Can Internet reduce the income gap caused by the education gap? An empirical study based on education and urban-rural heterogeneity

Xin Wen

School of Agricultural and Forestry Economic Management, Huazhong Agriculture University, Wuhan, China

wenx_7@webmail.hzau.edu.cn

Keywords: Internet use, income gap, education, urban and rural areas

Abstract: In the context of the continuous development of the Internet, To explore whether the Internet can bridge the income gap caused related to education, this paper selected data from 2018 China Family Panel Studies (CFPS), basing on the perspective of heterogeneity, using the OLS and treatment effect model (TEM) to analyze whether Internet use can bridge the income gap caused by education gap. The study found that Internet use has a significant impact on reducing the income gap caused by the inequality of educational resources. But the Internet is widening the income gap as well as narrowing it, because there are educational benefits to Internet use, only at a certain level of education, Internet use will bring more income. After excluding the self-selection of Internet use, the results show that the income increase effect of Internet use by rural residents is obviously better than that of urban residents. Therefore, on the basis of guaranteeing residents' right to complete compulsory education, the state should encourage Chinese residents to continue to receive higher education. At the same time, the government is supposed to increase investment in infrastructure in rural areas to guide farmers to use it and make good use of the Internet.

1. Introduction

Since the reform and opening up, great breakthroughs have been made in China's economic development. By the end of 2019, China's per capital GDP had reached 9,915 dollars, which makes China into the rank of the upper middle-income countries. Li Shi et al. (2018) pointed out that China's income gap has undergone two major stages of change. The first stage froms 1978 to 2008, the income gap continued to expand. The other stage is the income gap remained high and hovering stage after 2008.China's Gini coefficient expanded from 0.31 in 1981 to 0.491 in 2008.After 2008, China's Gini Coefficient generally showed a decreasing trend, but it has picked up again in recent years. The Gini coefficients from 2015 to 2017 were 0.462, 0.465 and 0.467 respectively. The Gini coefficient is partly due to the rural-urban income gap. Therefore, it is urgent to explore the key factors affecting income and alleviate the large income gap.

Studies have shown that there are many reasons for the income gap. According to Shen Ping et al. (2009), the reasons for the widening of income gap can be summarized as policy preference, property, marketization, reform of employment system, rule of law, human capital difference and statistical methods. Li Shi et al. (2018) also pointed out that the income gap is caused by the structure of income distribution, the unfair tax burden, the difference of public services and the lag of the reform of the household registration system. In addition, Sun Jingshui (2013) found that the basic characteristics of household head, human capital, regional difference and urban-rural difference have significant influence on household income. The difference in the educational level of household heads contributes the most to the income gap between urban and rural residents.

Previous papers have confirmed the important impact of education on human capital investment, and human capital is the main source of income, therefore, the income gap caused by literacy gap cannot be ignored. Tu Bin et.al.(2020) found that a education philanthropy program increases well-being via human capital accumulation for needy children significantly. Yang Juan et al. (2015) put forward that an individual's natural endowment and acquired education are the two most

  DOI: 10.25236/emss.2021.064

important factors affecting income inequality. Among them, compulsory education plays a more important role. Families' educational choices and public education policies can aggravate disparities in human capital and income, thereby enhancing the intergenerational correlation of income. Urban-rural education inequality will aggravate the urban-rural income gap. Meanwhile, the urban-rural income gap will also aggravate the urban-rural education inequality (Lv Wei et al., 2015). Education inequality will further widen the income gap, which has a direct impact on the educational level of individuals. In other words, people with better education earn more, and people with less education earn less. If this situation is not improved, it will form the ‘Matthew Effect’, which will lead to social instability, reduce people's happiness and further hinder social development.

With the advent of the Internet era, while technological developments may create a new digital divide, they are also helping to narrow the income gap between residents. In the context of continuous historic breakthroughs in the development of Internet foundation, Internet application and Internet government affairs in China, the Internet market is increasingly active. Survey data show that by June 2020, the number of Chinese netizens had reached 940 million, and the Internet penetration rate had reached 67.0 percent. The Internet coverage in rural areas is constantly increasing, among which the number of rural Internet users reached 285 million, accounting for 30.4% of the total Internet users, an increase of 30.63 million compared with March 2020.

With the continuous optimization of rural Internet infrastructure and the acceleration of rural Internet popularization, the Internet has become a new approach to create wealth in rural areas. The significance of the Internet for increasing farmers' income and rural revitalization has become increasingly prominent, and "Internet + agriculture" has also gained more widespread attention. In recent years, there has been an increasing amount of papers on Internet is effective in helping farmers increase their incomes. Bauer Johannes-M demonstrated that ICTS (Information and communication technologies) could influence income directly and indirectly. After overcoming the problem of sample self-selection, Internet use can still bring 41.2% ~ 51.1% of farmers' non-agricultural income (Yang Ningze et al., 2019).The promotion and use of the Internet in rural areas can significantly promote rural non-agricultural employment and enrich farmers' income sources, thus promoting rural development (Zhou Dong, 2016).Han Changgen et al.(2017) analyzed that although the income-increasing effect brought by the Internet is different in the eastern and western regions, the influence of Internet popularization on rural residents' income is still significantly positive .

The Internet has a huge impact on people's lives today, influenced by the difference in resource allocation, people with different educational backgrounds use the Internet to achieve different income growth effects: The technical progress of the Internet, which is a type of skill bias, naturally has a greater impact on highly educated workers (Ma Junlong et al., 2017).He Yaping(2019) found that the exacerbation of the urban-rural income gap is caused by the difference in the level of Internet penetration between urban and rural areas, which changes with the different levels of human capital and economic development in different regions. So will Internet use narrow the income gap? Is there a significant difference in the effect of Internet use on income growth due to different educational backgrounds, which will exacerbate the income gap instead? In China, the urban-rural income gap has drawn wide attention from all walks of life, are people in urban areas necessarily better off using the Internet than people in rural areas? This paper will explore the role of the Internet in reducing the income gap caused by the education gap from the perspective of the heterogeneity of different educational backgrounds and areas.

2. Data and methodology

2.1 Data

Based on availability and timeliness, this paper uses 2018 China Family Panel Studies (CFPS) for empirical analysis. The data is provided by the Chinese Social Science Survey Center of Peking University, through the nationally representative sample of village residents, families and family members tracking survey, reflecting the development and changes of Chinese society, economy, population, education and health. The database has covered 32,669 individuals in 31 municipalities, autonomous regions and provinces .It was divided into five parts mainly, according to the research needed in this paper, relevant variables in the personal self-answering questionnaire were selected mainly. Firstly, the individuals with missing values in the control variables were eliminated. For the group without annual income but they have monthly income, their monthly income was multiplied by 12 instead of their annual income. After excluding the individual data with abnormal data, 13,563 sample data were finally obtained. Among them, 6,498 people use the Internet, accounting for 47.91% of the sample number; 7,065 people don't use the Internet, accounting for 52.09% of the sample.

2.2 Variable selection

In order to research the impact of Internet use on income gap, variables were selected as follows:

1) Explained variable

To reduce the absolute value of the data and avoid the heteroscedasticity of income variables interfering with the calculation, the logarithm of personal income from work was selected as the explained variable.

2) Core explanatory variable

The core explanatory variables were education level and whether the residents use Internet. The educational level is shown as the number of years of education completed. People with different educational levels was divided into four groups according to the years of education, which are primary school and below, junior high school, senior high school and junior college and above. There are two corresponding questions on the database questionnaire about whether to use the Internet variable, which are "Do you use mobile devices, such as mobile phones and tablets PC to surf the Internet? "And" Do you use computer to surf the Internet?"

3) Control variables

There are many factors that affect residents' income. By referring to relevant papers and starting from the perspective of residents themselves and their residence, their residents' age, age squared, gender, current marital status, political identity, physical health status, intelligence level, province and residence were selected as urban or rural control variables. The above variables and their descriptions are shown in Table 1.

Table 1 Definition of Variables

Category and variable	Definition
lnincome	Logarithm of work
edu	Numbers of years of education, Continuous variable
webuse	Use Internet=1;do not use Internet=0
age	Continuous variable
squage	Age squared
gender	Male=1;Female=0
marriage	Married=1;otherwise=0
partymember	Partymember=1;otherwise=0
health	health condition, very poor=1;very good=7
intelligence	Intelligence level, very low=1;very high=7
urban	Urban=1;rural=0
provcd	Province

2.3 Descriptive statistics

The descriptive statistics of the variables are shown in Table 2.The average age of interviewees is 48.31 years old, and they are in the stage of relatively stable income. The average of respondents' use of the Internet in 2018 was 0.479, a difference from the standard deviation of 0.500, but not a big difference. In terms of the number of years of education completed, the average number of years of education received by residents is about 7 years, which is equivalent to only receiving the level of one year of middle school education. With the strong support of national policies, residents may still not complete the nine-year compulsory education due to various reasons. With the continuous improvement of China's urbanization, people living in rural areas still account for the majority of the population, with only 45.9% living in urban areas.

Table 2 Descriptive Statistic of Variables

Variable	Mean	Standard deviation	Min	Max
lnincome	0.852	8.478	-6.908	13.64
edu	7.037	4.831	0	22
webuse	0.479	0.500	0	1
age	48.31	13.54	16	87
squage	2517	1310	256	7569
gender	0.506	0.500	0	1
marriage	0.869	0.337	0	1
partymember	0.007	0.085	0	1
health	5.554	1.290	1	7
intelligence	4.943	1.436	1	7
urban	0.459	0.498	0	1
provcd	38.67	15.03	11	65

2.4 Model building

This paper used the CFPS2018 data set, from the perspective of the heterogeneity of different educational backgrounds and living areas, and used the Treatment effect model to further explore whether the Internet can narrow the income gap caused by the education gap.

Because the logarithm of personal income from work is a continuous variable, in order to study the impact of years of education on personal income, OLS was chosen for regression. Therefore, the first baseline model was set as follows:

$$\ln income_i = \alpha + \beta edu_i + \delta X_i + \varepsilon i \qquad (1)$$

Among Model(1), $\ln income_i$ is the logarithmic of personal income from work, and edu_i is the number of years of education completed by the individual. X_i is the control variable, including individual level characteristic variable and regional variable, εi Is the error term.

The first model only research the effect of years of education on a person's income, in order to explore the impact of years of education on personal income after the introduction of $webuse$, the second baseline model was set:

$$\ln income_i = \alpha + \beta edu_i + \gamma webuse_i + \delta X_i + \varepsilon i \qquad (2)$$

Model (2) builds on the first model by introducing $webuse$. $webuse$ is a dummy variable.‘1’ means the residents use the Internet while ‘0’means the residents do not use the Internet. The remaining variables have the same meaning as the variables in the first model.

Based on the second baseline model, the third baseline model continues to introduce the interaction term between Internet use and years of education. The continuous variable edu_i was taken as the core explanatory variable, and the dummy variable was $webuse$.

$$\ln income_i = \alpha + \beta edu_i + \gamma webuse_i + \varphi webedu_i + \delta X_i + \varepsilon i \qquad (3)$$

Subsequent models and heterogeneity analyses were conducted based on the above three baseline models.

In the second model, if the use of the Internet is regarded as an exogenous variable, OLS regression can be used to analyze whether the Internet can effectively reduce the income gap. But whether residents use the Internet is a kind of self-selection behavior. Lopez-sintas Jordi et.al.(2020) Analyzed that most people use Internet is that they have available resources around them, at the same time, educational qualifications, age and gender also affect their use Internet. And Internet use can be influenced by younger members of the family (Michailidis Anastasios et al., 2011). As we all know, residents in developed areas have more opportunities or needs to get in touch with the Internet than residents in less developed areas. In some underdeveloped areas, maybe They don't have access to the Internet. Generally speaking, residents can be divided into two categories according to their own situation: using the Internet (Webuse =1) and do not use the Internet (Webuse =0).Therefore, the Treatment Effects Model proposed by Maddala in 1983 was adopted in this paper for more accurate processing, and the results were compared with the results of OLS regression.

$$\ln income_i = \alpha edu_i + \beta D_i + \gamma X_i + \varepsilon i \tag{4}$$

The explained variable is $\ln income_i$, D_i is dummy variable, indicating whether the Internet will be used. If the resident uses the Internet, $D_i = 1$;Conversely, if the residents do not use the Internet, $D_i = 0$. edu_i is the number of years of education completed, X_i is the exogenous explanatory variable to measure personal characteristics such as age, gender, political status, physical health level, etc. α , β and γ is the coefficient to be estimated, εi is the random error term.

In Model (4), if D_i is an exogenous variable, OLS regression can be used to analyze the impact of Internet use on individual total income from work. However, as mentioned above, whether to use the Internet or not is a self-selection behavior, so OLS regression cannot be used directly.

Internet use is a processing variable. Assume that the processing variables are determined by the following processing equation:

$$D_i = 1(z_i\delta + u_i) \tag{5}$$

In Model (5), $D_i = 1(\bullet)$ is the indicative function,and z_i is the observable control variables, but at least one variable is not in X_i , and it is assumed that the disturbance term (u_i, ε_i) obeies the two-dimensional normal distribution:

$$\begin{pmatrix} \mu_i \\ \varepsilon_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\mu^2 & \rho\sigma_\mu \\ \rho\sigma_\mu & 1 \end{pmatrix}\right] \tag{6}$$

In Model (6), ρ is the correlation coefficient of (u_i, ε_i) .If there is a correlation between the two error terms,that’s to say $\rho \neq 0$, proving that the model is endogeneous, therefore OLS regression cannot be used directly to estimate Model (4).If $\rho = 0$, it means that there is no endogeneity in the model, and OLS can be directly used for regression. Treatment effects model estimates Average Treatment Effect on the Treated(ATT). ATT can be calculated by the following equation:

$$ATT = E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 1) \tag{7}$$

In Model (7), $E(y_{1i} \mid D_i = 1)$ represents the income of residents who now use the Internet after using the Internet; And $E(y_{0i} \mid D_i = 1)$ shows the pre-Internet income of residents who use the Internet now. The estimation bias caused by observable and unobservable factors can be controlled by Treatment effect model.

3. Results

3.1 Analysis of baseline model regression results

Table 3 shows the baseline model regression results. According to the regression results of Model (1), a positive correlation was found between the degree of education and income. That is to say, for each additional year of education, the logarithm of personal total income from work will increase by 0.056.The use of the Internet also has a significant positive effect on income. The estimated coefficient is 0.296, which is significant at the p=0.01 level. At the same time, the contribution of completed years of education to the logarithm of individual total income from work decreased from 0.056 to 0.049, and it was significant at the statistical level of 1%.

Table 3 Baseline model regression results

Variable	Model(1)	Model(2)	Model(3)
lnincome	0.056*** (0.002)	0.049*** (0.002)	0.042*** (0.003)
edu	-	0.296*** (0.022)	0.186*** (0.038)
webuse	-	-	0.015*** (0.004)
age	0.054*** (0.004)	0.059*** (0.004)	0.062*** (0.004)
squage	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
gender	0.521*** (0.017)	0.519*** (0.017)	0.526*** (0.017)
marriage	0.196*** (0.027)	0.192*** (0.027)	0.196*** (0.027)
partymember	0.191* (0.100)	0.187* (0.099)	0.618* (0.099)
health	0.053*** (0.008)	0.051*** (0.008)	0.051*** (0.008)
intelligence	0.002 (0.008)	-0.004 (0.008)	-0.004 (0.008)
urban	0.278*** (0.018)	0.259*** (0.018)	0.253*** (0.018)
provcd	controlled	controlled	controlled
R-squared	0.3935	0.4013	0.4019
F-statistic	250.79***	251.82***	245.59***
Sample	13563	13563	13563

Note(s): Standard errors in parentheses ***p<0.01, **p<0.05,*p<0.1

In conclusion, the significant relationship between degree of education and income is actually affected by the Internet. For those who cannot stay in school for some reasons, other resources and information provided by the Internet can also bridge the income gap with those with higher qualifications.

The results show that the years of education, whether to use the Internet or not, and their interaction items have a significant positive impact on income. The coefficient of interaction term is 0.015, indicating that with the increase of years of education, individuals who use the Internet have higher income than those who do not use the Internet. In order to further analyze the use of Internet benefits from education, this paper continued to divide the educational level into four different groups: primary school and below, junior high school, senior high school and junior college and above, and conducted regression analysis by group.

Table 4 Regression results of different educational levels

Variable	Grouped by years of education			
	Primary school and below	Junior high school	Senior high school	College degree or above
webuse	0.306*** (0.036)	0.282*** (0.035)	0.283*** (0.055)	0.375*** (0.115)
Control variable	controlled	controlled	controlled	controlled
Sample	6452	4110	1786	1215

Note(s): Standard errors in parentheses ***p<0.01, **p<0.05,*p<0.1

As described in Table 4, people with higher education backgrounds benefit more from the use of the Internet. The highly educated have more knowledge, more skills, more vision, more access to more resources, and their incomes will double for the help with Internet. For the people with low education, the Internet provides more information and opportunities, giving the people more possibilities. Therefore, people with low education should pay more attention to the investment in their human capital, and also can learn more knowledge related to the Internet, so that the Internet can become a tool to help them realize wealth. Governments and companies should also focus on training people or employees with only a low level of education in some skills, making the Internet play a more important role in increasing incomes and narrowing income gaps.

3.1.1 Heterogeneity analysis

(1) Area disparity

As is known to everyone, urban education resources and infrastructure construction are far better than rural areas, so the Internet penetration rate in urban areas is generally higher than that in rural areas at the same time. There are also differences in the level of education and Internet use among different income groups. Therefore, from the perspective of heterogeneity, this paper continues to explore the role of the Internet plays in urban and rural areas with huge disparities in educational resources, as well as the impact of educational qualifications and Internet use on residents' income.

Table 5 Regression results of different areas

Variable	Model(1)		Model(2)		Model(3)	
	urban	rural	urban	rural	urban	rural
edu	0.061*** (0.003)	0.049*** (0.003)	0.053*** (0.003)	0.043*** (0.003)	0.044*** (0.004)	0.037*** (0.004)
webuse	—	—	0.288*** (0.031)	0.277*** (0.032)	0.154*** (0.056)	0.163*** (0.053)
The interaction between edu and webuse	—	—	—	—	0.017*** (0.006)	0.017*** (0.006)
Control variable	controlled		controlled		controlled	
Sample	6225	7338	6225	7338	6225	7338

Note(s): Standard errors in parentheses ***p<0.01, **p<0.05,*p<0.1

Models (3), (4) and (5) are further subdivided into urban and rural areas on the basis of models (1), (2) and (3).According to the regression results in Table 5, compared to the people with low education who live in rural areas,it’s more useful for people with low education who live in urban areas using Internet to reduce income gap between them and highly educated people. This result is not surprising, it reflects why most people prefer to move to big cities. Because big cities are rich in opportunities and resources, the use of Internet may be more effective in big cities than in rural areas.

There are also income gaps, education gaps in rural areas: Internet use and non-use. Model(5) is set to further explore the results, the interaction term between Internet use and years of education was continued to be introduced .In this model, the explanatory variable is the interaction term mentioned above, and the explained variable is the logarithm of personal total income from work. The results of the interaction term indicate that under the condition of Internet use, whether the residents live in city or country, the higher the level of education, the greater the income difference will be caused. Even if the information obtained is the same as that obtained by others, but they cannot make the best use of the information due to their lack of ability, which is a kind of waste of resources and also widens the income gap within the region.

(2)Income disparity

For purpose of exploring the impact of educational qualifications and Internet use on different income groups, quantile regression was further introduced. The results are shown in Table 6.

Table 6 Quantile regression results

β	0.25	0.50	0.75
edu	0.064*** (0.004)	0.052*** (0.003)	0.039*** (0.003)
webuse	0.359*** (0.036)	0.240*** (0.024)	0.201*** (0.015)
age	0.067*** (0.008)	0.067*** (0.004)	0.058*** (0.004)
squage	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
gender	0.579*** (0.025)	0.524*** (0.015)	0.458*** (0.016)
marriage	0.236*** (0.044)	0.149*** (0.025)	0.081*** (0.026)
partymember	0.123* (0.070)	0.097 (0.098)	0.142 (0.141)
health	0.053*** (0.010)	0.049*** (0.009)	0.041*** (0.009)
intelligence	-0.009 (0.012)	-0.001 (0.010)	0.006 (0.011)
urban	0.356*** (0.032)	0.226*** (0.026)	0.161*** (0.018)
Pseudo R2	0.2746	0.2600	0.2139
Observation	13563	13563	13563

Note(s): Standard errors in parentheses ***p<0.01, **p<0.05,*p<0.1

As can be seen from Table 6, as the distribution of the logarithm of personal total income from work increases from the low quantile to the high quantile, the influence of educational qualifications on it shows a gradually decreasing trend as a whole. Meanwhile, Table 6 shows the same trend about Internet use, it just declining at different rates across different quantiles.

Specifically, the effect of years of education on the logarithm of personal total income from work decreased from 0.064 to 0.052 on the range of 25 to 50 quantiles, the 50 to 75 quantiles, the coefficient continued to drop to 0.039.The results are both significant at the statistical level of 1%.This means that the marginal impact of education on income is greater for people on lower and middle incomes than for people on higher incomes. In other words, those who continue to improve in education at low - and middle-income levels will have higher returns than those who has already stayed at higher income levels. This may because high-income groups have more years of schooling. There are also various costs associated with receiving education. Continuing education with a relatively high level of education does not have a significant impact on increasing income. For low - and middle-income groups, those groups may not be very educated or even less educated, if they would continue their education and invest more in human capital. They will get a higher return.

A similar conclusion can be drawn from Table 6 about the marginal impact of the Internet on groups of different income groups. From 25 to 75 quantiles, the coefficient decreased from 0.359 to 0.201, and the parameter on the 50 quantile was 0.240, both of which were significant at the 1% statistical level. The decline from the middle to the top was more gradual than the decline from the low to the middle. This could mean that the Internet has opened up a wider range of possibilities for low - and middle-income groups. Their current jobs may not pay very well, but using the Internet to continue learning can increase their human capital and enable them to move up in the workplace, or they can use the Internet to do some other work to improve their income.

3.2 Treatment effect model

3.2.1 Endogeneity examination

In above models, the problem of endogeneity cannot be ignored. There are two reasons for endogeneity. Most endogeneity is caused by missing variables. In this paper, personal ability can affect both the level of education and income, Therefore, when estimating the impact of education on income, endogeneity due to missing variables is possible. To avoid this problem, individual’s health condition, intelligence and other key variables may affect education and income were controlled, as far as possible to control the endogeneity. What’s more, Internet use and personal income interact as both cause and effect. That means it's not clear whether Internet use is driving income growth, or whether it's because increased income makes people use the Internet. To solve the endogeneity, instrumental variable will be introduced to solve the problem. Appropriate IVs should be selected under two major criteria as required. One is that IVs should be highly correlated to the endogenous variable webuse, the other is that IV should be uncorrelated to the explained variable. In this paper, the selection of instrumental variables refers to the practice of Shan Depeng et al. (2020), "the importance of the Internet as an information channel" is chosen as the instrumental variable to deal with the endogeneity problem.

To ensure the rationality of the selection of instrumental variable, it is necessary to check the instrumental variable. According to the 2SLS regression test, the robust F statistic of the first stage regression is 449.33, which is far greater than the commonly used critical value of 10.In addition, compared with the OLS regression results of the Internet use and logarithm of total income from work, the regression results of 2SLS in Table 7 are significantly improved.In other words, models without instrumental variables may underestimate the impact of Internet use on an individual's total work income.

Table 7 2SLS regression results

Variable	Webuse	lnincome
Importance of the Internet as an information channel	0.126*** (0.002)	-
webuse	-	0.602*** (0.052)
Control variable	controlled	controlled
R-squared	0.5446	0.3930
F-statistic	449.33	8928.05
Observation	13563	13563

Note(s): Standard errors in parentheses ***p<0.01,**p<0.05,*p<0.1

3.2.2 Results of Treatment effect model

To avoid the errors caused by self-selection mentioned above, this paper restudied the relationship between Internet use and personal income by using the treatment effect model.

In order to explore the impact of the Internet on residents with different education levels and living areas, this paper continued analysis by grouping education and area.

Table 8 Treatment Effect Model for different educational levels

	Primary school and below	Junior high school	Senior high school	College degree or above
ATT	0.633***	0.473***	0.436***	0.652***
Std.Error	0.073	0.073	0.100	0.196
z	8.62	6.50	4.38	3.33
P>|z|	0.000	0.000	0.000	0.001
Control variable	controlled	controlled	controlled	controlled
Observation	6452	4110	1786	1215

Note(s): Standard errors in parentheses ***p<0.01, **p<0.05,*p<0.1

After overcoming the self-selection problem, we found that the p-value of ATT was statistically significant at the 1% level for participants in the primary school group and below, the middle school group and the junior college group and above. Comparing the results in Table 8 with those in Table 4, the coefficient of Internet use on personal income in OLS is significantly smaller than that in TEM.OLS underestimates this educational benefit compared to Timbit there is something in common between the two approaches: the impact of Internet use on income in the primary school and below group and in the college and above group is more obvious compared with the middle school group. At the same time, the income of people with higher education is higher than that of people with low education. Therefore, on the whole, both the results of OLS regression and the treatment effect model indicate that there are educational benefits in the use of the Internet, but there are also certain barriers to such educational benefits. In other words, when the education level reaches a certain level, the income increase effect brought by residents' use of the Internet will be more obvious.

Table 9 Treatment Effect Model for different areas

	Urban	Rural
ATT	0.513***	0.527***
Std.Error	0.056	0.066
z	9.13	8.00
P>|z|	0.000	0.000
Control variable	controlled	Controlled
Observation	6225	7338

Note(s): Standard errors in parentheses ***p<0.01, **p<0.05,*p<0.1

If basing on the heterogeneity of different areas, the results in Table 9 are relative to Table 5, the effect of Internet on income in TEM is also more significant than that in OLS, and is very significant. That indicates OLS also underestimates the positive effect of Internet use on revenue. TEM results show that residents in rural areas are more likely to increase their income by using the Internet than those in urban areas. The use of the Internet by residents in urban districts increased the logarithm of personal income from work by 0.513 units. The use of the Internet by residents in rural areas can increase the logarithmic of personal income from work by 0.527 units, and they are both significant at the statistical level of 1%.The survey conducted by Mora-rivera Jorge et al.(2021) found that the Internet has been more effective in reducing extreme income poverty and extreme multidimensional poverty in rural areas than urban areas. This may because the Internet has a more pronounced effect on changing the lifestyles of rural residents. Presumably there are the following reasons: First, the Internet has greatly remedied the problem of information asymmetry in rural areas, giving rural residents the opportunity to acquire knowledge and improve their skills, because a large portion of the income gap comes from the asymmetry of information and resources. On the

Internet, most of the information available to people is the same. Second, the use of the Internet can change the way farmers work. Rural residents can access the latest policy information through the Internet and make development plans in advance. Farmers can also learn about the use of modern agricultural technologies on the Internet, they can buy and use agricultural machinery and other tools to improve productivity if they can afford. Third, the Internet enables residents in rural areas to keep pace with The Times and thus find new jobs and lifestyles. For example, through the use of the Internet, rural residents can join the e-commerce market at a relatively low cost and benefit from it. Internet access can significantly promote the rate of rural migrant workers return to their hometowns to start up business, and at the same time greatly change the lifestyle of rural residents, making the ways more efficient for them to increase their income(Yuan Fang et al.).They can use new skills through Internet to make a living, earning higher incomes, work more flexibly and increase happiness (Fahmi Fikri-Zul et al.2020).As Canh Nguyen-Phuc et al.(2020)said: the Internet should actually be a new tool for reducing income inequality.

4. Conclusion

4.1 Summary and discussion

In this paper, CFPS2018 database, OLS and treatment effects model were used, basing on the perspective of heterogeneity, analyzing the influence of Internet use on residents with different educational levels and living areas. The results show that Internet use has a significant impact on reducing the income gap caused by the inequality of educational resources. But at the same time, the Internet is widening the income gap while narrowing it. Because the use of the Internet has educational benefits, only with a certain level of education, the use of the Internet will bring better income. After adjusting for the self-selection of Internet use, the results showed that rural residents use the Internet to bring about an increase in income effect is obviously better than urban residents.

Most of the current literature explored the relationship between the changes in farmers' income caused by the Internet, but their own educational level will also affect their ability to use the Internet to obtain information. Moreover, the data that former scholars used was mostly concentrated on the provincial panel data. The national situation is complex and there is a significant gap between the rich and the poor in different regions, so the overall conclusion cannot be made, or they based on a survey of rural residents in a specific province to analyze, so the conclusion is not universal. Secondly, scholars mostly focused on the wealth gap between urban and rural areas, ignoring the changes in income within rural areas. Finally, the current domestic literature mainly focuses on the analysis of theoretical factors in domestic, while the empirical study is still insufficient.

But at the same time, there are also some limitations in this paper. Due to the limitation of data availability, this paper only selected data from 2018CFPS, and did not include relevant data of other years including future years in this study.As a result, the conclusion of this paper cannot prove the trend that the Internet can make up the income gap caused by the education gap, which has a certain timeliness. Secondly, provincial variables are controlled in the empirical study of this paper. However, there are obvious differences in the development of different regions in China, so the research results of this paper can not be generalized for different provinces and regions.

Future research can continue to take the development of the Internet as the background and use relevant tracking data to study whether the conclusion of this paper is still valid at the time series level.

4.2 Policy Suggestions and Prospects

The above conclusions mainly have three policy suggestions: First, the state should make efforts to guarantee the right of residents to receive compulsory education, and encourage more residents to receive higher education. Due to the limitation of data availability, the results still show that many Chinese residents are poorly educated, most of them have not completed their junior middle school education. This will not only make them miss the opportunity due to their lack of ability, but also make the development process of the whole society become slow. Second, the country should increase investment and speed up the construction of infrastructure in rural areas. The 5G era is coming, but according to the data, there are still up to 50 percent of residents who do not use the Internet, including many rural residents. The inequality caused by the inequality of educational resources should be made up in the aspect of network resources. Third, the relevant government departments should actively train rural residents to use the Internet. For rural residents, the Internet is not only a means of entertainment, but also a way to become rich.

According to the results mentioned above, it is hoped that China can speed up the construction of infrastructure in rural areas, implement the basic education work in China, ensure that the majority of rural residents could use and make good use of the Internet, making the Internet become a powerful tool for people to become rich.

References

[1] Li Shi, Zhu Mengbing. Changes of Residents' Income Inequality in the 40 Years of China's Economic Transformation [J].Management World, 2018(v.34): 19-28.

[2] Shen Ping, Zhu Chunkui. Literature Review on the Income Gap of Chinese Residents [J]. Journal of the Party School of the CPC Ningbo Municipal Committee, 2009(v.31; No.167): 56-61.

[3] Sun Jingshui, Huang Qiuhong. Research on the Main Influential Factors and Contribution Rate of the Income Gap between Urban and Rural Residents in China: Based on the Data Analysis of 6937 Household Questionnaires Survey in 31 Provinces of China [J].Economic Theory and Management, 2013(270): 5-20.

[4] Tu Bin,Huang Chien-Chung,Sorensen Jennifer. Effects of education philanthropy on well-being of low-income and gifted students in China[Z]: Children and Youth Services Review, 2020: 0-0.

[5] Yang Juan, Lai Desheng, Qiu Muyuan. How to reduce income inequality through education?[J]. Economic Research Journal, 2015(v.50;No.576): 86-99.

[6] Lv Wei, Yang Mo, Wang Yan. Urban-rural Income Gap, Urban-rural Educational Inequality and Government Investment in Education [J].Comparison of Economic and Social Systems, 2015(179): 20-33.

[7] Bauer Johannes-M.. The Internet and income inequality: Socio-economic challenges in a hyperconnected society[Z]: Telecommunications Policy, 2018: 0-0.

[8] Yang Ningze, Zhou Jing. Can Internet Use Increase Farmers' Non-farm Income?-- Empirical Analysis Based on the Data of Chinese General Social Survey (CGSS) in 2015 [J].Economic Jingwei, 2019(v.36;No.192): 41-48.

[9] Zhou Dong. Research on the effect of Internet coverage driving rural employment [J]. World Economy Review, 2016(232): 76-90.

[10] Han Changgen, Zhang Li. The Impact of Internet Popularization on Urban and Rural Income Distribution -- Based on the System GMM Analysis of China's Provincial Panel Data [J].Research on Economic Issues, 2017(No.421): 18-27.

[11] Ma Junlong, Ning Guangjie. The Internet and China's Rural Labor Force's Non-agricultural Employment [J]. Science of Finance and Economics, 2017(No.352): 50-63.

[12] He Yaping, Xu Kangning. The Impact of the Internet on the Urban-Rural Income Gap: A Test Based on China's Ffacts [J].Economic Jingwei, 2019(v.36;No.189): 25-32.

[13] Lopez-sintas Jordi,Lamberti Giuseppe,Sukphan Jakkapong. The social structuring of the digital gap in a developing country. The impact of computer and internet access opportunities on internet use in Thailand[Z]: Technology in Society, 2020: 0-0.

[14] Michailidis Anastasios,Partalidou Maria,Charatsari Chrysanthi. Who goes online? Evidence of

internet use patterns from rural Greece[Z]: Telecommunications Policy, 2011: 0-0.

[15] Liu Tongshan. Does farmers' willingness to withdraw contracted land affect grain production? China Rural Economy, 2017(No.385): 68-81, 95-96.

[16] Zhang Jinhua, Liu Jing. The Migration Effect and Heterogeneity of Migrant Workers' Educational Returns: An Empirical Study Based on the Processing Effect Model [J].Agrotechnical Economics, 2018(273): 73-83.

[17] Shan Depeng, Zhang Yongqi. Internet Use and Farmers' Personal Income: Microscale Evidence and Influence Mechanism [J].North China Finance, 2020(No.524): 82-89.

[18] Mora-rivera Jorge,García-mora Fernando. Internet access and poverty reduction: Evidence from rural and urban Mexico[Z]: Telecommunications Policy, 2021: 0-0.

[19] Fahmi Fikri-Zul,Sari Ivanie-Destila. Rural transformation, digitalisation and subjective wellbeing: A case study from Indonesia[Z]: Habitat International, 2020: 0-0.

[20] Yuan Fang, Shi Qinghua. From Returning to Entrepreneurship -- An Empirical Analysis of the Impact of Internet Access on the Decision Making of Migrant Workers [J].Southern Economy, 2019(No.361): 61-77.

[21] Canh Nguyen-Phuc,Schinckus Christophe, Ling Felicia-Chong-Hui. Effects of the internet, mobile, and land phones on income inequality and The Kuznets curve: Cross country analysis[Z]: Telecommunications Policy, 2020: 0-0.