $\tilde{\rho}^2$ that solves

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

Under the null that $\boldsymbol{\gamma} = \mathbf{0}$, and conditional on the initial estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_1$, the Trace test is now given by

$$(T - m_x) Tr \left[\tilde{\mathbf{S}}(\hat{\boldsymbol{\theta}}_1) \right] \stackrel{a}{\sim} \chi_{(m_x-1)^2}^2, \quad (26)$$

where

$$\tilde{\mathbf{S}}(\hat{\boldsymbol{\theta}}_1) = \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \hat{\mathbf{H}}_{xx}^{-1}(\hat{\boldsymbol{\theta}}_1) \mathbf{S}_{xy}.$$

The estimate of θ used for the estimation of $\mathbf{H}_{xx}(\theta)$ can be iterated upon as required until convergence is achieved, subject to the normalization restriction, $\theta' \mathbf{S}_{yy} \theta = 1$.⁹ The heteroskedasticity and autocorrelation consistent t -statistic corresponds to the GMM test recently proposed by Harding and Pagan (2004) in the context of serially dependent binary variables ($m = 2$).

5 Monte Carlo Simulations

The previous section introduced a range of test statistics and derived their asymptotic distribution. However, it is also important to understand the finite sample properties of these tests, so in this section we undertake some Monte Carlo experiments to shed light on this question. We use the following setup. To capture serial dependence, Y_t was simulated from a first-order autoregressive process with parameter, φ , and Gaussian increments.¹⁰ Autocorrelations of $\varphi = 0.0, 0.5$ and 0.8 were considered, corresponding to zero, medium and strong serial correlation scenarios, respectively. To allow for different degrees of dependence between Y_t and X_t , we consider 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 thousands replications were carried out for each experiment. We report results for sample sizes of $T = 20, 50, 100, 500$ and 1000 .

The simulations consider the three trace statistics (19), (24) and (26) and the corresponding maximum canonical correlation tests. In each case we assume a critical level of five percent, using the finite sample critical values from Table 1. For the dynamically augmented reduced rank regression (23) that includes lags of \mathbf{X}_t and \mathbf{Y}_t , we consider up to four lags, selected in each case using the Akaike Information Criterion. The number of lags used in the HAC procedure is proportional to $T^{1/3}$.

We also consider the performance of the PT test for market timing referred to in Section 2 which tests whether the diagonal elements in a contingency table differ from their values implied under the null of independence,

$$H_0^* : \sum_{i=1}^m (P_{ii} - P_{i.} P_{.i}) = 0.$$

Notice that when $m > 2$, the null hypothesis for the PT test differs from the null of independence, whereas the two coincide when $m = 2$. In the multicategory case (7) can be generalized to

$$\sqrt{m} V_s^{-1/2} S_m \sim N(0, 1), \quad (27)$$

where

$$\begin{aligned} S_m &= \sum_{i=1}^{m_x} (\hat{P}_{ii} - \hat{P}_{i.} \hat{P}_{.i}) \\ \hat{P}_{ii} &= \frac{n_{ii}}{n}, \quad \hat{P}_{i.} = \frac{n_{i.}}{n}, \quad \hat{P}_{.i} = \frac{n_{.i}}{n}, \\ V_s &= \left(\frac{\partial f(P_0)}{\partial P} \right)' (\Psi - P_0 P_0') \left(\frac{\partial f(P_0)}{\partial P} \right) \end{aligned}$$

⁹ Alternatively one could follow a simultaneous non-linear approach to the estimation of θ , as in the continuous updating procedure employed in GMM estimation (see, for example, Hall (2004)).

¹⁰ To consider the effect of higher order dynamics, we also simulated from a second-order autoregressive process. The simulation results were very similar to those reported for the AR(1) process and are hence omitted.

and

$$\frac{\partial f(P_0)}{\partial P} = \begin{cases} 1 - P_{.i} - P_i & \text{for } i = j \\ -P_j - P_i & \text{for } i \neq j \end{cases}.$$

This is the form of the PT statistic used in the simulations.

5.1 Size

Table 2 reports the size of the test statistics assuming a zero correlation between Y_t and X_t ($r_{yx} = 0$), whilst 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 test statistics generally have the right size. Exceptions to this only occur when m is large relative to the sample size (i.e. when $m = 3$ or 4 and $T = 20$)—cases where one would hesitate to rely on statistical tests. In the latter case the dynamically augmented and iterated canonical correlation tests tend to be somewhat oversized, reflecting the effect of estimation error in very small samples. The PT test also tends to be oversized in the smaller samples. Similar findings are well known in the time series literature on hypothesis testing with serial adjustment in cases where the sample size is small relative to the number of estimated parameters.

Turning to the case with serially correlated outcomes (but still no cross-correlation, i.e. $r_{yx} = 0$), under mild serial correlation ($\varphi = 0.5$ in Panel B), the static canonical correlation test statistic is somewhat oversized with rejection rates of 7-10 percent. A similar but somewhat larger over-rejection rate is observed for the PT test. The size of the dynamically augmented test statistics are generally reasonably close to the 5% nominal level. The iterated test is undersized in the smaller samples when $m = 2$ or $m = 3$. Small sample size distortions disappear for the dynamically augmented and iterated tests as the sample grows. In contrast, the size distortions grow as a function of the sample size for the static canonical correlation test that assumes serially uncorrelated outcomes.

When the serial correlation gets even stronger ($\varphi = 0.8$, in Panel C), size distortions become much more serious for the static canonical correlation test and the PT test. At this level of persistence, rejection rates around 20-30% are common for either test. In contrast, the dynamically augmented and iterated test that allow for serially correlated outcomes generally have the right size. Furthermore, they appear to converge to the right limits in the largest sample sizes and hence properly adjust for serial dependencies. In contrast, the static canonical correlation test is clearly over-sized in both the small and the largest samples with a bias that grows with the sample size.

5.2 Power

Table 3 presents values of the power when the correlation between innovations to X and Y is $r_{yx} = 0.2$. In the absence of any serial correlation in the outcomes ($\varphi = 0$), the power of the static and dynamically augmented canonical correlation statistics is around 10 percent in samples with 50 observations. This rises to 25 and 80 percent when the sample size is increased to 100 and 500 observations, respectively. Power is close to 100% in the largest sample with 1000 observations. For the static and dynamically augmented canonical correlation statistics, the power does not vary too much across different values of m except for in the very smallest sample with $T = 20$, where the size distortions of the dynamically augmented test leads to over-rejections. The iterated test generally has weak nominal power as a result of its tendency to be undersized in the smaller samples.

Introducing mild serial persistence ($\varphi = 0.5$ in Panel B) does not change the results very much although the rejection rate is now generally marginally higher for the static canonical correlation

test than for the dynamically augmented test due to the tendency of the former to over-reject as documented in Panel B of Table 2. This same point also explains the much higher rejection rates for the static canonical correlation test in the high persistence scenario covered in Panel C.

Table 4 shows results when the correlation between innovations to X and Y is raised to $r_{yx} = 0.8$. Not surprisingly, the power of the static and dynamically augmented test is generally very high for this case—close to 70% in the smallest sample with 20 observations and close to unity in the larger samples. Once again, the iterated procedure tends to have lower power in the smaller samples due to the fact that it is undersized for these cases.

The power of the tests should of course be seen in relation to the size-distortions discussed previously. If size distortions grow as a result of serial correlation, the power of some tests may appear to increase as a result of stronger serial correlation, even though this is not really the case. To demonstrate this point, Figures 1-3 plot the rejection rates as a function of the cross-sectional correlation between innovations to X and Y under no persistence ($\varphi = 0$, in Figure 1), mild persistence ($\varphi = 0.5$, in Figure 2) and strong persistence ($\varphi = 0.8$, in Figure 3). The figures assume $m = 2$, $T = 100$ and uses 2,000 simulations. Figure 1 shows that in the absence of serial persistence, the iterated test is undersized and the static canonical correlation test has marginally higher power than the dynamically augmented test. This is to be expected since the static test uses the correct assumption of zero coefficients on lagged values of X and Y and hence estimates fewer parameters. Figure 2 shows that the static canonical correlation test tends to get oversized in the presence of serial correlation. It is clear from Figure 3 that this problem grows very severe as the degree of persistence of the outcome variable goes up. Under mild serial persistence the dynamically augmented and iterated canonical correlation methods tend to have the right size whereas the latter is a bit oversized under strong serial persistence.

5.3 Alternative Adjustments for Heteroskedasticity and Serial Correlation

When $m_x = m_y = 2$, we considered a standard least-squares t -test without correction for serial correlation in addition to t -tests that account for serial correlation using either the Newey-West (1987) or Kiefer-Vogelsang (2002) corrections. We also corrected for serial correlation by means of Andrews' plug-in bandwidth (Andrews (1991), Andrews and Monahan (1992)) and the Phillips, Sun and Jin (2005) approach. These approaches were applied separately for Bartlett, Parzen and QS (quadratic spectral) kernels used to adjust for serial correlation. We consider two regressions, namely one that does not include lagged values of the dependent and independent variables (regression I), along with a regression that includes lags of X and Y (regression II) selected by the AIC.

Table 5 presents the results. To save space we report results only for least squares, Bartlett and QS kernels, but very similar results were obtained using other kernels. In the case without serial correlation, the tests generally have the correct size, albeit with a tendency of the dynamically augmented regressions to be slightly oversized in the smallest sample with $T = 20$ observations. Results are robust across kernels used to compute the standard errors. Turning to the case with mild serial persistence (panel B), the results now suggest that the static OLS method is oversized in the largest samples, while the other methods control size much better. Finally, for the case with strong serial persistence, the OLS method rejects in close to 25% of the cases, indicating again that serial correlation is a big problem for this method and leads to significant size distortions. Correcting for serial correlation by using Bartlett or QS weights only partially resolves this problem as the tests based on these methods still tend to be oversized with a rejection rate closer to 10-15% in the smaller samples. The best method seems again to be to dynamically augment the regression and use HAC standard errors. Under this method the size

of the tests is close to the correct value of 5% except for in the smallest sample, $T = 20$, where the rejection rate is closer to 10%.

In the case without serial correlation, the power is around 10% in the smallest samples, growing to 25% when $T = 100$ and to 80% and close to 100% in the two largest samples. In the case with strong serial correlation and mild correlation between innovations to X and Y , the power of the dynamically augmented methods that use HAC standard errors declines to around 15%, 40% and 75% in the samples with 100, 500 and 1,000 observations. While the power appears to be higher for the simple OLS method, this only reflects the size distortions described earlier. Power is uniformly high in the experiments with a correlation between innovations to X and Y of 0.8 even in the presence of strong serial correlation.

6 Survey Forecasts of GDP Growth

Our empirical application considers the Survey of Professional Forecasters which reports current- and next-year forecasts of real GDP growth in the form of probability values spread across various outcome ranges. Our sample covers the period between 1992Q1 and 2005Q2, the longest recent period during which the definition of the outcome ranges was not subject to change. We study real GDP growth forecasts and realizations in the ranges below 2%, between 2% and 3%, between 3% and 4% and above 4%. The range with the highest probability among survey participants—the most likely range—is assigned unity, while the other ranges get a zero. The same ranges are then used for the ‘realized’ time series which we take to be the real-time vintages of output growth. By using real-time data, we avoid distortions due to subsequent data revisions that could not have been known by the survey participants. All data is published by the Philadelphia Federal Reserve.

Table 6 reports the value of the PT, static, dynamically augmented and iterated maximum and trace canonical correlation tests using $m = 2$ categories (based on the ranges $(-\infty, 3\%)$, $[3\%, \infty)$), $m = 3$ categories (based on the ranges $(-\infty, 2\%)$, $[2\%, 4\%)$, $[4\%, \infty)$) and $m = 4$ categories (based on the ranges $(-\infty, 2\%)$, $[2\%, 3\%)$, $[3\%, 4\%)$, $[4\%, \infty)$).¹¹

For the current-year predictions, the PT test and the static canonical correlation test indicate strong evidence of predictability irrespective of the number of categories used. However, it should be kept in mind that these tests tend to be seriously over-sized in situations with significant serial correlation. Adjusting for serial correlation is important here: The autocorrelation of the first linear combination of the dependent variable is 0.8, while that of the second and third linear combination is 0.40.

In view of the serial correlation in the data, it is therefore not surprising that the dynamically augmented and iterated canonical correlation tests are far smaller than the static canonical correlation test. Moreover, these test statistics generally fall some distance below the 5% finite-sample critical values reported in Table 1 (assuming a sample size of 50). The only case with robust evidence of current-year predictability is when $m_x = m_y = 4$, where both the dynamically augmented and iterated canonical correlation statistics generate p -values far below 5%. This suggests that survey participants have the ability to predict the finer ranges of output growth and that information is lost by aggregating the data too much.

Turning to the year-ahead forecasts, most results indicate lack of forecasting ability. The PT test is now insignificant and even slightly negative. The static canonical correlation test generates a p -value around 5% using $m = 3$ categories. However, when account is taken of

¹¹Although these ranges are not of equal length, they account for the fact that, much of the time, GDP growth falls in the range between two and four percent per annum.

persistence in the data, once again the test statistic declines substantially and is now far from being significant.

These results indicate no ability to predict the correct range of real GDP growth one-year ahead, whereas current-year GDP growth is predictable. The analysis also demonstrates that, in practice, it can make a sizeable difference whether or not persistence is accounted for when calculating test statistics for multi-category data.

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 need for such statistics arises in a variety of applications in areas such as business cycle research, market timing analysis and in the analysis of survey data.

In the absence of serially dependent outcomes we show that the traditional contingency table test statistic is identical to a trace tests based on reduced rank regressions of categorical variables. Casting the problem in this manner allows us to develop methods to control for serial dependence in the context of dynamically augmented regressions. It is worth emphasizing that the reduced rank regression approach is extremely easy to implement and only requires computing a multivariate regression of $(m - 1)$ categorized variables on an intercept, $(m - 1)$ explanatory variables and lags of these and then using standard HAC estimates for standard errors.

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 proposed maximum and trace canonical correlation statistics perform well in small samples 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. To establish the result we first write the various moment matrices in the trace expression in terms of n_{ij} notations. Since the events in the various categories are mutually exclusive, $\mathbf{Y}'\mathbf{Y}$ and $\mathbf{X}'\mathbf{X}$ will be diagonal matrices with their i^{th} diagonal elements given by

$$n_{i.} = \sum_{t=1}^T y_{it} \text{ and } n_{.j} = \sum_{t=1}^T x_{jt}.$$

Also the (i, j) element of $\mathbf{Y}'\mathbf{X}$ is n_{ij} , and

$$\begin{aligned} \boldsymbol{\tau}'_T \mathbf{Y} &= \mathbf{h}'_y = (n_{1.}, n_{2.}, \dots, n_{m_y-1.})', \quad \boldsymbol{\tau}'_T \mathbf{X} = \mathbf{h}'_x = (n_{.1}, n_{.2}, \dots, n_{.m_x-1})', \\ (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{h}_y &= \boldsymbol{\tau}_{m_y-1}, \quad (\mathbf{X}'\mathbf{X})^{-1} \mathbf{h}_x = \boldsymbol{\tau}_{m_x-1}, \end{aligned}$$

where $\boldsymbol{\tau}_{m_y-1}$ is a $(m_y - 1) \times 1$ vector of ones. Hence

$$\mathbf{Y}'\mathbf{M}_\tau \mathbf{Y} = \mathbf{Y}'\mathbf{Y} - T^{-1} \mathbf{h}_y \mathbf{h}'_y, \quad \mathbf{Y}'\mathbf{M}_\tau \mathbf{X} = \mathbf{Y}'\mathbf{X} - T^{-1} \mathbf{h}_y \mathbf{h}'_x.$$

But

$$(\mathbf{Y}'\mathbf{M}_\tau \mathbf{Y})^{-1} = (\mathbf{Y}'\mathbf{Y})^{-1} + \frac{T^{-1} (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{h}_y \mathbf{h}'_y (\mathbf{Y}'\mathbf{Y})^{-1}}{1 - T^{-1} \mathbf{h}'_y (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{h}_y},$$

and noting that

$$\mathbf{h}'_y (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{h}_y = \sum_{i=1}^{m_y-1} n_{i.} = T - n_{m_y}.$$

we have

$$\begin{aligned} (\mathbf{Y}'\mathbf{M}_\tau \mathbf{Y})^{-1} &= (\mathbf{Y}'\mathbf{Y})^{-1} + \frac{\boldsymbol{\tau}_{m_y-1} \boldsymbol{\tau}'_{m_y-1}}{n_{m_y}}, \\ (\mathbf{X}'\mathbf{M}_\tau \mathbf{X})^{-1} &= (\mathbf{X}'\mathbf{X})^{-1} + \frac{\boldsymbol{\tau}_{m_x-1} \boldsymbol{\tau}'_{m_x-1}}{n_{.m_x}}. \end{aligned}$$

Therefore

$$\begin{aligned} (\mathbf{Y}'\mathbf{M}_\tau \mathbf{Y})^{-1} \mathbf{Y}'\mathbf{M}_\tau \mathbf{X} &= \left[(\mathbf{Y}'\mathbf{Y})^{-1} + \frac{\boldsymbol{\tau}_{m_y-1} \boldsymbol{\tau}'_{m_y-1}}{n_{m_y}} \right] (\mathbf{Y}'\mathbf{X} - T^{-1} \mathbf{h}_y \mathbf{h}'_x) \\ &= (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X} - T^{-1} (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{h}_y \mathbf{h}'_x \\ &\quad + \frac{\boldsymbol{\tau}_{m_y-1} \boldsymbol{\tau}'_{m_y-1} \mathbf{Y}'\mathbf{X}}{n_{m_y}} - T^{-1} \frac{\boldsymbol{\tau}_{m_y-1} \boldsymbol{\tau}'_{m_y-1} \mathbf{h}_y \mathbf{h}'_x}{n_{m_y}}. \end{aligned}$$

It is easily seen that

$$\begin{aligned} (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{h}_y \mathbf{h}'_x &= \boldsymbol{\tau}_{m_y-1} \mathbf{h}'_x, \\ \boldsymbol{\tau}'_{m_y-1} \mathbf{Y}'\mathbf{X} &= (\mathbf{h}_x - \mathbf{q}_x)', \\ \boldsymbol{\tau}_{m_y-1} \boldsymbol{\tau}'_{m_y-1} \mathbf{h}_y \mathbf{h}'_x &= \boldsymbol{\tau}_{m_y-1} \mathbf{h}'_x \left(\boldsymbol{\tau}'_{m_y-1} \mathbf{h}_y \right) \\ &= \left(\boldsymbol{\tau}_{m_y-1} \mathbf{h}'_x \right) (T - n_{m_y}), \end{aligned}$$

where $\mathbf{q}_x = (n_{m_y 1}, n_{m_y 2}, \dots, n_{m_y, m_x-1})'$. Using these results, after some algebra we have

$$\begin{aligned} (\mathbf{Y}'\mathbf{M}_\tau\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{M}_\tau\mathbf{X} &= (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{X} - \frac{\tau_{m_x-1}\mathbf{q}'_x}{n_{m_y}}, \\ (\mathbf{X}'\mathbf{M}_\tau\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_\tau\mathbf{Y} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \frac{\tau_{m_x-1}\mathbf{q}'_y}{n_{m_x}}, \end{aligned}$$

where $\mathbf{q}_y = (n_{1m_x}, n_{2m_x}, \dots, n_{m_y-1, m_x})'$. Hence

$$\begin{aligned} & Tr \left[(\mathbf{Y}'\mathbf{M}_\tau\mathbf{Y})^{-1} (\mathbf{Y}'\mathbf{M}_\tau\mathbf{X}) (\mathbf{X}'\mathbf{M}_\tau\mathbf{X})^{-1} (\mathbf{X}'\mathbf{M}_\tau\mathbf{Y}) \right] \\ &= Tr \left[(\mathbf{Y}'\mathbf{Y})^{-1} (\mathbf{Y}'\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{Y}) \right] - \frac{\mathbf{q}'_y (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X} \tau_{m_x-1}}{n_{m_x}} \\ &\quad - \frac{\mathbf{q}'_x (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \tau_{m_y-1}}{n_{m_y}} + \frac{(\mathbf{q}'_x \tau_{m_x-1}) (\mathbf{q}'_y \tau_{m_y-1})}{n_{m_y} \cdot n_{m_x}}. \end{aligned}$$

Consider now the various terms in this expression. First, since $\mathbf{Y}'\mathbf{Y}$ and $\mathbf{X}'\mathbf{X}$ are diagonal matrices and the typical element of $\mathbf{Y}'\mathbf{X}$ is n_{ij} , it readily follows that

$$Tr \left[(\mathbf{Y}'\mathbf{Y})^{-1} (\mathbf{Y}'\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{Y}) \right] = \sum_{i=1}^{m_y-1} \sum_{j=1}^{m_x-1} \frac{n_{ij}^2}{n_i \cdot n_j}.$$

Also,

$$\mathbf{Y}'\mathbf{X} \tau_{m_x-1} = \mathbf{h}_y - \mathbf{q}_y, \quad \mathbf{X}'\mathbf{Y} \tau_{m_y-1} = \mathbf{h}_x - \mathbf{q}_x,$$

and (recalling that $(\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{h}_y = \tau_{m_y-1}$)

$$\begin{aligned} \mathbf{q}'_y (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{X} \tau_{m_x-1} &= \mathbf{q}'_y (\mathbf{Y}'\mathbf{Y})^{-1} (\mathbf{h}_y - \mathbf{q}_y) \\ &= \mathbf{q}'_y \tau_{m_y-1} - \mathbf{q}'_y (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{q}_y \\ \mathbf{q}'_x (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \tau_{m_y-1} &= \mathbf{q}'_x (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{h}_x - \mathbf{q}_x) \\ &= \mathbf{q}'_x \tau_{m_x-1} - \mathbf{q}'_x (\mathbf{X}'\mathbf{X})^{-1} \mathbf{q}_x \end{aligned}$$

Finally,

$$\begin{aligned} \mathbf{q}'_x \tau_{m_x-1} &= \sum_{j=1}^{m_x-1} n_{m_y j} = n_{m_y} - n_{m_y m_x}, \\ \mathbf{q}'_y \tau_{m_y-1} &= \sum_{i=1}^{m_y-1} n_{i m_x} = n_{m_x} - n_{m_y m_x}, \end{aligned}$$

with

$$\begin{aligned} \mathbf{q}'_y (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{q}_y &= \sum_{i=1}^{m_y-1} \frac{n_{im_x}^2}{n_i}, \\ \mathbf{q}'_x (\mathbf{X}'\mathbf{X})^{-1} \mathbf{q}_x &= \sum_{j=1}^{m_x-1} \frac{n_{m_y j}^2}{n_j}. \end{aligned}$$

Therefore, it follows that

$$\begin{aligned}\frac{\mathbf{q}'_y(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{X}\tau_{m_x-1}}{n_{\cdot m_x}} &= \frac{\mathbf{q}'_y\tau_{m_y-1}}{n_{\cdot m_x}} - \frac{\mathbf{q}'_y(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{q}_y}{n_{\cdot m_x}} \\ &= 1 - \frac{n_{m_y m_x}}{n_{\cdot m_x}} - \sum_{i=1}^{m_y-1} \frac{n_{im_x}^2}{n_i \cdot n_{\cdot m_x}}, \\ \frac{\mathbf{q}'_x(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\tau_{m_y-1}}{n_{m_y \cdot}} &= \frac{\mathbf{q}'_x\tau_{m_x-1}}{n_{m_y \cdot}} - \frac{\mathbf{q}'_x(\mathbf{X}'\mathbf{X})^{-1}\mathbf{q}_x}{n_{m_y \cdot}} \\ &= 1 - \frac{n_{m_y m_x}}{n_{m_y \cdot}} - \sum_{j=1}^{m_x-1} \frac{n_{m_y j}^2}{n_{m_y \cdot} \cdot n_{\cdot j}},\end{aligned}$$

and

$$\begin{aligned}\frac{(\mathbf{q}'_x\tau_{m_x-1})(\mathbf{q}'_y\tau_{m_y-1})}{n_{m_y \cdot} \cdot n_{\cdot m_x}} &= \frac{(n_{m_y \cdot} - n_{m_y m_x})(n_{\cdot m_x} - n_{m_y m_x})}{n_{m_y \cdot} \cdot n_{\cdot m_x}} \\ &= 1 - \frac{n_{m_y m_x}}{n_{m_y \cdot}} - \frac{n_{m_y m_x}}{n_{\cdot m_x}} + \frac{n_{m_y m_x}^2}{n_{m_y \cdot} \cdot n_{\cdot m_x}}.\end{aligned}$$

Substituting these terms in the expression for trace, we have

$$\begin{aligned}&Tr \left[(\mathbf{Y}'\mathbf{M}_\tau\mathbf{Y})^{-1} (\mathbf{Y}'\mathbf{M}_\tau\mathbf{X}) (\mathbf{X}'\mathbf{M}_\tau\mathbf{X})^{-1} (\mathbf{X}'\mathbf{M}_\tau\mathbf{Y}) \right] \\ &= \sum_{i=1}^{m_y-1} \sum_{j=1}^{m_x-1} \frac{n_{ij}^2}{n_i \cdot n_{\cdot j}} - 1 + \sum_{i=1}^{m_y-1} \frac{n_{im_x}^2}{n_i \cdot n_{\cdot m_x}} + \sum_{j=1}^{m_x-1} \frac{n_{m_y j}^2}{n_{m_y \cdot} \cdot n_{\cdot j}} + \frac{n_{m_y m_x}^2}{n_{m_y \cdot} \cdot n_{\cdot m_x}} \\ &= \sum_{i=1}^{m_y} \sum_{j=1}^{m_x} \frac{n_{ij}^2}{n_i \cdot n_{\cdot j}} - 1,\end{aligned}$$

as required. ■

Proof of Proposition 2. To prove the proposition, we show that nothing is gained by using “optimal scores”, i.e. by optimizing over the values assigned to the variables in each category - see, e.g., Kendall and Stuart (1979, pp. 597-606). Accordingly, suppose that the problem of interest is to assign scores a_i ($i = 1, 2, \dots, m_y$) and b_j ($j = 1, 2, \dots, m_x$) to the categories of an $m_y \times m_x$ table, such that a maximal correlation between $y_t^* = \sum_{i=1}^{m_y} a_i y_{it}$ and $x_t^* = \sum_{j=1}^{m_x} b_j x_{jt}$ is obtained. Without loss of generality the scores are standardized so that y_t^* and x_t^* have zero means and unit variances. In the case of serially independent outcomes, y_{it} and x_{jt} , and for n sufficiently large we have

$$\begin{aligned}E(y_t^*) &= n^{-1} \sum_{i=1}^{m_y} a_i n_i = 0, \quad E(x_t^*) = n^{-1} \sum_{j=1}^{m_x} b_j n_{\cdot j} = 0, \\ Var(y_t^*) &= n^{-1} \sum_{i=1}^{m_y} a_i^2 n_i = 1, \quad Var(x_t^*) = n^{-1} \sum_{j=1}^{m_x} b_j^2 n_{\cdot j} = 1 \\ Cov(y_t^*, x_t^*) &= Corr(y_t^*, x_t^*) = n^{-1} \sum_{i=1}^{m_y} \sum_{j=1}^{m_x} n_{ij} a_i b_j.\end{aligned}$$

The first order conditions for this optimization problem are given by

$$\begin{aligned}\sum_{j=1}^{m_x} n_{ij} b_j - \lambda_y n_i a_i &= 0, \text{ for } i = 1, 2, \dots, m_y \\ \sum_{i=1}^{m_y} n_{ij} a_i - \lambda_x n_{.j} b_j &= 0, \text{ for } j = 1, 2, \dots, m_x,\end{aligned}$$

where λ_y and λ_x are Lagrange multipliers associated with the constraints $Var(y_t^*) = 1$ and $Var(x_t^*) = 1$, respectively. In matrix notations

$$\mathcal{Y}' \mathcal{X} \mathbf{b} = \lambda_y (\mathcal{Y}' \mathcal{Y}) \mathbf{a}, \quad \mathcal{X}' \mathcal{Y} \mathbf{a} = \lambda_x (\mathcal{X}' \mathcal{X}) \mathbf{b},$$

$\mathbf{a}' (\mathcal{Y}' \mathcal{Y}) \mathbf{a} = 1$, and $\mathbf{b}' (\mathcal{X}' \mathcal{X}) \mathbf{b} = 1$, where \mathcal{Y} and \mathcal{X} are $T \times m_y$ and $T \times m_x$ matrices on y_{it} , $i = 1, 2, \dots, m_y$ and x_{jt} , $j = 1, 2, \dots, m_x$, for $t = 1, 2, \dots, T$, $\mathbf{a} = (a_1, a_2, \dots, a_{m_y})'$, and $\mathbf{b} = (b_1, b_2, \dots, b_{m_x})'$. In the earlier notations $\mathcal{X} = (\mathbf{X}, \mathbf{x}_{m_x})$ and $\mathcal{Y} = (\mathbf{Y}, \mathbf{y}_{m_y})$. Note that $\mathcal{Y}' \mathcal{Y}$ and $\mathcal{X}' \mathcal{X}$ are diagonal matrices with diagonal elements given by $n_i > 0$ and $n_{.j} > 0$, respectively, and $\mathcal{Y}' \mathcal{X}$ is an $m_y \times m_x$ matrix with its (i, j) element given by n_{ij} . Also $R = n^{-1} \sum_{i=1}^{m_y} \sum_{j=1}^{m_x} n_{ij} a_i b_j = n^{-1} \mathbf{a}' (\mathcal{Y}' \mathcal{X}) \mathbf{b}$. It is now easily seen that $\lambda_y = \lambda_x = R$, and hence

$$\begin{aligned}\left[\mathcal{Y}' \mathcal{X} (\mathcal{X}' \mathcal{X})^{-1} \mathcal{X}' \mathcal{Y} - R^2 (\mathcal{Y}' \mathcal{Y}) \right] \mathbf{a} &= \mathbf{0}, \\ \left[\mathcal{X}' \mathcal{Y} (\mathcal{Y}' \mathcal{Y})^{-1} \mathcal{Y}' \mathcal{X} - R^2 (\mathcal{X}' \mathcal{X}) \right] \mathbf{b} &= \mathbf{0}.\end{aligned}$$

Since $\mathcal{Y}' \mathcal{Y}$ and $\mathcal{X}' \mathcal{X}$ are non-singular and $m_x \leq m_y$, R^2 can be obtained by computing the eigenvalues of $(\mathcal{X}' \mathcal{X})^{-1} \mathcal{X}' \mathcal{Y} (\mathcal{Y}' \mathcal{Y})^{-1} \mathcal{Y}' \mathcal{X}$, that are the same as the non-zero eigenvalues of $(\mathcal{Y}' \mathcal{Y})^{-1} \mathcal{Y}' \mathcal{X} (\mathcal{X}' \mathcal{X})^{-1} \mathcal{X}' \mathcal{Y}$. Since $\mathcal{X}' \boldsymbol{\tau}_{m_x} = \boldsymbol{\tau}_T$ and $\mathcal{Y}' \boldsymbol{\tau}_{m_y} = \boldsymbol{\tau}_T$, it is easily seen that one of the eigenvalues is unity.¹² This leaves $m_x - 1$ non-zero eigenvalues, $R_1^2 \geq R_2^2 \geq \dots \geq R_{m_x-1}^2$. The solutions to \mathbf{a} and \mathbf{b} are given by the eigenvectors associated with R_1^2 . It also follows that

$$Tr \left[(\mathcal{X}' \mathcal{X})^{-1} \mathcal{X}' \mathcal{Y} (\mathcal{Y}' \mathcal{Y})^{-1} \mathcal{Y}' \mathcal{X} \right] = \sum_{i=1}^{m_y} \sum_{j=1}^{m_x} \frac{n_{ij}^2}{n_i n_{.j}} = 1 + \sum_{j=1}^{m_x-1} R_j^2,$$

which is the result obtained in Proposition 1. ■

¹²Let $\mathbf{a} = \boldsymbol{\tau}_{m_y}$ and note that since $\mathcal{Y}' \boldsymbol{\tau}_{m_y} = \boldsymbol{\tau}_T$, and $\mathcal{X} (\mathcal{X}' \mathcal{X})^{-1} \mathcal{X}' \boldsymbol{\tau}_T = \mathcal{X}' \boldsymbol{\tau}_{m_x} = \boldsymbol{\tau}_T$, we have

$$\left[\mathcal{Y}' \mathcal{X} (\mathcal{X}' \mathcal{X})^{-1} \mathcal{X}' \mathcal{Y} - R^2 (\mathcal{Y}' \mathcal{Y}) \right] \boldsymbol{\tau}_{m_y} = (1 - R^2) \boldsymbol{\tau}_T = \mathbf{0}.$$

But $\boldsymbol{\tau}_T = (n_{1.}, n_{2.}, \dots, n_{m_y.}) > 0$ and it must be that $R^2 = 1$.

References

- [1] Anderson, T.W., 2003, An Introduction to Multivariate Statistical Analysis, Third Edition, Wiley Series in Probability and Statistics, Wiley, New York.
- [2] Andrews, D.W.K., 1991, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica* 60, 953-966.
- [3] Andrews, D.W.K., and J.C. Monahan, 1992, An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica* 60, 953-966.
- [4] Andrews, D.W.K. and W. Ploberger, 1994, Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative. *Econometrica* 62, 1383-1414.
- [5] Cumby, R.E. and D.M. Modest, 1987, Testing for Market Timing Ability. A Framework for Forecast Evaluation. *Journal of Financial Economics* 19, 169-189.
- [6] Davies, R., 1977, Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative, *Biometrika*, 64, 247-254.
- [7] Favero, C.A. and F. Giavazzi, 2002, Is the International Propagation of Financial Shocks Non-linear? Evidence from the ERM. *Journal of International Economics*, 57, 231-246
- [8] Hall, A. 2004, Generalized Method of Moments. Oxford University Press, Oxford.
- [9] Hamilton, J.D., 1989, A New Approach to the Economic Analysis of Non-Stationary Time Series and the Business Cycle. *Econometrica* 57, 357-384.
- [10] Hansen, B., 1996, Inference when a Nuisance Parameter is not Identified under the Null Hypothesis. *Econometrica* 64, 413-430.
- [11] Harding, D. and A.R. Pagan, 2004. Synchronization of Cycles. Forthcoming in *Journal of Econometrics*.
- [12] Henriksson, R.D. and R.C. Merton, 1981, On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills. *Journal of Business* 54, 513-533.
- [13] Hirschfeld, H.O., 1935, A Connection between Correlation and Contingency, *Proceedings of the Cambridge Philosophical Society*, 31, 520-524.
- [14] Hotelling, H., 1935, The Most Predictable Criterion. *Journal of Educational Psychology*, 26, 139-142.
- [15] Hotelling, H., 1936, Relations Between Two Sets of Variates. *Biometrika*, 28, 321-377.
- [16] Johansen, S., 1995, Likelihood Based Inference on Cointegration in the Vector Autoregressive Model. Oxford University Press, Oxford, 2nd ed..
- [17] Kendall M. and A. Stuart, 1979, *The Advanced Theory of Statistics (Fourth Edition)*, Volume 2, Charles Griffin, London.
- [18] Kiefer, N.M. and T.J. Vogelsang, 2002, Heteroskedasticity and Autocorrelation Robust Standard Errors using the Bartlett Kernel without Truncation. *Econometrica* 70, 2093-2095.
- [19] Lancaster, H.O., 1958, The Structure of Bivariate Distributions, *The Annals of Mathematical Statistics*, 29, 719-736.

- [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] Newey, W.K. and K.D. West, 1987, A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 703-708.
- [22] Pagan, A.R., 2005, Some Econometric Analysis of Constructed Binary Time Series. Manuscript, Australian National University.
- [23] Pagan, A.R. and Sossounov, K.A., 2003, A Simple Framework for Analyzing Bull and Bear Markets. *Journal of Applied Econometrics* 18, 23-46.
- [24] Pesaran, M.H., 1981, Pitfalls of Testing Non-nested Hypotheses by the Lagrange Multiplier Method, *Journal of Econometrics*, 17, 323-331.
- [25] Pesaran, M.H. and A. Pick, 2006, Econometric Issues in the Analysis of Contagion. Forthcoming in *Journal of Economic Dynamics and Control*.
- [26] Pesaran, M.H. and A. Timmermann, 1992, A Simple Nonparametric Test of Predictive Performance. *Journal of Business and Economic Statistics* 10, 461-465.
- [27] Phillips, P.C.B., Y. Sun and S. Jin, 2005, Spectral Density Estimation and Robust Hypothesis Testing Using Steep Origin Kernels without Truncation. Forthcoming in *International Economic Review*.

Table 1. Finite Sample Critical Values for Maximum Canonical Correlation and Trace Canonical Correlation Tests

A. Maximum Canonical Correlation Test									
sample size	90%			95%			99%		
	<i>m</i> = 2	<i>m</i> = 3	<i>m</i> = 4	<i>m</i> = 2	<i>m</i> = 3	<i>m</i> = 4	<i>m</i> = 2	<i>m</i> = 3	<i>m</i> = 4
20	2.73	5.90	8.39	3.69	7.01	9.40	6.31	9.26	11.54
50	2.71	6.62	10.09	3.79	7.99	11.61	6.30	11.07	14.78
100	2.63	6.80	10.71	3.87	8.27	12.43	6.60	11.64	16.09
500	2.67	7.00	11.14	3.85	8.59	12.99	6.70	12.10	17.01
1000	2.71	6.97	11.20	3.84	8.50	13.03	6.66	12.11	17.07

B. Trace Canonical Correlation Test									
sample size	90%			95%			99%		
	<i>m</i> = 2	<i>m</i> = 3	<i>m</i> = 4	<i>m</i> = 2	<i>m</i> = 3	<i>m</i> = 4	<i>m</i> = 2	<i>m</i> = 3	<i>m</i> = 4
20	2.73	6.63	11.52	3.69	7.84	12.98	6.31	10.52	16.16
50	2.71	7.36	13.46	3.79	8.88	15.34	6.30	12.26	19.39
100	2.63	7.57	14.12	3.87	9.18	16.19	6.60	12.71	20.66
500	2.67	7.77	14.55	3.85	9.48	16.72	6.70	13.32	21.44
1000	2.71	7.74	14.60	3.84	9.40	16.84	6.66	13.28	21.45

Notes: The table is based on 100,000 Monte Carlo simulations under the null of no serial correlation in the data, *m* is the number of categories.

Table 2. Size Properties of the Tests (rejection frequencies under no cross-correlation, $r_{xy} = 0$)

A. No Serial Correlation ($\varphi = 0$)								
m	Sample Size	PT Test	Trace Canonical Correlation			Maximum Canonical Correlation		
			Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated
2	20	0.042	0.042	0.061	0.009	0.042	0.061	0.009
2	50	0.077	0.053	0.059	0.035	0.053	0.059	0.035
2	100	0.068	0.068	0.060	0.048	0.068	0.060	0.048
2	500	0.053	0.053	0.051	0.047	0.053	0.051	0.047
2	1000	0.051	0.043	0.046	0.044	0.043	0.046	0.044
3	20	0.106	0.061	0.141	0.024	0.054	0.132	0.018
3	50	0.073	0.052	0.077	0.017	0.050	0.067	0.003
3	100	0.070	0.054	0.065	0.016	0.056	0.065	0.011
3	500	0.048	0.050	0.053	0.036	0.051	0.053	0.034
3	1000	0.060	0.066	0.065	0.053	0.059	0.057	0.047
4	20	0.129	0.048	0.304	0.174	0.057	0.281	0.194
4	50	0.068	0.045	0.104	0.045	0.049	0.095	0.036
4	100	0.065	0.054	0.074	0.028	0.052	0.070	0.012
4	500	0.062	0.051	0.056	0.033	0.052	0.058	0.024
4	1000	0.054	0.054	0.058	0.044	0.051	0.051	0.033
B. Medium Serial Correlation ($\varphi = 0.5$)								
m	Sample Size	PT Test	Trace Canonical Correlation			Maximum Canonical Correlation		
			Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated
2	20	0.072	0.072	0.063	0.015	0.072	0.063	0.015
2	50	0.105	0.090	0.053	0.045	0.090	0.053	0.045
2	100	0.103	0.101	0.055	0.043	0.101	0.055	0.043
2	500	0.086	0.086	0.050	0.050	0.086	0.050	0.050
2	1000	0.097	0.090	0.060	0.062	0.090	0.060	0.062
3	20	0.149	0.074	0.141	0.036	0.077	0.137	0.025
3	50	0.112	0.085	0.077	0.013	0.079	0.074	0.004
3	100	0.097	0.081	0.063	0.018	0.085	0.057	0.016
3	500	0.088	0.077	0.048	0.041	0.076	0.050	0.042
3	1000	0.073	0.086	0.057	0.049	0.087	0.058	0.046
4	20	0.142	0.077	0.338	0.198	0.071	0.326	0.219
4	50	0.092	0.070	0.094	0.053	0.068	0.088	0.047
4	100	0.078	0.082	0.071	0.035	0.071	0.076	0.019
4	500	0.075	0.074	0.052	0.035	0.073	0.052	0.027
4	1000	0.081	0.073	0.050	0.035	0.064	0.051	0.029
C. High Serial Correlation ($\varphi = 0.8$)								
m	Sample Size	PT Test	Trace Canonical Correlation			Maximum Canonical Correlation		
			Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated
2	20	0.153	0.153	0.068	0.034	0.153	0.068	0.034
2	50	0.234	0.202	0.049	0.081	0.202	0.049	0.081
2	100	0.235	0.224	0.058	0.089	0.224	0.058	0.089
2	500	0.245	0.245	0.046	0.073	0.245	0.046	0.073
2	1000	0.246	0.234	0.056	0.080	0.234	0.056	0.080
3	20	0.231	0.167	0.161	0.045	0.171	0.151	0.035
3	50	0.207	0.232	0.082	0.020	0.225	0.074	0.007
3	100	0.203	0.267	0.053	0.025	0.263	0.057	0.017
3	500	0.206	0.306	0.050	0.047	0.304	0.050	0.050
3	1000	0.207	0.307	0.054	0.051	0.297	0.051	0.050
4	20	0.196	0.135	0.349	0.267	0.141	0.324	0.298
4	50	0.183	0.255	0.126	0.085	0.238	0.111	0.076
4	100	0.165	0.275	0.076	0.049	0.268	0.070	0.032
4	500	0.177	0.286	0.057	0.041	0.275	0.055	0.024
4	1000	0.165	0.290	0.050	0.039	0.285	0.047	0.033

Notes: Results are based on 2,000 Monte Carlo replications, using the finite-sample critical values from Table 1. When $m = 2$, the PT statistic is a test of independence, while when $m > 2$, the null of this test is that the diagonal cell probabilities in the $m \times m$ contingency table take their values under the null of independence between the row and column variables - m is the number of categories.

Table 3. Power of the Tests (under the alternative of medium cross-correlation, $r_{xy} = 0.2$)

A. No Serial Correlation ($\varphi = 0$)									
m	Sample	PT	Trace Canonical Correlation			Maximum Canonical Correlation			
	Size	Test	Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated	
2	20	0.082	0.082	0.089	0.013	0.082	0.089	0.013	
2	50	0.154	0.123	0.134	0.081	0.123	0.134	0.081	
2	100	0.266	0.266	0.238	0.195	0.266	0.238	0.195	
2	500	0.828	0.828	0.815	0.811	0.828	0.815	0.811	
2	1000	0.984	0.980	0.981	0.978	0.980	0.981	0.978	
3	20	0.134	0.078	0.156	0.029	0.080	0.156	0.021	
3	50	0.138	0.117	0.152	0.029	0.118	0.143	0.011	
3	100	0.228	0.212	0.231	0.098	0.208	0.219	0.068	
3	500	0.755	0.829	0.830	0.774	0.837	0.837	0.786	
3	1000	0.960	0.994	0.993	0.988	0.995	0.994	0.992	
4	20	0.101	0.063	0.313	0.180	0.067	0.283	0.201	
4	50	0.102	0.099	0.154	0.063	0.107	0.160	0.053	
4	100	0.163	0.172	0.208	0.068	0.169	0.191	0.032	
4	500	0.648	0.790	0.793	0.645	0.820	0.819	0.662	
4	1000	0.914	0.988	0.989	0.976	0.990	0.991	0.982	
B. Medium Serial Correlation ($\varphi = 0.5$)									
m	Sample	PT	Trace Canonical Correlation			Maximum Canonical Correlation			
	Size	Test	Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated	
2	20	0.096	0.096	0.085	0.017	0.096	0.085	0.017	
2	50	0.215	0.175	0.141	0.104	0.175	0.141	0.104	
2	100	0.290	0.278	0.202	0.178	0.278	0.202	0.178	
2	500	0.793	0.793	0.703	0.701	0.793	0.703	0.701	
2	1000	0.972	0.966	0.941	0.942	0.966	0.941	0.942	
3	20	0.173	0.105	0.181	0.035	0.106	0.175	0.024	
3	50	0.164	0.144	0.132	0.024	0.143	0.132	0.009	
3	100	0.251	0.256	0.204	0.063	0.252	0.199	0.053	
3	500	0.719	0.798	0.750	0.603	0.802	0.752	0.615	
3	1000	0.933	0.984	0.971	0.931	0.986	0.970	0.937	
4	20	0.129	0.090	0.347	0.237	0.095	0.328	0.263	
4	50	0.150	0.136	0.150	0.079	0.137	0.140	0.067	
4	100	0.161	0.203	0.175	0.060	0.209	0.162	0.030	
4	500	0.663	0.762	0.716	0.441	0.781	0.735	0.446	
4	1000	0.899	0.968	0.970	0.845	0.970	0.979	0.864	
C. High Serial Correlation ($\varphi = 0.8$)									
m	Sample	PT	Trace Canonical Correlation			Maximum Canonical Correlation			
	Size	Test	Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated	
2	20	0.172	0.172	0.081	0.033	0.172	0.081	0.033	
2	50	0.326	0.300	0.106	0.141	0.300	0.106	0.141	
2	100	0.381	0.360	0.135	0.168	0.360	0.135	0.168	
2	500	0.724	0.724	0.430	0.473	0.724	0.430	0.473	
2	1000	0.908	0.900	0.736	0.764	0.900	0.736	0.764	
3	20	0.240	0.181	0.178	0.042	0.186	0.170	0.033	
3	50	0.256	0.309	0.118	0.033	0.308	0.108	0.013	
3	100	0.357	0.422	0.154	0.059	0.422	0.158	0.043	
3	500	0.661	0.761	0.518	0.321	0.763	0.528	0.328	
3	1000	0.839	0.921	0.845	0.596	0.921	0.849	0.604	
4	20	0.205	0.148	0.350	0.285	0.160	0.323	0.321	
4	50	0.230	0.311	0.153	0.103	0.298	0.147	0.090	
4	100	0.262	0.390	0.143	0.071	0.387	0.142	0.047	
4	500	0.594	0.751	0.515	0.188	0.757	0.534	0.187	
4	1000	0.824	0.917	0.869	0.427	0.921	0.887	0.464	

Notes: Results are based on 2,000 Monte Carlo replications, using the finite-sample critical values from Table 1. When $m = 2$, the PT statistic is a test of independence, while when $m > 2$, the null of this test is that the diagonal cell probabilities in the $m \times m$ contingency table take their values under the null of independence between the row and column variables - m is the number of categories.

Table 4. Power of the Tests (under the alternative of high cross-correlation, $r_{xy} = 0.8$)

A. No Serial Correlation ($\varphi = 0$)								
m	Sample	PT	Trace Canonical Correlation			Maximum Canonical Correlation		
	Size	Test	Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated
2	20	0.732	0.732	0.709	0.330	0.732	0.709	0.330
2	50	0.994	0.989	0.991	0.972	0.989	0.991	0.972
2	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	20	0.819	0.727	0.771	0.297	0.740	0.774	0.247
3	50	0.993	0.998	0.996	0.758	0.997	0.996	0.636
3	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	20	0.721	0.630	0.812	0.640	0.681	0.790	0.651
4	50	0.989	0.997	0.995	0.840	0.996	0.996	0.793
4	100	1.000	1.000	1.000	0.994	1.000	1.000	0.992
4	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
B. Medium Serial Correlation ($\varphi = 0.5$)								
m	Sample	PT	Trace Canonical Correlation			Maximum Canonical Correlation		
	Size	Test	Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated
2	20	0.708	0.708	0.639	0.302	0.708	0.639	0.302
2	50	0.992	0.986	0.980	0.970	0.986	0.980	0.970
2	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	20	0.802	0.717	0.737	0.282	0.736	0.723	0.236
3	50	0.986	0.996	0.991	0.662	0.996	0.992	0.514
3	100	1.000	1.000	1.000	0.999	1.000	1.000	0.999
3	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	20	0.705	0.615	0.776	0.589	0.656	0.751	0.611
4	50	0.981	0.997	0.994	0.765	0.998	0.996	0.704
4	100	1.000	1.000	1.000	0.973	1.000	1.000	0.965
4	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
C. High Serial Correlation ($\varphi = 0.8$)								
m	Sample	PT	Trace Canonical Correlation			Maximum Canonical Correlation		
	Size	Test	Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated
2	20	0.675	0.675	0.516	0.288	0.675	0.516	0.288
2	50	0.975	0.958	0.864	0.887	0.958	0.864	0.887
2	100	1.000	0.999	0.986	0.993	0.999	0.986	0.993
2	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	20	0.735	0.688	0.626	0.277	0.710	0.609	0.239
3	50	0.963	0.977	0.942	0.536	0.980	0.943	0.389
3	100	0.999	1.000	1.000	0.954	1.000	1.000	0.960
3	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	20	0.638	0.600	0.735	0.571	0.636	0.706	0.589
4	50	0.936	0.976	0.958	0.667	0.976	0.960	0.628
4	100	0.997	1.000	1.000	0.868	1.000	1.000	0.814
4	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: Results are based on 2,000 Monte Carlo replications, using the finite-sample critical values from Table 1. When $m = 2$, the PT statistic is a test of independence, while when $m > 2$, the null of this test is that the diagonal cell probabilities in the $m \times m$ contingency table take their values under the null of independence between the row and column variables - m is the number of categories.

Table 5. Alternative Procedures for Allowing for Serial Correlation

A. Size ($r_{xy} = 0.0$)							
sample size	serial correl.	Static regressions			Dyn. Augm. regressions		
		OLS	Barlett	QS	OLS	Barlett	QS
20	0.0	0.042	0.045	0.051	0.078	0.091	0.098
50	0.0	0.077	0.050	0.053	0.069	0.067	0.068
100	0.0	0.068	0.064	0.062	0.066	0.064	0.068
500	0.0	0.053	0.049	0.048	0.052	0.052	0.051
1000	0.0	0.051	0.047	0.047	0.046	0.046	0.046
20	0.5	0.072	0.084	0.101	0.081	0.101	0.111
50	0.5	0.105	0.074	0.080	0.058	0.057	0.059
100	0.5	0.103	0.071	0.066	0.060	0.061	0.062
500	0.5	0.086	0.057	0.055	0.052	0.051	0.053
1000	0.5	0.097	0.069	0.065	0.060	0.061	0.061
20	0.8	0.153	0.150	0.179	0.083	0.095	0.101
50	0.8	0.234	0.129	0.124	0.055	0.051	0.057
100	0.8	0.235	0.120	0.113	0.062	0.058	0.059
500	0.8	0.245	0.076	0.072	0.047	0.051	0.052
1000	0.8	0.246	0.081	0.070	0.057	0.058	0.059
B. Power ($r_{xy} = 0.2$)							
sample size	serial correl.	Static regressions			Dyn. Augm. regressions		
		OLS	Barlett	QS	OLS	Barlett	QS
20	0.0	0.082	0.088	0.100	0.110	0.130	0.141
50	0.0	0.154	0.114	0.115	0.144	0.140	0.137
100	0.0	0.266	0.249	0.243	0.254	0.248	0.245
500	0.0	0.828	0.822	0.821	0.819	0.819	0.820
1000	0.0	0.984	0.982	0.982	0.982	0.981	0.982
20	0.5	0.096	0.102	0.113	0.102	0.117	0.118
50	0.5	0.215	0.161	0.171	0.150	0.149	0.152
100	0.5	0.290	0.238	0.232	0.212	0.207	0.213
500	0.5	0.793	0.724	0.717	0.706	0.704	0.704
1000	0.5	0.972	0.946	0.944	0.941	0.940	0.939
20	0.8	0.172	0.170	0.192	0.100	0.106	0.112
50	0.8	0.326	0.220	0.218	0.111	0.098	0.099
100	0.8	0.381	0.221	0.203	0.147	0.141	0.140
500	0.8	0.724	0.480	0.459	0.434	0.435	0.436
1000	0.8	0.908	0.759	0.750	0.738	0.737	0.740
C. Power ($r_{xy} = 0.8$)							
sample size	serial correl.	Static regressions			Dyn. Augm. regressions		
		OLS	Barlett	QS	OLS	Barlett	QS
20	0.0	0.732	0.731	0.733	0.748	0.726	0.726
50	0.0	0.994	0.989	0.990	0.992	0.990	0.992
100	0.0	1.000	1.000	1.000	1.000	1.000	1.000
500	0.0	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.0	1.000	1.000	1.000	1.000	1.000	1.000
20	0.5	0.708	0.704	0.712	0.684	0.627	0.640
50	0.5	0.992	0.985	0.984	0.983	0.975	0.974
100	0.5	1.000	1.000	1.000	1.000	1.000	1.000
500	0.5	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.5	1.000	1.000	1.000	1.000	1.000	1.000
20	0.8	0.675	0.671	0.690	0.554	0.476	0.485
50	0.8	0.975	0.935	0.933	0.875	0.811	0.813
100	0.8	1.000	0.995	0.994	0.988	0.979	0.980
500	0.8	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.8	1.000	1.000	1.000	1.000	1.000	1.000

Notes: Results are based on 2,000 Monte Carlo replications and assume $m = 2$ categories. The static regressions do not include lags of x and y , while the dynamically augmented regression includes a number of lags of each variable selected by the Akaike Information Criterion.

Table 6. Empirical Results from the Survey of Professional Forecasters Predictions of GDP Growth

m	Trace Canonical Correlation			Maximum Canonical Correlation		
	Static	Dyn. Augm.	Iterated	Static	Dyn. Augm.	Iterated
A. Current Year results						
2	4.928*	2.009	2.640	4.928*	2.009	2.640
3	13.347 [†]	3.124	5.946	13.27 [†]	2.979	5.888
4	44.886 [‡]	21.03 [‡]	41.398 [‡]	31.07 [‡]	14.185*	35.634 [‡]
B. Next year results						
2	1.578	0.084	1.234	1.578	0.084	1.234
3	9.512*	6.649	7.311	7.557 [†]	6.568	6.182
4	13.230	8.450	9.046	7.754	6.873	7.091

Notes: [†], * and [‡] indicate significance at the 10%, 5% and 1% levels, respectively.

Figure 1. Power Curves for Maximum Canonical Correlation Test ($\varphi = 0, m = 2, T = 100$)

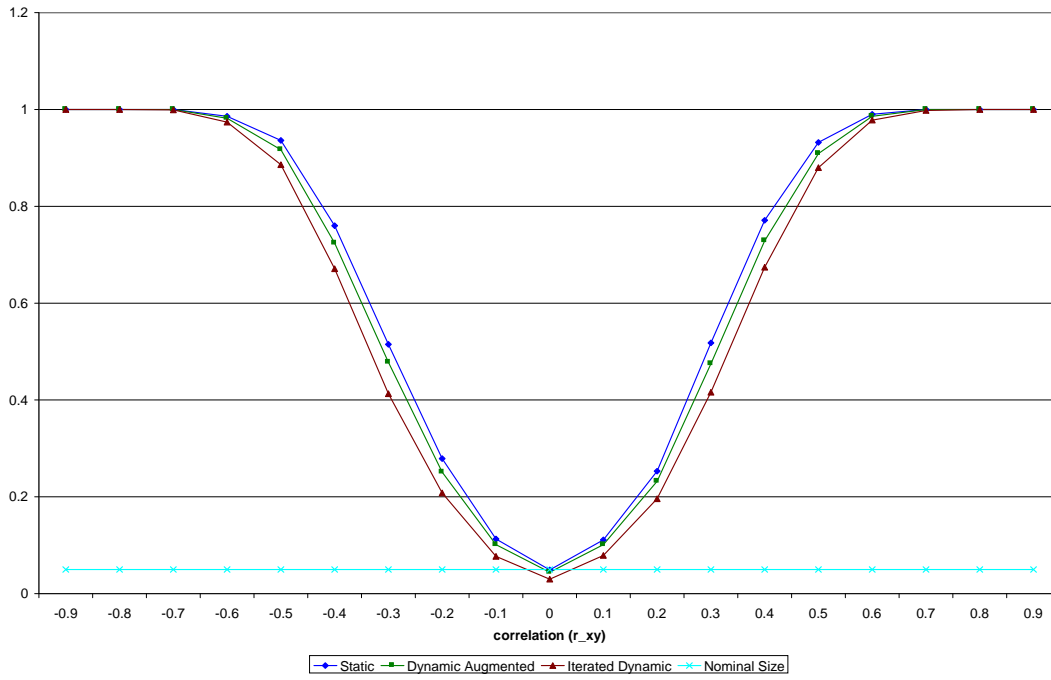


Figure 2. Power Curves for Maximum Canonical Correlation Test ($\varphi = 0.5, m = 2, T = 100$)

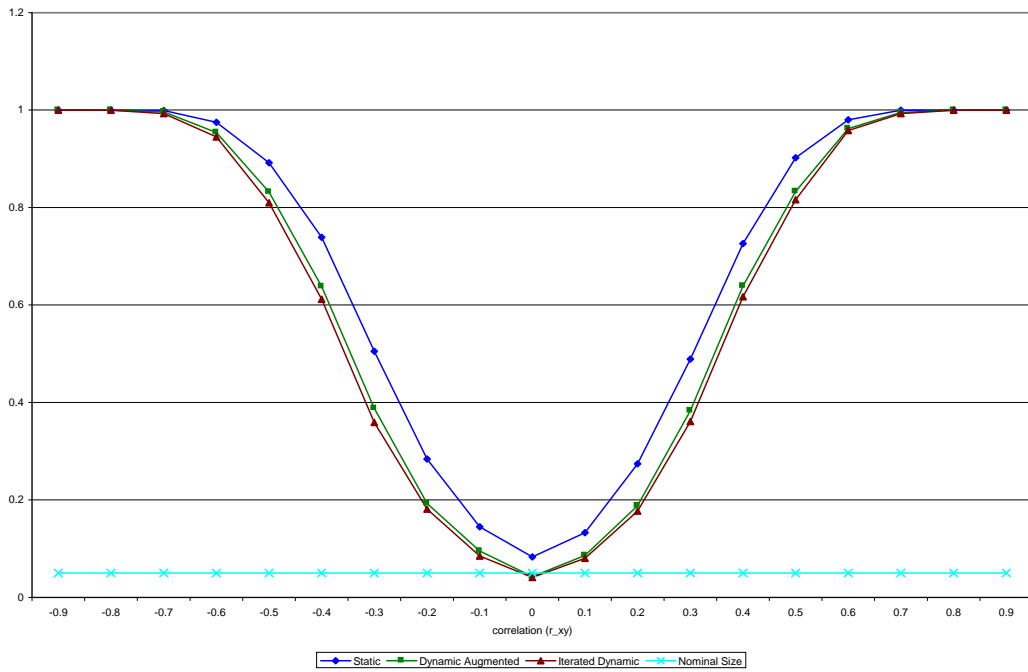


Figure 3. Power Curves for Maximum Canonical Correlation Test ($\varphi = 0.8, m = 2, T = 100$)

