

An Empirical Study of China's E-business Economy from the perspective of Internet Penetration

Bing LI⁺

School of Economics and Business Administration, Shanxi University

Abstract. With the development of global economy, China has been standing out as one of the largest emerging economies especially from the outbreak of the financial crisis in the year of 2008. E-business is rapidly developing worldwide and thus it will also impact China's economy. This paper attempts to evaluate how China's economy is affected by e-business. Some important data or indices related to Internet penetration are used to make empirical analysis from the national and provincial levels. The regression results show that China's economy is greatly related to Internet penetration.

Keywords: E-business, Internet penetration, Regression

1. Introduction

The information and communication technologies are changing the world. Internet is one of the important technologies and it is changing the world from various perspectives such as socially, economically and politically. With the Internet penetration, e-business economy has amounted for a proportion of national economies, not only for developed countries but also for the developing countries. The e-business trading volume of China in 2009 has reached 3,800 billion RMB [1]. From 2007 to 2010, China's B2B trading volume has increased with the average percent of 37.6% [2]. According to the latest survey carried out by China Internet Network Information Center (CNNIC) by the end of 2011, China's Internet population has reached 513 million, with a yearly increase of 55.8 million. The penetration rate is 38.3% with a yearly increase of 4% [3].

The above data from the China's economy and the Internet penetration give us a direct perception of the rapid development in e-business. This paper attempts to empirically analyze the China's e-business economy. The national and provincial data will be studied separately.

2. Research Data and Methodology

Research data are mostly selected from the reports published by CNNIC and National Bureau of Statistics of China (NBS). Considering the integrity and availability of the historical statistical data, the time range is from 1997 to 2011. The semi-annual data under analysis include the numbers of the domain names with the ending of .cn (i.e., ndomain_cn), the web sites (web) and the netizens (popn). Some percentages are also included, such as Internet penetration rate (rpen); the age structure reflected by the proportion in the thirties' (pop_age_thirty), in the forties' (pop_age_forty) and below the thirties' (pop_age_young) among the netizens; the education structure reflected by the proportion of the netizens who have received the education of senior high school (pop_edu_shs)/junior college (pop_edu_jco)/college and above (pop_inc_col); the income structure reflected by the proportion of low level (pop_inc_low) and middle level (pop_inc_mid) among the netizens. The economic data is represented by the GDP. Since the direct e-business data in China such as B2B and B2C trading volume cannot form a time series with enough samples, this direct relation between gdp and e-business trading will be represented by the relation between the GDP and the Internet-related indices. These Internet-related statistical data are considered as indices representing the extent of Internet penetration.

Furthermore, the provincial-level analysis is made. The used data are the yearly data in 2011 and 31 provincial areas are selected, covering most parts of China. These data include the numbers of IPv4 addresses (i.e., ipv4), the domain names (ndomain), the domain names with the ending of .cn (ndomain_cn),

⁺ Email: libingxjtu@hotmail.com

the web sites (web), the web pages (webpage), the data volume of all the web pages (page_data_vol), the netizens (popn) and the Internet penetration rate (rpen). The provincial GDP is used to indicate the economic yield in each province.

From the above selected data, the research is conducted in the statistical way. Factor analysis is initially employed to disclose the underlying relationship among those numbers or indices; then the significant factors along with the corresponding indices are used to construct the measured model and thus the confirmatory factor analysis is made. Since the data sample scale is very limited and the variables are collinear, the results of the factor analysis are not well and cannot show us a strong confidence. But the correlations are obvious and thus the simple regression is used to illustrate the strong correlation between the gdp and Internet-related indices. As follows, the regression results are explained at the national and provincial levels.

3. National Analysis

The above mentioned data work as an input for regression analysis in SPSS. The independent variables include 12 indices related to Internet penetration; the dependent variable is gdp. The stepwise method is used to make regression analysis. In the resulted model, the variable popn is entered while the others are removed, shown in Table 1-1, 1-2, 1-3.

Table.1-1 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.821	.674	.662	64993.3361	.674	55.866	1	27	.000

Table.1-2 Anov

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	235986722098.247	1	235986722098.247	55.	.000
	Residual	114051611012.021	27	4224133741.186		
	Total	350038333110.268	28			

Table.1-3 Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
			Std. Error	Beta		
1	(Constant)	63354.055	16854.793		3.759	.001
	popn	5.612	.751	.821	7.474	.000

The evaluation statistics show the close relation between the gdp and popn. By removing the variable of popn, the other variables emerge as the predicting variables for the gdp. The univariate linear regression results are shown in Table 2. As shown in Table 2, the R squares and adjusted R squares decrease, meaning the decreasing capability of explaining the growth of gdp by using these indices. The absolute values of standard coefficients also decrease, meaning that the influences from these indices decrease.

Table.2 Univariate Linear Regression Model

	R Square	Adj. R Square	Stand. Coef.	t	Sig.
popn	.674	.662	.821	7.474	.000
rpen	.674	.662	.821	7.469	.000
pop_edu_jco	.603	.589	-.777	-6.407	.000
pop_age_forty	.583	.568	.764	6.144	.000
pop_age_thirty	.537	.519	.733	5.592	.000
web	.530	.512	.728	5.515	.000
pop_edu_col	.475	.456	-.689	-4.942	.000
pop_edu_shs	.303	.277	.550	3.423	.002
ndomain_cn	.297	.271	.545	3.381	.002
pop_inc_low	.258	.231	-.508	-3.066	.005

Table.3 Multivariate Linear Regression Model

	R Square	Adj. R Square	Stand. Coef.	t	Sig.
pop_age_forty	.665	.639	.671	5.621	.000
pop_inc_low			-.301	-2.525	.018
pop_age_thirty	.637	.610	.641	5.215	.000
pop_inc_low			-.331	-2.689	.012
pop_edu_col	.691	.667	-2.474	-5.708	.000
pop_edu_shs			-1.844	-4.255	.000
pop_edu_shs	.511	.473	.505	3.664	.001
pop_inc_low			-.459	-3.327	.003

Table 3 shows the results for multivariate regression where the predicting variables are listed and the gdp is the dependent variable. The first model in Table 3 shows the relationship between the gdp and the age structure and income structure in the netizens. Besides, the models listed in Table 2 and Table 3 are evaluated by using augmented Engle-Granger test (AEG) for cointegration. The predicting variable rpen shows the best statistics (AEG= -2.8989) in Table 2. The variable group of pop_age_forty and pop_inc_low shows the best result (AEG= -3.2535) in Table 3. The corresponding critical values are -2.90 and -3.36 at the significance level of 0.1. That is, the regression results cannot pass the AEG test, meaning that other models should be studied in the future work.

But the regression models can still give us some explanation about China's economy. The increasing population of netizens and Internet penetration rate shows a positive effect on the GDP. For the education structure represented by population percentages of senior high school, junior college, college and above in the netizens, the senior high school percentage showing a positive effect while the junior college and the above showing negative effects. The age structure analysis shows that the percentage of the netizens in the forties' and thirties' show positive effects. The numbers of the websites and domain names show positive effects. The low income percentage shows a negative effect. Although the above explanations are not accurate, further research can be conducted on these interesting points.

4. Provincial Analysis

The data related to the 31 provinces of China are analyzed, including the numbers of IPv4 addresses (ipv4), domain names (ndomain), domain names with the ending of .cn (ndomain_cn), the web sites (web), the web pages (webpage), the population of netizens (popn), total data volume of web pages (page_data_vol) and Internet penetration rate (rpen) in each province. These Internet-related data are independent variables while the provincial GDP is the dependent variable. Similar with the national analysis, the following linear regression models can be obtained, shown in Table 4. The population of the citizens is the significant index that can predict the provincial GDP.

Table.4 Univariate Linear Regression Model
(Provincial Level-31)

	R Square	Adj. R Square	Stand. Coef.	T	Sig.
Popn	.914	.911	.956	17.550	.000
Ndomain	.456	.438	.676	4.934	.000
ndomain_cn	.395	.374	.628	4.350	.000
Web	.390	.369	.625	4.307	.000
Rpen	.151	.122	.388	2.270	.031
ipv4	.148	.119	.385	2.248	.032
Webpage	.106	.075	.325	1.851	.074
page_data_vol	.089	.057	.298	1.682	.103

Table.5 Univariate Linear Regression Model
(Provincial Level - 27)

	R Square	Adj. R Square	Stand. Coef.	T	Sig.
popn	.931	.928	.965	18.301	.000
ipv4	.930	.928	.965	18.273	.000
page_data_vol	.671	.657	.819	7.134	.000
web	.661	.648	.813	6.983	.000
ndomain_cn	.636	.622	.798	6.614	.000
ndomain	.625	.610	.790	6.451	.000
webpage	.616	.600	.785	6.329	.000
rpen	.062	.025	.249	1.287	.210

Furthermore, factor analysis is made to the provincial Internet-related data. The 31 provinces are divided into three groups: Beijing forms Group 2 by itself; three provinces, including Guangdong, Shanghai and Zhejiang, form Group 3; the other 27 provinces form Group 1. These three groups represent the extent to which the Internet-related indices are high, i.e., the better or less developed Internet penetration. For the limitation of sample scale, the 27 provinces of Group1 are further analyzed. The regression results are shown in Table 5. Obviously, the Adjusted R squares are larger than the former results in the analysis of 31 provinces in Table 4. Meanwhile standard coefficients are also larger, which shows that the Internet has greater influences on these less developed provinces.

Similarly, the models listed in Table 4 are evaluated by AEG test. The predicting variable popn gives the statistics (AEG= -5.0948) while the corresponding critical value is -4.12 at the significance level of 0.01. That is, the population of netizens can explain the GDP well at the provincial level.

5. Conclusion

This paper attempts to make study of e-business economy in China by using some Internet-related statistical data or indices and to show the importance of e-business to China's economic development. Since the sample scale is not large enough and most of the selected data are collinear, the results of factor analysis are not well. Thus the linear regression is employed to make national and provincial analysis.

Some obvious conclusions can be made. At the national level, the population of netizens predicts the GDP well and then is the Internet penetration rate. One percent increase in these variables results in about 0.8 percent of increase in GDP. The other Internet-related data also show great influences on the GDP positively or negatively. Regretfully, the results for cointegration test fail to reject spurious regression at the national level. But since the increasing proportion of e-business trading in the GDP, the closed relatedness between these Internet-related indices and e-business can obviously explain their effects on the GDP. Further research should be conducted to provide more evidence.

At the provincial level, similar Internet-related indices show the significant contribution to the provincial GDP except for the extent and the priority are a little different. Furthermore, the 27 provinces with less developed Internet penetration are studied. The analysis results show a better model fit and greater influences that these variables have on the provincial GDP.

In conclusion this research empirically shows the importance of e-business economy in China from the perspective of Internet penetration that can be reflected by some statistical Internet related data or indices. Actually this paper is just a launch for further study of e-business in China. The future work can cover a variety of perspectives, such as further study about the influences of the age structure, education structure, and income structure on the e-business and GDP. Besides, direct data about e-business trading should be collected and analyzed. Provincial level analysis such as credit risk distribution and evaluation can be conducted.

6. Acknowledgements

This work was supported in part by the Humanities and Social Science Fund of Shanxi University (No.1009011). The author can be contacted by libingsxu@sxu.edu.cn.

7. References

- [1] Ministry of Commerce of China (MOC), "China's E-commerce 2008-2009", Tsinghua University Press, 2010.
- [2] iResearch Consulting Group.<http://www.iresearch.com.cn/>
- [3] The 29th statistical report on Internet development in China, China Internet Network Information Center (CNNIC), 2012.
- [4] The 1st ~28th statistical report on Internet development in China. CNNIC, 1997~ 2011.
- [5] China Internet Network Information Center (CNNIC). <http://www.cnnic.cn/>

Appendix

The following tables contain the Internet-related statistical data used for national analysis in this paper. They are collected and processed mostly from the semi-annual reports published by CNNIC.

Table.6-1 Internet-related Indices in China from 1997-2011

OBV	ndomain_cn (10, 000)	web (10, 000)	rpen	popn (10, 000)	GDP (100, 000, 000)
1997-10	0. 41	0. 15	0. 05%	62. 0	78973. 0
1998-06	0. 94	0. 37	0. 09%	117. 5	37222. 7
1998-12	1. 84	0. 53	0. 17%	210. 0	84402. 3
1999-06	2. 90	0. 99	0. 32%	400. 0	39554. 9
1999-12	4. 87	1. 52	0. 71%	890. 0	89677. 1
2000-06	9. 97	2. 73	1. 34%	1690. 0	43748. 2
2000-12	12. 21	26. 54	1. 78%	2250. 0	99214. 6

2001-06	12.84	24.27	2.08%	2650.0	48950.9
2001-12	12.73	27.71	2.64%	3370.0	109655.2
2002-06	12.61	29.32	3.58%	4580.0	53341.0
2002-12	17.95	37.16	4.60%	5910.0	120332.7
2003-06	25.07	47.39	5.30%	6800.0	59868.9
2003-12	34.00	59.56	6.20%	7950.0	135822.8
2004-06	38.22	62.66	6.72%	8700.0	70405.9
2004-12	43.20	66.89	7.20%	9400.0	159878.3
2005-06	62.25	68.00	7.90%	10300.0	81913.0
2005-12	110.00	69.00	8.50%	11100.0	184937.4
2006-06	119.06	79.00	9.40%	12300.0	95429.0
2006-12	180.00	84.00	10.50%	13700.0	216314.4
2007-06	615.00	131.00	12.30%	16200.0	115999.0
2007-12	900.00	150.00	16.00%	21000.0	265810.3
2008-06	1190.00	192.00	19.10%	25300.0	140478.0
2008-12	1357.00	288.00	22.60%	29800.0	314045.4
2009-06	1296.00	306.00	25.50%	33800.0	148204.0
2009-12	1346.00	323.00	28.90%	38400.0	340902.8
2010-06	725.00	279.00	31.80%	42000.0	174878.8
2010-12	435.00	191.00	34.30%	45730.0	401512.8
2011-06	350.00	183.00	36.20%	48500.0	204459.0
2011-12	353.00	230.00	38.30%	51300.0	471563.7

Data Source: CNNIC, The 1st ~ 29th statistical report on Internet development in China.
GDP data from National Bureau of Statistics of China (NBS).

Table.6-2 The Netizens' Age/Education/Income Structure in China from 1997-2011

OBV	pop_age_ thirty	pop_age_ forty	pop_age_ young	pop_edu_ shs	pop_edu_ jco	pop_edu_ col	pop_inc_ low	pop_inc_ mid
1997-10	17.5%	6.8%	70.90%	10.0%	30.0%	59.0%	65.0%	32.0%
1998-06	14.9%	3.5%	80.40%	10.0%	30.0%	58.9%	49.6%	32.4%
1998-12	16.2%	4.0%	78.50%	10.0%	30.0%	59.0%	42.0%	33.0%
1999-06	16.5%	4.4%	77.50%	12.0%	27.0%	59.0%	50.0%	28.0%
1999-12	15.9%	4.5%	78.0%	13.0%	32.0%	52.0%	36.0%	36.0%
2000-06	15.6%	5.1%	77.6%	12.8%	32.8%	51.9%	36.5%	37.8%
2000-12	16.0%	5.7%	74.9%	23.5%	29.0%	41.1%	57.6%	23.4%
2001-06	20.1%	8.0%	68.0%	28.8%	26.7%	35.8%	58.8%	25.3%
2001-12	20.3%	7.6%	67.8%	30.0%	26.9%	32.9%	65.4%	13.2%
2002-06	18.8%	6.6%	70.4%	30.5%	26.3%	31.7%	61.9%	24.7%
2002-12	17.6%	6.8%	71.9%	30.6%	26.1%	30.4%	60.9%	23.6%
2003-06	17.7%	6.0%	73.4%	30.9%	27.1%	28.1%	58.6%	25.7%
2003-12	19.7%	6.4%	70.1%	29.3%	27.4%	29.8%	49.8%	27.8%
2004-06	18.8%	6.7%	70.5%	30.6%	26.0%	30.8%	55.3%	22.8%
2004-12	19.0%	7.6%	69.4%	29.3%	27.0%	30.7%	53.2%	27.4%
2005-06	17.7%	7.4%	70.9%	31.3%	25.6%	28.9%	53.4%	26.7%
2005-12	18.7%	6.8%	71.0%	30.2%	24.4%	29.2%	45.7%	25.2%
2006-06	17.6%	7.0%	72.2%	31.6%	23.0%	27.6%	50.2%	27.8%
2006-12	18.6%	6.2%	76.0%	31.1%	23.3%	28.5%	47.6%	24.8%
2007-06	18.5%	7.2%	70.6%	34.2%	20.1%	23.8%	51.7%	25.3%
2007-12	20.5%	8.1%	66.7%	36.0%	18.7%	17.5%	45.3%	28.7%
2008-06	19.7%	7.8%	68.6%	39.0%	15.9%	15.3%	43.7%	23.8%
2008-12	17.6%	9.6%	67.1%	39.4%	13.9%	13.2%	43.7%	29.8%
2009-06	20.7%	9.9%	63.7%	41.0%	12.7%	12.4%	44.2%	28.2%
2009-12	21.5%	10.7%	61.5%	40.2%	12.2%	12.1%	42.5%	27.1%
2010-06	22.8%	11.3%	59.1%	40.1%	12.0%	11.3%	42.1%	28.3%
2010-12	23.4%	12.6%	57.5%	35.7%	11.8%	11.4%	39.1%	27.7%
2011-06	23.2%	11.6%	42.0%	33.9%	10.5%	11.7%	39.2%	23.7%
2011-12	25.7%	11.4%	57.4%	33.3%	10.5%	11.9%	37.9%	22.0%

Data Source: CNNIC, The 1st ~ 29th statistical report on Internet development in China.