A Panel Date Model of Cell Phone Internet Influencing on Undergraduates' Score

Lirong Ai [1], Shipeng Chen[2] and Shengping Jin[2,#]

[1]*Wuhan Institute of Shipbuilding Technology, Wuhan, 430050, China*

[2]*Department of Statistics ,Wuhan University of Technology, Wuhan, 430070,China*

Abstract

Cell phone internet influence on undergraduate's examination score is an important research issue in the education field in universities. The panel data model of cell phone internet influencing on undergraduate's mark is built based on the theory of probability and statistics. Data of undergraduate studying on internet is obtained from QQ contact group by using the mobile phone Internet. According to the statistical method, the influence factors weight of the regression on score is calculated, and results show that Cell Phone Internet affects undergraduates' scores significantly in universities.

Keywords: cell phone internet, theory of probability, QQ contact group, panel data.

1. Introduction

Using internet by cell phones is becoming a part life for undergraduates in universities. Cell phone internet refer to internet can be used by cell phones, or mobile phones. Whether undergraduate's examination score affected by cell phone internet or not can be analyzed by the data collected from contact QQ group.

Panel data refers to selected a number of sample data on the section of specific factor sequence[1]. Panel data has many advantages such as it can control individual heterogeneity data, and can provide more information from sample, and reduce linear collinearity, and provide more degrees of data freedom, and adjustment process of individual effects and time effects. Nowadays the panel data research is quickly development on the theory and application. Since 1960, scholars began to pay attention to panel data econometrics. In early 1986, Hsiao[1] wrote a monograph about the research of panel data model, and published second edition in 2003. Wooldridge[2], Arellano[3], Baltag[4] described many aspects of the panel data model. In recent years, a lot of researchers around the world set up a great study on the panel data model, such

[#] Corresponding author: spjin@whut.edu.cn

as nonlinear panel data model, the nonconstrained stationary panel data [5] model, dynamic panel data model, and variable structure model[6] and so on.

This paper will research mobile phone internet which affects on undergraduates examining mark by using panel data theory. In Section 2, panel data model is introduced. Section 3 discribed the estimation of parameters in panel data. And in Section 4, the influencing of cell phone internet on undergraduates' score is analyzed by panel data model. And Section 5 gives some conclusion remarks.

2. Panel data model

Support y_{it} is repones variable, or observation for undergraduate examination score, and suppose $X_{it} = (x_{1it}, x_{2it}, ..., x_{kit})$ is $1\times k$ explanatory variable,or the observation probability of undergraduate using cell phone internet. $b = (b_{1i}, b_{2i}, ..., b_{ki})^{\mathrm{T}}$ is $k\times 1$ coefficient vector, u_{it} is random error term which satisfied mutually independent, zero mean, and same variances. The single equation panel data model can be written as:

$$y_{it} = a_i + x_{it}b + u_{it},\ i = 1,2,...,N,\ t = 1,2,...,T \tag{1}$$

According to the matrices terminology, let:

$$Y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \cdots \\ y_{iT} \end{pmatrix}, \quad X_i = \begin{pmatrix} x_{1i1} & x_{2i1} & \cdots & x_{ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{ki2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1iT} & x_{2iT} & \cdots & x_{kit} \end{pmatrix} = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \cdots \\ X_{iT} \end{pmatrix}, \quad U_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \cdots \\ u_{iT} \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}, X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}, U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{pmatrix}, a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, Z = I_N \otimes e_T ,$$

Where I_N is the identity matrix of order N , e_T is the vector of ones of dimension T , and $\otimes$ denotes Kronecker product.

The panel data can be represented by metrix equation:

$$Y = Za + Xb + U \tag{2}$$

The number of sample observation value of panel data is NT . If $N = 1$, the panel data model of Eq. 2 becomes one dimension observation data.

3. Estimation of the parameters in panel data

We consider short time panel data analyzed by using OLS estimation method[2]. The different value of panel data with its average data is calculated, and then estimate the parameters in the model for the probability of undergraduate using cell phone internet.

The average of Eq. 1 are:

$$\bar{y}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \tag{3}$$

So that the average formula are:

$$\bar{y}_{i\cdot} = a_i + \bar{x}_{i\cdot} b + \bar{u}_{i\cdot}, \quad i = 1, 2, ..., N \tag{4}$$

From above Eq. 3 and Eq. 4 , we have

$$y_{it} - \bar{y}_{i\cdot} = \left(x_{it} - \bar{x}_{i\cdot}\right) b + \left(u_{it} - \bar{u}_{i\cdot}\right) \tag{5}$$

By applying OLS estimation method[2], we have parameter estimator from Eq. 5 if b is a sclar:

$$\hat{b} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_{i\cdot}\right)\left(y_{it} - \bar{y}_{i\cdot}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_{i\cdot}\right)\left(x_{it} - \bar{x}_{i\cdot}\right)}$$

If b is a vector, the estimation of b can also be easy to calculated by the multivariate linear regression. It is called OLS estimative value, and is unbiased estimator, and estimative value of a_i can be from the formula as

$$\hat{a}_i = \bar{y}_{i\cdot} - \bar{x}_{i\cdot}\hat{b}\,.$$

Let

$$RSS_i = W_{yy,i} - W^{T}{}_{xy,i} W^{-1}{}_{xx,i} W_{xy,i}$$

be the residual sum of squares for i group, $i = 1, 2, \cdots, N$, where

$$W_{xx,i} = \sum_{t=1}^{T}\left(x_{it} - \bar{x}_{i\cdot}\right)^{T}\left(x_{it} - \bar{x}_{i\cdot}\right)$$

$$W_{xy,i} = \sum_{t=1}^{T}\left(x_{it} - \bar{x}_{i\cdot}\right)^{T}\left(y_{it} - \bar{y}_{i\cdot}\right), \; W_{yy,i} = \sum_{t=1}^{T}\left(y_{it} - \bar{y}_{i\cdot}\right)^{2}$$

And denote

$$W_{xx} = \sum_{i=1}^{N} W_{xx,i}, W_{xy} = \sum_{i=1}^{N} W_{xy,i}, W_{yy} = \sum_{i=1}^{N} W_{yy,i}\,,$$

Then the whole residual sum of squares for our model and other statistics are

$$S_1 = \sum_{i=1}^{N} RSS_i\ ,\ S_2 = W_{yy}\ ,\ S_3 = W_{yy} - W'_{xy} W_{xx}^{-1} W_{xy}$$

We can show below theorem.

Theorem 1[4] If x_{it} is panel data, we have below conclusions:

(1) $S_1/\sigma_u^2 \sim \chi^2\left[N\left(T-k-1\right)\right]$

(2) $S_3/\sigma_u^2 \sim \chi^2\left[NT-\left(k+1\right)\right]$

(3) $\left(S_3-S_1\right)/\sigma_u^2 \sim \chi^2\left[\left(N-1\right)\left(k+1\right)\right]$

(4) $\left(S_3-S_1\right)/\sigma_u^2$ is independent with S_1/σ_u^2

(5) $$F_2 = \frac{\left(S_3-S_1\right)/\left[\left(N-1\right)\left(k+1\right)\right]}{S_1/\left[N\left(T-k-1\right)\right]}$$

$$\sim F\left[\left(N-1\right)\left(k+1\right), N\left(T-k-1\right)\right]$$

(6) $$F_1 = \frac{\left(S_2-S_1\right)/\left[\left(N-1\right)k\right]}{S_1/\left[N\left(T-k-1\right)\right]} \sim F\left[\left(N-1\right)k, N\left(T-k-1\right)\right]$$

(7) $$LM = \frac{NT}{2\left(T-1\right)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T} e_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^{\ 2}} - 1\right] = \frac{NT}{2\left(T-1\right)} \cdot \left(\frac{e^T DD^T e}{e^T e} - 1\right)^2$$

$$LM \sim \chi^2(1)$$

where e_{it} is residual of linear regression, e and D are matrices of it.

4. Affect examination score model and calculation

Through the QQ contact group, a total of 2438 undergraduates, from freshmen to senior grades, are investigated. The undergraduates come from enginnering colleges, transportantion colleges or ship building colleges. Some of them come from the countryside, or mountainous area, which family income are not high. Some of them come from cities, or economically developed areas, which family income are relative higher. According to the following standards, the family income are divided into 5 categories.

0—2500,2501-5000,5001—75000,75001--100000, >100000,

From the undergraduate’s family income distribution, which is depicted in Fig.1, we find that most undergraduate’s family income is less than 2500.

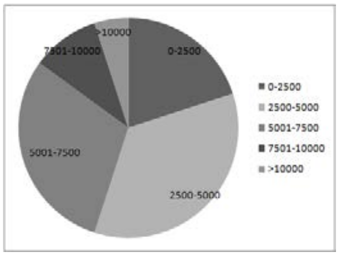

Fig.1 distribution for the student's family income

4.1 Estimation of the parameters

Applying Eviews software, we calculate Hausman estimative value shown in the Table 1.

Table 1 Hausman estimative value

Test Summary		Chi-Sq.	Chi-Sq. d.f.	Prob.
Cross-section random		14.554275	1	0.0000
Cross-section random effects test comparisons:				
Variable	Fixed	Random	Var(Diff.)	Prob.
PO	0.001914	0.000895	0.000000	0.0000

From Table 1, the Hausman estimative value is 14.554275, fixed individual parameter estimation value is 0.001914, random parameter estimation value is 0.000895.

Filling YD in the Cross section Specific coefficients of Eviews software, and choose fixed effects in the window of "intercept", we get coefficient of undergraduate's probability shown in the Table 2. In the table, Std. Error is parameter normal error, t-Statistic is t value of t statistic.

Table 2 fixed effects model（OLS）estimative value

Variable	Coefficient	Std. Error	t-Statistic
C	-25.1105	57.810	-7.63354
BJ--POBJ	0.02157	0.000121	17.78737
SH--POSH	0.01611	0.000119	13.49879
TJ--POTJ	0.01524	0.000195	7.822738
GZ--POGZ	0.09227	0.000796	11.59204
NJ--PONJ	0.08890	0.001062	8.368066
CS--POCS	0.06079	0.001280	4.749214
CD--POCD	0.03671	0.000584	6.285426
HF--POHF	0.01367	0.000290	4.713353
WH--POWH	0.06376	0.001169	5.454732
ZZ--POZZ	0.01005	0.000243	4.143437
XA--POXA	0.04200	0.000893	4.705909
TY--POTY	0.08906	0.001706	5.219903
LZ--POLZ	0.12275	0.003411	3.598143

The goodness of fitting regression equation, F statistic and DW statistic value index and other statistics are listed in the Table 3.

Table 3 fixed model（OLS）estimative results

R-squared	0.940960	Mean dependent var	52.803
Adjusted R-squared	0.928308	S.D. dependent var	33.205
S.E. of regression	89.4364	Akaike info criterion	16.60438
Sum squared resid	1.02E+04	Schwarz criterion	17.15655
Log likelihood	-120.537	Hannan-Quinn criter.	16.82867
F-statistic	74.37568	Durbin-Watson stat	0.992090

Two lays regression model is doubled in this paper as below:

$$z = a_1 y_1 + a_2 y_2 + a_3 y_3 + a_4 y_4 + \varepsilon$$

Where z is all grad student regression examination score, y_k is *k*th grade undergraduate examination score represented by:

$$y_1 = a_{10} + a_{11}x_{11} + a_{12}x_{12} + a_{13}x_{13} + a_{14}x_{14} + a_{15}x_{15} + \varepsilon_1$$

$$y_2 = a_{20} + a_{21}x_{21} + a_{22}x_{22} + a_{23}x_{23} + a_{24}x_{24} + a_{25}x_{25} + \varepsilon_2$$

$$y_3 = a_{30} + a_{31}x_{31} + a_{32}x_{32} + a_{33}x_{33} + a_{34}x_{34} + a_{35}x_{35} + \varepsilon_3$$

$$y_4 = a_{40} + a_{41}x_{41} + a_{42}x_{42} + a_{43}x_{43} + a_{44}x_{44} + a_{45}x_{45} + \varepsilon_4$$

By using data, we easy obtain regression equation of examination score z is

$$\begin{aligned} z &= 0.25y_1 + 0.23y_2 + 0.27y_3 + 0.25y_4 \\ y_1 &= 77 - 0.210x_{11} + 0.131x_{12} + 0.323x_{13} + 0.236x_{14} + 0.204x_{15} \\ y_2 &= 73 - 0.131x_{21} + 0.204x_{22} + 0.271x_{23} + 0.240x_{24} + 0.210x_{25} \\ y_3 &= 70 - 0.013x_{31} + 0.302x_{32} + 0.284x_{33} + 0.227x_{34} + 0.221x_{35} \\ y_4 &= 69 + 0.141x_{41} + 0.204x_{42} + 0.268x_{43} + 0.251x_{44} + 0.216x_{45} \end{aligned} \quad (6)$$

From above formula, we find that cell phone affects more on first year to third year undergraduate. For students from poor income families, the affective degree is decrease with the time of undergraduate in the university. For other undergraduates, the cell phone internet is beneficial to examination score. In this calculation, the sum square residue is 201.34, the R-square is 0.94121, see Table 4.

Table 4 estimative results of regression

R-squared	0.94090	Mean dependent var	57.803
Adjusted R-squared	0.92838	S.D. dependent var	339.205
S.E. of regression	29.4364	Akaike info criterion	6.60438
Sum squared residue	201.304	Schwarz criterion	7.15655
Log likelihood	-120.537	Hannan-Quinn criter.	6.82867
F-statistic	24.3568	Durbin-Watson stat	0.99090

4.2 Diagnosis of the model

The diagnosis of the first regression of Eq.(6) is as following by using the R software.

Coefficients:

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*	
(Intercept)	*75.02919*	*0.07039*	*1065.92*	*<2e-16*	*****
group2	*0.121261*	*0.08621*	*14.07*	*<2e-16*	*****
group3	*0.313797*	*0.09444*	*33.23*	*<2e-16*	*****
group4	*0.227017*	*0.12192*	*18.62*	*<2e-16*	*****
group5	*0.212814*	*0.15740*	*13.52*	*<2e-16*	*****

---

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*
Residual standard error: 0.9955 on 995 degrees of freedom
Multiple R-squared: 0.5535, Adjusted R-squared: 0.5517
F-statistic: 308.4 on 4 and 995 DF, p-value: < 2.2e-16

We can see from the diagnosis results that the affecting factors are all significants, and the whole model is significants also. So our inference in Section 4.1 is credible.

Fig. 2 is the graph of Model checking plots for the first equation of regression of Eq.(6). It shows that the model assumption is reasonable.

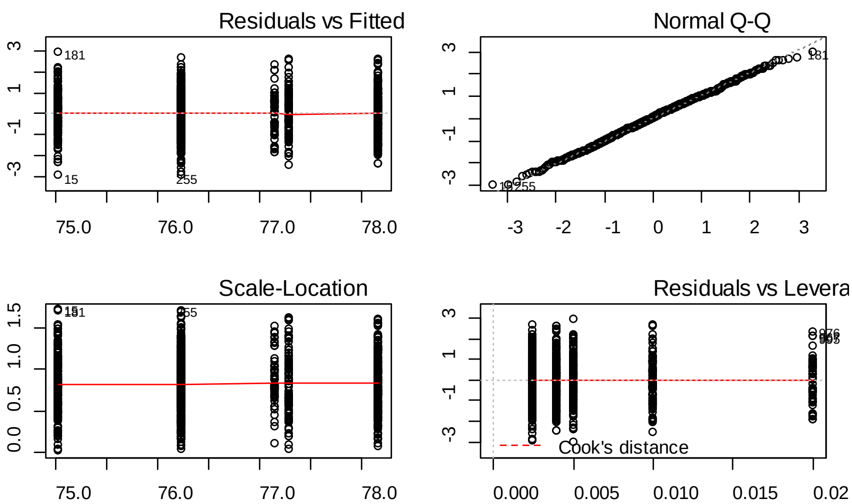

Fig.2 Model checking plots for the first equation of regression of Eq.(6)

The same diagnosis for other equations of Eq.(6) are similar to the above and is ommited, other checking tests by using *Theorem 1* are not given here for saving the space.

5. Conclusions

The mobile phone has been developed rapidly on using Internet, undergraduates are an important group of mobile phone access to the Internet, so the research on Mobile Phone Internet influence on undergraduates is inprotant project. According to the undergraduate's grade at the University and their family income, we have constructed a 2-D panel data model and calculate parameter in the OLS model.

The results show that the mobile phone access to the Internet has a great influence on undergraduate examination scores and is also beneficial to examination score.

Acknowledgements

The research work was supported by National Natural Science Foundation of China under Grant No.51279149, No.51179147, No.51139005 and China national education department doctoral foundation under Grant No. 20120143130002.

References

[1] Hsiao C. Analysis of Panel Data (second edition)[M]. Bei Jing: The PeKing University Press, (2003).

[2] Wooldridge J.M. Econometric Analysis of Cross Section and Panel Data [M], Massachusetts, The MIT Press, (2002).

[3] Arellano M. Panel data econometrics[M]. Oxford: Oxford University Press. (2003).

[4] Baltagi B.H. Econometric Analysis of Panel Data [M] .New York:Johu Wiley&Sons,(2005).

[5]Ye Xiaoqing, Estimator for nonlinear panel data model with interactive fixed effects[J], Statistical Research, Vol. 02, (2014), 97-101

[6] Hubrich K., Terasvirta T. Thresholds and smooth transitions in vetor autoregressive model[R], CREATSS research paper,(2013).