PAPER DETAILS

TITLE: Specification test for fixed effects in binary panel data model: a simulation study

AUTHORS: Gang YU,Weiguo WANG,Maozu LU,Shaoping WANG

PAGES: 755-760

ORIGINAL PDF URL: https://dergipark.org.tr/tr/download/article-file/655482

Hacettepe Journal of Mathematics and Statistics
Volume 44 (3) (2015), 755 – 760

Specification test for fixed effects in binary panel data model: a simulation study

Gang Yu [*] , Weiguo Wang [†] , Maozu Lu [‡] and Shaoping Wang [§]

Abstract

In this paper, we examine the specification tests which have been proposed for fixed effects in binary panel data model, using several different data generating processes to evaluate the performance of the specification test in different situations. By simulations, we find the specification test based on moment conditions is able to outperform the Lagrange multiplier test proposed by Gurmu [5].

2000 AMS Classification: 62F03, 62F05.

Keywords: Specification test, Fixed effects, Binary panel data.

Received 16/12/2013 : Accepted 30.06.2014 Doi : 10.15672/HJMS.2014227473

[*]Center for Econometrics Analysis and Forecasting, School of Mathematics and Quantitative Economics, Dongbei University of Finance and Economics, Liaoning, 116025, P.R.China. School of Economics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, P.R.China
Email: yugang543@163.com Corresponding Author.

[†]Center for Econometrics Analysis and Forecasting, School of Mathematics and Quantitative Economics, Dongbei University of Finance and Economics, Liaoning, 116025, P.R.China
Email:wwguodufe@163.com

[‡]Center for Econometrics Analysis and Forecasting, School of Mathematics and Quantitative Economics, Dongbei University of Finance and Economics, Liaoning, 116025, P.R.China
Email:maozulu@gmail.com

[§]School of Economics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, P.R.China
Email:wangspi@sina.com

1. Introduction

Binary panel data models remain of major interest in microeconometrics. This paper examines the specification test for fixed effects in binary panel data model. The binary panel data model is in the following form:

$$y_{it} = 1(x'_{it}\beta + \eta_i + v_{it} \geq 0), \qquad i = 1, \cdots, N; t = 1, \cdots, T, \tag{1}$$

where $1(\cdot)$ denotes the indicator function that equals one if $\cdot$ is true and zero otherwise, y_{it} is an observed dependent variable, x_{it} is a $k \times 1$ vector of exogenous regressors, η_i denotes the individual's fixed effects and v_{it} is unobservable error term which is independently identical distributed with cdf $F(x)$ across units and time periods, where $F(x)$ is known and symmetric around 0.

In the binary panel data model (1), fixed effects estimation suffers from inconsistency under the incidental parameters problem, first considered by Neyman and Scott [9]. The incidental parameters problem persists in the binary panel data case because the nuisance parameters η_i can not be separated from estimators of coefficients of interest. As both N and T increase, the increasing number of parameters to estimate means that the coefficients will have an asymptotic bias.

Baltagi [1] proposes an open problem in Econometric Theory, i.e. the following test for fixed effects in binary panel data model (1):

$$H_0 : \eta_i = 0 \quad \text{for} \quad i = 1, \cdots, N. \tag{2}$$

If H_0 is not reject, the estimation procedure is simple and utilizes the usual logit and probit procedures. However, if H_0 is rejected, the maximum likelihood procedure is complicated by the presence of the incidental parameters problem. Furthermore, Gurmu [5] solves the open problem and proposes the lagrange multiplier (LM) test for the test problem H_0 by artificial regression, which is analogous to those used for tests in binary response model regression(BRMR) proposed by Davidson and MacKinnon [4], and shows LM $\sim \chi^2(N)$ under the null hypothesis. Some discussions about test for fixed effects in binary panel data also can be found in Baltagi [2]. Both Gurmu [5] and Baltagi [2] do not present the Monte Carlo simulations studies, LM test's small sample performance is unknown and will be tested in this paper through the use of Monte Carlo simulations.

For test problem (2) in binary panel data proposed by Baltagi [1], this paper also derives a test based on moment conditions, which asymptotic null distribution is the $\chi^2(1)$ distribution. The test is applied to Monte Carlo simulations and its power is compared with LM test proposed by Gurmu [5].

The structure of the paper is organized as follows. Section 2 introduces the test statistic based on moment conditions and its large sample properties. In section 3, we report some Monte Carlo simulation results. Section 4 concludes the paper.

2. Specification test based on moment conditions

The framework of deriving the test statistic is similar to Mora and Moro-Egido [8]. We assume that independent and identically distributed (i.i.d) observations $(y_{it}, x'_{it})'$ are available, where $i = 1, \cdots, N; t = 1, \cdots, T$. The following notation will be used: $p_{1,it}(\theta) \equiv \Pr(y_{it} = 1|x'_{it}) = F(x'_{it}\beta + \eta_i), p_{0,it}(\theta) \equiv \Pr(y_{it} = 0|x'_{it}) = 1 - p_{1,it}(\theta), p_{it} \equiv [p_{1,it}(\theta)]^{y_{it}} \times [p_{0,it}(\theta)]^{1-y_{it}}$, where $\theta = (\beta', \eta')'$ and $\eta = (\eta_1, \cdots, \eta_N)'$. Conditioning on the observations, the MLE of θ, $\hat{\theta} = (\hat{\beta}', \hat{\eta}')'$ maximizes the following log-likelihood function $l(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \ln p_{it}$.

Define $m_{it}(\theta) \equiv y_{it} - F(x'_{it}\beta + \eta_i)$. From binary panel data model (1), we have $\mathrm{E}m_{it}(\theta) = 0$. To derive the test statistic, we consider the random variable $\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta})$, where $\hat{\theta} = (\hat{\beta}', 0')'$ is a well-behaved maximum likelihood estimator of $\theta_0 = (\beta', 0')'$ and $\hat{\beta}$ is the vector of ML estimate subject to the restriction $H_0 : \eta_i = 0 \quad \text{for} \quad i = 1, \cdots, N$.

2.1. Theorem. *Consider model (1), assuming the following regularity conditions hold,*
(i) In the neighborhood of true value θ_0, $\partial \ln p_{it}/\partial\theta, \partial^2 \ln p_{it}/\partial\theta^2, \partial^3 \ln p_{it}/\partial\theta^3$ *exist;*
(ii) In the neighborhood of true value θ_0, $|\partial^3 \ln p_{it}/\partial\theta^3| \leq H(x)$, *and* $EH(x) < \infty$;
(iii) At the true value θ_0, $E_{\theta_0}[\partial \ln p_{it}/\partial\theta] = 0, E_{\theta_0}[p''_{it}/p_{it}] = 0$,
$I(\theta_0) = Var_{\theta_0}[\partial \ln p_{it}/\partial\theta] > 0$.
Under the null hypothesis given in equation (2), when $N, T \longrightarrow \infty$, *the* C_{NT}^{M} *statistics*

$$C_{NT}^{M} = (NT)^{-1}[\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta})]^2/\hat{V} \tag{3}$$

converges to a chi squared distribution with one degree, where

$$\hat{V} = (NT)^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}^2(\hat{\theta}) - [\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta})g_{it}(\hat{\theta})]^2/\sum_{i=1}^{N}\sum_{t=1}^{T} g_{it}^2(\hat{\theta})\}. \tag{4}$$

3. Monte Carlo simulation study

In this section, we present a small Monte Carlo study to illustrate the performance of the above test statistic (3) proposed in Section 2. For comparison, we also report the finite sample sizes and powers of LM test proposed by Gurmu [5].

The simulation is based on the logit model

$$y_{it} = 1(x_{it}\beta + \eta_i + v_{it} \geq 0), \qquad i = 1, \cdots, N; t = 1, \cdots, T, \tag{5}$$

where the true parameter value is $\beta = 1$, x_{it} is an exogenous variable and independently identical distributed with distribution $N(0, 1)$, v_{it} is independently identical distributed with logistic distribution $P\{v_{it} < x\} = F(x) = e^x/(1+e^x)$, and $\eta_i = (\sum_{t=1}^{T} z_{it})/T$, z_{it} is an exogenous variable and independently identical distributed with distribution $N(\mu, \sigma^2)$, so that the fixed effects η_i are generated from normal distribution. In model (5), we use statistics (3) to test $H_0 : \eta_i = 0 \quad \text{for} \quad i = 1, \cdots, N$. Parameter β is estimated by ML estimate assuming that H_0 holds. Values of both μ and σ^2 different from 0 allow us to examine the ability of the test statistic to detect misspecification in binary panel data model.

Table 1a Empirical sizes for logit design with different N and T.

N	Test	T=5			T=10			T=15		
		1%test	5%test	10%test	1%test	5%test	10%test	1%test	5%test	10%test
50	LM	0.013	0.052	0.103	0.008	0.044	0.114	0.002	0.051	0.093
	C_{NT}^{M}	0.004	0.060	0.117	0.021	0.062	0.118	0.016	0.057	0.113
100	LM	0.008	0.031	0.090	0.001	0.049	0.102	0.011	0.052	0.103
	C_{NT}^{M}	0.020	0.050	0.108	0.015	0.061	0.103	0.017	0.063	0.124
200	LM	0.003	0.034	0.078	0.006	0.040	0.112	0.010	0.052	0.083
	C_{NT}^{M}	0.012	0.062	0.108	0.015	0.066	0.091	0.019	0.062	0.119

Table 1b Empirical powers for logit design with different T when N=50.

μ	σ	Test	T=5			T=10			T=15		
			1%test	5%test	10%test	1%test	5%test	10%test	1%test	5%test	10%test
0.2	0.2	LM	0.010	0.061	0.169	0.030	0.139	0.201	0.045	0.164	0.275
		C_{NT}^{M}	0.148	0.334	0.450	0.335	0.602	0.709	0.527	0.753	0.837
	0.4	LM	0.011	0.093	0.163	0.035	0.145	0.242	0.056	0.180	0.298
		C_{NT}^{M}	0.177	0.346	0.460	0.365	0.587	0.682	0.518	0.715	0.826
	0.6	LM	0.016	0.101	0.210	0.032	0.176	0.289	0.072	0.249	0.335
		C_{NT}^{M}	0.171	0.360	0.449	0.340	0.582	0.668	0.531	0.738	0.824
	0.8	LM	0.023	0.130	0.230	0.061	0.207	0.354	0.101	0.289	0.399
		C_{NT}^{M}	0.155	0.322	0.437	0.340	0.540	0.675	0.527	0.722	0.819
0.4	0.2	LM	0.031	0.202	0.320	0.202	0.432	0.610	0.387	0.642	0.760
		C_{NT}^{M}	0.677	0.839	0.893	0.965	0.989	0.997	0.995	0.999	0.999
	0.4	LM	0.045	0.213	0.350	0.221	0.471	0.609	0.365	0.609	0.777
		C_{NT}^{M}	0.677	0.834	0.896	0.948	0.987	0.994	0.993	0.994	1.000
	0.6	LM	0.068	0.222	0.356	0.224	0.506	0.626	0.495	0.692	0.820
		C_{NT}^{M}	0.689	0.826	0.892	0.955	0.984	0.992	0.988	0.996	1.000
	0.8	LM	0.081	0.275	0.437	0.287	0.555	0.712	0.509	0.746	0.838
		C_{NT}^{M}	0.646	0.827	0.874	0.935	0.980	0.989	0.992	1.000	0.997

Table 2a Empirical sizes for probit design with different N and T.

N	Test	T=5			T=10			T=15		
		1%test	5%test	10%test	1%test	5%test	10%test	1%test	5%test	10%test
50	LM	0.002	0.025	0.067	0.008	0.052	0.086	0.009	0.038	0.086
	C_{NT}^{M}	0.012	0.064	0.106	0.007	0.052	0.095	0.011	0.056	0.111
100	LM	0.002	0.045	0.068	0.006	0.037	0.087	0.014	0.053	0.091
	C_{NT}^{M}	0.012	0.065	0.107	0.014	0.053	0.097	0.011	0.063	0.116
200	LM	0.001	0.034	0.069	0.003	0.036	0.092	0.006	0.050	0.074
	C_{NT}^{M}	0.007	0.053	0.107	0.011	0.042	0.101	0.006	0.050	0.104

Table 2b Empirical powers for probit design with different T when N=50

μ	σ	Test	T=5			T=10			T=15		
			1%test	5%test	10%test	1%test	5%test	10%test	1%test	5%test	10%test
0.2	0.2	LM	0.011	0.072	0.139	0.053	0.210	0.298	0.132	0.320	0.487
		C_{NT}^{M}	0.333	0.586	0.676	0.710	0.876	0.930	0.899	0.963	0.981
	0.4	LM	0.022	0.116	0.234	0.086	0.248	0.393	0.179	0.431	0.525
		C_{NT}^{M}	0.359	0.578	0.695	0.677	0.876	0.921	0.883	0.965	0.976
	0.6	LM	0.035	0.190	0.333	0.170	0.342	0.516	0.270	0.509	0.646
		C_{NT}^{M}	0.347	0.545	0.622	0.670	0.878	0.908	0.871	0.955	0.982
	0.8	LM	0.092	0.291	0.481	0.274	0.512	0.685	0.396	0.661	0.770
		C_{NT}^{M}	0.358	0.554	0.690	0.635	0.819	0.886	0.829	0.938	0.963
0.4	0.2	LM	0.108	0.328	0.501	0.606	0.815	0.898	0.916	0.980	0.993
		C_{NT}^{M}	0.957	0.995	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	0.4	LM	0.146	0.396	0.575	0.651	0.861	0.918	0.940	0.983	0.993
		C_{NT}^{M}	0.945	0.989	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	0.6	LM	0.215	0.480	0.657	0.720	0.884	0.954	0.940	0.989	0.996
		C_{NT}^{M}	0.958	0.984	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	0.8	LM	0.285	0.594	0.736	0.816	0.938	0.967	0.973	0.995	0.997
		C_{NT}^{M}	0.919	0.970	0.988	0.999	1.000	1.000	1.000	1.000	1.000

The simulation results of our test based on moment(C_{NT}^{M}), and Gurmu's test (LM) are reported in Table 1a and Table 1b based on 1000 simulations, where the nominal sizes are set to be 0.01, 0.05 and 0.10. From Table 1a, the empirical sizes for both tests are very close to the nominal sizes, with the LM test having less size distortion in most cases. From Table 1b, the proposed test C_{NT}^{M} is more powerful than Gurmu's LM test in all designs, and the powers significantly increase demonstrated by increasing the panel length T.

On the above data generating process (DGP), if assuming $F(x)$ is the Normal cdf, we also report the simulation results for a probit model in Table 2a and Table 2b. The results are qualitatively similar to those for the logit model in the previous Table 1a and Table 1b.

4. Conclusion

Specification test is an important part of panel data econometrics. This paper focuses on examining the specification test for fixed effects in binary panel data model by Monte Carlo simulations. The simulation results of this paper, along with the earlier work, show that the proposed test C_{NT}^{M} is more powerful than Gurmu's LM test.

In economics, it is more realistic to consider dynamic binary panel data model with fixed effects, for example, Hsiao [6], Bartolucci and Nigro [3], Yu, Gao and Shi [11] . As a possible area of further research it would be interesting to investigate the specification test for fixed effects in dynamic binary panel data model by using the proposed test C_{NT}^{M}.

5. Appendix A: Proof of results

Proof of Theorem 2.1: Using a first-order Taylor expansion for $\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator under H_0, we have

$$(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta}) = (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\theta_0) + B_0 \times (NT)^{1/2}(\hat{\theta}-\theta_0) + o_p(1), \tag{A.1}$$

where $B_0 = \mathrm{E}\{\partial m_{it}(\theta_0)/\partial\theta'\}$.

Under the conditions (i)-(iii) of Theorem 2.1, see the detailed proof of Theorem 2.3 in page 415, Lehmann [7], the ML estimator $\hat{\theta}$ satisfies that

$$(NT)^{1/2}(\hat{\theta}-\theta_0) = A_0^{-1} \times (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} g_{it}(\theta_0) + o_p(1), \tag{A.2}$$

where $g_{it}(\theta_0) = \partial \ln p_{it}/\partial\theta = 1(y_{it}=0) \times \frac{-f(x_{it}'\beta)x_{it}}{1-F(x_{it}'\beta)} + 1(y_{it}=1) \times \frac{f(x_{it}'\beta)x_{it}}{F(x_{it}'\beta)}$, $f(x)$ denotes the first derivative of $F(x)$, and $A_0 = E\{-\partial g_{it}(\theta_0)/\partial\theta'\} = I(\theta_0)$.

Inserting the asymptotic expansion of $(NT)^{1/2}(\hat{\theta}-\theta_0)$ into the Taylor expansion of $(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta})$, we have

$$(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta}) = (NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[m_{it}(\theta_0) + B_0A_0^{-1}g_{it}(\theta_0)] + o_p(1),$$

and we know that random variables $m_{it}(\theta_0)+B_0A_0^{-1}g_{it}(\theta_0)$ is independent and identically distributed, $\mathrm{E}(m_{it}(\theta_0) + B_0A_0^{-1}g_{it}(\theta_0)) = 0$, by central limit theorem (CLT),

$$(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta}) \xrightarrow{d} N(0,V), \tag{A.3}$$

where $V = \mathrm{Var}(m_{it}(\theta_0) + B_0A_0^{-1}g_{it}(\theta_0)$.

We can find that $m_{it}(\theta_0) + B_0A_0^{-1}g_{it}(\theta_0) = (1 : B_0A_0^{-1})(m_{it}(\theta_0), g_{it}(\theta_0))'$, so

$$V = (1 : B_0A_0^{-1})\begin{pmatrix} \mathrm{E}m_{it}^2(\theta_0) & \mathrm{E}[m_{it}(\theta_0)g_{it}(\theta_0)] \\ \mathrm{E}[m_{it}(\theta_0)g_{it}(\theta_0)] & \mathrm{E}g_{it}^2(\theta_0)\end{pmatrix}(1 : B_0A_0^{-1})'$$
$$= \mathrm{E}\{m_{it}^2(\theta_0)\} - \mathrm{E}^2\{m_{it}(\theta_0)g_{it}(\theta_0)\}/\mathrm{E}\{g_{it}^2(\theta_0)\},$$

where $\mathrm{E}[g_{it}^2(\theta_0)] = A_0, \mathrm{E}[m_{it}(\theta_0)g_{it}(\theta_0)] = -B_0$.

To obtain the a test statistic,a consistent estimator V must be proposed. The natural candidate for estimating V is

$$\hat{V} = (NT)^{-1}\{\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}^2(\hat{\theta}) - [\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta})g_{it}(\hat{\theta})]^2 / \sum_{i=1}^{N}\sum_{t=1}^{T} g_{it}^2(\hat{\theta})\}, \tag{A.4}$$

we replace population moments by sample moments, it is a standard estimate of V following Newey-Tauchen methodology, detailed discussion can be found in Orme [10].

Based on (A.3), (A.4) and Slutsky' Theorm, the test statistic proposed

$$C_{NT}^M = (NT)^{-1}[\sum_{i=1}^{N}\sum_{t=1}^{T} m_{it}(\hat{\theta})]^2/\hat{V} \xrightarrow{d} \chi^2(1), \tag{A.5}$$

what justifies the use of C_{NT}^M as an asymptotically valid test statistics. This completes the proof.

Acknowledgments

The authors would like to thank the editor Cem kadilar and two anonymous referees for their constructive comments and suggestions. This research is supported by the MOE (Ministry of Education in China) Youth Foundation Project of Humanities and Social Sciences (No.13YJC790185) , National Natural Science Foundation of China under Grants (Nos.71171035,71471030), China Postdoctoral Science Foundation (Nos.2012M511598, 2013T60710), Liaoning Educational Committee Special Project of Humanities and Social Science Key Research Base(No.ZJ2013039) and Dongbei University of Finance and Economics Project of Excellent Scientific Research Innovative Talents (No.DUFE2014R11).

References

[1] Baltagi B.H. *Testing for fixed effects in logit and probit models using an artificial regression*, Problem 95.5.4, Econometric Theory, **11**, 1179-1179, 1995.

[2] Baltagi B.H. *Specification tests in panel data models using artificial regressions*, Annals of Economics and Statistics, 277-297, 1999.

[3] Bartolucci F., Nigro V. *A dynamic model for binary panel data with unobserved heterogeneity admitting a-$\sqrt{N}$ consistent conditional estimator*, Econometrica, **78**, 719-733, 2010.

[4] Davidson R., MacKinnon J. *Convenient specification tests for logit and probit models*, Journal of Econometrics, **25**, 241-262, 1984.

[5] Gurmu S. *Testing for fixed effects in logit and probit models using an artificial regression*, Solution 95.5.4, Econometric Theory, **12**, 872-874, 1996.

[6] Hsiao C. *Analysis of Panel Data*, (Cambridge University Press, New York, 2003).

[7] Lehmann E.L. *Theory of Point Estimation*, (Wiley and Sons, Inc, 1983).

[8] Mora J, Moro-Egido A.I. *On specification testing of ordered discrete choice models*, Journal of Econometrics, **143**, 191-205, 2008.

[9] Neyman J., Scott E.L. *Consistent estimates based on partially consistent observations*, Econometrica, **16**, 1-32, 1948.

[10] Orme C.D. *The small sample performance of the information matrix test*, Journal of Econometrics, **46**, 309-331, 1990.

[11] Yu G., Gao W., Shi N.-Z. *A note on the estimation problem of dynamic binary panel data model with fixed effects*, Pakistan Journal of Statistics, **28**(2), 271-278, 2012.