The Determinants of Internet Penetration:

A Cross-city Analysis of China

Gong Chen, Xiding Chen, and Fangfang Zhang [+]

Abstract. To identify the determinants of cross-city disparities in Internet penetration, we examine a panel of 343 cities of China over the 2001-2007 period. Our strongest findings are that Internet use is associated with high levels of human capital, telephone density, and electricity consumption. We also find considerable evidence that Internet use is enhanced by high levels of income per capita. There is also some evidence for a negative role of the share of agriculture and the share of manufacturing in GDP. Our results suggest that public investment in human capital, telecommunications infrastructure can mitigate the gap in Internet use.

1. Introduction

Information and communication technologies (ICTs) are viewed as the engine of productivity and economy [1]. ICTs could increase knowledge diffusion, improve communication efficiency, decrease transaction cost and allow developing countries to "leapfrog" traditional methods of increasing productivity (Steinmueller, 2003). Many economists also ascribe the acceleration of productivity growth in the United States since 1995 to the greater investments in ICTs (Autor, et al., 1998; Brynjolfsson and Hitt, 2003).

As the striking international differences in information and communication technology (ICT) diffusion[2], China's ICTs experienced high rates of growth accompanied by huge digital divide. The wide disparities are viewed as a serious threat to China's future prosperity (World Bank country report, 2003). Many landlocked west cities of China have Internet penetration rates that are less than two percent of the rates found in east cities. For example, there are less than 15 Internet users per 1000 people in Longnan city of Gansu Province, whereas nearly 900 Internet users per 1000 people in Shenzhen city of Guangdong Province in 2007(China Statistical Yearbook, 2008).

Although these differences in technology diffusion may have substantial economic consequences, the empirical literature aimed at identifying the causes is limited. We remedy this deficiency in the literature with a comprehensive econometric analysis of the determinants of Internet use in China. In this paper, we use macrodata of 343 cities of China over the 2001-2007 period to examine the determinants of Internet use.

[+] School of Economics and Department of Public Economics, Xiamen University, China, 361005 (e-mail: cg@xmu.edu.cn, xidingchen@qq.com, xmuzff@gmail.com)

[1] "All sectors of Europe's economy depend on ICTs. We must continue to invest heavily in research and in bringing innovations to market". Viviane Reding Commissioner for Information Society and Media.

[2] It is often referred to as the "Global Digital Divide".

Our strongest findings are that Internet use is associated with high levels of human capital, telephone density, and electricity consumption. We also find considerable evidence that Internet use is enhanced by high levels of income per capita. There is also some evidence for a negative role of the share of agriculture and the share of manufacturing in GDP.

2. Related Literature

Caselli and Coleman (2001) examine a cross-country dataset encompassing up to 89 countries over the years 1970-1990, focusing on the ratio of computer imports to worker, as a proxy measure for the investment in ICTs. In attempting to explain the variation in this variable, they rely on a large set of variables, including income per worker, investment per worker, the sectoral shares of agriculture, the sectoral shares of manufacturing, human capital, imports and exports from and to the OECD. They also include an institutional variable, in this case an index of property rights. Openness to imports from OECD countries, the level of educational attainment are two notable variables with statistical significance.

Dasgupta et al. (2001) examine Internet use in a sample of 44 countries, spanning both OECD and developing countries, from 1990 to 1997. Their variable of focus is the Internet to telephone mainlines ratio as a dependent variable. They find that differentials in income, human capital, regulatory effectiveness, and telecommunications infrastructure are the most important determinants of the global digital divide.

Kiiski and Pohjola (2001) use the Internet hosts per 1000 inhabitants as the variable of Internet use. In a broad sample encompassing about sixty OECD and developing countries from 1995 to 2000, they find that the five year growth in this variable is related to income per capita, telephone access costs, and the average years of schooling.

Pohjola (2003) studies observed investment in information and communication technology in 49 countries from 1993 to 2000. He regresses ICTs investment per capita on income per capita, the relative price of ICT equipment, human capital measures, the share of agriculture and openness to international trade. He finds that a positive impact of human capital and a negative impact of agriculture share in GDP on ICT diffusion. Chinn and Fairlie (2007) examine a panel of 161 countries over the 1999-2001 period and find that the global digital divide is mainly accounted for by income differentials.

In a related line of research, a few recent studies have explored the determinants of the digital divide within countries. using microdata, researchers find that income and education are important determinants of Internet use (for example, Primrose (2003) for Australia, Ono and Zavodny 2003 for Japan, and Singh (2004) for Canada).

3. Data and Descriptive Statistics

Data on technology use and telecommunications are from China Economic Information Network's (CEIT) Telecommunication Indicators Database. The data are obtained primarily through annual report of China National Bureau of Statistics (CNBS). Supplemental information and data is obtained from China Internet Development Statistics Report (CIDSR). Detailed data are available on telephone, Internet, and

Table 1—Descriptive Statistics

	All cities	High Internet Hosts cities	low Internet hosts cities
	(1)	(2)	(3)
Log Internet hosts per 1000 people	4.09	4.75	3.39
	(1.10)	(0.78)	(0.95)
Log GDP per capita (constant in 2001Yuans)	9.69	10.11	9.24
	(0.74)	(0.59)	(0.59)
Log education	2.95	3.40	2.42
	(1.06)	(0.89)	(0.99)
Log electricity consumption per capita	7.69	8.16	7.22
	(0.92)	(0.66)	(0.91)
Log telephone number per 1000 people	5.75	6.08	5.39
	(0.59)	(0.44)	(0.51)
Log agriculture share in GDP	1.71	1.09	2.35
	(1.05)	(0.84)	(0.85)
Log manufacturing share in GDP	3.87	3.94	3.79
	(0.29)	(0.28)	(0.32)
Log Openess	-4.10	-3.69	-4.59
	(1.39)	(1.26)	(1.39)
Observations	1656	842	814
cities	237	142	126

Notes: Values are averages during sample period, with standard deviations in parentheses. Column 1 refers to the full sample, and columns 2 and 3 split the sample in column 1 by the median Internet hosts in 2007 in the sample of column 1. The number of observations refers to the total number of observations in the unbalanced panel. The number of cities refers to the number of cities for which we use observations. The number of Internet hosts and telephone is from China Economic Information Network. Other variables are from China Statistical Yearbook (2002-2008). Education is the number of higher school students per 1000 people. Openess to foreign countries is the ratio of FDI to GDP.

other telecommunications and electronics use. The primary source for data on demographics, income, FDI, human capital, and other variables included in the regression analysis is from the annual China Statistical Yearbook (2002-2008).

Internet penetration rate analyzed below are derived from the number of Internet hosts per thousand people and spans the period 2001 to 2007. GDP per capita is in constant 2001 Yuans. We choose the number of higher school students per 1000 people as the index of human capital and the ratio of FDI to GDP as the index of openness to foreign countries. Then we construct the base sample comprises 343 cities[3] and spans the period 2001 to 2007 annually.

Table 1 contains descriptive statistics for the main variables. The table shows these statistics for all cities and also for high- and low-Internet hosts cities, split according to median Internet hosts in 2007. In each case,

[3] In China, according to administrative levels, the cities can be divided into Municipalities, Provincial Cities, Prefecture-level City and County-level City four class. In our base sample, there are 4 Municipalities, 15 Provincial Cities, and 324 Prefecture-level City. In order to increase the comparability, all cities' statistical data only cover the municipal districts.

we report means, standard deviations, and also the total number of cities for which we have data and the total number of observations.

4. The Determinants of Internet Penetration

Our strategy to investigate the determinants of differences in Internet penetration is to look at a variety of regression results using specifications of the form:

$$I_{it} = \alpha + \delta_t \beta + X_{it} \gamma + n_i + u_{it} \quad (1)$$

where I_{it} is the log Internet hosts per 1000 people in city i and year t, X_{it} is a set of explanatory variables, δ_t is a set of year dummies, n_i is a city effect, and u_{it} is independently and identically distributed among cities and years. All the variables we will include in the vector X are available at annual frequency.

In panel data estimation of this kind there is considerable controversy regarding the appropriate estimation technique, and in particular regarding the treatment of the city-specific term, n_i. The basic choice is between random effects (RE) and fixed effects (FE). For the purpose of estimating equation (1), the major difference between these two techniques is the information utilized to calculate the coefficients. The fixed effects estimates are calculated from differences within each city across time, but the city dummies absorb a lot of the variation in the data and make the estimator relatively inefficient.; the random effects estimates are more efficient, since they incorporate information across individual cities as well as across periods. The major drawback with random effects is that it is consistent only under the most stringent assumptions (i.e., that city-specific effect n_i is uncorrelated with the vector X_{it}). A Hausman specification test can evaluate whether this independence assumption is satisfied.

We treat the vector X_{it} as exogenous for I_{it}. Reverse causation is unlikely to be a problem. Because Internet penetration is limited between 2001 and 2007, it is unlikely to have caused changes in any of the macroeconomic variables on the right-hand side. For example, it is unrealistic that Internet penetration may have impacted the education. Of course, that reverse causation is not a major concern does not rule out the possibility that we may omit some important explanatory variable and induce bias.

We report representative results in Table 2. All standard errors in the paper are fully robust against arbitrary heteroskedasticity and serial correlation at the city level (i.e., they are clustered at the city level; see Wooldridge 2002). The specification is the same in all six columns, but every two columns use a different sample. Columns 1 and 2 use the full sample; Columns 3 and 4 use the sample of high Internet hosts cities; Columns 5 and 6 use the sample of low Internet hosts cities. These regressions include the following set of explanatory variables: the log of real per capita income; the log of education, as measured by the number of higher school students per 1000 people; the log of electricity consumption per capita; the log of telephone number per 1000 people; the (log) share of agriculture in GDP; the (log) share of manufacturing in GDP; and (log) openness to foreign countries, as measured by the ratio of FDI to GDP. To conserve space, in Table 2 we do not report the coefficients on the year dummies or on the regional dummies. They are both highly significant, as expected.

Table 2—Determinants of Internet Penetration

Independent variable	All cities		High Internet Hosts cities		low Internet hosts cities	
	Random Effects	Fixed Effects	Random Effects	Fixed Effects	Random Effects	Fixed Effects
	(1)	(2)	(3)	(4)	(5)	(6)
Log GDP per capita	0.39***	0.04	0.37***	0.21	0.25**	-0.16
	(0.07)	(0.15)	(0.09)	(0.17)	(0.12)	(0.25)
Log education	0.16***	0.14***	0.10***	0.17**	0.12***	0.11*
	(0.03)	(0.05)	(0.04)	(0.07)	(0.04)	(0.06)
Log electricity consumption per capita	0.11**	0.12*	0.05	-0.002	0.09	0.19*
	(0.04)	(0.06)	(0.06)	(0.082)	(0.06)	(0.10)
Log telephone number per 1000 people	0.51***	0.28***	0.26***	0.08	0.59***	0.48**
	(0.06)	(0.08)	(0.08)	(0.10)	(0.09)	(0.12)
Log agriculture share in GDP	-0.09**	-0.05	-0.03	-0.004	-0.10*	-0.15
	(0.04)	(0.06)	(0.04)	(0.070)	(0.06)	(0.11)
Log manufacturing share in GDP	-0.29***	-0.16	-0.31	-0.10	-0.32**	-0.36
	(0.11)	(0.17)	(0.14)	(0.24)	(0.16)	(0.26)
Log Openess	0.03**	0.01	0.05**	0.04	0.02	-0.01
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
R-squared	0.64	0.41	0.48	0.48	0.47	0.39
Observations	1410	1410	798	798	612	612
Cities	268	268	142	142	126	126

Notes: The dependent variable is the log of Internet hosts per 1000 people from 2001 to 2007. Random effects OLS regression in columns 1, 3, and 5, with robust standard errors clustered by city in parentheses. Fixed effects OLS regressions in columns 2, 4, and 6, with city dummies and robust standard errors clustered by city in parentheses.

Columns 1 and 2 is the full sample; Columns 3 and 4 is the sample of high Internet hosts cities; Columns 5 and 6 is the sample of low Internet hosts cities. Year dummies are included in all regressions. R^2 is the within- R^2 for fixed effects and the overall- R^2 for random effects.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 Percent level.

In the full sample, we test the validity of the assumptions about random-effects estimates and fixed-effects estimates. Hausman specification test comparing the random-effects estimates of column 1 with the fixed-effects estimates of column 2 rejects the assumption required for random effects[4], so we rely primarily on the results of fixed-effects estimates. According to column 2, high levels of education is important determinant of Internet penetration, even after controlling for a variety of other macroeconomic variables, including per capita income. The results of high Internet hosts sample and low Internet hosts sample also confirm the findings that education boosts the Internet penetration.

[4] The test statistic is $\chi^2(12)$ =68.77. This rejects the null hypothesis at any standard level of significance.

Column 2 also shows the conventional infrastructure variables (electricity consumption per capita and telephone lines per 1000 people) play positive and important role in Internet penetration. A one percentage point increase in these variables is associated with a 0.4 (0.12+0.28) percentage point increase in Internet penetration. Telephone lines may be important for users accessing the Internet. And well developed communication infrastructure is also likely to have other unobservable attributes that encourage Internet use.

One surprising result is the importance of income per capita. In column 2, the income per capita has the positive sign, but it isn't significant. We don't think it is a big problem. According to Caselli and Coleman (2001), a fully successful case study of Internet penetration should lead to specifications in which income per capita is not statistically significant. And income per capita continued significance may signal that those additional determinants for which income per capita is a stand-in have not been fully identified. Column 2 also shows that agriculture share in GDP and manufacturing share in GDP are negative with Internet penetration.

In high Internet hosts sample, we also rely primarily on the results of fixed-effects estimates (column 4)[5]. We only find that education significantly boosts the Internet penetration. In low Internet hosts sample, we rely primarily on the results of random-effects estimates (column 5)[6]. Column 5 shows that in low Internet hosts cities education, telephone line, and income per capita are important determinants promoting Internet penetration. One puzzling results is that column 5 shows manufacturing share in GDP could hinder the Internet penetration. One possibility is that China is such a huge manufacturing country[7] that providing more service rather than more manufacturing products could promote Internet penetration. As Caselli and Coleman (2001) and Pohjola (2003), Column 5 also shows high agriculture share in GDP have a negative impact on Internet penetration.

In none of the three samples is there any evidence that openess to foreign countries are important determinants of Internet penetration.

To conserve space, we have not included in the table the coefficients on the city and year dummy variables. Yet such coefficients are of some interest in themselves. We find many coast city dummy variables are significantly positive but many landlocked city dummy variables are significantly negative, which suggests geography position really matters. In terms of year dummy variables, we find 2005 year[8] dummy is significantly positive (in all columns).

5. Conclusions

To identify the determinants of cross-city disparities in Internet penetration, we examine a panel of 343 cities of China over the 2001-2007 period. First, unsurprisingly, we confirm the importance of human capital, telephone density, and electricity consumption in explaining the gap in Internet use. Second, We

---

[5] the test statistic is $\chi^2(12)$ =26.89, which rejects the null hypothesis at any standard level of significance.
[6] the test statistic is $\chi^2(12)$ =17.07, which can't reject the null hypothesis.
[7] In many cities the manufacturing share in GDP over fifty percent.
[8] According to China Internet Development Statistics Report (CIDSR), the number of Internet hosts reaches 100 million in 2005.

find considerable evidence that Internet use is enhanced by high levels of income per capita. But evidently it isn't the dominant factor. Third, surprisingly, we find there is also some evidence for a negative role of the share of manufacturing in GDP.

The conclusions of this paper also have implications for the public policy of China. As illustrated in paper, greater human capital and ICTs infrastructure imply higher Internet penetration. So the public investment in human capital, telecommunications infrastructure can mitigate the gap in Internet use.

6. References

[1] Autor, D., Lawrence Katz, and Alan Krueger, 1998, "Computing inequality: Have computers changed the labor market?" *Quarterly Journal of Economics* 113(4): 1169-1213.

[2] Brynjolfsson, Erik and Loren Hitt, 2003, "Computing productivity: Firm-level evidence," *Review of Economics and Statistics* 85(4): 793-808.

[3] Casseli, Francesco, and Wilbur John Coleman II, 2001, "Cross-country technology diffusion: The case of computers,"*The American Economic Review*, Vol. 91, No. 2, pp. 328-335

[4] Menzie D. Chinn and Robert W. Fairlie, 2007. "The determinants of the global digital divide: a cross-country analysis of computer and internet penetration,"*Oxford Economic Papers*, Oxford University Press, vol. 59(1), pages 16-44, January

[5] Dasgupta, Susmita, Somik Lall, and David Wheeler, 2001, "Policy reform, economic growth and the digital divide: An econometric analysis," World Bank Working Paper No. 2567 (March).

[6] Kiiski, Sampsa, and Matti Pohjola, 2001, "Cross-country diffusion of the internet," WIDER Discussion Paper No. 2001/11 (Helsinki: World Institute for Development Economics Research).

[7] Ono, Hiroshi, and Madeline Zavodny, 2003, "Gender differences in information technology usage: A U.S.-Japan comparison," mimeo (Stockholm School of Economics).

[8] People's Republic of China, State Statistical Bureau, China Statistics Yearbook (Zhongguo Tongji Nianjian) (Beijing: China Statistical Publishing House, annual issues).

[9] Pohjola, Matti, 2003, "The adoption and diffusion of ICT across countries: Patterns and determinants," The New Economy Handbook (Academic Press).

[10] Primrose, Jane. 2003. "2001 Census: Computer and Internet Use," Census Paper No. 03/03, Australian Bureau of Statistics.

[11] Singh, Vik, 2004, "Factors associated with household internet use in Canada, 1998-2000," Agriculture and Rural Working Paper Series No. 66 (Ottawa: Statistics Canada).

[12] Steinmuller, W. Edward, 2001, "ICTs and the possibilities of leapfrogging by developing countries," *International Labour Review*, 140(2): 193-210.

[13] Wooldridge, Jeffrey M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

[14] World Bank. China: Promoting Growth with Equity. Country Economic Memorandum, Poverty Reduction and Economic Management Unit, East Asia and Pacific Region, Report No. 24169CHA, 2003.