Economics Letters 220 (2022) 110850

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Tests for independence between categorical variables

Juan Sentana

Department of Economics, ICADE, Universidad Pontificia Comillas, Madrid, Spain

ARTICLE INFO

Article history:
Received 28 March 2022
Received in revised form 12 June 2022
Accepted 6 September 2022
Available online 16 September 2022

JEL classification:
C25
C35
C57

Keywords:
Linear Probability Model
Logit
Overidentifying Restrictions
Probit

ABSTRACT

I prove the numerical equivalence between Pearson's independence test statistic for categorical variables and the Lagrange Multiplier and overidentifying restrictions test statistics in several popular linear and non-linear regression models. I also show that its asymptotically equivalent Likelihood Ratio test is numerically identical in the non-linear regression models, and that the heteroskedasticity-robust Wald test statistic in the multivariate linear probability model and the moment condition model coincide with the Wald test statistic in the conditional multinomial model. Finally, I show that all these equivalences also apply to serial independence tests in discrete Markov chains.

1. Introduction

Many important economic theories imply the independence between two categorical variables. Examples include the sign of price movements and excess demand/supply (Bouissou et al. (1986)), financial market timing (Pesaran and Timmermann (1994)), blood donations and monetary compensation (Mellström and Johannesson (2008)) and the Minimax theorem (Brown and Rosenthal (1990), Chiappori et al. (2002) and Palacios-Huerta (2003)).

There are multiple procedures to test independence. For example, Palacios-Huerta (2003) used Pearson's test in a contingency table, Chiappori et al. (2002) used Wald tests in a Linear Probability Model (LPM) and Brown and Rosenthal (1990) relied on Likelihood Ratio (LR) tests in a logit model.

Anatolyev and Kosenok (2009) showed the asymptotic equivalence between Pearson's test statistic and the Wald test statistic in a multivariate LPM under *i.i.d* sampling. However, this equivalence does not prevent reaching different conclusions in finite samples with the same dataset even if p -values are computed in analogous ways.

In this paper I prove the numerical equivalence between many seemingly unrelated independence test statistics. Table 1 summarizes the results. ◯ corresponds to Pearson's test statistic and its numerically equivalent versions, △ and □ represent the LR and Wald test statistics in the multinomial model, and ▽ stresses asymptotic equivalences. All these equivalences also apply to serial independence tests for discrete Markov chains.

Table 1
Equivalence results.

Models\Test statistics	LM	LR	Wald	Wald-Robust	J-test
Multivariate LPM	◯	▽	▽	□	–
Unconditional multinomial model	◯	△	▽	▽	–
Conditional multinomial model	◯	△	□	▽	–
Multinomial probit	◯	△	▽	▽	–
Multinomial logit	◯	△	▽	▽	–
Moment condition model	◯	◯	▽	□	◯

2. Testing methods

Let x be the $K \times 1$ categorical variable $(A_1, \ldots, A_K)$ and $\tilde{y}$ the $H \times 1$ categorical variable $(B_1, \ldots, B_H)$. Both A_k and B_h, for $k = 1, \ldots, K$ and $h = 1, \ldots, H$, are dummy variables equal to 1 if its corresponding categorical value is equal to its kth or hth value, respectively.

Let the contingency table be

$\tilde{y}\backslash x$	A_1	$\cdots$	A_K	Sum
B_1	n_{11}	$\cdots$	n_{1K}	$n_{1\diamond}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_H	n_{H1}	$\cdots$	n_{HK}	$n_{H\diamond}$
Sum	n_{*1}	$\cdots$	n_{*K}	n

where n_{hk}, for $h = 1, \ldots, H$ and $k = 1, \ldots, K$, denotes the observed joint frequency, $n_{h\diamond} = \sum_{k=1}^{K} n_{hk}$ the number of times that B_h is 1, $n_{*k} = \sum_{h=1}^{H} n_{hk}$ the number of times A_k is 1 and $n = \sum_{k=1}^{K} n_{*k} = \sum_{h=1}^{H} n_{h\diamond}$ the sample size.

E-mail address: jsentana@icade.comillas.edu.

https://doi.org/10.1016/j.econlet.2022.110850

2.1. Contingency test

Pearson's statistic is:

$$Pearson = \sum_{k=1}^{K}\sum_{h=1}^{H}\left(n_{hk}-(n_{*k}n_{h\diamond}/n)\right)^2\left(n/n_{*k}n_{h\diamond}\right). \quad (1)$$

Under the null hypothesis of independence between $\tilde{y}$ and x, (1) asymptotically follows a chi-squared distribution with $(H-1)\times(K-1)$ degrees of freedom under appropriate regularity conditions (see Mood et al. (1974)).[1,2]

2.2. Multivariate LPM

Consider

$$\tilde{y} = \Delta x + u, \quad (2)$$

where $u = (u_1, \ldots, u_H)'$ and $\Delta = \begin{pmatrix} \delta_{11} & \cdots & \delta_{1K} \\ \vdots & \ddots & \vdots \\ \delta_{H,1} & \cdots & \delta_{H,K} \end{pmatrix}$, with $\delta = vec\left(\Delta'\right)$ and $\Sigma_U = E(u_i u_i') = Var(u_i)$.

Given that A_k and B_h are dummy variables, the coefficients of the explanatory variables are

$$\begin{aligned}\delta_{hk} &= E(B_h|A_1 = 0, \ldots, A_k = 1, \ldots, A_K = 0)\\ &= \Pr(B_h = 1|A_1 = 0, \ldots, A_k = 1, \ldots, A_K = 0).\end{aligned}$$

Their sum is equal to 1 for all the columns in the regression coefficient matrix, so I can cross out the last equation without loss of generality to avoid a singular covariance matrix because $B_H = 1 - \sum_{h=1}^{H-1} B_h$ (see Judge et al. (1985)).

δ can be estimated by OLS equation by equation without loss of efficiency relative to the seemingly unrelated regression (SUR) estimator because the regressors in all the $H-1$ equations are identical. Thus, $\hat{\delta}_h^{SUR} = \hat{\delta}_h^{OLS} = (n_{h1}/n_{*1}, \ldots, n_{hK}/n_{*K})'$ provides the natural estimator of the conditional probabilities δ_{hk}, which are always non-negative and add up to 1. Therefore, it avoids a common criticism of the LPM motivated by the fact that this model does not necessarily imply conditional probabilities between 0 and 1 (see Wooldridge (2002)).

Under H_0, $\delta_{h1} = \cdots = \delta_{hK} = \delta_h$ for $h = 1, \ldots, H-1$, so the conditional and unconditional probabilities of $B_h = 1$ are the same. As in the unrestricted model, $\tilde{\delta}_h^{SUR} = \tilde{\delta}_h^{OLS} = n_{h\diamond}/n$, for $h = 1, \ldots, H-1$, under H_0.

Under the alternative, the multivariate LPM violates the homoskedasticity assumption because $Var(u|x)$ will change depending on the values of A_k (see Wooldridge (2002)), which justifies the use of heteroskedasticity-robust Wald tests. However, $\Sigma_R = Cov(u|x)$ is constant under H_0, implying that the non-robust regression tests remain asymptotically valid.

2.3. Conditional multinomial model

Let $P_{hk} = \Pr(B_h = 1|A_1 = 0, \ldots, A_k = 1, \ldots, A_K = 0)$ denote the conditional probabilities. The joint probability of $B_h = 1$ and $A_k = 1$ is $\pi_{hk} = P_{hk} \times \pi_{*k}$, where $\pi_{*k} = \Pr(A_k = 1)$, so the parameters of interest become P_{hk} and π_{*k}.

Under the alternative, $\hat{P}_{hk} = n_{hk}/n_{*k}$ and $\hat{\pi}_{*k} = n_{*k}/n_{*K}$, so $\hat{P}_{hk} = \hat{\delta}_{hk}$ for $h = 1, \ldots, H-1$ and $k = 1, \ldots, K$. Under H_0, $P_{hk} = P_{h\diamond}$ for $k = 1, \ldots, K$ and $h = 1, \ldots, H-1$, so $\tilde{P}_{h\diamond} = n_{h\diamond}/n$ and $\hat{\pi}_{*k} = n_{*k}/n_{*K}$, which results in $\tilde{P}_{h\diamond} = \tilde{\delta}_h$ for $h = 1, \ldots, H-1$. In contrast, $\hat{\pi}_{*k}$ coincides under the null and alternative.

It is worth mentioning that the information matrix evaluated under H_0 is block diagonal between P_{hk} and π_{*k} (see Online Appendix A.1.3).

2.4. Unconditional multinomial model

Let $\pi_{hk} = \Pr(B_h = 1; A_k = 1)$. The null hypothesis states that $\pi_{hk} = \pi_{h\diamond} \times \pi_{*k}$, $h = 1, \ldots, H$ and $k = 1, \ldots, K$, where $\pi_{h\diamond} = \Pr(B_h = 1)$. Therefore, it is convenient to write the joint probabilities under the alternative as the product of two sets of parameters: (i) $\pi_{h\diamond}$, for $h = 1, \ldots, H-1$, and π_{*k}, for $k = 1, \ldots, K-1$, which denote the marginal probability distribution for $\tilde{y}$ and x respectively, and (ii) $(K-1)\times(H-1)$ additional parameters ϑ which should be 0 under H_0 (see Mood et al. (1974) section 3.5.4 and Online Appendix A.1.4).

As in Section 2.3, the estimators of $\pi_{h\diamond}$ and π_{*k} are the same under the null and alternative, and the information matrix evaluated under H_0 is block diagonal between $\pi_{h\diamond}$, π_{*k} and the ϑ's.

2.5. Multinomial probit model

Consider the following "random utilities" model:

$$\left.\begin{aligned} B_{1i}^* &= \alpha_{11}A_{1i} + \cdots + \alpha_{1K}A_{Ki} + \varepsilon_{1i}\\ &\vdots\\ B_{Hi}^* &= \alpha_{H1}A_{1i} + \cdots + \alpha_{HK}A_{Ki} + \varepsilon_{Hi}\end{aligned}\right\}, \quad (3)$$

where $\varepsilon_h|x \sim i.i.d.\ N(0, \omega)$ (see section 27.3 of Ruud (2000)).

Let $B_{hi} = 1\left\{B_{hi}^* = \max_{j=1,\ldots,H} B_{ji}^*\right\}$, where $1\{\}$ is the indicator function, so that $B_{hi} = 1$ if h is the preferred choice. This implies $P_{hk} = \Pr(B_h = 1|x)$, which is a normal cumulative distribution function of dimension $H-1$. Under H_0, $\alpha_{h1} = \cdots = \alpha_{hK} = \alpha_h$, for $h = 1, \ldots, H$.

2.6. Multinomial logit model

The multinomial logit model is obtained if in (3), ε_{hi}, instead of being normal, is drawn from an *i.i.d.* extreme value distribution (see section 27.4 of Ruud (2000)). This model ensures $P_{hk} \geq 0$, for all h, k, as well as $\sum_{h=1}^{H} P_{hk} = 1$ by assuming that

$$\left.\begin{aligned}\Pr(B_h = 1 \mid A_1, \ldots, A_K) &= (1+D)^{-1}\exp\left(\textstyle\sum_{k=1}^{K}\gamma_{hk}A_{ki}\right),\\ \Pr(B_H = 1 \mid A_1, \ldots, A_K) &= (1+D)^{-1}\end{aligned}\right\},$$

where $D = \sum_{h=1}^{H-1}\exp\left(\sum_{k=1}^{K}\gamma_{hk}A_{ki}\right)$, and γ_{hk}, for $h = 1, \ldots, H-1$ and $k = 1, \ldots, K$, are the model parameters. Under H_0, $\gamma_{h1} = \ldots = \gamma_{hK} = \gamma_h$ for $h = 1, \ldots, H-1$.

2.7. Moment condition model

We can express all the conditional probabilities P_{hk} in terms of the following set of moment conditions

$$E\left[(y - \Pi x)\otimes x\right] = 0, \quad (4)$$

where $\otimes$ denotes Kronecker product, y is a categorical variable that coincides with the first $H-1$ elements of $\tilde{y}$ and Π contains the corresponding elements of Δ. These moment conditions coincide with the first order conditions of the multivariate LPM, as well as with the scores of the conditional multinomial model. Under H_1, Π is unrestricted while under H_0, $\Pi = \upsilon l_K'$, where l_K is a vector of K ones, but one can write $\Pi'(\upsilon) = l_K\upsilon' I_{H-1}$, which implies that $\delta(\upsilon) = vec(\Pi'(\upsilon)) = (I_{H-1}\otimes l_K)\upsilon$.

---

1 Apart from random sampling, all joint population frequencies must be strictly positive and fixed, so that n_{hk} increases asymptotically with n.

2 If $n_{h\diamond}$ and n_{*k} were fixed in repeated samples, the finite sample distribution of (1) would coincide with that of Fisher's (1922) exact test. In general, its finite sample distribution is unknown.

The GMM estimator is defined as:

$$\hat{\upsilon} = \arg\min_{\upsilon} \left(\frac{1}{n} \sum_{i=1}^{n} \{[y_i - \Pi(\upsilon)x_i] \otimes x_i\} \right)' \times \Upsilon^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \{[y_i - \Pi(\upsilon)x_i] \otimes x_i\} \right),$$

where Υ is a symmetric positive definite $[K \times (H-1)] \times [K \times (H-1)]$ weight matrix.

With random sampling, the optimal GMM estimator is the one which minimizes the GMM criterion function when $\Upsilon = [\Sigma_R \otimes \sum_{i=1}^{n} (x_i x_i')]$, where Σ_R is defined in Section 2.2.

The J-test statistic is just the value of the GMM objective function evaluated at the efficient GMM estimator (see Hansen (1982)). Algebraically, $J = n \times \bar{g}(\hat{\upsilon})' \Upsilon^{-1} \bar{g}(\hat{\upsilon})$, where $\bar{g}(\hat{\upsilon}) = n^{-1} \sum_{i=1}^{n} \{[y_i - \Pi(\hat{\upsilon})x_i] \otimes x_i\}$.

3. Numerical equivalence results

The following proposition, which I prove in Online Appendix A.1, contains the main result:

Proposition 1. *For general H and K, the LM test statistic for independence in a multivariate LPM, multinomial logit, multinomial probit and the conditional and unconditional multinomial models, computed using the information matrix, are numerically identical to Pearson's contingency table test statistic and the GMM J-test statistic in the moment condition model. Additionally, the same numerical equivalence result holds if one exchanges regressors and regressands in all those models.*

This means that, for any sample size and sampling scheme, researchers will reach the same conclusions if they use any of those test statistics as long as p-values are computed in a similar manner. This numerical result is substantially different from the famous inequality $Wald \geq LR \geq LM$ in the multivariate LPM (see Berndt and Savin (1977)) because it shows that the LM test statistics are numerically identical across multiple linear and non-linear models.

Computationally, the easiest test is Pearson's statistic due to its simple closed-form expression (1). In contrast, the multinomial logit and especially probit models should be avoided because they require numerical optimization.

Another implication of Proposition 1 is that the Monte Carlo experiments previously reported in the literature will apply to all those tests because there will only be one finite sample distribution, so they could be combined in a meta study.

Proposition 1 also says that if we exchange regressors and regressands the corresponding test statistics will not change. For example, one obtains numerically the same LM statistic if one regresses $\tilde{y}_i$ on x_i or x_i on $\tilde{y}_i$ in the multivariate LPM. Similarly, imposing independence on $\Pr(B_h = 1 \mid A_1,\ldots, A_K)$ for all h yields the same LM statistic in a conditional multinomial model as imposing it on $\Pr(A_k = 1 \mid B_1,\ldots, B_H)$ for all k.

Four of the models in Section 2 are essentially the same. Specifically, the log-likelihood function under the null and alternative of the multinomial logit and probit models are analogous to the conditional component of the log-likelihood of the multinomial model. In addition, the unconditional multinomial model can be regarded as an alternative reparametrization of the joint probabilities. Therefore, I prove in Online Appendix A.2 the following equality:

Proposition 2. *For general H and K, the LR test statistic for independence in the multinomial logit, multinomial probit and the conditional and unconditional multinomial models are numerically identical.*

Although the Wald test statistics in all those models will generally differ, the numerical equivalence between the OLS estimator in the multivariate LPM, the ML estimators of the conditional probabilities and the unrestricted GMM estimators suggest a close relationship. The crucial difference is the homoskedasticity assumption in the Wald test of the multivariate LPM. Specifically, if a robust test was carried out, the following numerical equality would hold (see Online Appendix A.3):

Proposition 3. *For general H and K, the heteroskedasticity-robust Wald test statistic for independence in the multivariate LPM and the moment condition model is numerically identical to the Wald test statistic of the conditional multinomial model.*

Given that the LM test statistic is numerically equivalent in all the models in Table 1, all the other statistics will also be asymptotically equivalent (see section 17.3 of Ruud (2000)) even though they will be numerically different. For example, $LR_{LPM} \neq LR_{Multinomial}$ because the true conditional distribution of the LPM is not normal, so the (pseudo) likelihood function of the multivariate LPM is different from the multinomial model one even under H_0 (see Online Appendix A.1). Similarly, the Wald test statistic of the multinomial logit and probit is different from the one in the conditional multinomial model because Wald statistics are not invariant to non-linear transformations of the restrictions, despite having the same log-likelihood functions under the null and alternative.

3.1. Serial independence tests for Markov chains

Propositions 1–3 can be extended to serial independence tests for discrete Markov chains.

Let x_t summarize the K variables $(A_1,\ldots, A_K)$ at time t, which has the Markov property if for all $k \geq 1$ and all t

$$\Pr(x_{t+1}|x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-k}) = \Pr(x_{t+1}|x_t).$$

The Markov chain is fully characterized by the $K \times K$ transition matrix

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1K} \\ \vdots & \ddots & \vdots \\ P_{K1} & \cdots & P_{KK} \end{pmatrix},$$

where $P_{hk} = \Pr(x_{t+1} = x_h|x_t = x_k)$ are the one step transition probabilities with states $k = 1, \ldots, K$, where $P_{Kk} = 1 - \sum_{h=1}^{K-1} P_{hk}$, for all k and $h = 1, \ldots, K-1$.

If the Markov chain is serially independent, the matrix P will be:

$$P = l_K \times \begin{pmatrix} \pi_1 & \cdots & \pi_{K-1} & 1 - \sum_{k=1}^{K-1} \pi_k \end{pmatrix}.$$

The main difference with the conditional model in Section 2.3 is that the marginal model of the Markov chain is based on a single observation while the conditional model is recursive. Nevertheless, serial independence can still be assessed by Wald, LR and LM test statistics. Not surprisingly, I can easily show that the numerical equivalence results in Propositions $1-3$ also apply in this context.

In summary, the only reason why researchers might reach different conclusions in empirical applications is because they compute p-values differently or use Wald or LR versions.

Acknowledgments

I would like to thank Dante Amengual, Abhimanyu Gupta, Gordon Kemp, Tatiana Komarova, Friederike Mengel, Simon Weidenholzer, and an anonymous reviewer for their useful and insightful comments, as well as the audience of the 2019 Econometric Society Summer Meeting in Manchester. I have also benefited from many useful conversations with Enrique Sentana. Any remaining errors are my own. I gratefully acknowledge the financial support from Fundacion Ramon Areces, Spain and the University of Essex, United Kingdom.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econlet.2022.110850.

References

Anatolyev, S., Kosenok, G., 2009. Tests in contingency tables as regression tests. Econom. Lett. 105, 189–192.
Berndt, E.R., Savin, N.E., 1977. Conflict among criteria for testing hypotheses in the multivariate linear regression model. Econometrica 45, 1263–1277.
Bouissou, M.B., Laffont, J., Vuong, Q.H., 1986. Tests of noncausality under Markov assumptions for qualitative panel data. Econometrica 54, 395–414.
Brown, J., Rosenthal, R., 1990. Testing the minimax hypothesis: A reexamination of O'Neill's experiment. Econometrica 58, 1065–1081.
Chiappori, P.-A., Levitt, S., Groseclose, T., 2002. Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. Amer. Econ. Rev. 92, 1138–1151.
Fisher, R.A., 1922. On the interpretation of $\mathcal{X}^2$ from contingency tables, and the calculation of P. J. R. Stat. Soc. 85, 87–94.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Judge, G., Griffiths, W., Lütkepohl, H., Lee, T.-C., Hill, R.C., 1985. The Theory and Practice of Econometrics. Wiley.
Mellström, C., Johannesson, M., 2008. Crowding out in blood donation: Was titmuss right? J. Eur. Econom. Assoc. 4, 845–863.
Mood, A., Graybill, F-A., Boes, D., 1974. Introduction to the Theory of Statistics, third ed. McGraw-Hill.
Palacios-Huerta, I., 2003. Professionals play minimax. Rev. Econom. Stud. 70, 395–415.
Pesaran, M.H, Timmermann, A., 1994. A generalization of the non-parametric Henriksson-Merton test of market timing. Econom. Lett. 44, 1–7.
Ruud, P., 2000. An Introduction to Classical Econometric Theory. Oxford University Press.
Wooldridge, J., 2002. Introductory Econometrics: A Modern Approach, second ed. South-Western.