

Does Internet Use Improve the Income of Residents? —Empirical Evidence from CGSS2017

Xiaoxiao Peng, Jie Zhang, Gang Peng*

The Internet penetration rate rises sharply in recent years in China. This change has had a significant impact on residents' income. By using Chinese General Social Survey (CGSS) data set, this paper investigates whether residents who use the Internet earn a higher income than similar residents who do not use the Internet by using propensity score matching. The results show that there is a premium associated with Internet use. Estimates suggest that a premium for residents who use the Internet is around twice as much for residents who do not use the Internet. Additionally, this paper finds that the income differences between using the Internet and not using the Internet for groups of middle-aged and elderly resident and agricultural household registration residents are more significant. Based on the research results, several relevant policy implications are presented to improve resident's income.

Keywords: propensity score matching, Internet use, income

1. Introduction

In recent years, the Chinese government issued a series of policies—including those for new infrastructure—committing to improving Internet penetration, and promoting economic growth. Today, China has the largest number of Internet users worldwide and this number continues to increase. According to *The 47th China Statistical Report on Internet Development*, the number of Internet users in China had reached 989 million by the end of 2020, an increase of 85.4 million from March 2020. This translates to an Internet penetration rate of 70.4%, and 55.9% for rural areas. In the era of the digital economy, the Internet has played an extremely important role, affecting almost all aspects of the social economy and people's lives. The introduction and penetration of broadband has a significant role in promoting economic growth in China (Han and Zhu, 2014). For instance, the Internet has promoted the rapid development of e-commerce: numerous online shopping platforms, such as Taobao, Jingdong, and Suning, make purchases faster and more convenient. Further, an increasing number of

* Xiaoxiao Peng, Doctoral Candidate at School of Statistics of Southwestern University of Finance and Economics, China; Jie Zhang, Doctoral Candidate at School of Statistics of Southwestern University of Finance and Economics, China; Gang Peng (email: penggang2016@swufe.edu.cn), Associate Professor at School of Statistics of Southwestern University of Finance and Economics, China.

Chinese people are choosing to use online payment methods.

Nevertheless, now, with China's economic development having reached a new normal, the growth rate of the residents' income is again gradually slowing down, especially since the onset of the COVID-19 pandemic. Many industries were impacted by the pandemic in 2020. Thus, improving the income of Chinese residents and maintaining the steady growth of the national economy is a national concern, and promoting internet use may be part of the solution. Internet consumption is a new driver of growth and its use is reported to affect residents' income and consumption significantly. For instance, the impact of Internet popularization on Residents' income is significantly positive (Han and Zhang, 2017).

2. Literature Review

The literature on Internet use is rich in recent year. A growing number of studies have explored the role of Internet use in influencing various aspects of people's lives, agricultural and social development. Scholars have examined the impact of Internet use on economic well-being (Choi and Yi, 2018; Ma *et al.*, 2020), e-commerce and online banking development (Li *et al.*, 2021; Peng *et al.*, 2021), sustainable agricultural development (Zheng and Ma, 2021; Zhu *et al.*, 2021), people's perceptions of social fairness (Asongu and Roux, 2017; Zhu *et al.*, 2020), and subjective well-being (Long *et al.*, 2019; Zhu *et al.*, 2020). For example, from the perspective of the e-commerce, Li *et al.* (2021) examined the effect of e-commerce adoption on household income using survey data of 1030 households in China by PSM model. From the perspective of people's perceptions of social fairness, Zhu *et al.* (2020) estimated the impact of Internet use on social fairness perception among Chinese farms. From the perspective of subjective well-being, Long and Yi (2019) found that Internet use had no significant impact on the well-being of residents compared to non-use, but frequency of Internet use did significantly improve subjective well-being. Yang (2021) pointed out that Internet use had a significant positive effect on the health of Chinese residents, but the role of employment performance in the process of improving the health of Chinese residents through Internet use is very limited.

A growing body of study has also estimated the association between Internet use and income (Krueger, 1993; DiNardo and Pischke, 1997; Lee and Kim, 2004). Krueger (1993) focused on the productivity-enhancing effect of computer use at work relying on cross-sectional microdata in US and suggested that employees who directly used a computer at work earned 10 to 15 percent higher wages than those who did not. This result was considered an overestimation as it is difficult to control for the effect of unobserved skills on computer use wage premiums (DiNardo and Pischke, 1997). Replicating Krueger's analysis, DiNardo and Pischke (1997) examined this issue with large cross-sectional survey data on German workers, which contained more detailed information on the tools used by workers in their jobs. They pointed out that positive

selection into computer use resulted in wages differences, so wage premiums tended to disappear as one controls for unobserved individual characteristics. Further, in addition to computer use, researchers found the employees having average computer experience earned a return for their skills (Pabilonia and Zoghi, 2005). With the development of computer networks, the percentage of workers using the Internet at work has increased. Metcalfe's law predicts that network value is proportional to the square of the number of users. Individuals acquire considerable benefits from Internet as well. Further, estimates from newer research suggested that workers who used the Internet at work earned 8% higher than those who did not (Lee and Kim, 2004). In addition, several researchers explored whether wage premiums are associated with Internet use based on cross-sectional data, and estimated premiums are approximately 15% (Benavente *et al.*, 2011). In China, studies have documented that Internet use had a significant positive effect on individual income, which could bring 8 to 20 percent extra income return (Hua, 2018). The Internet may significantly increase the wage income of residents and that the premium for Internet use in the city (25.7%) is lower than the premium in the countryside (29.3%) (Liu and Jin, 2015). In addition, the relation between Internet use and rural income has been highly focused (Zhou *et al.*, 2020). Zhou *et al.* (2020) considered that Internet use had a direct effect on rural residents' income growth by using multiple linear regression analysis and mediation effect analysis. Ma *et al.* (2020) analyzed the effect of internet use on household income and expenditure using a sample of rural household from China, and the result showed that Internet use increased household income and expenditure significantly. There is a strong consensus in the numerous existing literatures that internet use has a positive impact on residents' income. The reasons for using the Internet to improve income are attributed to the following points. People using the Internet can reduce the time of searching for employment. In addition to Internet use significantly improving the employment rate and productivity of workers, the Internet has an advantage in facilitating the obtaining of information and resources (Xiao *et al.*, 2010). The Internet can provide a great deal of the latest information and convenient social methods which can reduce the cost of residents' entrepreneurship and increase their income from entrepreneurship, and this is particularly significant in rural areas (Zhang *et al.*, 2015).

The empirical methods of previous literature could be divided to two strands according to the model they use. Several studies adopted parametric models using ordinary least squares (OLS) (Krueger, 1993; DiNardo and Pischke, 1997; Lee and Kim, 2004). The other is non-parametric models using propensity score matching techniques (PSM). PSM techniques have also been widely used in the literature that estimated the impact of Internet use on environment and agriculture (Rahman *et al.*, 2021; Zheng and Ma, 2021), People's well-being and behaviors (Yang *et al.*, 2021; Zhang *et al.*, 2021), technical efficiency of production (Zheng *et al.*, 2021), mental health problems (Kim *et al.*, 2018), consumption structure (Wang and Deng, 2020) and individual income (Navarro, 2010; Li

et al., 2021; Zhang *et al.*, 2021). For instance, Zhang *et al.* (2021) operationalized Internet use as one of indicators of an informed citizen, and tested whether they are predictive of confidence in the police by PSM and regression models. Besides, based on PSM approach, Mora-Rivera and García-Morab (2021) estimated the impact of Internet access on the multidimensional and income poverty of rural and urban Mexico.

China is currently in a period of steady economic development and technological transformation, and whether using the Internet significantly increases residents' income will provide further direction for the future development and improvement of the Internet. Thus, this paper attempts to clarify the impact of using the Internet on the income of Chinese residents. Although numerous studies have explored various effect of Internet use, few is about the relationship between Internet use and residents' income by PSM in China. We fill the gap by analyzing the influence of internet use on residents' income in China, using the PSM models to address selection bias. Previous literature mainly focused on the impact of Internet use on employee wages and heterogeneity analysis in various characteristics and many were limited by a sample self-selection problem in the regression process, which led to deviation between the estimated results and the actual situation. To address the sample selection bias with Internet use, this study estimates the impact of Internet use on income under different characteristics based on propensity score matching model (Li *et al.*, 2021). Therefore, this study considers the endogeneity effects in the research of the impact of Internet use on Chinese residents' income, and uses propensity score matching techniques for empirical analysis.

3. Empirical Design

3.1. Data

The data in this study are selected from the Chinese General Social Survey (CGSS) in 2017, in which the effective sample size of the survey is 12582. The CGSS started in 2003 and is carried out by the China Survey and Data Center of Renmin University of China. It is China's earliest nationwide, comprehensive, and continuous academic survey project. The CGSS system comprehensively collects data on multiple levels of society, communities, families, and individuals. This system creates a precedent for the opening and sharing of large-scale academic survey data in China. The CGSS module contains the question of residents' Internet use, which is currently rare in China. It is also one of the most representative Internet use data in Chinese Internet research. Regarding age, we select the samples whose birth year is between 1951 and 1998, from 18 to 65 years old, according to adult and retirement ages in China. In terms of sample types, this study chooses the "agricultural household registration" and "non-agricultural household registration" categories. For work experience, the "currently engaged in non-agricultural work" and "currently working in agriculture" groups are

used. Observations with missing data are eliminated. The effective sample size is 5818. Compared with previous studies (Li, *et al.*, 2021; Rahman *et al.*, 2021; Yang *et al.*, 2021; Zhang *et al.*, 2021), this paper retains more valid observations, which further efficiently support the empirical results and findings from the respective of statistics.

3.2. Variables and Descriptive Analysis

Residents' *income* is selected as the dependent variable and *Internet use* as the independent variable. *Income* represents residents' total income in 2016, which is measured by the survey results of the question related to "individual income in the last year (2016)" in the CGSS. *Lnincome* is the logarithm of income. *Internet use* refers to people either use the Internet or not, which is measured by the survey results for the question related to "Internet (including mobile Internet) use in the past year". In the CGSS, the options under the question of whether one uses the Internet are "never", "rarely", "sometimes", "frequently", "very frequent", "don't know" and "refuse to answer". To facilitate variable selection and model setting, the sample of individuals who "don't know" and "refuse to answer" are eliminated, and the above options are converted into two categories based on the practices of previous literature. Those who selected the answer "never" constitute the category of residents that do not use the Internet, while those who selected "rarely", "sometimes", "frequently" and "very frequent" are placed under the category of residents who use the Internet. The form of dummy variables is added to the model, with values of 0 and 1, respectively.

In order to estimate the impact of other individual characteristics on income, the variables of *gender*, *age*, *age square*, *education*, *political status*, *physical health*, *marital status*, *employment status*, *region*, *household registration*, are added to reflect the impact of using the Internet on residents' income (Krueger, 1993; Lee and Kim 2004; Chen and Wu, 2008; Forman *et al.*, 2012). The *gender* variable equals one if the resident is male and zero if the resident is female. The base category of *education* includes five types: (1) no educational experience(equal to 1), (2) private schools, literacy classes, elementary schools (equal to 2), (3) middle school (equal to 3), (4) vocational high school, general high school, technical secondary school, and technical school (equal to 4),and (5) college and above (equal to 5). The *political status* variable equals 1 if the resident is a communist, and 0 otherwise. The *physical condition* variable has three categories: (1) very unhealthy and relatively unhealthy (equal to 0), (2) generally healthy (equal to 1), and (3) relatively healthy and very healthy (equal to 2). *Marital status* equals 0 if the resident is unmarried, and 1 otherwise. *Occupation* is a resident's type of work, including engaged in agricultural work (equal to 0) and engaged in non-agricultural work (equal to 1). The *region* variable, where residents live, includes three areas: the eastern region (equal to 0), central region (equal to 1), and western region (equal to 2). *Household registration* status equals 0 if the resident

is registered as an agricultural household, and 1 if the resident is registered as a non-agricultural household.

The relationship between Internet use and residents' income is drawn using the kernel density graph. Figure 1 shows that there is a significant income difference between residents who use the Internet and those who do not. Residents who use the Internet have a higher density in the higher logarithmic income range. Intuitively, the graph shows that Internet use and residents' income present a positive correlation.

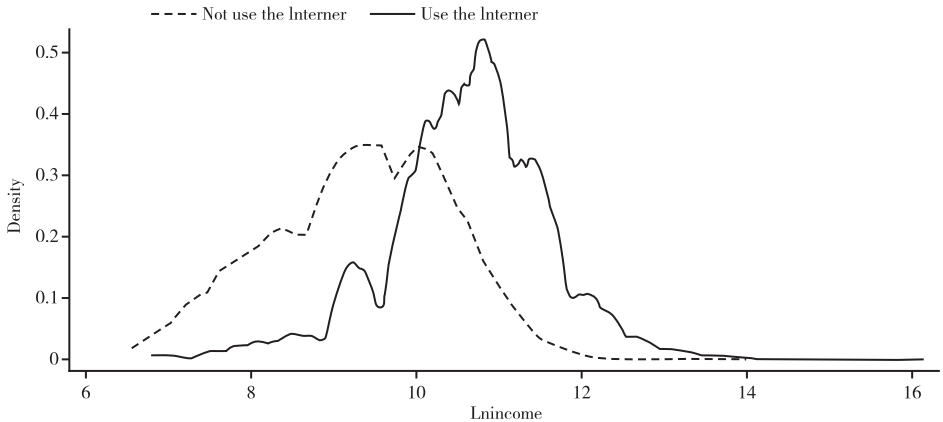


Figure 1. Internet Use and Logarithm of Resident's Income

The descriptive statistics of the selected sample are presented in Table 1. The table shows that 72.57% (sample size with using Internet/sample size = 4222/5818) of individuals use the Internet, and their average income is nearly 4 times higher than that of individuals who do not use the Internet. The percentage of male residents who reported not use the Internet is 54.51%, and that of male residents who use it is 56.18%, both of which are higher than the females' statistics. From other perspectives, residents who use the Internet are younger, more educated, and healthier on average. Among those who do not use the Internet, the percentage of those who engage in agricultural work is 53.86% higher than that of those who do use the Internet. Compared with central and western region, more residents use the Internet in eastern region. Central region has more residents who do not use the Internet in sample. From the perspective of household registration, among the groups using the Internet, agricultural and non-agricultural households are relatively even. However, among those who do not use the Internet, the percentage of agricultural households was relatively high, reaching 86.84% (1–13.16%). Thus, residents who are younger, male, more educated, healthier, and employed in non-agricultural jobs use the Internet more than their respective counterparts. Non-agricultural and agricultural households have similar proportions using the Internet in the sample.

Table 1. Descriptive Statistics of Variables

Variable	Sample size (n=5818)		Sample size with not using Internet (n=1596)		Sample size with using Internet (n=4222)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
<i>Income</i>	55315.26	201343.10	18044.46	35862.04	69404.36	233789.4
<i>Lnincome</i>	10.2011	1.2351	9.1681	1.1639	10.5916	1.0170
<i>Internet use</i>	0.7257	0.4462	0.0000	—	1.0000	—
<i>Gender</i>	0.5572	0.4968	0.5451	0.4981	0.5618	0.4962
<i>Age</i>	42.3800	11.8083	53.0965	7.9305	38.3289	10.4190
<i>Age square</i>	1935.48	1007.265	2882.09	808.7706	1577.64	827.13
<i>Education</i>	3.4261	1.2442	2.2964	0.8863	3.8532	1.0824
<i>Political status</i>	0.1114	0.3146	0.0370	0.1887	0.1395	0.3465
<i>Physical health</i>	1.5249	0.7015	1.1742	0.8303	1.6575	0.5947
<i>Marital status</i>	0.8817	0.3229	0.9812	0.1359	0.8441	0.3628
<i>Employment Status</i>	0.7374	0.4401	0.3465	0.4760	0.8851	0.3189
<i>Region</i>	1.7114	0.7720	2.0050	0.7537	1.6004	0.7495
<i>Household Registration</i>	0.4323	0.4954	0.1316	0.3381	0.5459	0.4979

Source: Estimated by CGSS2017 Survey Data.

3.3. Empirical Methods

This study uses Krueger's (1993) model setting to test the impact of Internet use on residents' income by constructing a Mincer equation that includes the dummy variable of "whether to use the Internet" (Mincer, 1991).

$$\ln w_i = \alpha + \beta I_i + \gamma X_i + \varepsilon_i \quad (1)$$

where $\ln w_i$ represents the Logarithm of income for the resident i , I_i is a dummy variable that equals 1 if the resident uses the Internet, and 0 otherwise. X_i represents a vector of the observed characteristics, and ε_i is the error term. α , β and γ are parameters to be estimated. Reviewing the previous literature, this paper gradually adds *employment status*, *regions*, *household registration* into the model to obtain the true rate of return brought about by using the Internet.

It is difficult to identify the direction of causality by simply observing the data. The residents' Internet use decisions are not random and it is impossible to observe the unused Internet income of residents who use the Internet, which would incur a selection bias. Propensity score matching can be used to solve sample selection problems. Propensity score matching techniques has been widely used in the literature

(Minah, 2021; Shimada and Sonobe, 2021;), including the impact of Internet use on residents' income (Li *et al.*, 2021). Propensity score matching (Rosenbaum and Rubin, 1983) could be based on counterfactual inference to calculate the net value of the impact of using the Internet on individual income.

$$ATT = E(\ln w_1 - \ln w_0 | T = 1) = E(\ln w_1 | T = 1) - E(\ln w_0 | T = 1) \quad (2)$$

where $\ln w_0$ represents the logarithm of annual income of residents who do not use the Internet and $\ln w_1$ is the logarithm of income for the residents who use the Internet.

To eliminate the interference of other factors, the average treatment effect on the treated group limits the sample to residents who use the Internet ($T=1$). The net impact of using the Internet on income is the difference between residents' income from using and not using the Internet. In this method, the income of residents who use the Internet is observable, but the unused Internet income of residents who use the Internet are not observable ($E(\ln w_0 | T=1)$ is unobservable), which is a counterfactual result. Propensity score matching can deal with the self-selection bias in the sample. Matched residents differ in their use of the Internet, while other characteristics remain the same. This paper uses data that do not use the Internet to simulate the "counterfactual results of residents who use the Internet". Given the personal characteristic variable X and Internet use variable $T=1$, the conditional probability of residents using the Internet is

$$p(X_i) = Pr(T = 1 | X_i) E(T | X_i) \quad (3)$$

where X_i represents the individual characteristics that influence the Internet use behavior. $p(X_i)$ is the propensity score, which is the probability of Internet use conditional on a set of observable characteristics X_i . This can be estimated from the model. The matching balance test is performed based on the propensity score to investigate whether there is a significant deviation in the match between the treatment and control groups. If there is, then the matching effect is not accurate. According to the actual situation of the sample, this study uses the nearest neighbor matching to match the treatment group and the control group.

4. Empirical Results and Discussion

4.1. Ordinary Least Square Analysis

Here, we use Stata15 to perform ordinary least squares regression analysis on the sample data. The results are presented in Table 2.

Table 2. OLS of Impact of Internet Use on Resident's Income

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Does use the Internet (yes=1)	1.4234*** (0.3112)	0.5903*** (0.0371)	0.3354*** (0.0360)	0.3079*** (0.0353)	0.2949*** (0.0353)
<i>age</i>		0.0815*** (0.0083)	0.0691*** (0.0078)	0.0681*** (0.0077)	0.0633*** (0.0077)
<i>age</i> ²	—	−0.0010*** (0.0001)	−0.0009*** (0.0001)	−0.0009*** (0.0001)	−0.0008*** (0.0001)
<i>male</i>	—	0.3641*** (0.0247)	0.3673*** (0.0233)	0.3755*** (0.0228)	0.3814*** (0.0228)
<i>private school</i>		0.3815 (0.0557)	0.0299 (0.0523)	0.0424 (0.0512)	0.0433 (0.0511)
<i>middle school</i>	—	0.2883*** (0.0559)	0.1297** (0.0528)	0.1353*** (0.0517)	0.1258** (0.0516)
<i>high school etc.</i>	—	0.7108*** (0.0605)	0.4198*** (0.0577)	0.4088*** (0.0565)	0.3656*** (0.0571)
<i>college and above</i>	—	1.2721*** (0.0616)	0.9480*** (0.0589)	0.8868*** (0.0578)	0.8030*** (0.0602)
<i>communist</i>	—	0.07189* (0.0412)	0.0773** (0.0387)	0.0822** (0.0378)	0.0774** (0.0378)
<i>generally healthy</i>	—	0.4443*** (0.0437)	0.3345*** (0.0412)	0.2788*** (0.0405)	0.2769*** (0.0404)
<i>relatively healthy, very healthy</i>	—	0.6036*** (0.0404)	0.4440*** (0.0384)	0.3920*** (0.0377)	0.3909*** (0.0377)
<i>married</i>	—	0.1736*** (0.0472)	0.2063*** (0.0443)	0.2277*** (0.0434)	0.2333*** (0.0433)
<i>non-agricultural work</i>	—	—	0.9260*** (0.0331)	0.8020*** (0.0333)	0.7596*** (0.0344)
<i>central region</i>	—	—	—	−0.2661*** (0.0267)	−0.2588*** (0.0267)
<i>western region</i>	—	—	—	−0.4957*** (0.0317)	−0.4866*** (0.0317)
<i>household registration status</i>	—	—	—	—	0.1433*** (0.0295)
intercept	9.1681*** (0.0265)	6.8859*** (0.1648)	6.8704*** (0.1547)	7.2697*** (0.1534)	7.3969*** (0.1554)
Adj R-squared	0.2643	0.4500	0.5153	0.5359	0.5377

Note: Values in parentheses are standard errors of the independent variables. *, ** and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Source: calculated by CGSS2017 Survey Data.

All models use the logarithm of residents' 2016 income as the dependent variable for ordinary least squares regression. In model 1, the *Internet use* dummy variable is the only right-hand side variable. Model 1 shows the difference in annual income between residents who use the Internet and those who do not is 315.12% ($\exp(1.423)-1$). In model 2, several variables of the residents are added, including *age*, *age squared*, *gender*, *education*, *political status*, *physical health*, and *marital status*. After adding the resident's observable characteristics, using the Internet reduces the Internet premium on income to 80.45%. The Internet dummy variable continues to have a significant and sizable effect on income, with a *t*-statistic of 15.91. Models 3, 4, and 5 are based on Model 2, adding

employment status, *region*, *household registration*, sequentially as independent variables. In these three models, the adjusted *R*-square gradually increases. This indicates that the model fits more. After including these variables, the significant influences of residents' income by Internet use are 39.85%, 36.06%, and 34.30%, respectively. This shows that using the Internet can increase residents' income. Compared with several scholars' findings (Krueger, 1993; DiNardo and Pischke, 1997; Goss and Phillips, 2002; Lee and Kim, 2004; Mossberger *et al.*, 2006), the impact on improving the residents' income is more significant in China. Currently, the Internet is developing rapidly, and practical skills and technologies can be learned from the Internet to improve employment and work efficiency. Various online application modes provide residents with the latest employment information and job opportunities, as well as employment consulting and training. With the development of the e-commerce, the live broadcast industry has grown stronger and is increasing employment opportunities in various industries and improving income levels.

As shown in Table 2, *age* has a significant positive effect on annual income, while the square of residents' ages has a negative effect. This indicates that the residents' annual income have an inverted U-shape correlation with age. *education* has a significant positive impact on residents' annual income: the higher the education level, the greater the impact on annual income. This means that a higher education level can result in a higher annual income. *political status*, *physical health*, and *marital status* also have a positive impact on income. After adding *employment status*, it is found that non-agricultural employment status has a more significant impact on income. Furthermore, the partial regression coefficients of the central and western regions are negative, indicating that the income of residents in the central and western regions is lower than that of the eastern region, and the western region has the lowest income. *Household registration* dummy variables have a positive effect on annual income, indicating that the annual income of non-agricultural households will be higher.

4.2. Propensity Score Matching Analysis

This study explores the impact of Internet use on the annual income of residents by using propensity score matching. Since this paper focuses on the impact of Internet use on residents' income and Rosenbaum and Rubin (1983) suggested using the logit model as well, this study selects the logit model to analyze the propensity score matching to reflect the impact of Internet use on residents' income. In the balancing test between the independent variables under the nearest neighbor matching, the absolute values of the standard deviations of other variables are less than 30%, except for the four variables including *age*, *age squared*, *employment status*, and *household registration*. The difference in the mean of the variables of *gender*, *education*, *physical health*, and *marital status* is not significant after matching, and the coefficients of *education* and *physical health* decreased. The variable coefficients of *political status*,

employment status, and *household registration* all become smaller. In summary, the estimated results of the propensity score matching are reasonable. This can weaken the problem of sample self-selection bias, and the results can reflect causality to a certain extent. The propensity score matching results are presented in Table 3.

Table 3. ATT of Nearest Neighbor Matching

Method	Sample	Use the Internet	Not use the Internet	ATT	Standard error	<i>t</i> -statistic
—	Unmatched	10.5916	9.1681	1.4234***	0.0311	45.73
Nearest neighbor	Matched	10.0953	9.1681	0.9271***	0.0389	23.82

Note: Nearest neighbor matching is *k*-nearest neighbor matching when *k* equals 1. *, ** and *** indicate statistical significance at 10%, 5% and 1% levels, respectively.

The unmatched result shows that the logarithms of the annual income of residents who use the Internet and those who do not are 10.5916 and 9.1681, respectively, and the average treatment effect on the treated group (ATT) is 1.4234, which is significant at the 1% significance level. Table 3 shows that the logarithms of the annual income of residents who use the Internet and those who do not are 10.0953 and 9.1681 under the nearest neighbor matching, and the ATT is 0.9271, which is smaller than unmatched results. In summary, the use of the Internet has significantly increased the income of residents by 152.72% and has a positive impact on the annual income of residents after correcting for sample self-selection bias.

Compared with previous literature, the impact of Internet use on the income of Chinese residents is far greater than that in other countries, for instance, US, Mexico, Brazil and Ghana (Freeman, 2002). In China, residents who use the Internet earn about 153% income higher income than those who do not. These figures are much higher than those obtained for other countries using similar data. This is because that Internet technology has developed rapidly in recent years in China, resulting in an Internet coverage and penetration boom. Almost all residents enjoy the convenience brought about by the Internet. Besides, our result is consistent with the finding of Liu and Jin (2015) and Hua (2018) who showed that Internet use is positively associated with residents' income in China, but the estimated Internet use returns for rural residents are below the returns estimated for rural and urban residents in this paper. One reason for high returns in rural and urban residents may be the higher popularization and usage of Internet use in urban area compared with the rural area. Another reason should be considered is that the unmatched differences are so large between the treated and untreated. There are still two types of people who do not use the Internet: the elderly and the residents living in remote areas which are not covered by the Internet. The income of this group is relatively low, leading to a large difference in income between people who use the Internet and those who do not.

4.3. Robustness Test

4.3.1. Alternative Variables

First, we replace the dependent variable. To demonstrate the impact of Internet use on residents' income, this study uses the income of previous year (2016) as the dependent variable. Residents' income is replaced with the earning of previous year (2016) to test the robustness of the estimated results. The OLS and ATT results are listed in Table 4.

Table 4. Robustness Test of Variable Substitution: Total Occupational Income

Method	Sample	Use the Internet	Not use the Internet	ATT	standard error	t-statistic
—	Unmatched	10.5864	9.2231	1.3634***	0.0347	39.26
Nearest neighbor	Matched	10.1044	9.2231	0.8813***	0.0462	19.06

Note: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Compared with Table 3, Internet use increases individuals' full-time/labor income by 290.95% in the OLS, which is significant at the 1% level. This result is similar to the initial OLS result. ATT is positive and the estimated result is the same as the initial model, which uses current year income as the dependent variable in the nearest neighbor matching procedure.

Second, we replace the independent variable. The dependent variable *Internet use* from the initial "Internet usage in the past year" is replaced with "the main source of information", and represents a dummy variable that equals 1 if the *i*th individual uses the Internet as the main source of information, and 0 otherwise. The other variables and models do not change. The estimated results are presented in Table 5. Result shows that using the Internet as the main source of information can increase income by 215.88% in OLS, which is significant at the 1% significance level. The ATT of using the Internet as an information source is positively matched and statistically significant at the 1% significance level. The result is similar to the ATT when handling *Internet use in the past year* as a dummy variable. In summary, the regression results and matched results are relatively robust.

Table 5. Robustness Test of Variable Substitution: Main Source of Information

Method	Sample	Use the Internet	Not use the Internet	ATT	Standard error	t-statistic
—	Unmatched	10.7116	9.5614	1.1502***	0.0289	39.28
Nearest neighbor	Matched	10.6242	9.5614	1.0629***	0.0307	34.57

Note: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

4.3.2. Alternative Model

Nearest neighbor matching was selected to estimate the ATT in matched results.

However, even the closest individual with a matching propensity score in the sample may be far away from it and incomparable. Therefore, this study re-estimates ATT using the radius estimator and kernel estimator, and judges the robustness by comparing the estimated results. The estimated results are presented in Table 6.

Only the absolute value of the standard deviation of the *political status* variable is greater than 30% in the radius matching, and the *employment status* and *household registration* variables cannot pass the 10% significance level test. Only the absolute value of the standard deviation of *education* and *marital status* is greater than 30% in the kernel matching, and all variables are significant at the 1% level of significance, except for the *education* variable, which is significant at the 5% level of significance. In summary, these can reduce sample self-selection bias, and the estimated results can reflect causality to a certain extent.

The ATT of the radius and kernel estimators are 0.3306 and 0.3378, respectively, which are significant at the 10% and 5% levels, respectively. This result is similar to that of the nearest neighbor estimator. In summary, after correcting the sample self-selection bias, Internet use significantly increased the income of residents by 39.18% and 40.19% in radius and kernel matching, and Internet use still has a positive impact on the residents' income. This result is similar to the nearest neighbor matching result. This indicates that the impact of Internet use on residents' income is robust.

Table 6. Robustness Test: Radius and Kernel

Method	Sample	Use the Internet	Not use the Internet	ATT	Standard error	t-statistic
Nearest neighbor	Matched	10.0953	9.1681	0.9271***	0.0389	23.82
Radius	Matched	10.5916	10.2609	0.3306*	0.2269	1.46
Kernel	Matched	10.5916	10.2538	0.3378**	0.1328	2.54

Note: Nearest neighbor matching with $n = 1$, radius matching with $r = 0.01$, and kernel matching equal to 0.06. Sample size is 5818. *, **, and *** indicate statistical significance at 10%, 5% and 1% levels, respectively.

Besides, we compare the estimated results of the logit model with the probit and tobit models to judge the robustness of the model. The results of the three regression models are presented in Table 7. Internet use still significantly improves the income of residents after using different regression models. This result is consistent with the propensity score matching results in the OLS and logit models. This indicates that the initial model is reliable.

Table 7. Model Robustness Test: Probit Model and Tobit Model

	Logit ($n=5818$)	Probit ($n=5818$)	Tobit ($n=5818$)
Using the Internet	1.0183*** (0.0397)	1.0075*** (0.0398)	0.3816*** (0.0352)
Pseudo R-squared	0.4856	0.4863	0.2283

Note: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

4.4. Endogenous Discussion

There is a bias in the estimation coefficient arising from the endogenous relationship between Internet use and residents' income. Here, we select the instrumental variable (IV) to eliminate the effect of endogeneity based on the two-stage least squares (2SLS) (Theil, 1953; Basmann, 1957) and limited information maximum likelihood (LIML) estimations. Two variables, *Whether there is a mobile phone for your personal use* and *Whether WeChat payment has been used* are selected as the IVs of whether to use the Internet. The possession of a mobile phone or a computer is the basic premise for using the Internet (Shi and Wang, 2017; Zhu, 2020). Additionally, WeChat payments are based on the Internet. This shows a direct correlation between the two IVs and Internet use. According to *Communications Industry Statistical Bulletin in 2017* from Ministry of Industry and Information Technology, the total number of mobile phone users is 1.42 billion with 102.5 mobile phones per 100 people on average in China. Mobile broadband users (3G and 4G users) account for 79.8% of mobile phone users. There has been a mobile phone per people and considerable people use mobile broadband by the end of 2017. Additionally, residents can easily buy and use mobile phones regardless of their income as the price of mobile phones continues to decline in China. Therefore, whether a resident owns a possession of a mobile phone or not does not have a direct impact on their income. WeChat is one of the most prevalent instant communication apps in China. Residents who have mobile phone usually install this app and select WeChat payment as a main payment method. Similarly, whether a resident uses WeChat payments or not does not directly affect their income. The two IVs are not directly related to income. In these cases, the IVs meet the exogenous requirements.

The overidentification test result shows that Sargan statistic is 1.1402 (p value = 0.2856). This means the IVs are exogenous (5% significance level). The coefficients of the LIML and 2SLS are similar, which proves the absence of weak IVs. The underidentification test (p -value of underidentification test statistic is 0.0000) and weak instruments test (Cragg-Donald Wald F statistic is 461.624; Kleibergen-Paaprk Wald F statistic is 330.439) show that IV is correlative with Internet use. In Hausman specification test, Hausman specification test statistic (71.75; p -value = 0.0000) shows that whether the Internet is used is an endogenous variable. The regression coefficient is significantly positive at 1% significant level in the first stage. That means there is a correlation between the IVs and Internet use. In the second stage, the coefficient regression of Internet use is significantly positive (1% significant level). Test results of endogeneity remain consistent, which indicates that the regression results are robust.

4.5. Heterogeneity Analyses

In this subsection, we examine the impact of Internet use on residents' annual

income by limiting the variables' characteristics. Although the model is robust, the estimated results of OLS and ATT are slightly different owing to the differences in the models.

4.5.1. Age

We divide the sample data (made up of 18~65-year olds) into young residents (aged 18~44 years) and middle-aged and elderly residents (aged 45~65 years). The PSM results of the different age groups are listed in Table 8. Compared with the impact on residents aged 18 to 44 years, Internet use has a more significant promotion effect on middle-aged and elderly people (45~65 year olds) in the unmatched sample, and the effect is significant at the 1% significance level. The results are presented in Table 8. Even after estimation by the nearest neighbor matching, the ATT of young the residents is estimated to be 0.1747, which is not significant (t -statistic = 1.65). Conversely, the ATT of residents aged 45 to 65 is estimated to be 1.2259, which is significant at the 1% level. Thus, the impact of Internet use has a more observable promotional effect on middle-aged and elderly residents' annual income. The estimated result can be explained by the fact that young people's use of the Internet differs from that of middle-aged and elderly people. Most young people prefer to use the Internet for entertainment and leisure, while middle-aged and elderly people are more inclined to learn new skills and knowledge. The estimated results show that middle-aged and elderly people can increase their income more significantly through Internet use.

Table 8. ATT of Internet Use by Age

Method	Sample	18~44		45~65	
		ATT	t -statistic	ATT	t -statistic
—	Unmatched	1.2429***	17.61	1.2986***	30.23
nearest neighbor	Matched	0.1747	1.65	1.2259***	29.34

Note: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

4.5.2. Household Registration

Here, we divide the type of *household registration* into agricultural and non-agricultural, with most agricultural household residents being those concentrated in agricultural production and planting. The results are presented in Table 9. The impact of Internet use on agricultural and non-agricultural households is significant at the 1% significance level for both, whether unmatched or matched. Nonetheless, Internet use has a stronger promotional impact on residents with an agricultural household registration. This may be because that with the rapid development of China's e-commerce setting off a live broadcast of agricultural products on the Internet—which presented opportunities for reduction of transaction time and additional agricultural

transportation costs—the income of agricultural households rapidly increased over a short period. The Internet is an effective way to obtain information quickly and Internet use can reduce the cost of farmers to obtain information (Aker and Mbiti 2010). Compared with the agricultural household residents, non-agricultural household residents are inclined to learn new technologies and obtain new knowledge on the Internet. Therefore, the impact of Internet use on the annual income of agricultural household residents can be slower than those of non-agricultural household residents. Additionally, among the agricultural household residents, agricultural household residents who use the Internet can obtain more information and learn new technologies compared to those who do not use the Internet, which could eventually lead to higher income. Among non-agricultural household residents, however, the income difference resulting from Internet use is relatively small since these residents already have more access to information.

Table 9. ATT of Internet Use for Household Registration

Method	Sample	Agricultural		Non-agricultural	
		ATT	<i>t</i> -statistic	ATT	<i>t</i> -statistic
—	Unmatched	1.1464***	30.06	0.9644***	15.14
Nearest neighbor	Matched	0.9947***	23.91	0.3366***	4.00

Note: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

5. Conclusions and Policy Implications

This study uses data from the 2017 Chinese General Social Survey to investigate the impact of Internet use on residents' income through ordinary least squares regression and propensity score matching. The empirical results show that Internet use can significantly increase residents' income after adding control variables. The ordinary least squares analysis shows that Internet use can significantly increase the income of residents by 34.30%, while the nearest neighbor matching, which eliminates the bias caused by sample self-selection, shows that Internet use can significantly increase residents' income by 152.72 %. These results imply that Internet use has a much higher impact on income in China than in other countries. Furthermore, by restricting some control variables and examining the heterogeneity for age and household registration, it is found that the impact of using the Internet on income is more significant among middle-aged and elderly residents and agricultural household residents. The estimation results of the substituted variables and model are consistent with those of the initial method. This means that not only is the impact of Internet use on residents' income positive, but also the result itself is robust.

Therefore, the findings of this study suggest that to improve residents' income in China, the size and penetration rate of the Internet need to be improved further. The research results provide the following policy inspirations for promoting residents'

income. First, more Internet-related infrastructure should be constructed and Internet use should be promoted to improve penetration rates. Governments need to increase the construction of network infrastructure and commit to such constructions in remote areas. On the one hand, local governments need to formulate preferential policies for network construction to stimulate enterprises to build basic infrastructure and to consolidate the foundation of Internet use. On the other hand, governments need to increase financial investment in subsidizing rural residents' Internet installations and ensuring the use and application of local resident networks with low barriers. Second, the basic Internet skills of residents, especially middle-aged and elderly people, should be improved through training. Communities should organize Internet skills training programs for middle-aged and elderly residents to provide them with basic Internet use and information processing capabilities.

References

- Aker, J. C., & Mbiti, I. M. (2010). Mobile Phones and Economic Development in Africa. *Journal of Economic Perspectives*, 24 (3), 207–232.
- Asongu, S. A., & Roux, S. L. (2017). Enhancing ICT for Inclusive Human Development in Sub-Saharan Africa. *Technological Forecasting and Social Change*, 118 (May), 44–54.
- Basmann, R. L. (1957). A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica*, 25 (1), 77–83.
- Benavente, J. M., Bravo, D., & Montero, R. (2011). Wages and Workplace Computer Use in Chile. *The Developing Economies*, 49 (4), 382–403.
- Chen, Y., & Wu, Y. (2008). The Impact of IT on Labor Markets: An Estimation of the Returns to Computer Uses. *China Economic Quarterly (Jingjixue Jikan)*, 4, 1149–1166.
- Choi, C., & Yi, M. H. (2018). The Internet, R&D Expenditure and Economic Growth. *Applied Economics Letters*, 25(4/6), 264–267.
- DiNardo, J. E., & Pischke, J. S. (1997). The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too? *Quarterly Journal of Economics*, 112, 291–303.
- Forman, C., Goldfarb, A., & Greenstein, S. (2012). The Internet and Local Wages: A Puzzle. *American Economic Review*, 102 (1), 556–575.
- Freeman, R. (2002). The Labour Market in the New Information Economy. *Oxford Review of Economic Policy*, 18 (3), 288–305.
- Han, B., & Zhu, F. (2014). Empirical Analysis on Effect of Broadband on Economic Growth in China. *Statistical Research(Tongji Yanjiu)*, 31 (10), 49–54.
- Han, C., & Zhang, L. (2017). The Influence of Internet Popularization on Urban and Rural Income Distribution: GMM Analysis Based on the Data of Provincial Panel

- in China. *Inquiry into Economic Issues (Jingji Wenti Tansuo)*, 8, 18–27.
- Hua, Y. (2018). The Income Growth Effect of Internet Use: Theoretical Mechanism and Empirical Test. *Jianghai Academic Journal (Jianghai Xuekan)*, 3, 219–224.
- Kim, Y. J., Jang, H., & Lee, Y., *et al.* (2018). Effects of Internet and Smartphone Addictions on Depression and Anxiety Based on Propensity Score Matching Analysis. *International Journal of Environmental Research and Public Health*, 15 (5), 859.
- Krueger, A. B. (1993). How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984–1989. *Quarterly Journal of Economics*, 108, 33–60.
- Lee, S. H., & Kim, J. (2004). Has the Internet Changed the Wage Structure Too? *Labour Economics*, 11 (1), 119–127.
- Li, X., Guo, H., & Jin, S., *et al.* (2021). Do Farmers Gain Internet Dividends from E-Commerce Adoption? Evidence from China. *Food Policy*, 101, 102024.
- Liu, Z., & Jin, W. (2015). The Analysis of Salary Income Premium Effect from the Computer Network. *Review of Industrial Economics (Chanye Jingji Pinglun)*, 1, 67–78.
- Long, C., Yi, C., & Li, C. (2019). The Impact of Internet Use on Residents' Subjective Well-Being: An Empirical Analysis Based on National Data. *Social Sciences in China (Zhongguo Shehui Kexue)*, 40(4), 106–128.
- Ma, W. L., Nie, P., & Zhang, P., *et al.* (2020). Impact of Internet Use on Economic Well-Being of Rural Households: Evidence from China. *Review of Development Economics*, 24 (2), 503–523.
- Minah, M. (2021). What Is the Influence of Government Programs on Farmer Organizations and Their Impacts? Evidence from Zambia. *Annals of Public & Cooperative Economics*, 1, 1–25.
- Mincer, J. (1991). Human Capital, Technology, and The Wage Structure: What Do Time Series Show? National Bureau of Economic Research Working Papers 3581.
- Mora-Rivera, J., & García-Mora, F. (2021). Internet Access and Poverty Reduction: Evidence from Rural and Urban Mexico. *Telecommunications Policy*, 45 (2), 102076.
- Pablonia, S. W., & Zoghi, C. (2005). Returning to the Returns to Computer Use. *American Economic Review*, 95 (2), 314–317.
- Peng, C., Ma, B., & Zhang, C. (2021). Poverty Alleviation through E-Commerce: Village Involvement and Demonstration Policies in Rural China. *Journal of Integrative Agriculture*, 20 (4), 998–1011.
- Rahman, M. S., Majumder, M. K., & Sujan, M. K. (2021). Adoption Determinants of Biogas and Its Impact on Poverty in Bangladesh. *Energy Reports*, 7, 5026–5033.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70 (1), 41–45.
- Shi, J., & Wang, W. (2017). The Impact of Internet Use on Entrepreneurship: An Empirical Analysis Base on the Micro-Level Data. *Journal of Zhejiang University (Engineering Science) (Zhejiang Daxue Xuebao (Renwen Shehui Kexueban))*, 47 (4), 159–175.
- Shimada, G., & Sonobe, T. (2021). Impacts of Management Training on Workers:

- Evidence from Central America and the Caribbean Region. *Review of Development Economics*, 25 (3), 1492–1514.
- Theil, H. (1953). Repeated Least Squares Applied to Complete Equation System. *The Hague*, Central Planning Bureau, Mimeo.
- Xiao, Q., Marino, L. D., & Zhuang, W. (2010). A Situated Perspective of Entrepreneurial Learning: Implications for Entrepreneurial Innovation Propensity. *Journal of Business & Entrepreneurship*, 22 (1), 69–89.
- Yang, K. (2021). Internet Use, Employment Performance and the Health of Chinese Residents. *International Health*, 0, 1–13.
- Zhang, B., Hu, J.Y., & Fan, C. C. (2015). Social Network, Information Acquisition and Households' Entrepreneurial Income: An Empirical Research Based on Perspective of Urban-Rural Differences in China. *Economic Review (Jingji Pinglun)*, 2, 52–67.
- Zhang, F., Sarkar, A., & Wang, H. (2021). Does Internet and Information Technology Help Farmers to Maximize Profit: A Cross-Sectional Study of Apple Farmers in Shandong, China. *Land*, 10 (4), 390.
- Zheng, H., & Ma, W. (2021). Smartphone-Based Information Acquisition and Wheat Farm Performance: Insights from a Doubly Robust IPWRA Estimator. *Electronic Commerce Research*, 4, 1–26.
- Zheng, H., Ma, W., & Wang, F., *et al.* (2021c). Does Internet Use Improve Technical Efficiency of Banana Production in China? Evidence from a Selectivity-Corrected Analysis. *Food Policy*, 102, 102044.
- Zhu, Z. (2020). Do Internet Skills Contribute to the Consumption Upgrading of Rural Residents: Evidence From Chinese Social Survey 2015. *Statistical Research (Tongji Yanjiu)*, 37(9), 68–81.
- Zhu, Z. K., Ma, W. L., & Sousa-Poza, A., *et al.* (2020). The Effect of Internet Usage on Perceptions of Social Fairness: Evidence from Rural China. *China Economic Review*, 62, 101508.
- Zhu, Z., Bai, Y., & Dai, W., *et al.* (2021). Quality of E-Commerce Agricultural Products and the Safety of the Ecological Environment of the Origin Based on 5G Internet of Things Technology. *Environmental Technology & Innovation*, 22, 101462.