Centre for
Economic
and Financial
Research
at
New Economic
School

December 2006

Tests in Contingency Tables as Regression Tests

Stanislav Anatolyev
Grigory Kosenok

Working Paper No 75

CEFIR / NES Working Paper series

Tests in contingency tables as regression tests

by

Stanislav Anatolyev* and Grigory Kosenok

New Economic School

Abstract

Applied researchers often use tests based on contingency tables in preliminary data analysis and diagnostic testing. We show that many of such tests may be alternatively implemented by testing for coefficient restrictions in linear regression systems (as a rule, employing the Wald test). This unifies the theories of regression analysis and contingency tables, sheds more light on intuitive contents of contingency table tests, and provides a more convenient and familiar tool for practitioners.

Key Words and Phrases: Contingency table; Linear regression, χ^2-test, Wald test, Ranks. **JEL codes:** C12, C22, C32, C53.

*Corresponding author. Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru.

1 Introduction

Often in their research applied economists and financiers use tests associated with contingency tables (see a review in Kendall and Stuart 1973, chapter 33). Such tests are designed for verifying independence or homogeneity properties of original data or regression residuals, and are heavily used in preliminary data analysis and diagnostic testing. For example, the phrase "contingency table" leads to 184 and 58 hits in the "advance search" (performed in 2006) in JSTOR economics (24 journals) and finance (5 journals) collections, respectively. Some of associated tests are even more frequently mentioned.

The tests related to contingency tables are performed by utilizing particular often complex formulas and a table of (usually) the χ^2 critical values. In this paper, we show that typically they may be alternatively implemented via a system of linear regressions. The χ^2 tests are asymptotically equivalent to Wald tests, while standard normal tests (which is a rarer situation) are asymptotically equivalent to t tests, performed in certain regression systems. Such reformulation is useful for a number of reasons, both for econometric theory and econometric practice. First, this unifies the regression analysis with the theory of contingency tables. Second, the bridge between the two theories sheds more light on intuitive contents of contingency table tests and test statistics. Third, running regressions may be more convenient and familiar for econometric practitioners.

Some remarks on notation used now follow. The sample size is denoted by n. Bars denote taking sample averages, i.e., for example, $\overline{a}_{ij} = n^{-1} \sum_{t=1}^{n} a_{ij}$. By $\|a_i\|_{i=1}^{\ell}$ we mean a column $\ell \times 1$ vector with i^{th} element a_i. By $\|a_{i,j}\|_{i=1}^{\ell_1} {}_{j=1}^{\ell_2}$ we mean an $\ell_1 \times \ell_2$ matrix whose i^{th}, j^{th} element equals $a_{i,j}$.

The paper is structured as follows. Section 2 describes notation specific for contingency tables. The most widely used test for independence and its variations are presented in section 3. Tests for accordance with distributions including the famous Pearson goodness-of-fit test are discussed in section 4. More rarely used tests for a symmetry in contingency tables are examined in section 5. Section 6 is devoted to popular tests based on ranks, such as the Kruskal–Wallis and Spearman tests. All proofs are relegated to appendix A, while appendix B contains some supplementary material.

2 Two-way contingency tables

We consider a two-way $(\ell_x+1)\times(\ell_y+1)$ contingency table. The state space Ω_x of one variable, x, is partitioned into the cover $\{K_i\}_{i=1}^{\ell_x+1}$; similarly, the state space Ω_y of the other variable, y, is partitioned into the cover $\{\Lambda_j\}_{j=1}^{\ell_y+1}$. Let us denote

$$\mathbb{I}_{i,\cdot}=\mathbb{I}\left(x\in K_i\right), \quad \mathbb{I}_{\cdot,j}=\mathbb{I}\left(y\in\Lambda_j\right), \quad \mathbb{I}_{i,j}=\mathbb{I}\left(x\in K_i\right)\mathbb{I}\left(y\in\Lambda_j\right)$$

where $\mathbb{I}\left(\cdot\right)$ denotes the indicator function. Let

$$\pi_{i,\cdot}=\Pr\left\{x\in K_i\right\}, \quad \pi_{\cdot,j}=\Pr\left\{y\in\Lambda_j\right\}, \quad \pi_{i,j}=\Pr\left\{x\in K_i, y\in\Lambda_j\right\},$$

and define

$$\pi_x=\|\pi_{i,\cdot}\|_{i=1}^{\ell_x}, \quad \pi_y=\|\pi_{\cdot,j}\|_{j=1}^{\ell_y} \quad \pi=\|\pi_{i,j}\|_{i=1}^{\ell_x}{}_{j=1}^{\ell_y}.$$

We assume that $\pi_{i,j}>0$ for all i and j.

The contingency table looks as follows:

y

		Λ_1	Λ_2	$\cdots$	Λ_{ℓ_y}	Λ_{ℓ_y+1}	Ω_y
	K_1	$p_{1,1}$	$p_{1,2}$	$\cdots$	p_{1,ℓ_y}	p_{1,ℓ_y+1}	$p_{1,\cdot}$
	K_2	$p_{2,1}$	$p_{2,2}$	$\cdots$	p_{2,ℓ_y}	p_{2,ℓ_y+1}	$p_{2,\cdot}$
x	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
	K_{ℓ_x}	$p_{\ell_x,1}$	$p_{\ell_x,2}$	$\cdots$	p_{ℓ_x,ℓ_y}	p_{ℓ_x,ℓ_y+1}	$p_{\ell_x,\cdot}$
	K_{ℓ_x+1}	$p_{\ell_x+1,1}$	$p_{\ell_x+1,2}$	$\cdots$	p_{ℓ_x+1,ℓ_y}	p_{ℓ_x+1,ℓ_y+1}	$p_{\ell_x+1,\cdot}$
	Ω_x	$p_{\cdot,1}$	$p_{\cdot,2}$	$\cdots$	$p_{\cdot,\ell_y}$	$p_{\cdot,\ell_y+1}$	1

The figures in the tables are empirical probabilities of falling into corresponding cells

$$p_{i,j} = \overline{\mathbb{I}}_{i,j}$$

and marginal empirical probabilities

$$p_{i,\cdot} = \overline{\mathbb{I}}_{i,\cdot}, \quad p_{\cdot,j} = \overline{\mathbb{I}}_{\cdot,j}.$$

In many applications, each K_i is an interval $[\kappa_{i-1}, \kappa_i)$ and each Λ_j is an interval $[\lambda_{j-1}, \lambda_j)$, where $-\infty = \kappa_0 < \kappa_1 < \cdots < \kappa_{\ell_x} < \kappa_{\ell_x+1} = +\infty$ and $-\infty = \lambda_0 < \lambda_1 < \cdots < \lambda_{\ell_y} < \lambda_{\ell_y+1} = +\infty$. When $\ell_x = \ell_y$ and $K_i = \Lambda_i$ for all i, the contingency table is referred to as one with identical categorizations. However, rows and columns of a contingency table need not correspond to partitionings of a real axis, and categorizations need not be identical.

3 Tests for independence

The classical χ^2-test statistic for independence between the variables x and y (to be more precise, for no association between x and y) is equal to

$$X^2 = n \sum_{i=1}^{\ell_x+1} \sum_{j=1}^{\ell_y+1} \frac{\left(p_{i,j} - p_{i,\cdot}p_{\cdot,j}\right)^2}{p_{i,\cdot}p_{\cdot,j}}, \tag{3.1}$$

and is asymptotically distributed as $\chi^2\left(\ell_x \ell_y\right)$.

Theorem 1 *The χ^2 test (3.1) is asymptotically equivalent to an OLS-based Wald test for the nullity of all slope coefficients in a linear multiple regression of $\mathbb{I}_{\cdot,j}$ on $\mathbb{I}_{i,\cdot}$ with a constant in each equation, i.e. for the null*

$$H_0 : \beta_{ji} = 0,\ i = 1, \cdots, \ell_x,\ j = 1, \cdots, \ell_y$$

in the regression system

$$\mathbb{I}_{\cdot,j} = \alpha_j + \sum_{i=1}^{\ell_x} \beta_{ji} \mathbb{I}_{i,\cdot} + \eta_j, \quad j = 1, \cdots, \ell_y. \tag{3.2}$$

Alternatively, $\mathbb{I}_{i,\cdot}$ may be regressed on $\mathbb{I}_{\cdot,j}$ rather than $\mathbb{I}_{\cdot,j}$ are regressed on $\mathbb{I}_{i,\cdot}$.

When the contingency table is 2×2, the test for independence can be run using only one bivariate regression. In the economics and finance literatures, the test for independence is usually applied to 2×2 tables, for example, in Knetsch and Sinden (1984), Greenwood et al (1991), Brown et al (1996), Artis et al (1997), Pecorino and van Boening (2001). In some studies, however, larger tables are considered, e.g., and 3×2 tables in Battalio et al (2001), 2×5 and 2×4 tables in Eckel and Grossman (1998), 5×3 in Russo et al (2001), and 5×2 and 5×4 tables in Bouckaerta and Dhaene (2004).

Pesaran and Timmermann (1992, Section 2; 1994) suggest another test of independence in an $(\ell + 1) \times (\ell + 1)$ contingency table with identical categorizations (i.e. when $\{K_i\}_{i=1}^{\ell} = \{\Lambda_j\}_{j=1}^{\ell}$) in the context of testing for directional time series predictability. Their test statistic is the suitably normalized quantity

$$S_n = \sum_{i=1}^{\ell+1} \left(p_{i,i} - p_{i,\cdot} p_{\cdot,i}\right), \tag{3.3}$$

and is asymptotically standard normal under no association between x and y. Pesaran and Timmermann (1992) point out that their test is particularly appropriate when the focus is on predicting overall changes in y by x rather than on statistical independence between x and y. This test is applied, aside from Pesaran and Timmermann (1992, 1994), in Lane et al (1996).

The Pesaran–Timmermann test can be made regression-based as follows.

Theorem 2 *The Pesaran–Timmermann S_n test is asymptotically equivalent to an OLS-based t test for the null*

$$H_0 : \alpha = 0$$

in the simple regression

$$\sum_{i=1}^{\ell+1} \left(\mathbb{I}_{i,i} - \mathbb{I}_{i,\cdot}\bar{\mathbb{I}}_{\cdot,i} - \mathbb{I}_{\cdot,i}\bar{\mathbb{I}}_{i,\cdot}\right) = \alpha + \eta. \tag{3.4}$$

Note to the editor and referees: Appendix B contains more straightforward versions of the regression-based S_n test which however have certain drawbacks spelled out in appendix B. If this material is not, we will gladly remove it.

4 Tests for accordance with distribution

The previous χ^2 test may be also interpreted as a test for homogeneity of $\ell_x + 1$ subsamples $y|x \in K_i$, i.e. that the conditional distribution of y does not depend on x. Suppose now that the marginals $\{\pi_{\cdot,j}\}_{j=1}^{\ell_y+1}$ are known a priori. Then the χ^2 test, whose statistic can be modified to take account of this knowledge, may be interpreted as a test for accordance of $\ell_x + 1$ independent subsamples with a known multinomial distribution. The modified test statistic is equal to

$$X^2 = n \sum_{i=1}^{\ell_x+1} \sum_{j=1}^{\ell_y+1} \frac{\left(p_{i,j} - p_{i,\cdot}\pi_{\cdot,j}\right)^2}{p_{i,\cdot}\pi_{\cdot,j}}, \tag{4.5}$$

and is asymptotically distributed as $\chi^2\left(\left(\ell_x + 1\right)\ell_y\right)$.

Theorem 3 *The χ^2 test (4.5) is asymptotically equivalent to an OLS-based Wald test for the nullity of all coefficients in a linear multiple regression of $\mathbb{I}_{\cdot,j} - \pi_{\cdot,j}$ on $\mathbb{I}_{i,\cdot}$ with a constant in each equation, i.e. for the null*

$$H_0 : \alpha_j = \beta_{ji} = 0,\ i = 1, \cdots, \ell_x,\ j = 1, \cdots, \ell_y$$

in the regression system

$$\mathbb{I}_{\cdot,j} - \pi_{\cdot,j} = \alpha_j + \sum_{i=1}^{\ell_x} \beta_{ji}\mathbb{I}_{i,\cdot} + \eta_j, \quad j = 1, \cdots, \ell_y. \tag{4.6}$$

Remark 1 *The regression from theorem 3 is the same as that from theorem 1. However, the set of restrictions is expanded by additional ℓ_y restrictions of equality of the intercepts in (3.2) to the known a priori y-marginals. This explains the additional ℓ_y degrees of freedom in the asymptotic χ^2 distribution.*

When $\ell_x = 0$, the contingency table essentially becomes one-way, and there is only one subsample $y|x \in \Omega_x$. Then the test is called the Pearson (1900) test for goodness of fit, and it simply verifies if the sample is drawn from a given multinomial distribution. The test statistic becomes Pearson's

$$X^2 = n \sum_{j=1}^{\ell_y+1} \frac{\left(p_{\cdot,j} - \pi_{\cdot,j}\right)^2}{\pi_{\cdot,j}}, \tag{4.7}$$

and is asymptotically distributed as $\chi^2(\ell_y)$.

Corollary 1 *The Pearson test (4.7) is asymptotically equivalent to an OLS-based Wald test for the nullity of all coefficients in a linear multiple regression of* $\mathbb{I}_{\cdot,j} - \pi_{\cdot,j}$ *on a constant in each equation, i.e. for the null*

$$H_0 : \alpha_j = 0, \; j = 1, \cdots, \ell_y$$

in the regression system

$$\mathbb{I}_{\cdot,j} - \pi_{\cdot,j} = \alpha_j + \eta_j, \quad j = 1, \cdots, \ell_y. \tag{4.8}$$

Applications of the Pearson test in economics can be divided into two categories. In the first category, frequencies of simulated values from an estimated model falling into pre-specified bins are compared to a multinomial distribution implied by an assumed continuous distribution, the latter possibly having shape parameters estimated. For example, Hsieh (1989), compares his model's residuals to normal, Student's and generalized error distributions; Alaouze (1987) compares them to one- and two-parameter beta distributions using 10 bins. Most often, however, the reference distribution is uniform so that $\pi_{\cdot,j} = 1/(\ell_y + 1)$, and the leading application is evaluation of conditional density forecasts (see, for example, Chan and Maheu 2002 who use 100 and 70 bins; Bauwens et al. 2004 who use 20 bins). In the second category of applications, one compares frequencies of model-generated predictions falling into pre-specified bins to a multinomial distribution implied by an empirical density. Examples are Merlo (1997) who uses 18 and 10 bins, Moffitt (1989) who uses 101 and 81 bins, Keane and Moffitt (1998) who use 112 and 24 bins, and Hyslop (1999) who uses 148 bins.

When the reference distribution has $p \geq 1$ nuisance shape parameters, applied researchers follow the recommendation of the statistical literature which yields that the asymptotic distribution of the Pearson statistic is unknown, but is bounded between two chi-squared distributions, one with ℓ_y, and the other with $\ell_y - p$ degrees of freedom. Below we show (i) that it is not necessary to estimate the nuisance parameters beforehand, and (ii) how to modify the regression (4.8) to avoid their preliminary estimation. The resulting Wald test statistic on the modified regression will have a chi-squared distributions with $\ell_y - p$ degrees of freedom.

Note that the regression system (4.8) can be rewritten as

$$\mathbb{I}_{\cdot,j} = \alpha_j + \pi_{\cdot,j} + \eta_j, \quad j = 1, \cdots, \ell_y.$$

Let us rewrite this system in a vector form as

$$\|\mathbb{I}_{\cdot,j}\|_{j=1}^{\ell_y} = \|\alpha_j\|_{j=1}^{\ell_y} + \|\pi_{\cdot,j}\|_{j=1}^{\ell_y} + \left\|\eta_j\right\|_{j=1}^{\ell_y},$$

and the associated test as $H_0 : \|\alpha_j\|_{j=1}^{\ell_y} = 0$. Suppose now that there is parameterization $\pi_{\cdot,j} = \pi_{\cdot,j}(\theta)$, where θ is a $p \times 1$ vector of unknown parameters indexing the shape of the reference distribution.

Assume first that $\pi_{\cdot,j}(\theta)$ is linear in θ, i.e. $\|\pi_{\cdot,j}(\theta)\|_{j=1}^{\ell_y} = \Pi_0 + \Pi_1\theta$, where Π_0 is $\ell_y \times 1$ known vector, and Π_1 is $\ell_y \times p$ known matrix with full rank p. Substituting this into the regression system yields

$$\|\mathbb{I}_{\cdot,j}\|_{j=1}^{\ell_y} = \|\alpha_j\|_{j=1}^{\ell_y} + \Pi_0 + \Pi_1\theta + \left\|\eta_j\right\|_{j=1}^{\ell_y}. \tag{4.9}$$

Estimation of α_j's and θ jointly is not possible because they are not separately identified. Let us denote by Φ the $(\ell_y - p) \times \ell_y$ matrix whose rows are the basis of the kernel of Π_1, and pre-map both sides of (4.9) by Φ. Then the system (4.9) becomes

$$\Phi\left(\|\mathbb{I}_{\cdot,j}\|_{j=1}^{\ell_y} - \Pi_0\right) = \Phi\left\|\alpha_j\right\|_{j=1}^{\ell_y} + \Phi\left\|\eta_j\right\|_{j=1}^{\ell_y},$$

because Φ annihilates $\Pi_1\theta$, and the null becomes $\Phi \left\|\alpha_j\right\|_{j=1}^{\ell_y} = 0$, which contains $\ell_y - p$ restrictions placed on α_j's. The associated Wald statistic therefore is distributed as $\chi^2(\ell_y - p)$. In a simple example with $\ell_y = 2$ and $p = 1$, let $\Pi_0 = (0,0)'$, $\Pi_1 = (1,1)'$. Then θ is just added to right sides of both equations of the system, and obviously α_1, α_2 and θ are not separately identified. The matrix Φ is $(1,-1)$, which pre-maps by subtracting the second equation from the first, so the left side of the only emerging equation becomes $\mathbb{I}_{\cdot,1} - \mathbb{I}_{\cdot,2}$, while the right side consists only of the (identified) intercept $\alpha_1 - \alpha_2$, so the null restriction is $\alpha_1 - \alpha_2 = 0$. The associated Wald test is asymptotically $\chi^2(1)$. To summarize, when $\pi_{\cdot,j}(\theta)$ is linear in θ, one needs to exclude θ from the system by taking a suitable linear transformation of its equations and the null restriction, losing p degrees of freedom in the way.

Now consider the general case when $\pi_{\cdot,j}(\theta)$ may be nonlinear in θ, and its Jacobian Π_1 is of full rank at θ_0, where θ_0 is the true value of θ. Substituting the known functional form of $\left\|\pi_{\cdot,j}(\theta)\right\|_{j=1}^{\ell_y}$ into the regression system yields

$$\left\|\mathbb{I}_{\cdot,j}\right\|_{j=1}^{\ell_y} = \left\|\alpha_j\right\|_{j=1}^{\ell_y} + \left\|\pi_{\cdot,j}(\theta)\right\|_{j=1}^{\ell_y} + \left\|\eta_j\right\|_{j=1}^{\ell_y}. \tag{4.10}$$

Because $\pi_{\cdot,j}(\theta)$ is nonlinear, α_j's and θ are identified and may be estimated jointly from (4.10) by using Nonlinear least squares (NLLS). The null hypothesis is still $\left\|\alpha_j\right\|_{j=1}^{\ell_y} = 0$, which contains ℓ_y restrictions. The associated Wald test, however, asymptotically behaves as a $\chi^2(\ell_y - p)$ random variable rather than a $\chi^2(\ell_y)$ one, because $\hat{\alpha}_j$'s and $\hat{\theta}$ are asymptotically linearly dependent so that the asymptotic variance has a rank of only $\ell_y - p$. This happens because the linearization of $\left\|\pi_{\cdot,j}(\theta)\right\|_{j=1}^{\ell_y}$ in a vicinity of θ_0 yields $\left\|\pi_{\cdot,j}(\theta)\right\|_{j=1}^{\ell_y} = \left\|\pi_{\cdot,j}(\theta_0)\right\|_{j=1}^{\ell_y} + \Pi_1(\theta - \theta_0) + o(\theta - \theta_0)$. The rest of the logic remains as in the linear case. This requires that the Wald statistic be constructed using the generalized inverse of the asymptotic variance of estimates. To summarize, when $\pi_{\cdot,j}(\theta)$ is nonlinear in θ, one needs to jointly estimate $\hat{\alpha}_j$'s and $\hat{\theta}$ and test the original null, remembering the loss of p degrees of freedom.

5 Tests of symmetry

In this subsection we analyze two tests for symmetry in contingency tables with identical categorizations. We have not found any examples of their use in applied econometric practice, so we consider these tests partly for completeness, partly in hope that they will eventually be used by economists.

Stuart (1955) suggested a test for homogeneity of the marginal distributions of x and y. Formally, the null is

$$H_0 : \pi_{i,\cdot} = \pi_{\cdot,i}, \ i = 1, \cdots, \ell,$$

which automatically implies also $\pi_{\ell+1,\cdot} = \pi_{\cdot,\ell+1}$. The test is based on the $\ell \times 1$ vector of differences $p_{i,\cdot} - p_{\cdot,i}, \ i = 1, \cdots, \ell$. Let $d_n = \|p_{i,\cdot} - p_{\cdot,i}\|_{i=1}^{\ell}$. The test statistic is

$$Q_n = n d_n' V^{-1} d_n,$$

where $V = \|V_{i,j}\|_{i=1}^{\ell}{}_{j=1}^{\ell}$, and $V_{i,i} = p_{i,\cdot} + p_{\cdot,i} - 2p_{i,i} - (p_{i,\cdot} - p_{\cdot,i})^2$, $V_{i,j} = -p_{i,j} - p_{j,i} - (p_{i,\cdot} - p_{\cdot,i})(p_{j,\cdot} - p_{\cdot,j})$, $i \neq j$. Under H_0, Q_n is asymptotically distributed as $\chi^2(\ell)$.

Theorem 4 *The Stuart Q_n test is asymptotically equivalent to an OLS-based Wald test for the nullity of all intercepts in a linear multiple regression of $\mathbb{I}_{i,\cdot} - \mathbb{I}_{\cdot,i}$ on a constant, i.e. for the null*

$$H_0 : \alpha_i = 0, \ i = 1, \cdots, \ell$$

in the regression system

$$\mathbb{I}_{i,\cdot} - \mathbb{I}_{\cdot,i} = \alpha_i + \eta_i, \quad i = 1, \cdots, \ell. \tag{5.11}$$

Bowker (1948) suggested a test for complete symmetry of the contingency table. Such symmetry implies a stronger equivalence between the two classifications than equality of marginal distributions. In fact, it is the two conditional distributions that are compared. Formally, the null is

$$H_0 : \pi_{i,j} = \pi_{j,i}, \ i = 2, \cdots, \ell + 1, \ j = 1, \cdots, i - 1.$$

The test is based on $\ell(\ell+1)/2$ differences $p_{i,j}-p_{j,i},\ i=2,\cdots,\ell+1,\ j=1,\cdots,i-1$. The test statistic is

$$U_n = n\sum_{i=2}^{\ell+1}\sum_{j=1}^{i-1}\frac{(p_{i,j}-p_{j,i})^2}{p_{i,j}+p_{j,i}}.$$

Under H_0, U_n is asymptotically distributed as $\chi^2(\ell(\ell+1)/2)$.

Theorem 5 *The Bowker U_n test is asymptotically equivalent to an OLS-based Wald test for the nullity of all intercepts in a linear multiple regression of $\mathbb{I}_{i,j}-\mathbb{I}_{j,i}$ on a constant, i.e. for the null*

$$H_0:\alpha_{ij}=0,\ i=2,\cdots,\ell+1,\ j=1,\cdots,i-1$$

in the regression system

$$\mathbb{I}_{i,j}-\mathbb{I}_{j,i}=\alpha_{ij}+\eta_{ij},\quad i=2,\cdots,\ell+1,\ j=1,\cdots,i-1. \tag{5.12}$$

6 Tests based on ranks

Often researchers carry out testing for independence or homogeneity using rank transformed data rather than the original data, the idea being to compare more objective "ordinal" data characteristics instead of "cardinal" ones. Below we review two class of tests based on ranks – the Kruskal–Wallis test and the Spearman rank test. Both are used in a number of economic and financial applications: for example, Bizjak and Coles (1995), Theodossiou (1996), Krigman et al (1999), Moel and Tufano (2002) use the Kruskal–Wallis test, and Selten et al (1997), Attanasio et al (2000), Dickens (2000), Chance and Hemler (2001) and many others use the Spearman rank correlation coefficient.

Suppose that k random samples of size $n_1,...,n_k$ are tested for identity of distributions they come from. Let the vector of ranks $(r_{1,1},...,r_{1,n_1},...,r_{k,1},...,r_{k,n_k})'$ correspond to the pooled sample (of length $n=\sum_{j=1}^{k}n_j$). Let j index samples, while i index observations within a sample. The sums of ranks for the separate samples is denoted by $R_j=\sum_{i=1}^{n_j}r_{j,i}$.

The Kruskal–Wallis test statistic is

$$KW = \frac{12}{(n-1)\,n}\sum_{j=1}^{k}\frac{1}{n_j}\left(R_j - \frac{n+1}{2}n_j\right)^2.$$

Let asymptotically $n \to \infty$ and $\min_j n_j \to \infty$ so that $\lambda_j \equiv \lim_{\min_j n_j \to \infty} n_j/n \neq 0$ for $j = 1, ..., k$. Under these circumstances, KW is asymptotically distributed as $\chi^2(k-1)$.

The Kruskal–Wallis test may be run via a linear regression.

Theorem 6 *The Kruskal–Wallis KW test is asymptotically equivalent to an OLS-based Wald test for the nullity of all intercepts in a linear multiple regression of* $r_{j,i} - \frac{n+1}{2}$, $j = 1, \cdots, k-1$, *on a constant, i.e. for the null*

$$H_0 : \alpha_j = 0,\ j = 1, \cdots, k-1$$

in the regression system

$$r_{j,i} - \frac{n+1}{2} = \alpha_j + \eta_{ij}, \quad j = 1, \cdots, k-1, \tag{6.13}$$

with observations running from $i = 1$ *to* $i = n_j$ *for equation* j.

Remark 2 *A similar, but different, situation is considered in statistical literature. A fixed number* s *of products are ranked by a fixed number* k *of experts. Denote by* $K_{i,j}$ *the ranking that the expert* j *gave the product* i *(*$K_{i,j}$ *varies from 1 to* s*), and by* $N_{i,j}$ *the number of times the product* i *received ranking* j.

The Friedman test statistic

$$F = \frac{12}{ks\,(s+1)}\sum_{j=1}^{s}\left(\sum_{j=1}^{k}K_{i,j} - \frac{s+1}{2}k\right)^2$$

is used to test for homogeneity of products. It is asymptotically distributed as $\chi^2(s-1)$ *as the number of experts increases. It is possible to show by following the same steps as in the proof of Theorem 6 that* F *is asymptotically equivalent to an OLS-based Wald test for*

$$H_0 : \alpha_i = 0,\ i = 1, \cdots, s-1$$

in the regression system

$$K_{i,j} - \frac{s+1}{2} = \alpha_i + \eta_{ij}, \quad i = 1, \cdots, s-1, \tag{6.14}$$

with observations running from $j = 1$ to $j = k$.

The Anderson test statistic

$$A = \frac{s}{k} \sum_{i=1}^{s} \sum_{j=1}^{s} \left(N_{i,j} - \frac{k}{s} \right)^2,$$

asymptotically distributed as $\chi^2\left((s-1)^2\right)$ as the number of experts increases, is used to test for homogeneity of products. It is possible to show by following the same steps as in the proof of Theorem 1 that A is asymptotically equivalent to an OLS-based Wald test for

$$H_0 : \alpha_{i,j} = 0, \ i, j = 1, \cdots, s-1$$

in the regression system

$$N_{i,j} - \frac{k}{s} = \alpha_{i,j} + \eta_{ij}, \quad i, j = 1, \cdots, s-1, \tag{6.15}$$

with one observation per equation.

Suppose the vector of ranks $(R_1, ..., R_n)'$ corresponds to a random sample of length n. The Spearman rank statistic ρ is defined as

$$\rho = \frac{12}{(n-1)\, n} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) \left(R_i - \frac{n+1}{2} \right).$$

Theorem 7 *The Spearman rank statistic ρ is equal to $n+1$ times the OLS slope coefficient in a linear regression of R_i on a constant and observation number i*

$$R_i = \alpha + \beta i + \eta, \tag{6.16}$$

with observations running from $i = 1$ to $i = n$.

As a consequence, an OLS-based t test for the null $H_0 : \beta = 0$ may be used to test for independence of elements in a given sample, which is often realized in practice by informally comparing ρ to zero.

The pairwise Spearman rank correlation coefficient ρ between two vectors of ranks $(R_1, ..., R_n)'$ and $(S_1, ..., S_n)'$ corresponding to two random samples of length n is defined as

$$\rho = \frac{\sum_{i,j=1}^{n} (S_j - S_i)(R_j - R_i)}{\sqrt{\sum_{i,j=1}^{n} (S_j - S_i)^2 \sum_{i,j=1}^{n} (R_j - R_i)^2}}.$$

Theorem 8 *The Spearman rank correlation coefficient ρ is equal the OLS slope coefficient in a linear regression of $R_j - R_i$ on $S_j - S_i$*

$$R_j - R_i = \beta (S_j - S_i) + \eta, \tag{6.17}$$

with n^2 observations for all possible pairs (i, j) where each index runs from 1 to n. Alternatively, one may switch $S_j - S_i$ and $R_j - R_i$ in the regression (6.17). A constant term may be innocuously introduced into the regression (6.17).

As a consequence, an OLS-based t test for the null $H_0 : \beta = 0$ may be used to test for independence of two given random samples.

Remark 3 *In the context of the previous remark, the Umbrella test statistic*

$$U = \frac{12}{\sqrt{k(s-1)s(s+1)}} \left(\sum_{i=1}^{s} i \sum_{j=1}^{k} K_{i,j} - \frac{1}{2} ks(s+1)^2 \right),$$

asymptotically distributed as $N(0, 1)$ as the number of experts increases, is used to test for homogeneity of products. It is possible to show that U is asymptotically equivalent to an OLS-based t test for

$$H_0 : \beta = 0$$

in the regression

$$K_{i,j} = \alpha + \beta i + \eta_{ij}, \tag{6.18}$$

with observations running from $i, j = 1$ to $i = s, j = k$.

7 References

Alaouze, C.M. (1987) Empirical Evidence on the Sign of the Slope of the Hazard Rate from Unemployment from a Fixed Effects Model. *Journal of Applied Econometrics* 2, 159–168.

Artis, M.J., Z.G. Kontolemis, and D.R. Osborn (1997) Business Cycles for G7 and European Countries. *Journal of Business* 70, 249–279.

Attanasio, O.P., L. Picci, and A.E. Scorcu (2000) Saving, Growth, and Investment: A Macroeconomic Analysis Using a Panel of Countries. *Review of Economics and Statistics* 82, 182–211.

Battalio, R., L. Samuelson, and J.V. Huyck (2001) Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games. *Econometrica* 69, 749–764.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004) A comparison of financial duration models via density forecasts. *International Journal of Forecasting* 20, 589–609.

Bizjak, J.M. and J.L. Coles (1995) The Effect of Private Antitrust Litigation on the Stock-Market Valuation of the Firm . *American Economic Review* 85, 436–461.

Bouckaerta, J. and G. Dhaene (2004) Inter-ethnic trust and reciprocity: results of an experiment with small businessmen. *European Journal of Political Economy* 20, 869–886.

Bowker, A.H. (1948) A Test for Symmetry in Contingency Tables. *Journal of the American Statistical Association* 43, 572–574.

Brown, K.C., W.V. Harlow, and L.T. Starks (1996) Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry. *Journal of Finance* 51, 85–110.

Chan, W.H. and J.M. Maheu (2002) Conditional Jump Dynamics in Stock Market Returns. *Journal of Business and Economic Statistics 20, 377–389.*

Chance, D.M. and M.L. Hemler (2001) The performance of professional market timers: daily evidence from executed strategies. *Journal of Financial Economics* 62, 377–411.

Dickens, R. (2000) Caught in a Trap? Wage Mobility in Great Britain: 1975–1994. *Economica* 67, 477–498.

Greenwood, M.J.; G.L. Hunt, D.S. Rickman, and G.I. Treyz (1991) Migration, Regional Equilibrium, and the Estimation of Compensating Differentials. *American Economic Review* 81, 1382–1390.

Hsieh, D.A. (1989) Testing for Nonlinear Dependence in Daily Foreign Exchange Rates. *Journal of Business* 62, 339–368.

Hyslop, D.R. (1999) State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women. Econometrica 67, 1255–1294.

Eckel, C.C. and P.J. Grossman (1998) Are Women Less Selfish Than Men?: Evidence from Dictator Experiments. *Economic Journal* 108, 726–735.

Keane, M. and R. Moffitt (1998) A Structural Model of Multiple Welfare Program Participation and Labor Supply. International Economic Review 39, 553–589.

Kendall, M.G. and A. Stuart (1973) *The Advanced Theory of Statistics.* Griffin: London.

Knetsch, J.L. and J.A. Sinden (1984) Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. *Quarterly Journal of Economics* 99, 507–521.

Krigman, L., W.H. Shaw, and K.L. Womack (1999) The Persistence of IPO Mispricing and the Predictive Power of Flipping. *Journal of Finance* 54, 1015–1044.

Lane, J.A., D.A. Peel, and E.J. Raeburn (1996) Some Empirical Evidence on the Time-Series Properties of Four UK Asset Prices. *Economica* 63, 405–426.

Merlo, Antonio (1997) Bargaining over Governments in a Stochastic Environment. Journal of Political Economy 105, 101–131.

Moel, A. and P. Tufano (2002) When Are Real Options Exercised? An Empirical Study of Mine Closings. *Review of Financial Studies* 15, 35–64.

Moffitt, R. (1989) Estimating the Value of an In-Kind Transfer: The Case of Food Stamps.

Econometrica 57, 385–409.

Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series* 50, 157–175.

Pecorino, P. and M. van Boening (2001) Bargaining and Information: An Empirical Analysis of a Multistage Arbitration Game. *Journal of Labor Economics* 19, 922–948.

Pesaran, M.H. and A. Timmermann (1992) A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* 10, 561–565.

Pesaran, M.H. and A. Timmermann (1994) A generalization of the nonparametric Henriksson–Merton test of market timing. *Economics Letters* 4, 1–7.

Russo, G., C. Gorter, and R. Schettkat (2001) Searching, hiring and labour market conditions. *Labour Economics* 8, 553–571.

Selten, R., M. Mitzkewitz, and G.R. Uhlich (1997) Duopoly Strategies Programmed by Experienced Players. *Econometrica* 65, 517–555.

Stuart, A. (1955) A Test for Homogeneity of the Marginal Distributions in a Two-Way Classification. *Biometrika* 42, 412–416.

Theodossiou, I. (1996) Promotions, Job Seniority, and Product Demand Effects on Earnings. *Oxford Economic Papers* 48, 456–472.

A Appendix: proofs

Proof. [of Theorem 1] Let us represent X^2 as

$$\begin{aligned}\frac{X^2}{n} &= \sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y}\frac{(p_{i,j}-p_{i,\cdot}p_{\cdot,j})^2}{p_{i,\cdot}p_{\cdot,j}} + \sum_{i=1}^{\ell_x}\frac{\left(p_{i,\ell_y+1}-p_{i,\cdot}p_{\cdot,\ell_y+1}\right)^2}{p_{i,\cdot}p_{\cdot,\ell_y+1}} \\ &\quad + \sum_{j=1}^{\ell_y}\frac{(p_{\ell_x+1,j}-p_{\ell_x+1,\cdot}p_{\cdot,j})^2}{p_{\ell_x+1,\cdot}p_{\cdot,j}} + \frac{\left(p_{\ell_x+1,\ell_y+1}-p_{\ell_x+1,\cdot}p_{\cdot,\ell_y+1}\right)^2}{p_{\ell_x+1,\cdot}p_{\cdot,\ell_y+1}}.\end{aligned} \tag{A.1}$$

Let us analyze the numerators in the second, third and fourth terms in (A.1). Note that

$$\begin{aligned}p_{i,\ell_y+1}-p_{i,\cdot}p_{\cdot,\ell_y+1} &= \left(p_{i,\cdot}-\sum_{j=1}^{\ell_y}p_{i,j}\right) - p_{i,\cdot}\left(1-\sum_{j=1}^{\ell_y}p_{\cdot,j}\right) \\ &= -\sum_{j=1}^{\ell_y}\left(p_{i,j}-p_{i,\cdot}p_{\cdot,j}\right).\end{aligned}$$

Similarly,

$$p_{\ell_x+1,j}-p_{\ell_x+1,\cdot}p_{\cdot,j} = -\sum_{i=1}^{\ell_x}\left(p_{i,j}-p_{i,\cdot}p_{\cdot,j}\right).$$

Finally,

$$\begin{aligned}p_{\ell_x+1,\ell_y+1}-p_{\ell_x+1,\cdot}p_{\cdot,\ell_y+1} &= \left(1-\sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y}p_{i,j}-\sum_{i=1}^{\ell_x}p_{i,\cdot}-\sum_{j=1}^{\ell_y}p_{\cdot,j}\right) \\ &\quad -\left(1-\sum_{i=1}^{\ell_x}p_{i,\cdot}\right)\left(1-\sum_{j=1}^{\ell_y}p_{\cdot,j}\right) \\ &= -\sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y}\left(p_{i,j}-p_{i,\cdot}p_{\cdot,j}\right).\end{aligned}$$

The denominators in all terms in (A.1) asymptotically have probability limits

$$\begin{aligned}&p_{i,\cdot}p_{\cdot,j} \xrightarrow{p} \pi_{i,\cdot}\pi_{\cdot,j}, \\ &p_{i,\cdot}p_{\cdot,\ell_y+1} \xrightarrow{p} \pi_{i,\cdot}\left(1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}\right), \\ &p_{\ell_x+1,\cdot}p_{\cdot,j} \xrightarrow{p} \left(1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}\right)\pi_{\cdot,j}, \\ &p_{\ell_x+1,\cdot}p_{\cdot,\ell_y+1} \xrightarrow{p} \left(1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}\right)\left(1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}\right).\end{aligned}$$

Therefore, the X^2 statistic is asymptotically equivalent to

$$\begin{aligned}\frac{X^2}{n} \overset{A}{=} \frac{\tilde{X}^2}{n} &= \sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y}\frac{(p_{i,j}-p_{i,\cdot}p_{\cdot,j})^2}{\pi_{i,\cdot}\pi_{\cdot,j}} + \sum_{i=1}^{\ell_x}\frac{\left(\sum_{j=1}^{\ell_y}(p_{i,j}-p_{i,\cdot}p_{\cdot,j})\right)^2}{\pi_{i,\cdot}\left(1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}\right)} \\ &+ \sum_{j=1}^{\ell_y}\frac{\left(\sum_{i=1}^{\ell_x}(p_{i,j}-p_{i,\cdot}p_{\cdot,j})\right)^2}{\left(1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}\right)\pi_{\cdot,j}} + \frac{\left(\sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y}(p_{i,j}-p_{i,\cdot}p_{\cdot,j})\right)^2}{\left(1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}\right)\left(1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}\right)}.\end{aligned}$$

Note that

$$p_{i,j}-p_{i,\cdot}p_{\cdot,j} = \overline{\mathbb{I}_{i,\cdot}\mathbb{I}_{\cdot,j}} - \bar{\mathbb{I}}_{i,\cdot}\bar{\mathbb{I}}_{\cdot,j} = \overline{\left(\mathbb{I}_{i,\cdot}-\bar{\mathbb{I}}_{i,\cdot}\right)\left(\mathbb{I}_{\cdot,j}-\bar{\mathbb{I}}_{\cdot,j}\right)}.$$

Now we can represent $\tilde{X}^2$ as a quadratic form of the type

$$\tilde{X}^2 = n\xi'\Xi\xi,$$

where

$$\begin{aligned}\xi &= \overline{\xi_x \otimes \xi_y}, \\ \xi_x &= \left\|\mathbb{I}_{i,\cdot}-\bar{\mathbb{I}}_{i,\cdot}\right\|_{i=1}^{\ell_x}, \\ \xi_y &= \left\|\mathbb{I}_{\cdot,j}-\bar{\mathbb{I}}_{\cdot,j}\right\|_{j=1}^{\ell_y},\end{aligned}$$

i.e. ξ is $\ell_x\ell_y \times 1$ vector containing values of $\overline{\left(\mathbb{I}_{i,\cdot}-\bar{\mathbb{I}}_{i,\cdot}\right)\left(\mathbb{I}_{\cdot,j}-\bar{\mathbb{I}}_{\cdot,j}\right)}$, with index j running faster than index i. The $\ell_x\ell_y \times \ell_x\ell_y$ matrix Ξ contains weights implied by the formula for $\tilde{X}^2$, and is equal to

$$\begin{aligned}\Xi &= \operatorname{diag}\left\{\frac{1}{\pi_{i,\cdot}}\right\}_{i=1}^{\ell_x} \otimes \operatorname{diag}\left\{\frac{1}{\pi_{\cdot,j}}\right\}_{j=1}^{\ell_y} + \operatorname{diag}\left\{\frac{1}{\pi_{i,\cdot}}\right\}_{i=1}^{\ell_x} \otimes \frac{\iota_{\ell_y}\iota'_{\ell_y}}{1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}} \\ &+ \frac{\iota_{\ell_x}\iota'_{\ell_x}}{1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}} \otimes \operatorname{diag}\left\{\frac{1}{\pi_{\cdot,j}}\right\}_{i=1}^{\ell_y} + \frac{\iota_{\ell_x}\iota'_{\ell_x}}{1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}} \otimes \frac{\iota_{\ell_y}\iota'_{\ell_y}}{1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}} \\ &= \Xi_x \otimes \Xi_y,\end{aligned}$$

where

$$\begin{aligned}\Xi_x &= \operatorname{diag}\left\{\frac{1}{\pi_{i,\cdot}}\right\}_{i=1}^{\ell_x} + \frac{\iota_{\ell_x}\iota'_{\ell_x}}{1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}}, \\ \Xi_y &= \operatorname{diag}\left\{\frac{1}{\pi_{\cdot,j}}\right\}_{j=1}^{\ell_y} + \frac{\iota_{\ell_y}\iota'_{\ell_y}}{1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}}.\end{aligned}$$

The inverse of Ξ equals

$$\begin{aligned}
\Xi^{-1} &= V_x \otimes V_y, \\
V_x &= \operatorname{diag}\{\pi_{i,\cdot}\}_{i=1}^{\ell_x} - \pi_x \pi_x', \\
V_y &= \operatorname{diag}\{\pi_{\cdot,j}\}_{j=1}^{\ell_y} - \pi_y \pi_y',
\end{aligned}$$

which can be easily established by direct multiplication.

To summarize,

$$X^2 \overset{A}{=} n \left(\overline{\xi_x \otimes \xi_y}\right)' (V_x \otimes V_y)^{-1} \left(\overline{\xi_x \otimes \xi_y}\right),$$

which is (up to substitution of V_x and V_y by their sample analogs) a Wald test statistic for testing the null hypothesis of joint insignificance of all slope coefficients in a linear multiple regression of $\mathbb{I}_{\cdot,j}$ on $\mathbb{I}_{i,\cdot}$ with a constant in each equation. Indeed, the standardizing matrix $V_x \otimes V_y$ contains variances and covariances of "regressors" and "dependent variables" under their independence, because

$$\begin{aligned}
var\left(\mathbb{I}_{i,\cdot}\right) &= \pi_{i,\cdot} - \pi_{i,\cdot}^2, \\
cov\left(\mathbb{I}_{i_1,\cdot}, \mathbb{I}_{i_2,\cdot}\right) &= -\pi_{i_1,\cdot}\pi_{i_2,\cdot},
\end{aligned}$$

and similarly for entries of V_y.

The last conclusion follows from the symmetry between x and y. ■

Proof. [of Theorem 2] The residual variance from this regression is

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{i=1}^{\ell+1}\left(\mathbb{I}_{i,i} - \overline{\mathbb{I}}_{i,i} - \left(\mathbb{I}_{i,\cdot} - \overline{\mathbb{I}}_{i,\cdot}\right)\overline{\mathbb{I}}_{\cdot,i} - \left(\mathbb{I}_{\cdot,i} - \overline{\mathbb{I}}_{\cdot,i}\right)\overline{\mathbb{I}}_{i,\cdot}\right)\right)^2 \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{i=1}^{\ell+1}\left(\mathbb{I}_{i,i} - \pi_{i,i} - \left(\mathbb{I}_{i,\cdot} - \pi_{i,\cdot}\right)\pi_{\cdot,i} - \left(\mathbb{I}_{\cdot,i} - \pi_{\cdot,i}\right)\pi_{i,\cdot}\right)\right)^2 + o_p(1) \\
&\overset{p}{\to} var\left(f\left(\|\mathbb{I}_{i,j}\|_{i=1}^{\ell+1}{}_{j=1}^{\ell+1}\right)\right),
\end{aligned}$$

where

$$
\begin{aligned}
f\left(\|\mathbb{I}_{i,j}\|_{i=1}^{\ell+1}{}_{j=1}^{\ell+1}\right) &= \sum_{i=1}^{\ell+1}\left(\mathbb{I}_{i,i}-\mathbb{I}_{i,\cdot}\pi_{\cdot,i}-\mathbb{I}_{\cdot,i}\pi_{i,\cdot}\right) \\
&= \sum_{i=1}^{\ell+1}\left(\mathbb{I}_{i,i}-\left(\sum_{j=1}^{\ell+1}\mathbb{I}_{i,j}\right)\pi_{\cdot,i}-\left(\sum_{j=1}^{\ell+1}\mathbb{I}_{j,i}\right)\pi_{i,\cdot}\right).
\end{aligned}
$$

Note that $var\left(\|\mathbb{I}_{i,j}\|_{i=1}^{\ell+1}{}_{j=1}^{\ell+1}\right) = \mathrm{diag}\,(\Pi) - \Pi\Pi'$, where $\Pi = \|\pi_{i,j}\|_{i=1}^{\ell+1}{}_{j=1}^{\ell+1}$. As $f(\cdot)$ is linear in all elements,

$$
var\left(f\left(\|\mathbb{I}_{i,j}\|_{i=1}^{\ell+1}{}_{j=1}^{\ell+1}\right)\right) = \frac{\partial f(\Pi)}{\partial \Pi'}\left(\mathrm{diag}\,(\Pi) - \Pi\Pi'\right)\frac{\partial f(\Pi)}{\partial \Pi},
$$

which coincides with V_s in Pesaran and Timmermann (1992, p.463; 1994, formula (6)). ■

Proof. [of Theorem 3] The proof goes along the same lines as that of Theorem 1. Let us represent X^2 as

$$
\begin{aligned}
\frac{X^2}{n} &= \sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y}\frac{\left(p_{i,j}-p_{i,\cdot}\pi_{\cdot,j}\right)^2}{p_{i,\cdot}\pi_{\cdot,j}} + \sum_{i=1}^{\ell_x}\frac{\left(p_{i,\ell_y+1}-p_{i,\cdot}\pi_{\cdot,\ell_y+1}\right)^2}{p_{i,\cdot}\pi_{\cdot,\ell_y+1}} \\
&\quad + \sum_{j=1}^{\ell_y}\frac{\left(p_{\ell_x+1,j}-p_{\ell_x+1,\cdot}\pi_{\cdot,j}\right)^2}{p_{\ell_x+1,\cdot}\pi_{\cdot,j}} + \frac{\left(p_{\ell_x+1,\ell_y+1}-p_{\ell_x+1,\cdot}\pi_{\cdot,\ell_y+1}\right)^2}{p_{\ell_x+1,\cdot}\pi_{\cdot,\ell_y+1}}.
\end{aligned}
\tag{A.2}
$$

Let us analyze the numerators in the second, third and fourth terms in (A.2). Note that

$$
\begin{aligned}
p_{i,\ell_y+1}-p_{i,\cdot}\pi_{\cdot,\ell_y+1} &= \left(p_{i,\cdot}-\sum_{j=1}^{\ell_y}p_{i,j}\right) - p_{i,\cdot}\left(1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}\right) \\
&= -\sum_{j=1}^{\ell_y}\left(p_{i,j}-p_{i,\cdot}\pi_{\cdot,j}\right),
\end{aligned}
$$

$$
\begin{aligned}
p_{\ell_x+1,j}-p_{\ell_x+1,\cdot}\pi_{\cdot,j} &= \left(p_{\cdot,j}-\sum_{i=1}^{\ell_x}p_{i,j}\right) - \left(1-\sum_{i=1}^{\ell_x}p_{i,\cdot}\right)\pi_{\cdot,j} \\
&= -\sum_{i=1}^{\ell_x}\left(p_{i,j}-p_{i,\cdot}\pi_{\cdot,j}\right) + \left(p_{\cdot,j}-\pi_{\cdot,j}\right),
\end{aligned}
$$

$$\begin{aligned} p_{\ell_x+1,\ell_y+1} - p_{\ell_x+1,\cdot}\pi_{\cdot,\ell_y+1} &= \left(1 - \sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y} p_{i,j} - \sum_{i=1}^{\ell_x} p_{i,\cdot} - \sum_{j=1}^{\ell_y} p_{\cdot,j}\right) \\ &\quad - \left(1 - \sum_{i=1}^{\ell_x} p_{i,\cdot}\right)\left(1 - \sum_{j=1}^{\ell_y} \pi_{\cdot,j}\right) \\ &= -\sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y} \left(p_{i,j} - p_{i,\cdot}\pi_{\cdot,j}\right) - \sum_{j=1}^{\ell_y} \left(p_{\cdot,j} - \pi_{\cdot,j}\right). \end{aligned}$$

Therefore, the X^2 statistic is asymptotically equivalent to

$$\begin{aligned} \frac{X^2}{n} \overset{A}{=} \frac{\tilde{X}^2}{n} &= \sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y} \frac{\left(p_{i,j} - p_{i,\cdot}\pi_{\cdot,j}\right)^2}{\pi_{i,\cdot}\pi_{\cdot,j}} + \sum_{i=1}^{\ell_x} \frac{\left(\sum_{j=1}^{\ell_y}\left(p_{i,j} - p_{i,\cdot}\pi_{\cdot,j}\right)\right)^2}{\pi_{i,\cdot}\left(1 - \sum_{j=1}^{\ell_y}\pi_{\cdot,j}\right)} \\ &\quad + \sum_{j=1}^{\ell_y} \frac{\left(\sum_{i=1}^{\ell_x}\left(p_{i,j} - p_{i,\cdot}\pi_{\cdot,j}\right) - \left(p_{\cdot,j} - \pi_{\cdot,j}\right)\right)^2}{\left(1 - \sum_{i=1}^{\ell_x}\pi_{i,\cdot}\right)\pi_{\cdot,j}} \\ &\quad + \frac{\left(\sum_{i=1}^{\ell_x}\sum_{j=1}^{\ell_y}\left(p_{i,j} - p_{i,\cdot}\pi_{\cdot,j}\right) + \sum_{j=1}^{\ell_y}\left(p_{\cdot,j} - \pi_{\cdot,j}\right)\right)^2}{\left(1 - \sum_{j=1}^{\ell_y}\pi_{\cdot,j}\right)\left(1 - \sum_{i=1}^{\ell_x}\pi_{i,\cdot}\right)}. \end{aligned}$$

Note that

$$p_{i,j} - p_{i,\cdot}\pi_{\cdot,j} = \overline{\mathbb{I}_{i,\cdot}\left(\mathbb{I}_{\cdot,j} - \pi_{\cdot,j}\right)}$$

and

$$p_{\cdot,j} - \pi_{\cdot,j} = \overline{\mathbb{I}_{\cdot,j} - \pi_{\cdot,j}}.$$

Now we can represent $\tilde{X}^2$ as a quadratic form of the type

$$\tilde{X}^2 = n\xi'\Xi\xi,$$

where

$$\begin{aligned} \xi &= \overline{\xi_x \otimes \xi_y}, \\ \xi_x &= \begin{pmatrix} 1 \\ \|\mathbb{I}_{i,\cdot}\|_{i=1}^{\ell_x} \end{pmatrix}, \\ \xi_y &= \|\mathbb{I}_{\cdot,j} - \pi_{\cdot,j}\|_{j=1}^{\ell_y}. \end{aligned}$$

The $(\ell_x+1)\,\ell_y \times (\ell_x+1)\,\ell_y$ matrix Ξ contains weights implied by the formula for $\tilde{X}^2$, and is equal to

$$\Xi = \Xi_x \otimes \Xi_y,$$

where

$$\begin{aligned}
\Xi_x &= \begin{pmatrix} \dfrac{1}{1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}} & -\dfrac{\iota_{\ell_x}'}{1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}} \\ -\dfrac{\iota_{\ell_x}}{1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}} & \operatorname{diag}\left\{\dfrac{1}{\pi_{i,\cdot}}\right\}_{i=1}^{\ell_x} + \dfrac{\iota_{\ell_x}\iota_{\ell_x}'}{1-\sum_{i=1}^{\ell_x}\pi_{i,\cdot}} \end{pmatrix}, \\
\Xi_y &= \operatorname{diag}\left\{\frac{1}{\pi_{\cdot,j}}\right\}_{i=1}^{\ell_y} + \frac{\iota_{\ell_y}\iota_{\ell_y}'}{1-\sum_{j=1}^{\ell_y}\pi_{\cdot,j}}.
\end{aligned}$$

The inverse of Ξ equals

$$\begin{aligned}
\Xi^{-1} &= E_x \otimes E_y, \\
E_x &= \begin{pmatrix} 1 & \pi_x' \\ \pi_x & \operatorname{diag}\{\pi_{i,\cdot}\}_{i=1}^{\ell_x} \end{pmatrix}, \\
E_y &= \operatorname{diag}\{\pi_{\cdot,j}\}_{i=1}^{\ell_y} - \pi_y\pi_y',
\end{aligned}$$

which can be easily established by direct multiplication.

To summarize,

$$X^2 \overset{A}{=} n\left(\overline{\xi_x \otimes \xi_y}\right)'\left(E_x \otimes E_y\right)^{-1}\left(\overline{\xi_x \otimes \xi_y}\right),$$

which is (up to substitution of E_x and E_y by their sample analogs) a Wald test statistic for testing the null hypothesis of joint insignificance of all coefficients in a linear multiple regression of $\mathbb{I}_{\cdot,j} - \pi_{\cdot,j}$ on $\mathbb{I}_{i,\cdot}$ with a constant in each equation. Indeed, the standardizing matrix $E_x \otimes E_y$ contains expectations of cross-products of different "regressors" and expectations of cross-products of different "dependent variables" under independence of "regressors" and "dependent variables". ■

Proof. [of Theorem 4] The conclusion is obvious, because Q_n already has the form of the Wald test, and V is a sample analog to $var\left(\|\mathbb{I}_{i,\cdot} - \mathbb{I}_{\cdot,i}\|_{i=1}^{\ell+1}\right)$. ■

Proof. [of Theorem 5] The pivotization in the Wald test statistic equals the variance matrix of the $\ell(\ell+1)/2 \times 1$ vector of differences $\mathbb{I}_{i,j} - \mathbb{I}_{j,i},\ i = 1, \cdots, \ell+1,\ j = 1, \cdots, i-1$, i.e. $var\left(\|\mathbb{I}_{i,j} - \mathbb{I}_{j,i}\| \left|_{i=1}^{\ell+1}\right|_{j=1}^{i-1}\right)$, where the variance of each element $\mathbb{I}_{i,j} - \mathbb{I}_{j,i}$ equals

$$var(\mathbb{I}_{i,j}) + var(\mathbb{I}_{j,i}) + cov(\mathbb{I}_{i,j}, \mathbb{I}_{j,i}) = \pi_{i,j}(1-\pi_{i,j}) + \pi_{j,i}(1-\pi_{j,i}) + 2\pi_{i,j}\pi_{j,i}$$

which under H_0 is asymptotically

$$\pi_{i,j} + \pi_{j,i}.$$

The covariance between element $\mathbb{I}_{i,j} - \mathbb{I}_{j,i}$ and $\mathbb{I}_{m,k} - \mathbb{I}_{k,m}$ where $k \neq j$ or $i \neq m$ equals

$$\begin{aligned} & cov(\mathbb{I}_{i,j}, \mathbb{I}_{m,k}) - cov(\mathbb{I}_{i,j}, \mathbb{I}_{k,m}) - cov(\mathbb{I}_{j,i}, \mathbb{I}_{m,k}) + cov(\mathbb{I}_{j,i}, \mathbb{I}_{k,m}) \\ = \;& 2(\pi_{i,j}\pi_{m,k} - \pi_{i,j}\pi_{k,m} - \pi_{j,i}\pi_{m,k} + \pi_{j,i}\pi_{k,m}) \\ = \;& (\pi_{i,j} - \pi_{j,i})(\pi_{m,k} - \pi_{k,m}) = 0. \end{aligned}$$

Hence, $var\left(\|\mathbb{I}_{i,j} - \mathbb{I}_{j,i}\| \left|_{i=1}^{\ell+1}\right|_{j=1}^{i-1}\right)$ equals $\text{diag}\left\{\{\pi_{i,j} + \pi_{j,i}\}_{i=1}^{\ell+1}\right\}_{j=1}^{i-1}$, and its inverse is $\text{diag}\left\{\left\{(\pi_{i,j} + \pi_{j,i})^{-1}\right\}_{i=1}^{\ell+1}\right\}_{j=1}^{i-1}$, which is exactly the weighting system in U_n. ■

Proof. [of Theorem 6] The vector containing all $\hat{\alpha}_j$ is equal to

$$\hat{\alpha}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(r_{j,i} - \frac{n+1}{2}\right) = \frac{1}{n_j}\left(R_j - n_j\frac{n+1}{2}\right).$$

To compute the asymptotics for this vector, we need the covariance matrix for the vector $\|R_j/n_j\|_{j=1}^{k-1}$.

Under the hypothesis of independence, each rank $r_{j,i}$ is distributed multinomially on $1, 2, ..., n$ with equal probabilities, therefore

$$var[r_{j,i}] = E\left[r_{j,i}^2\right] - E\left[r_{j,i}\right]^2 = \sum_{i=1}^{n}\frac{i^2}{n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}.$$

Now, for $j_1 \neq j_2$ or $i_1 \neq i_2$

$$\begin{aligned} cov[r_{j_1,i_1}, r_{j_2,i_2}] &= E\left[E\left[r_{j_1,i_1}|r_{j_2,i_2}\right]r_{j_2,i_2}\right] - E\left[r_{j,i}\right]^2 \\ &= \sum_{l=1}^{n}\frac{1}{n}\left(\sum_{i\neq l}\frac{i}{n-1}\right)l - \left(\frac{n+1}{2}\right)^2 = -\frac{n+1}{12}. \end{aligned}$$

Then also

$$\begin{aligned}
var\left[\frac{R_j}{n_j}\right] &= var\left[\frac{1}{n_j}\sum_{i=1}^{n_j} r_{j,i}\right] = \frac{1}{n_j}var[r_{j,i}] + \frac{n_j^2 - n_j}{n_j^2}cov[r_{j,i_1}, r_{j,i_2}] \\
&= \frac{n+1}{12}\left(\frac{n}{n_j} - 1\right)
\end{aligned}$$

and for $j_1 \neq j_2$

$$cov\left[\frac{R_{j_1}}{n_{j_1}}, \frac{R_{j_2}}{n_{j_2}}\right] = cov\left[\frac{1}{n_{j_1}}\sum_{i_1=1}^{n_{j_1}} r_{j_1,i_1}, \frac{1}{n_{j_2}}\sum_{i_2=1}^{n_{j_2}} r_{j_2,i_2}\right] = -\frac{n+1}{12}.$$

Then it follows that

$$\frac{\hat{\alpha}}{\sqrt{n}} = \frac{1}{\sqrt{n}}\left\|\frac{R_j}{n_j} - \frac{n+1}{2}\right\|_{j=1}^{k-1} \xrightarrow{d} \frac{1}{\sqrt{12}}N\left(0, \operatorname{diag}\left\{\frac{1}{\lambda_j}\right\}_{j=1}^{k-1} - \iota_{k-1}\iota_{k-1}'\right).$$

Hence, the Wald test statistic for H_0 is

$$\begin{aligned}
W &= \frac{12}{n}\hat{\alpha}'\left(\operatorname{diag}\left\{\frac{1}{\lambda_j}\right\}_{j=1}^{k-1} - \iota_{k-1}\iota_{k-1}'\right)^{-1}\hat{\alpha} \\
&= \frac{12}{n}\hat{\alpha}'\left(\operatorname{diag}\{\lambda_j\}_{j=1}^{k-1} + \frac{\|\lambda_j\|_{j=1}^{k-1}\left(\|\lambda_j\|_{j=1}^{k-1}\right)'}{1-\sum_{j=1}^{k-1}\lambda_j}\right)\hat{\alpha} \\
&= \frac{12}{n}\sum_{j=1}^{k-1}\frac{\lambda_j}{n_j^2}\left(R_j - n_j\frac{n+1}{2}\right)^2 + \frac{12}{n\lambda_k}\left(\sum_{j=1}^{k-1}\frac{\lambda_j}{n_j}\left(R_j - n_j\frac{n+1}{2}\right)\right)^2.
\end{aligned}$$

Obviously, the first term is asymptotically equivalent to

$$\frac{12}{(n-1)\,n}\sum_{j=1}^{k-1}\frac{1}{n_j}\left(R_j - \frac{n+1}{2}n_j\right)^2,$$

which is the first $k-1$ terms in KW. The last term is asymptotically equivalent to

$$\begin{aligned}
&\frac{12}{(n-1)\,n}\frac{1}{n-\sum_{j=1}^{k-1}n_j}\left(\sum_{j=1}^{k-1}R_j - \frac{n+1}{2}\sum_{j=1}^{k-1}n_j\right)^2 \\
&= \frac{12}{(n-1)\,n}\frac{1}{n-\sum_{j=1}^{k-1}n_j}\left(\frac{n(n+1)}{2} - R_k - \frac{n+1}{2}(n-n_k)\right)^2 \\
&= \frac{12}{(n-1)\,n}\frac{1}{n-\sum_{j=1}^{k-1}n_j}\left(R_k - \frac{n+1}{2}n_k\right)^2,
\end{aligned}$$

which is the last, k^{th}, term in KW. ■

Proof. [of Theorem 7] The sample slope coefficient is in the regression (6.16)

$$\hat{\beta}=\frac{\sum_{i=1}^{n}(i-\bar{\imath})\left(R_i-\bar{R}\right)}{\sum_{i=1}^{n}(i-\bar{\imath})^2},$$

where $\bar{\imath}$ is the average over observation number, i.e. $\bar{\imath}=(n+1)/2$, and $\bar{R}$ is a sample average rank, i.e. $\bar{R}=\bar{\imath}$. Then $\sum_{i=1}^{n}(i-\bar{\imath})^2=(n-1)n(n+1)/12$, and

$$\hat{\beta}=\frac{12}{(n-1)n(n+1)}\sum_{i=1}^{n}\left(i-\frac{n+1}{2}\right)\left(R_i-\frac{n+1}{2}\right)=\frac{\rho}{n+1}.$$

■

Proof. [of Theorem 8] The sample slope coefficient is in the regression (6.17)

$$\hat{\beta}=\frac{\sum_{i,j=1}^{n}(S_j-S_i)(R_j-R_i)}{\sum_{i,j=1}^{n}(S_j-S_i)^2},$$

because by construction the averages of regressors and regressands over the n^2 observations are zero. Taking into account that $\sum_{i,j=1}^{n}(S_j-S_i)^2=\sum_{i,j=1}^{n}(R_j-R_i)^2$ by construction of rank vectors, we get $\hat{\beta}=\rho$. When an additional constant term is included in the regression, its estimate is exactly zero because the sample averages of both sides equal zero. ■

B Appendix: supplimentary material on the Pesaran–Timmermann test

Let ι_q be $q\times 1$ vector of ones for an integer q, $\Upsilon=\iota_{\ell^2}-\text{vec}(I_\ell)$, $V_x=\text{diag}\{\pi_{i,\cdot}\}_{i=1}^{\ell}-\pi_x\pi_x'$, $v_x=\|\pi_{i,\cdot}(1-\pi_{i,\cdot})\|_{i=1}^{\ell+1}$, $V_y=\text{diag}\{\pi_{\cdot,j}\}_{j=1}^{\ell}-\pi_y\pi_y'$, and $v_y=\|\pi_{\cdot,j}(1-\pi_{\cdot,j})\|_{j=1}^{\ell+1}$. Let also B denote the matrix of slope coefficients in the system (3.2).

Theorem 9 *The Pesaran–Timmermann S_n test is asymptotically equivalent to an OLS-based t test for the null*

$$H_0:\Upsilon'\text{vec}(V_xB')=0$$

in the regression system 3.2. Alternatively, the Pesaran–Timmermann S_n test is asymptotically equivalent to an OLS-based t test for the null

$$H_0:\Upsilon'\text{vec}(V_yB')=0$$

in the regression system 3.2 where $\mathbb{I}_{i,\cdot}$ *are regressed on* $\mathbb{I}_{\cdot,j}$ *rather than* $\mathbb{I}_{\cdot,j}$ *are regressed on* $\mathbb{I}_{i,\cdot}$.

Proof. [of Theorem 9] As follows from the proof of Theorem 1,

$$\begin{aligned} S_n &= \sum_{i=1}^{\ell} (p_{i,i} - p_{i,\cdot}p_{\cdot,i}) + p_{\ell+1,\ell+1} - p_{\ell+1,\cdot}p_{\cdot,\ell+1} \\ &= \sum_{i=1}^{\ell} (p_{i,i} - p_{i,\cdot}p_{\cdot,i}) - \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} (p_{i,j} - p_{i\cdot}p_{\cdot j}) \\ &= -\Upsilon' \left(\overline{\xi_x \otimes \xi_y}\right). \end{aligned}$$

Let $\hat{B}$ be the OLS estimator of B. Because $\text{vec}\left(\hat{B}'\right) = \left(I_\ell \otimes \hat{V}_x^{-1}\right)\left(\overline{\xi_x \otimes \xi_y}\right)$, we have

$$S_n \overset{A}{=} -\Upsilon' \left(I_\ell \otimes V_x\right) \text{vec}\left(B'\right) = -\Upsilon' \text{vec}\left(V_x B'\right).$$

The last conclusion follows from the symmetry between x and y. ■

Theorem 10 *The Pesaran–Timmermann* S_n *test is asymptotically equivalent to an OLS-based t test for the null*

$$H_0 : v_x' \beta = 0,$$

where $\beta = \|\beta_i\|_{i=1}^{\ell+1}$, *in the regression system*

$$\mathbb{I}_{i,\cdot} = \alpha_i + \beta_i \mathbb{I}_{\cdot,i} + \eta_i, \quad i = 1, \cdots, \ell+1. \tag{B.1}$$

Alternatively, the Pesaran–Timmermann S_n *test is asymptotically equivalent to an OLS-based t test for the null*

$$H_0 : v_y' \beta = 0$$

in the regression system B.1 where $\mathbb{I}_{\cdot,i}$ *are regressed on* $\mathbb{I}_{i,\cdot}$ *rather than* $\mathbb{I}_{i,\cdot}$ *are regressed on* $\mathbb{I}_{\cdot,i}$.

Proof. [of Theorem 10] Consider the regression system (B.1). Let $\hat{\beta}$ be the OLS estimator of β. Because $\operatorname{diag}\left\{\overline{\left(\mathbb{I}_{\cdot,i}-\bar{\mathbb{I}}_{\cdot,i}\right)^2}\right\}_{i=1}^{\ell+1}\hat{\beta}=\left\|\overline{\left(\mathbb{I}_{i,\cdot}-\bar{\mathbb{I}}_{i,\cdot}\right)\left(\mathbb{I}_{\cdot,i}-\bar{\mathbb{I}}_{\cdot,i}\right)}\right\|_{i=1}^{\ell+1}$, we have

$$\begin{aligned} S_n &= \sum_{i=1}^{\ell+1}\left(p_{i,i}-p_{i,\cdot}p_{\cdot,i}\right)=\iota_{\ell+1}'\left\|\overline{\left(\mathbb{I}_{i,\cdot}-\bar{\mathbb{I}}_{i,\cdot}\right)\left(\mathbb{I}_{\cdot,i}-\bar{\mathbb{I}}_{\cdot,i}\right)}\right\|_{i=1}^{\ell+1} \\ &\overset{A}{=} v_x'\beta. \end{aligned}$$

The last conclusion follows from the symmetry between x and y. ■

One can see that if one wants to run the Pesaran–Timmermann regression-based test, one has to test a restriction that contains nuisance parameters (V_x, V_y, v_x or v_y). These, of course, are unknown and may be estimated before running the test, but their estimation generally has an impact on the asymptotic distribution of the test statistic and hence may distort the asymptotic size.