Exploring Internet Penetration Effects on China's Economy

Bing Li, Xiaoxia Zhang, and Haiyan Gao

***Abstract*—China's economy has emerged as one of the largest economies in the world. Meanwhile, e-business is playing an increasingly important role in China's economic system. This paper attempts to evaluate the e-business's impact on China's economy by empirically analyzing the relationship between China's GDP and some statistical data related to Internet penetration. National and provincial analyses are made and the results support that China's economy is influenced by Internet penetration.**

***Index Terms*—E-business, internet penetration, regression, factor analysis.**

I. Introduction

The emerging information and communication technologies are continuously changing the world. As one of the greatest innovations around the world in the 20th century, Internet continues to reform the people's work and life styles from the social, economic and political perspectives. Also with the Internet penetration, e-business economy has amounted for a great proportion of national economies, not only for developed countries but also for the developing countries. The e-business trading volume of China has reached 3,800 billion RMB, amounting for about 11.1% of China's GDP in 2009 [1]. From 2007 to 2010, China's B2B trading volume has increased with the average percentage 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 and the penetration rate is 38.3% with a yearly increase of 4% [3].

The above data about the China's economy and the Internet penetration give us direct perception of the rapid development in China's e-business. Thus it is meaningful to make an empirical study on China's e-business and economy. But from the literature review, only a little research has covered this topic. Yang et.al employed the Douglas Function method to study the relationship between Electronic commerce and economic growth [4]. For making comprehensive analyses at the national and provincial level, no work can be found. Accordingly, this paper attempts to study China's e-business economy from the perspective of Internet penetration. By employing the national and provincial statistical data, this paper will explore Internet penetration effects on China's economy.

Manuscript received October 10, 2012; revised November 3, 2012.

B. Li is with the School of Economics and Business Administration, Shanxi University, Taiyuan, CO 030006, P.R.China (e-mail: libingsxu@sxu.edu.cn).

X. X. Zhang and H. Gao are also with Shanxi University.

II. 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 longitudinal data, the time range is from 1997 to 2011 at the national level [3], [5]. The semi-annual data under analysis include some numbers or percentages listed in Table 1. For example, the variables of "ndomain_cn" and "web" reflect some infrastructural features of e-business while the "popn" and "rpen" reflect the direct Internet penetration. Furthermore, age structure, education structure and income structure can also be reflected by the corresponding variables in the data. Briefly, these data are called Internet penetration indices in this paper. On the other hand, the economy 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, the direct relationship between GDP and e-business can be studied by empirically analyzing the yearly GDP and the above mentioned Internet penetration indices.

TABLE I: The Data for National Analysis

Variable	Explanation
ndomain_cn	the number of the domain names with the ending of .cn
web	the number of the web sites
popn	the population of the netizens
rpen	Internet penetration rate, i.e., the netizens' proportion among the whole population
pop_age_thirty	the proportion of the netizens in their thirties'
pop_age_forty	the proportion of the netizens in their forties'
pop_age_young	the proportion of the netizens below their thirties'
pop_edu_shs	the proportion of the netizens who have received the education of senior high school
pop_edu_jco	the proportion of the netizens who have received the education of junior college
pop_edu_col	the proportion of the netizens who have received the education of college and above
pop_inc_low	the proportion of the netizens with the low level income
pop_inc_mid	the proportion of the netizens with the middle level income

DOI: 10.7763/IJTEF.2012.V3.236

TABLE II: THE DATA FOR PROVINCIAL ANALYSIS

Variable	Explanation
ipv4	the number of the IPv4 addresses
ndomain	the number of the domain names
ndomain_cn	the number of the domain names with the ending of .cn
web	the number of the web sites
webpage	the number of the web pages
page_data_vol	the number of the data volume of all the web pages
popn	the population of the netizens
rpen	Internet penetration rate, i.e., the netizens' proportion among the whole provincial population

At the provincial level, some Internet penetration indices can also be obtained, which can reflect the infrastructural features and direct Internet penetration, as shown as in Table 2. The used data are the cross section data at the end of the year of 2011 [3] and 31 provincial areas are selected, covering most parts of China. The provincial GDP is used to indicate the economic yield in each province.

From the above selected data in Table 1 and Table 2, the research is conducted in the statistical way by employing regression analysis and factor analysis. As follows, the results are separately shown and discussed at the national and provincial levels.

III. NATIONAL ANALYSIS

The data listed in Table I work as input for regression analysis in SPSS. The independent variables include twelve Internet penetration indices; the dependent variable is gdp. The stepwise method is used to make regression analysis. In the resulted model, the variable of "popn" is entered while the others are removed. 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 III. As shown in Table III, the R squares and adjusted R squares decrease, meaning the decreasing capability of predicting the gdp by using these indices. The absolute values of standard coefficients also decrease, meaning that the influences from these indices decrease.

Table IV shows the results for multivariate regression where the predicting variables are listed and the gdp is the dependent variable. The first model in Table IV shows the relationship between "pop_age_forty", "pop_inc_low" and gdp, reflecting the effects of the age structure and the income structure on the GDP. Besides, the models listed in Table III and Table IV are evaluated by using augmented Engle-Granger test (AEG) for cointegration [6]. The predicting variable of "rpen" shows the best statistics (AEG = -2.8989) in Table III. The variable group of "pop_age_forty" and "pop_inc_low" shows the best result

TABLE III: UNIVARIATE LINEAR REGRESSION MODEL (NATIONAL)

	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 IV: MULTIVARIATE LINEAR REGRESSION MODEL(NATIONAL)

	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 V: KMO AND BARTLETT'S TEST (NATIONAL)

KMO Measure of Sampling Adequacy.		0.794
Bartlett's Test of Sphericity	Approx. Chi-Square	806.659
	df	66
	Sig.	0.000

(AEG = -3.2535) in Table IV. The corresponding critical values are -2.90 and -3.36 at the significance level of 0.10. That is, the regression results cannot pass the AEG test, meaning that the regression models only have weak explanation capabilities for the gdp.

But the regression models can still show some points about China's economy. The population of netizens and Internet penetration rate show positive effects 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 shows a positive effect while the junior college and the above show negative effects. The age structure analysis shows that the percentages of the netizens in their forties' or 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.

Furthermore, the twelve Internet penetration indices are studied by factor analysis. Table V gives the results for KMO and Bartlett's Test. KMO is 0.79 and the Bartlett's test is significant at the level of 0.01, meaning that these indices are suitable for factor analysis.

TABLE VI: TOTAL VARIANCE EXPLAINED (FACTOR ANALYSIS - NATIONAL)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.533	71.105	71.105	8.533	71.105	71.105	8.509	70.912	70.912
2	1.795	14.955	86.06	1.795	14.955	86.06	1.818	15.148	86.06
3	0.878	7.317	93.377						
4	0.305	2.54	95.916						
5	0.223	1.855	97.771						
6	0.108	0.9	98.671						
7	0.085	0.711	99.382						
8	0.038	0.32	99.702						
9	0.023	0.19	99.892						
10	0.012	0.099	99.991						
11	0.001	0.009	100						
12	7.72E-06	6.43E-05	100						

*Extraction Method: Principal Component Analysis.

TABLE VII: ROTATED COMPONENT MATRIX (NATIONAL)

	Component	
	F1	F2
ndomain_cn	0.722	0.34
web	0.917	0.253
rpen	0.959	0.207
popn	0.957	0.209
pop_age_thirty	0.884	-0.167
pop_age_forty	0.951	-0.04
pop_age_young	-0.878	0.093
pop_edu_shs	0.856	-0.261
pop_edu_jco	-0.973	-0.132
pop_edu_col	-0.94	0.129
pop_inc_low	-0.392	-0.821
pop_inc__mid	-0.372	0.858

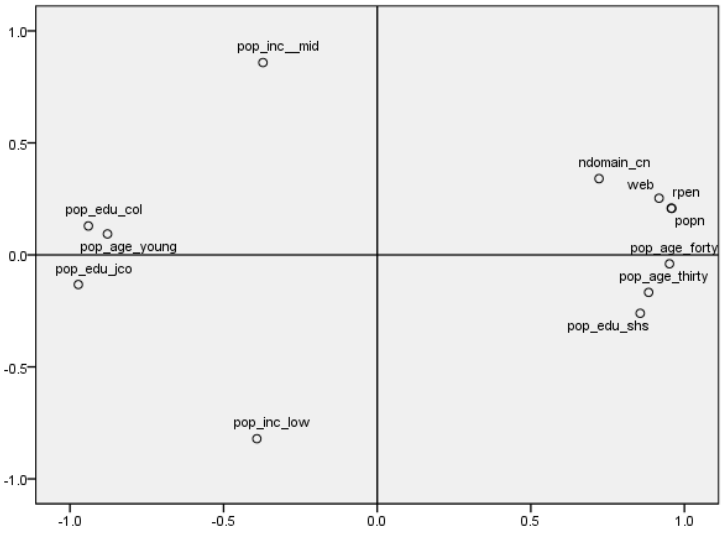

Fig. 1. Loading graph in rotated place (National)

Table VI shows that two eigenvalues are greater than I, meaning two significant latent factors exist. Table VII gives the rotated component matrix and Fig.1. shows the loading graph of the twelve indices. From Table 6, Table 7 and Fig.1, it can be shown that the two significant latent factors accumulatively explain the 86.1% of the total variance, ten of the indices have greater loading on the horizontal axis – Factor 1, which can explain 70.9% of the total variance; the other two indices, that is "pop_inc_mid" and "pop_inc_low", project more on the vertical axis – Factor 2, which can explain 15.1% of the total variance. Factor 2 is related to the income structure of the netizens and "pop_inc_low" has a negative score on this factor. This is similar with the negative effect that the "pop_inc_low" has on the gdp in Table 3 and 4. Factor 1 combines many indices, such as the age structure indices, the education structure indices and the direct Internet penetration indices. This combination actually makes it difficult to clarify this factor.

IV. PROVINCIAL ANALYSIS

The data related to the 31 provinces of China are analyzed as listed in Table 2. By contrast with national data, these provincial ones only cove the indices that reflect the infrastructural features such as "ipv4", "ndomain", and the direct Internet penetration indices such as "popn" and "rpen". These indices are considered as 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 8. The population of the citizens is the most significant index that can predict the provincial gdp.

Furthermore, cluster analysis is made to the provincial indices. 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 penetration indices are high, i.e., the better or less developed Internet penetration. For the limitation of sample scale, the 27 provinces of Group 1 are further analyzed. The regression results are shown in Table 9. Obviously, most of the Adjusted R squares are larger than the former results in the analysis of 31 provinces in Table 8. Meanwhile, standard coefficients are also larger, showing that the Internet has greater influences on these less developed provinces.

TABLE VIII: UNIVARIATE LINEAR REGRESSION MODEL (PROVINCIAL)

	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 IX: UNIVARIATE LINEAR REGRESSION MODEL (PROVINCIAL-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

TABLE X: TOTAL VARIANCE EXPLAINED (FACTOR ANALYSIS – PROVINCIAL)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.172	77.148	77.148	6.172	77.148	77.148	4.318	53.974	53.974
2	1.15	14.372	91.52	1.15	14.372	91.52	3.004	37.546	91.52
3	0.452	5.651	97.171						
4	0.17	2.119	99.291						
5	0.035	0.441	99.731						
6	0.02	0.249	99.981						
7	0.001	0.016	99.997						
8	0.00E+00	3.00E-03	100						

*Extraction Method: Principal Component Analysis.

TABLE XI: KMO AND BARTLETT'S TEST (PROVINCIAL)

KMO		0.667
Bartlett	Approx. Chi-Square	604.554
	df	28
	Sig.	0.000

TABLE XII: ROTATED COMPONENT MATRIX (PROVINCIAL)

	F1	F2
page_data_vol	.969	.206
webpage	.961	.241
ipv4	.926	.277
web	.710	.690
rpen	.665	.467
popn	.039	.940
ndomain_cn	.532	.796
ndomain	.607	.783

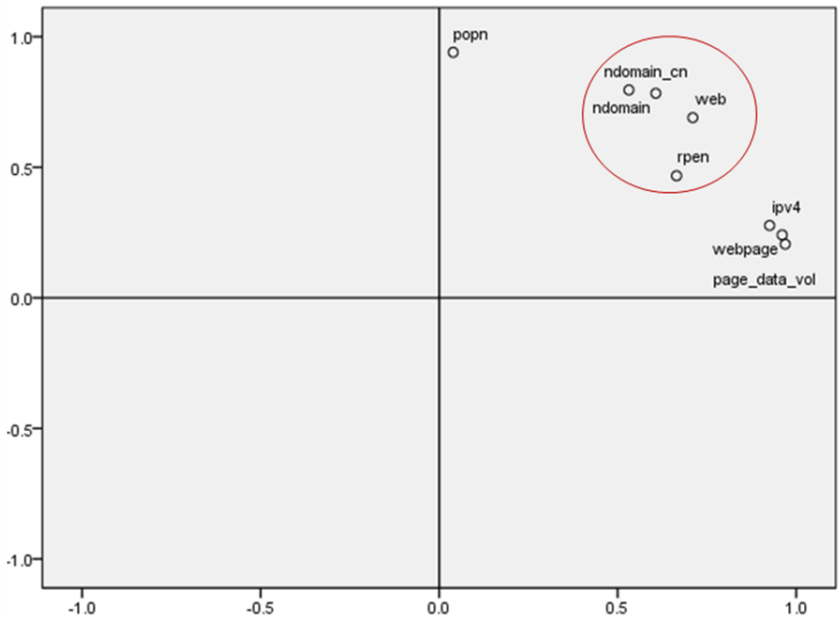

Fig. 2. Component Plot in Rotated Place (Provincial)

Similarly, the models listed in Table VIII 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 AEG test is passed and the population of netizens can explain the GDP well at the provincial level.

The factor analysis is also made for these provincial data. KMO is 0.667 and the Bartlett's test is significant at the level of 0.01 as shown in Table XI, meaning that these provincial indices are still basically suitable for factor analysis. Two factors emerge as significant ones that can accumulatively explain 91.5% of the total variance as shown in Table X. As shown in Fig.2, the loading graph illustrates the provincial indices' loading on the two significant factors (horizontal and vertical axes). Since several indices fall in the circular area, it is difficult to clarify the explanation of the two factors. But from the indices that are near the axes, Factor 1 is more related to infrastructural features while Factor 2 is more related to the population of the provincial netizens. Table XII is the loading scores of the indices on the two factors.

Table XIII, Table XIV, Table XV show the results for regression on the two significant factors emerged in factor

analysis. Model 1 is regression for the 31 provinces and Model 2 is for the 27 provinces in Group 1. As mentioned above, this group is the less developed provincial areas in Internet penetration. Table XIII and XIV show that the regression model for 27 provinces is more fit since Model 2 has a larger R square and F value.

TABLE XIII: MODEL SUMMARY FOR REGRESSION ON SIGNIFICANT FACTORS (PROVINCIAL)

Model	R	R Square	Adj. R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df 1	df 2	Sig. F Change	
1	0.866	0.750	0.733	6748.538	0.750	42.078	2	28	0.000	1.900
2	0.925	0.855	0.843	4576.263	0.855	70.854	2	24	0.000	1.897

* Predictors: (Constant), Factor 1, Factor 2 * Dependent Variable: gdp

TABLE XIV: ANOV (PROVINCIAL)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3832697222.933	2	1916348611.467	42.078	.000
	Residual	1275197446.768	28	45542765.956		
	Total	5107894669.702	30			
2	Regression	2967663371.544	2	1483831685.772	70.854	.000
	Residual	502612434.917	24	20942184.788		
	Total	3470275806.461	26			

As shown in Table XV, Factor 2 is significant at the level of 0.01. Factor 2 has a positive effect on the provincial gdp and this effect is greater for the less developed provinces, which is similar with the former results in Table 8 and Table 9. In Fig.2, the variable of "popn" shows greater loading on the vertical axis – Factor 2, that is, to increase the population of the netizons can help to increase the provincial gdp and this effect is greater for the less developed provinces. Thus, one of the meaningful public policies is to popularize the Internet technology and usage for the provincial governments.

The coefficient of Factor 1 is not significant. Besides, the effects of Factor 1 on the provincial gdp are different: in Model 1, Factor 1 has a positive effect; in Model 2, Factor 1 has a negative effect.

TABLE XV: COEFFICIENTS (PROVINCIAL)

Model		Unstd. Coef.		Std. Coef.	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16656.69742	1212.073217		13.742	0.000
	Factor 1	1060.735449	1232.108842	0.081291801	0.861	0.397
	Factor 2	11253.06245	1232.108842	0.86240326	9.133	0.000
2	(Constant)	18596.49517	1679.988164		11.069	0.000
	Factor 1	-1652.18823	5715.14785	-0.02268201	-0.289	0.775
	Factor 2	20167.32809	1705.71932	0.927662895	11.823	0.000

* Dependent Variable: gdp

TABLE A1: .INTERNET PENETRATION 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 A2: 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.

TABLE B: THE CHINA'S PROVINCIAL LEVEL INTERNET PENETRATION INDICES IN 2011

Province	IPv4 adress (10,000)	domain	cn_domain	web	webpage (10,000)	popn (10,000)	rpen	page_data _vol (GB)	GDP (￥100,000,000)
Anhui	561.7	93,898	36,098	25,805	124759	1585	0.266	41458	15110.3
Beijing	8459.3	1,061,328	471,979	384,881	2993042	1379	0.703	1172335	16000.4
Fujian	660.9	528,072	156,841	151,096	352463	2102	0.570	111167	17500.0
Gansu	165.2	16,104	6,310	4,505	11769	700	0.274	3829	5020.0
Guangdong	3172.2	1,401,965	783,362	383,928	880169	6300	0.604	306384	52673.6
Guangxi	462.6	50,518	21,159	13,342	61852	1353	0.294	27706	11714.0
Guizhou	132.2	24,068	8,385	6,071	11380	840	0.242	3544	5600.0
Hainan	165.2	35,825	9,794	9,987	56039	338	0.389	21845	2515.3
Hebei	958.3	216,685	46,473	65,749	218264	2597	0.361	78306	24228.2
Henan	892.2	197,583	58,669	63,128	480936	2582	0.275	158617	27000.0
Heilongjiang	396.5	67,775	35,743	16,668	41509	1206	0.315	15553	12503.8
Hubei	793.1	141,973	59,823	51,506	173466	2129	0.372	55904	19594.2
Hunan	793.1	124,135	52,586	39,855	104088	1936	0.295	32205	19635.2
Jilin	396.5	51,012	13,859	16,336	14408	966	0.352	5143	10400.0
Jiangsu	1586.1	406,578	146,310	126,298	522445	3685	0.468	180173	48000.0
Jiangxi	594.8	61,083	20,728	16,032	136356	1088	0.244	43152	11583.0
Liaoning	1123.5	142,505	44,869	47,744	97184	2092	0.478	35126	22025.9
Neimenggu	264.4	29,052	9,609	10,015	9840	854	0.346	3495	14000.0
Ningxia	66.1	13,668	5,099	3,074	3414	207	0.328	1170	2060.0
Qinghai	66.1	11,251	1,722	1,754	1317	208	0.369	457	1622.0
Shandong	1619.2	383,059	100,639	109,402	254667	3625	0.378	110437	45000.0
Shanxi	429.6	56,625	16,455	17,316	43595	1405	0.393	12464	11000.0
Shaanxi	561.7	93,282	31,002	30,041	75856	1429	0.383	28002	12391.3
Shanghai	1487.0	681,291	238,773	237,680	690155	1525	0.662	249834	19195.7
Sichuan	925.2	236,557	51,931	71,724	145321	2229	0.277	45748	21026.7
Tianjin	363.5	83,414	26,418	26,362	253232	719	0.556	94625	11191.0
Xizang	33.0	3,887	983	762	478	90	0.299	255	605.0
Xinjiang	198.3	24,541	8,266	4,263	15353	882	0.404	4567	6600.0
Yunnan	330.4	41,387	17,082	9,907	19460	1140	0.248	6299	8751.0
Zhejiang	1751.3	874,559	571,111	216,855	818251	3052	0.561	294707	31800.0
Chongqing	561.7	91,217	30,327	24,854	47160	1068	0.370	15523	10011.1

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

V. CONCLUSION

This paper attempts to make an empirical study of e-business economy in China by using Internet penetration indices and to show the importance of e-business to China's economic development. From the statistical data selected from the semi-annual reports published by CNNIC, the national and provincial analyses are carried out by using the tools of SPSS and Eviews. This paper can be concluded as follows.

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 penetration indices 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 relationship between these Internet penetration indices and e-business can obviously explain their effects on the GDP. Further research should be conducted to provide more evidence. For factor analysis, two significant factors emerge: Factor 1 combines the infrastructural indices, the age structural ones, and the education structural ones; Factor 2 combines the income structural indices.

At the provincial level, only the infrastructural indices and the direct Internet penetration indices can be obtained and studied. The analysis results show their significant contribution to the provincial GDP. Furthermore, cluster analysis is carried on the 31 provinces and the 27 of them form a group with less developed Internet penetration. By analyzing this group, the results show better model fit and greater effects that these indices have on the provincial GDP.

In conclusion this research empirically shows the relationship between China's economy and Internet penetration indices. From this perspective, it can also be seen the great importance of e-business economy in China that deserves more efforts and further study. Future research work can be done in several ways. Since e-business economy is a new economic formation, statistical data related to e-business can only form a short time series without enough observations, especially for the case of China. Thus one way is to study the China's e-business by using panel data analysis. A second way is to further study the structural influences on e-business, such as age, education and income.

APPENDIX

The following appendices contain the Internet penetration indices used for national and provincial analyses in this paper. They are collected and processed mostly from the semi-annual reports published by CNNIC. These data are provided for further reference and research use.

REFERENCES

[1] Ministry of Commerce of China (MOC), *China's E-commerce 2008-2009*, Tsinghua University Press, 2010.
[2] iResearch Consulting Group. [Online]. Available: http://www.iresearch.com.cn/
[3] The 29th statistical report on Internet development in China, China Internet Network Information Center (CNNIC), 2012.
[4] J. Yang, T. Zhou, and Q. Li, "An Empirical Analysis of Electronic Commerce on Economic growth," *World Economy Study*, no. 212, pp. 40-43, 2011(10).
[5] The 1st ～28th statistical report on Internet development in China, CNNIC, 1997～ 2011.
[6] R. F. Engle and B. S. Yoo, "Forecasting and Testing Co-integrated Systems," *Journal of Econometrics*, vol. 35, pp. 143-159, 1987.
[7] China Internet Network Information Center (CNNIC). [Online]. Available: http://www.cnnic.cn/

Bing Li was born in China in 1977. He got his doctor degree (2008) from the Department of Industrial Engineering and Management in Tokyo Institute of Technology, Japan. His master degree (2003) and bachelor degree (2000) was from Xi'an Jiaotong University, China.

He has some work experiences related to finance and information systems development in Japan from 2008 to 2010. He worked as a research student in Fuji Xerox Visiting Fellowship Program (2001-2002). Now he is a lecturer in the School of Economics and Business Administration, Shanxi University, China from September of 2010. His research interests include information systems and e-business, financial econometrics, organizational culture. He received the best paper award in the IFIP International Conference on Research and Practical Issues of Enterprise Information Systems (2007).

Xiaoxia Zhang is a Ph.D. candidate in the Graduate School of Economics in Rikkyo University, Japan. She is also a lecturer in the School of Foreign Languages, Shanxi University.

Haiyan Gao is now an undergraduate student in the School of Economics and Business Administration, Shanxi University.