Management Theory and Studies for Rural Business and Infrastructure Development
eISSN 2345-0355. 2023. Vol. 45. No. 4: 453-466
Article DOI: https://doi.org/10.15544/mts.2023.44

EXPLORING THE EFFECT OF INTERNET USAGE ON THE URBAN RURAL INCOME GAP: EMPIRICAL EVIDENCE FROM CHINA

Mingshuai Qin[1], Hao Dong[2], Hong Chen[3], Lijian Qin[4], Wenshuai Qin[5]

[1] *PhD Student, Technical University of Ostrava, Studentská 1770/1, Poruba-Ostrava, Czech Republic, Email address: qinmingshuai66@163.com*

[2] *Lect. University of Southampton, Building 2, 12 University Road, Highfield, Southampton, United Kingdom, Email address: H.Dong@soton.ac.uk*

[3] *Assoc. Prof., Anhui University of Finance and Economics, #962 Caoshan Road, Bengbu City, Anhui Province, China, Email address: chenhong.china@163.com*

[4] *Prof. Anhui University of Finance and Economics, #962 Caoshan Road, Bengbu City, Anhui Province, China, Email address: qinlj28@163.com*

[5] *Master's student, Philippine Christian University - Manila, 1648 Taft Ave.cor Pedro Cil, Malate Manila, Republic of The Philippines, Email address: wenshuai_qin@163.com*

Received 01 08 2023; Accepted 29 08 2023

Abstract
China has witnessed remarkable ongoing digitalization with the rapid spread and adoption of the Internet. However, this remarkable development remains uneven between urban and rural populations, and hence result in different impact on their income. Employing data China General Social Survey 2018, this study explores how internet usage affects income gap between the urban and rural China. Relying on the instrumental variables approach to regression analysis, we prove that internet usage contributes to higher increase in annual income for the urban employed compared to their rural counterparts. The RIF decomposition regression results then reveal the effects of differential urban-rural internet usage ratios, explaining the widened income gap between the urban and rural employed in various income levels. The difference in the returns to urban and rural internet usage narrowed the urban-rural income gap for low - and high-income employed, but further contributed to the urban-rural income gap for the middle-income employed.

Keywords*: China, Internet usage, RIF decomposition regression, Urban-rural income gap.*
JEL code: *R00, J31.*

Introduction

According to the latest official data, the per capita disposable income of urban residents in China increased from 21,810 CNY[1] in 2011 to 47,412 CNY in 2021; the per capita disposable income of rural residents in China increased from 6,977 CNY in 2011 to 18,931 CNY in 2021 (National Bureau of Statistics of China, 2022). It is clearly shown that the disposable income of both urban and rural residents has increased significantly over the past decade, but at the same time the income gap between urban and rural residents has also widened. Some existing studies have attributed this urban-rural income inequality to the inconvenience of population mobility, social welfare disparities, workplace discrimination, and differential access to public services associated with Chinese *hukou* system (Liu, 2005; Song 2014; Zhang et al., 2018; Colas et al., 2019; Song et al., 2019), which was originally established to serve China's planned economy by promoting urban citizens to support the development of heavy industry and by employing farmers to maintain surplus rural products (Donzuso, 2015).

[1] Onshore USD/CNY exchange rate mid-price of about 7.2 in July 2023.

Playing a leading role in the emerging Industry 4.0, China has witnessed remarkable ongoing digitalization with the rapid spread of the Internet and increasing adoption of mobile Internet. By the end of 2021, China has built the world's largest information and communication networks (the Central Government of China, 2022). In June 2022, it is reported that the nation's Internet penetration reached 74.4% (China Internet Network Information Center, 2022). On the one hand, China's digital economy is booming thanks to an extensive Information and Communication Technology (ICT) infrastructure, while the merging mobile technologies are both contributing to economic growth and providing employment opportunities. In 2021, China's mobile ecosystem alone created more than 6 million jobs (GSMA Intelligence, 2021). In the context of digital industrialization, China's industrial structure has undergone profound changes, and this structural change (Matthess et al., 2020) along with China's hukou system is affecting labor mobility and distribution across industries. On the other hand, as migrant workers have become an important labor supply for the Chinese economy (Han, et al., 2014), the role of migrant workers has been under scrutiny. In China, migration to cities for higher paying jobs has become a major way to increase rural households' income (Yang & Mukhopadhaya, 2022).

Hence, this study sets to investigate how internet usage shapes the income gap between the urban and rural employed. With empirical data from China, this research examines the impact of internet usage on individuals' and groups' income, the developing pattern, and associated factors. In the next section, the article will review the existing relevant studies. Section Three will introduce the data and their statistical descriptions and present the models and research methods used in this article. The empirical regression analysis is developed and discussed respectively in Sections Four and Five, and the latter interprets the decomposition regression results. Section Six summarize this paper and propose recommendations.

Literature review

Despite the various conceptualized terminological referents, the relationship between internet usage and differences in returns to the employed or similar themes have been widely studied. Studies on the United States described how rural communities lag urban areas in terms of income, wealth, and provision of public services during urbanization, albeit the large-scale adoption of ICT (Glasmeier, 2018). This can be partially explained by the fact that ICT availability, adoption, and use remain unevenly distributed in a way that favors cities and disadvantages rural areas (Braesemann et al., 2022). The impact of internet usage on income inequality was also explored in a study based on 87 economies around the world (Canh et al., 2020). The study found that in the short and long term, promoting the spread of the Internet and mobile communications appears to be a way to moderate income inequality, and its recommendations support that internet usage and mobile communications should in practice be part of an active economic policy aimed at reducing income inequality. Within the framework of the digital divide and income inequality, several studies have confirmed that at the individual level, all else being equal, efforts to increase internet usage will improve individuals' income opportunities (Van Deursen, 2017). At an aggregate level, however, the relationship between the digital divide and income inequality exhibited a stronger dynamic. Lower digital inequality was usually associated with lower income inequality. However, an opposite dynamic affects individuals and households, independent of their level of access and use (Bauer, 2018).

The diffusion of ICT is one of the drivers of economic growth (Lakhanpal, 2021; Arvin et al., 2021), so the application and dissemination of the Internet has been widely valued and studied in underdeveloped regions. An empirical study based on 48 African countries confirmed that the increase in internet penetration reduces the Gini and Palma ratios, and the study concluded that the

sciendo
Management Theory and Studies for Rural Business and Infrastructure Development
eISSN 2345-0355. 2023. Vol. 45. No. 4: 453-466
Article DOI: https://doi.org/10.15544/mts.2023.44

reduction of inequality through the design of policies that promote access to ICT is in line with the sustainable development goals pursued in the world today (Asongu & Odhiambo, 2019). Studies involving Southeast Asian countries also shown that internet penetration, as a proxy for technological change, has significantly reduced income inequality (Ningsih & Choi, 2018). Elements of the far-reaching impact of ICT include economic growth and income inequality, and the importance of connectivity was also reflected in the Indian government's ongoing efforts to promote internet penetration in the context of the existence of a clear digital divide between the rich and the poor (Maiti et al., 2020).

The focus returns to research on Chinese internet usage and income disparity. Income disparities associated with China's unique hukou system and the urban-rural dichotomy it underpins have been studied in high densities. Fong (2009) has been conducted to assess the developmental relationship between ICT adoption rates and the per capita income gap between urban and rural areas in China, and this income gap is strongly correlated with the developmental relationship between the adoption of the Internet, cell phones, personal computers, and telephones. Liu, et al. (2021) also discussed that internet penetration plays a crucial role in boosting the income growth of rural residents and narrowing the gap between urban and rural areas, and that internet penetration in rural areas should be increased to help rural economic and social development. Wang and Zhang (2019) examined the nonlinear impact of Internet penetration on the urban-rural income gap using provincial panel data from 2003-2016. The results shown that the impact of Internet penetration on the urban-rural income gap in China presents an inverted U-shaped trend, i.e., it first increases and then decreases. Moreover, their research further analyzed the reasons for the decreasing effect of Internet penetration and finds that, at the micro level, the income impact of internet usage is greater for rural residents than for urban residents. Based on longitudinal survey data from the China Family Panel Studies (CFPS), Ma (2022) found higher returns to internet usage for rural residents after simultaneously addressing heterogeneity and other endogeneity issues. Then, drawing on the Blinder-Oaxaca decomposition approach, verifies that differences in Internet access widen the income gap, while differences in returns to internet usage narrow the income gap. Also, the impact of internet usage on the income gap differs across populations, with the younger generation contributing a larger value to the income gap than their counterparts in terms of returns to internet usage. Mao et al. (2021) explored the relationship between internet usage and urban-rural income gap using RIF decomposition regression based on mixed China General Social Survey (CGSS) data, and the empirical study found that internet usage has a widening effect on the household wage gap, but it does not persist, and gradually decreases as the network skills and socioeconomic status of migrant workers improve.

Data and research methods

Data source. This paper employs the latest China General Social Survey (CGSS) data collected and collated under the leadership of Renmin University of China, i.e., CGSS 2018. Through stratified sampling, this survey data was obtained from households in 23 provinces, autonomous regions, and municipalities in mainland China. The data covers multi-dimensional information on internet usage, personal characteristics, income and job characteristics, which meet the needs of this research. CGSS data are widely used in academic studies on Chinese economy and are highly authoritative and widely representative. After filtering the data, the employed group between the ages of 18 and 59 with working experience and non-student status was retained for the research. Finally, a total of 6,319 valid samples were obtained, including 3,643 rural employees and 2,676 urban employees.

Regarding the basic demographic information of the respondents, the average age of the respondents in the entire sample was about 42.7 years, the percentage of Han Chinese was above 90%, and more than 80% of the respondents were married or cohabiting.

Rural employed respondents had about 1.5 children, while urban employed respondents had about 1 child, indicating that, on average, rural employed respondents raised more children than urban employed respondents. In addition, statistical description also introduces dependent variable, explanatory variable, and control variables in table 1.

Table 1. Statistical Description and Between-sample Difference Test

VARIABLES	Definition	Full sample	Rural *Hukou*	Urban *Hukou*	T-statistics
		Mean	Mean	Mean	
lnincome	Logarithmic form of annual income, ln(income+1).	9.059 (3.606)	8.479 (3.723)	9.849 (3.280)	15.234***
Internet usage	Using the internet as 1; otherwise, 0	0.823 (0.382)	0.746 (0.435)	0.928 (0.259)	19.262***
SoI	Whether to use the Internet as a primary source of information, yes as 1; otherwise, 0.	0.938 (5.364)	0.951 (6.274)	0.922 (3.789)	-0.212
Age	Respondent's date of interview minus the respondent's date of birth	42.738 (10.356)	42.741 (10.487)	42.734 (10.176)	-0.008
Gender	Male as 1; female as 0	0.473 (0.499)	0.465 (0.499)	0.485 (0.500)	1.600
Generation	The older generation was born before 1980 and the newer generation after that. New generation as1; Old generation as 0	0.366 (0.482)	0.362 (0.481)	0.371 (0.483)	0.743
Han	Han Chinese as 1; non-Han Chinese as 0	0.922 (0.268)	0.902 (0.297)	0.950 (0.218)	7.066***
Married	Married or cohabiting as 1; Single or divorced or widowed as 0	0.840 (0.366)	0.868 (0.339)	0.804 (0.397)	-6.942***
Children	Number of children	1.363 (0.970)	1.583 (1.057)	1.064 (0.737)	-21.779***
Hukou	Type of household register, Rural *hukou* as 1; Urban *hukou* as 0	0.576 (0.494)	-	-	-
Education					
Primary school and below	Education level of primary school and below as 1, others as 0	0.248 (0.432)	0.370 (0.483)	0.083 (0.275)	-27.659***
Junior high school	Education level of junior high school as 1, others as 0	0.312 (0.464)	0.392 (0.488)	0.204 (0.403)	-16.180***
Senior high school	Education level of senior high school as 1, others as 0	0.202 (0.402)	0.154 (0.361)	0.268 (0.443)	11.201***
College degree and above	Education level of college degree and above as 1, others as 0	0.237 (0.426)	0.085 (0.278)	0.445 (0.497)	36.697***
Father's highest education	Father's highest education. Primary school and below as 1; Junior high school as 2; High school as 3; College degree and above as 4	1.567 (0.865)	1.324 (0.627)	1.898 (1.021)	27.577***
Mather's highest education	Mather's highest education. Primary school and below as 1; Junior high school as 2; High school as 3; College degree and above as 4	1.362 (0.724)	1.149 (0.437)	1.652 (0.913)	29.087***
Physical health					
Very bad	Self-rated health, very bad as 1, others as 0	0.021 (0.144)	0.027 (0.163)	0.013 (0.114)	-3.851***
Bad	Self-rated health, bad as 1, others as 0	0.103 (0.304)	0.125 (0.331)	0.073 (0.261)	-6.653***

sciendo
Management Theory and Studies for Rural Business and Infrastructure Development
eISSN 2345-0355. 2023. Vol. 45. No. 4: 453-466
Article DOI: https://doi.org/10.15544/mts.2023.44

Moderate	Self-rated health, moderate as 1, others as 0	0.192 (0.394)	0.182 (0.386)	0.206 (0.404)	2.370**
Good	Self-rated health, good as 1, others as 0	0.433 (0.495)	0.413 (0.492)	0.459 (0.498)	3.635***
Very good	Self-rated health, very good as 1, others as 0	0.251 (0.434)	0.253 (0.435)	0.249 (0.432)	-0.373
Frequency of depression					
Always	Frequency of depression: Always as 1, others as 0	0.011 (0.105)	0.012 (0.109)	0.010 (0.100)	-0.744
Often	Frequency of depression: Often as 1, others as 0	0.063 (0.244)	0.074 (0.262)	0.049 (0.215)	-4.171***
Sometimes	Frequency of depression: Sometimes as 1, others as 0	0.232 (0.422)	0.250 (0.433)	0.208 (0.406)	-3.894***
Seldom	Frequency of depression: Seldom as 1, others as 0	0.389 (0.488)	0.394 (0.489)	0.383 (0.486)	-0.847
Never	Frequency of depression: Never as 1, others as 0	0.304 (0.460)	0.270 (0.444)	0.350 (0.477)	6.870***
Contract	Employment contract signed as 1; not signed as 0	0.273 (0.445)	0.139 (0.346)	0.455 (0.498)	29.717***
Medical insurance	Participation in public medical insurance as 1; otherwise, 0	0.915 (0.280)	0.900 (0.300)	0.934 (0.248)	4.814***
Pension insurance	Participation in basic pension insurance as 1; otherwise, 0	0.705 (0.456)	0.636 (0.481)	0.798 (0.401)	14.188***
Region					
Eastern region	Eastern region as 1, otherwise 0	0.454 (0.498)	0.340 (0.474)	0.610 (0.488)	19.458***
Central region	Central region as 1, otherwise 0	0.309 (0.462)	0.350 (0.477)	0.253 (0.435)	-7.810***
Western region	Western region as 1, otherwise 0	0.237 (0.425)	0.310 (0.463)	0.138 (0.344)	-16.291***
Observations		6319	3643	2676	-

**Source: Values in parentheses are standard deviations. *** p<0.01, ** p< 0.05, * p<0.1.*

Variables. Dependent variable

$\ln income$: the logarithmic form of annual wage income plus one, and the statistical description indicates that the income level of the rural employed is lower than that of the urban employed and the full sample. The corresponding t-statistic is significant at the 1% level, indicating a significant difference in income between urban and rural employees.

Key explanatory variable

Internet usage: Internet represents a binary dummy variable indicating whether the respondent has used the Internet, including using various devices such as computers, cell phones, and smart wears. The prevalence of internet usage was 74.6% for rural employed persons and 92.8% for urban employed persons. The t-statistic for internet usage between the two samples was significant at the 1% level, indicating that Internet penetration was significantly lower among the rural employed than their urban counterparts.

Control variables

Based on the existing literature, this paper considers the role of education, work, health, and regional characteristics on the annual incomes of employed individuals, transforming these factors into quantifiable dummy variables and incorporating them into the empirical study.

Models and methods. Referring to the set-up form of Mincer's wage equation (1974, 1997) and Krueger's (1993) econometric model of the relationship between the Internet and wage earnings, the benchmark regression model is designed in this paper (see equation 1).

$$\ln income = \alpha + \beta \, \mathrm{internet_i} + \gamma \, \mathrm{X_i} + \varepsilon_i \tag{1}$$

The dependent variable $\ln income$ is the logarithmic form of annual wage income plus one. Internet represents a binary dummy variable indicating whether the respondent has used the Internet (i.e., 1 if the respondent has used the Internet, or 0 otherwise). X_i indicates a series of demographic characteristics variables such as gender, age, education level, marital status, and health status. ε is an error term.

Internet usage may increase an individual's income level, but in turn, individuals with higher income levels are more likely to use the Internet. This endogeneity is likely to pose a challenge in examining the causal relationship between internet usage and individual income levels. Referring to other scholars' practices (Ma, 2022; Gong et al., 2020), this paper uses two instrumental variables: (1) the average of the Internet penetration rate in the respondent's province and the Internet penetration rate in the respondent's city at the time of the survey; (2) the length of long-distance broadband fiber lines in the respondent's province in 2015. These two instrumental variables were chosen because they are theoretically correlated with individual Internet usage at both the macro level and the historical level, while having no direct effect on individual annual income.

To mitigate bias, as part of the robustness test, this paper uses PSM to find employed individuals with otherwise consistent characteristics but different internet usage for matching, and then determines the relationship between internet usage and annual income by comparing the difference in annual income between the group that has used the Internet (i.e., the treatment group) and the group that has not used the Internet (i.e., the control group). PSM was first proposed by Rubin et al. (1983) and has the advantage of reducing the dependence on the correct setting of the functional model and alleviating the problem of biased estimation due to selection bias associated with observables (Shipman et al., 2017). In addition, the sample self-selection bias problem may also arise from unobservable or omitted variables that cause some employees to earn higher annual income than others regardless of whether they use the Internet or not, which in turn affects their return to internet usage. Therefore, the treatment effects model proposed by Maddala (1986) is used for robustness test as well.

Considering that the Internet is more prevalent and more frequently used among the new generation of peers compared to the older generation of employees, the article

Management Theory and Studies for Rural Business and Infrastructure Development
eISSN 2345-0355. 2023. Vol. 45. No. 4: 453-466
Article DOI: https://doi.org/10.15544/mts.2023.44

further analyzes the heterogeneity of the impact of Internet use on annual income across generations, and we introduce an interaction term between Internet use and generations to analyze the impact of Internet use on both old and new generations among urban and rural employed people. The basic form of the equation was then developed as Equation 2.

$$\ln income = \alpha + \beta\, \text{internet}_i + \delta\, \text{internet}_i * \text{generation}_i + \gamma X_i + \varepsilon_i \quad (2)$$

The examination of the impact of internet usage on wage or income inequality is usually carried out using Re-centered Influence Function (RIF). This decomposition method has two main steps: first, a consistent estimate of the target statistic is obtained by applying the Re-centered Influence Function, which can be the Gini coefficient, variance, or quantile, i.e., a RIF regression; second, the Oaxaca–Blinder decomposition approach is applied based on the results of the RIF regression to decompose the explanatory variables in different periods or on different target statistics and observe the impact of each explanatory variable on the target statistic. This paper draws on the unconditional quantile regression and RIF regression proposed by Firpo et al. (2009, 2018) to decompose the income gap at each quantile to obtain the effect of characteristic variables on the income gap at any quantile. The RIF regression equation for $\ln income$ located in the $Q\tau$quantile can be abbreviated as:

$$\text{RIF}(\ln income; Q_\tau) = Q_\tau + \text{IF}(\ln income; Q_\tau) \quad (3)$$

$Q\tau$is the unconditional distribution of $\ln income$ under theτth quantile. This function can be integrated into the target statistic:

$$\int \text{RIF}\left(\ln income; Q_\tau\right) \cdot d_F \ln income = \text{E}(\text{RIF}(\ln income; Q_\tau)) = \text{V(F)} \quad (4)$$

The RIF has an important property that underlies its decomposition regression (Firpo et al., 2009), i.e., the unconditional expectation of the RIF is the corresponding statistic itself, see equation 5.

$$\text{V(F)} = \text{E}_\text{Y}[\text{RIF}(\ln income; Q_\tau)] = \text{E}\left[X' \ \beta\right] + \text{E}[\varepsilon_i] = \bar{X}' \quad (5)$$

In addition, the generalized Oaxaca (1973) and Blinder (1973) decomposition methods allow for a quantitative decomposition of the income distributions of urban and rural employed persons. In equations 6 and 7, $\bar{Q}_{U\tau}(\ln income)$ and $\bar{Q}_{A\tau}(\ln income)$ denote the logarithmic form of annual income plus one for the urban and rural employed, respectively.

In equation 6, $\bar{Q}_{C\tau}(\ln income)$ then is the counterfactual annual income constructed based on the estimated coefficients of the characteristic variables of the rural employed and the urban employed, which corresponds to the income distribution function when the returns to the characteristic variables of the rural employed are the same as those of the urban employed.

In equation 7, $\bar{\beta}_{U\tau}$ and $\bar{\beta}_{A\tau}$ represent the expectation of the regression coefficients at different quantiles for urban and rural employed persons, respectively. $\bar{X}_U$ and $\bar{X}_A$ represents the mean of the characteristic vectors for the urban and rural employed, respectively. The first term on the right-hand side of equation 7 is the income gap triggered by the difference in the mean coefficients of the characteristic variables for the urban and rural employed, namely, the composition effect, which highlights the income gap caused by the difference in the proportion of internet usage for the urban and rural employed. The second term indicates the income gap generated by the difference in returns to the same characteristic variables for the urban and rural employed, namely, the income structure effect, which represents the effect of the difference in returns to internet usage on the income gap. In addition, equation 7 takes both specification error and reweighting error into account in the RIF decomposition process.

$$\bar{Q}_{U\tau}(\text{lnincome}) - \bar{Q}_{A\tau}(\text{lnincome}) = \left[\bar{Q}_{U\tau}(\text{lnincome}) - \bar{Q}_{C\tau}(\text{lnincome})\right] + \left[\bar{Q}_{C\tau}(\text{lnincome}) - \bar{Q}_{A\tau}(\text{lnincome})\right] \tag{6}$$

$$\bar{Q}_{U\tau}(\text{lnincome}) - \bar{Q}_{A\tau}(\text{lnincome}) = \left[(\bar{X}_U - \bar{X}_A)\bar{\beta}_{U\tau} + \bar{\varepsilon}_{UC}\right] + \left[(\bar{\beta}_{U\tau} - \bar{\beta}_{A\tau})\bar{X}_A + \bar{\varepsilon}_{CA}\right] \tag{7}$$

Empirical results

Impact of internet usage on individual annual income. Table 2 shows the regression results of internet usage on annual income for all employed, rural employed and urban employed. The validity of the instrumental variables largely determines whether credible conclusions can be drawn on the causal relationship between internet usage and annual income. In all three models, the p-values corresponding to the under-identification test are highly statistically significant, indicating that all three models do not suffer from under-identification. The endogeneity tests of instrumental variables all show F-values greater than 10 and statistically significant at the 1% level, which indicates that the selected instrumental variables are statistically correlated with the endogenous variables, endogeneity is satisfied, and there is no weak instrumental variable problem. Regarding the exogeneity of the instrumental variables, the p-values of the statistics corresponding to the sargan test in Models 1 and 3 are greater than 0.1, which indicates that the exogeneity of the instrumental variables is satisfied. For Model 2, the p-value of the statistic corresponding to the sargan test is statistically significant at the 10% level, and the exogeneity of the instrumental variables is diminished, and the quality of the regression results of Model 2 may be affected. In this regard, this will be remedied by a series of robustness tests in the following.

The coefficient value of internet usage is positive in all three models and is statistically significant at the 1% level. This means that, after overcoming endogeneity as much as possible, a clear conclusion is reached that an increase in the proportion of employed people using the Internet leads to an increase in the average annual income of employed people. When focusing on the subsample, the effect of internet usage on the annual income of urban employed persons is larger than their rural counterparts. There are several possible explanations behind this finding, and based on the available evidence, one of them may be due to the fact that the existence of the digital divide is multidimensional, such as geography, gender, industry, and generation (Luo and Lin, 2020; Várallyai, et al., 2015; Loges and Jung, 2001), differences in access to the Internet on these dimensions interact to make internet usage ultimately produce differential returns. Another reason may come from the sample self-selection problem. Therefore, in the robustness test section, it is necessary to overcome sample self-selection before observing the relationship between internet usage and annual income across models.

Management Theory and Studies for Rural Business and Infrastructure Development
eISSN 2345-0355. 2023. Vol. 45. No. 4: 453-466
Article DOI: https://doi.org/10.15544/mts.2023.44

Table 2. Results of Regression Analysis of internet usage on Annual Income of the Employed

		Model (1)	Model (2)	Model (3)
VARIABLES		Full sample	Rural *Hukou*	Urban *Hukou*
Internet usage characteristics				
	Internet usage	2.731 *** (0.502)	2.093 *** (0.538)	5.116 *** (1.555)
	SoI	-0.007 (0.008)	-0.010 (0.009)	0.002 (0.016)
Individual characteristics				
	Age	0.041 *** (0.008)	0.024 ** (0.011)	0.056 *** (0.010)
	Gender	1.687 *** (0.084)	2.226 *** (0.117)	0.986 *** (0.122)
	Han	0.014 (0.161)	0.185 (0.198)	-0.349 (0.284)
	Married	0.282 ** (0.133)	0.273 (0.196)	0.325 * (0.183)
	Children	-0.216 *** (0.051)	-0.193 *** (0.062)	-0.283 *** (0.097)
	Hukou	0.135 (0.115)	-	-
Educational characteristics				
Education (Reference group: Primary school and below)	Junior high school	-0.309 ** (0.152)	-0.200 (0.169)	-0.918 *** (0.348)
	Senior high school	0.160 (0.178)	0.238 (0.210)	-0.514 (0.409)
	College degree and above	0.785 *** (0.195)	0.506 * (0.279)	0.247 (0.430)
Father's highest education		-0.094 (0.064)	-0.138 (0.104)	-0.092 (0.081)
Mather's highest education		0.157 ** (0.078)	0.204 (0.148)	0.125 (0.091)
Healthy characteristics				
Physical health (Reference group: Very bad)	Bad	0.287 (0.318)	0.638 * (0.386)	-0.792 (0.604)
	Moderate	0.753 ** (0.317)	0.851 ** (0.386)	0.010 (0.603)
	Good	0.907 *** (0.314)	1.098 *** (0.380)	0.053 (0.603)
	Very good	0.813 ** (0.321)	0.852 ** (0.391)	0.085 (0.614)
Frequency of depression (Reference group: Always)	Often	-0.020 (0.428)	-0.131 (0.560)	-0.388 (0.694)
	Seldom	0.059 (0.409)	-0.150 (0.538)	-0.055 (0.652)
	Sometimes	0.295 (0.409)	0.134 (0.538)	0.061 (0.660)
	Never	0.127 (0.412)	0.073 (0.544)	-0.151 (0.650)
Job characteristics				
	Contract	1.177 *** (0.115)	1.283 *** (0.191)	1.098 *** (0.154)
	Medical insurance	0.104 (0.159)	0.243 (0.201)	-0.293 (0.274)
	Pension insurance	0.088 (0.103)	-0.040 (0.129)	0.378 ** (0.179)
Regional characteristics				
Region (Reference group: Central region)	Western region	0.227 * (0.117)	0.267 * (0.144)	0.153 (0.208)
	Eastern region	0.362 *** (0.106)	0.452 *** (0.146)	0.211 (0.153)
Constant		2.405 *** (0.738)	3.201 *** (0.941)	2.156 * (1.278)
Tests for Instrumental Variables				
Under identification test		416.283 (0.000)	307.059 (0.000)	71.240 (0.000)
Cragg-Donald Wald F statistic		221.833 (0.000)	166.419 (0.000)	36.225 (0.000)
Sargan statistic		0.743 (0.389)	3.194 (0.074)	0.926 (0.336)
Centered R^2		0.1742	0.1666	0.0995
Observations		6319	3643	2676

**Source: Standard errors for the corresponding parameters or the p-values of the corresponding statistics for the instrumental variable test are in parentheses. *** p<0.01, ** p< 0.05, * p<0.1.*

Robustness tests. In applying the PSM, a common support test was first conducted for the two samples of Internet use and non-Internet use, and most of the samples belonged to the common support domain after matching, and the matching effect was basically satisfactory. Then the balance test was conducted, and the proportion of standard deviation of most covariates was reduced to less than 20% after matching, and the overall mean deviation proportion was reduced from 50.4%-45.6% to 15.7%-7.5%. In addition, there were more covariates that were not significantly different between the two groups after matching compared to before matching. Therefore, it is reasonable to assume that the use of propensity score matching can effectively reduce the differences in the distribution of explanatory variables between the control and treatment groups to mitigate estimation bias due to sample self-selection. Comparing the coefficient values estimated by the instrumental variable method, the coefficient of the effect of Internet use on the annual income of employed persons was significantly lower. This suggests that the impact of Internet use on employed persons' income before matching may have been overestimated to some extent, but the direction of the impact of Internet use on employed persons' income after matching remains consistent with the previous estimates.

To mitigate endogeneity problems due to unobservable omitted variables, the treatment effect model was used to detect the presence of a statistically significant inverse Mills ratio (IMR). The results indicated that the regression coefficient values of IMR were statistically significant in all three models, implying a sample self-selection problem in the original model that did not account for omitted variables. And the relationship between Internet usage and annual income of employed persons remains statistically significant and positive in all three models. At the same time, the results of both the PSM and treatment effects models suggest that Internet usage generates different returns to annual incomes for urban and rural employed persons compared to the estimates from the instrumental variables approach, but the differences in the coefficients are reduced.

Heterogeneity Analysis. Table 3 shows the regression results after controlling for other covariates while adding the interaction term, internet usage multiple generation. In the three models, the results of the instrumental variable tests indicate that the selected instrumental variables basically satisfy the necessary conditions for endogeneity and exogeneity, and there is no weak instrumental variable problem. The regression coefficient of the interaction term is negative and statistically significant in Model 2. This demonstrates that the return to annual income from internet usage is lower for the new generation of rural employed than for the older generation of rural employed, while it is higher for the new generation of urban employed than for the older generation of urban employed. The likely reason for this is that the new generation of rural workers is relatively unskilled in their use of the Internet, which they use more for social and entertainment purposes and does not bring significant income returns.

Table 3. Impact of Internet usage on Different Generational Employment Groups

Variables	Full sample	Rural *hukou*	Urban *hukou*
Internet usage	2.743 *** (0.488)	2.092 *** (0.537)	5.011 *** (1.540)
Internet usage*generation	-0.216 (0.164)	-0.594 *** (0.230)	0.315 (0.177)
Tests for Instrumental Variables			
Under identification test	417.213 (0.000)	307.057 (0.000)	72.366 (0.000)
Cragg-Donald Wald F statistic	222.329 (0.000)	166.372 (0.000)	36.800 (0.000)
Sargan statistic	0.748 (0.387)	3.108 (0.078)	0.985 (0.321)
Centered R^2	0.1741	0.1682	0.1043
Observations	6319	3643	2676

**Source: Standard errors for the corresponding parameters are in parentheses, the p-values of the corresponding statistics for the instrumental variable test are in parentheses. *** p<0.01, ** p< 0.05, * p<0.1.*

Management Theory and Studies for Rural Business and Infrastructure Development
eISSN 2345-0355. 2023. Vol. 45. No. 4: 453-466
Article DOI: https://doi.org/10.15544/mts.2023.44

Impact of internet usage on urban-rural income gap

Following Fernando (2020), the RIF regression decomposition method is applied in combination with the bootstrap method here to further analyze how internet usage affects the income gap between the urban and rural employed. Tables 4 and 5 report the composition effect and income structure effect of the characteristic variables at the 25th, 50th, and 75th quantiles, respectively. In the explainable part, the total contribution of the characteristic effect in explaining the widening urban-rural wage gap is 85.9%, 73.7% and 60.5% at the 25th, 50th, and 75th quantiles, respectively, while the income structure effects are 14.1%, 26.3% and 39.5%, respectively. This indicates that the composition effect consistently dominates in explaining the widening income gap between the urban and rural employed, proving that the differences in the characteristics of urban and rural employed persons are the main cause of the widening wage gap. However, at the same time, the explanatory effectiveness of the composition effect gradually decreases as the wage level increases.

Table 4. Decomposition results of the composition effects at different quantiles

Variables	Composition effect					
	p25		p50		p75	
	Coefficient	Contribution Share	Coefficient	Contribution Share	Coefficient	Contribution Share
Internet usage	0.208 *** (0.036)	13.9%	0.130 *** (0.014)	16.5%	0.048 *** (0.009)	6.2%
Specification error	0.293 ** (0.124)	19.6%	-0.099 * (0.057)	-12.6%	-0.218 *** (0.075)	-28.2%
Total Explained	1.285 *** (0.123)	85.9%	0.582 *** (0.060)	73.7%	0.468 *** (0.083)	60.5%

**Source: Values in parentheses are bootstrapped standard errors (500 repetitions). *** p<0.01, ** p< 0.05, * p<0.1. Only the decomposition regression results for the core explanatory variables are shown, and the other covariates have been controlled for during the regression decomposition process.*

Table 5. Decomposition results of income structure effects at different quantiles

Variables	Income structure effect					
	p25		p50		p75	
	Coefficient	Contribution Share	Coefficient	Contribution Share	Coefficient	Contribution Share
Internet usage	-0.042 (0.309)	-2.8%	0.175 (0.139)	22.1%	-0.111 (0.108)	-14.4%
Reweighting error	0.047 (0.040)	3.1%	0.025 (0.023)	3.1%	0.052 ** (0.024)	6.7%
Total Unexplained	0.210 * (0.118)	14.1%	0.208 *** (0.061)	26.3%	0.306 *** (0.095)	39.5%

**Source: Values in parentheses are bootstrapped standard errors (500 repetitions). *** p<0.01, ** p< 0.05, * p<0.1. Only the decomposition regression results for the core explanatory variables are shown, and the other covariates have been controlled for during the regression decomposition process.*

In terms of composition effect, Internet usage has positive and statistically significant coefficient values at all three fractional levels, indicating that the difference in the proportion of internet usage between urban and rural employed people widens the income gap, especially for the middle- and low-income level cohorts. The difference in the proportion of internet usage between urban and rural areas explains 16.5% and 13.9% of the causes of the income gap for the middle-income group and the low-income group, respectively.

In terms of income structure effects, internet usage has negative coefficient values at the 25th and 75th quantiles and positive coefficient values at the 50th quantiles. This implies that for low-income and high-income employed persons, the difference in returns to internet usage between the urban and rural employed narrows the income gap. However, for the middle-income level of employment, the difference in returns to internet usage further contributes to the urban-rural income gap. It is worth considering that for the middle level income cohort, differential internet usage shares and different internet usage returns have opposite arguments for explaining the urban-rural income gap, which may be since internet usage returns are more widely heterogeneous within the middle-income group and contribute to the widening of the income gap.

Conclusion and Recommendation

The empirical analysis in this paper confirms that internet usage raises the annual income of both urban and rural employed people, but those from the urban areas are benefited with higher income returns. Among the factors other than the Internet, the age, gender, and labor contract of the employed contribute to the positive income effect. The results of the analysis also note the heterogeneous income effects of different levels of education and physical health status on urban and rural employed persons. Regarding the heterogeneity of the impact of internet usage on the income of urban and rural employed across generations, this study reveals that the annual income returns to internet usage are lower for the new generation of rural employed than for the older generation of rural employed, while the new generation of urban employed is higher than the older generation of urban employed.

This study demonstrates how internet usage shapes the income gap between the urban and rural employed: the difference in the proportion of internet usage between urban and rural employed widens the income gap between urban and rural employed; the extent of this widening income gap varies for employed cohorts at different income levels; and it has the most significant effect on groups of employed people distributed at middle-income and low-income levels. In addition, the study also identifies that the difference in the income return to internet usage between urban and rural employed reduces the income gap between them distributed at low- and high-income levels but renders the income gap at middle income level to be enlarged, and further research is expected on the urban-rural gap in middle-income groups shaped by Internet usage.

With urban and rural internet penetration rates of 82.9% and 58.8%, respectively, in June 2022 (China Internet Network Information Center, 2022), the existence of the digital divide between urban and rural China displays an urban-rural income gap. In order to close the income gap between urban and rural employed people, relevant policy recommendations are proposed based on the findings of this paper, which are expected to improve the income generating capacity of the rural employed and build a more digitally inclusive urban-rural development pattern.

Accelerate the improvement of rural digital infrastructure, promote the digital upgrade of rural infrastructure, continue to increase rural internet penetration, and stimulate the development of the digital economy in rural areas. Promote the rural expansion of “Internet+” in the fields of education, employment, health and public services, and provide accessible paths for rural employed people to improve their Internet skills, enjoy health services, acquire employment security and safeguard legal rights, and continuously enrich the human capital reserve of rural employed people. Rationalize income distribution and redistribution systems and social welfare policies, improve social underwriting mechanisms, and provide institutional guarantees for raising the income of rural employed persons and narrowing the income gap between urban and rural areas.

Acknowledgement

The authors would like to thank to the support of the major project of Social Science Foundation of Anhui Province of China (AHSKZD2019D04), the Social Science Foundation of Anhui Province of China (AHSKY2020D44) and the Key Project of Humanity and Social Science Supported by the Bureau of Education of Anhui Province (2022AH050561).

References

An, T. L., & Yang, C. (2020). How the Internet is reshaping China's economic geography: micro mechanism and macro effects. Economic Research Journal, 55(02): 4-19. (In Chinese).

Arvin, M. B., Pradhan, R. P., & Nair, M. (2021). Uncovering interlinks among ICT connectivity and penetration, trade openness, foreign direct investment, and economic growth: The case of the G-20 countries. *Telematics and Informatics*, 60, 101567. https://doi.org/10.1016/j.tele.2021.101567

Asongu S. A., & Odhiambo N. M. (2019). How enhancing information and communication technology has affected inequality in Africa for sustainable development: An empirical investigation. *Sustainable Development*, 27(4), 647–656. https://doi.org/10.1002/sd.1929

Bauer, J. M. (2018). The Internet and income inequality: Socio-economic challenges in a hyperconnected society. *Telecommunications Policy*, 42(4), 333-343. https://doi.org/10.1016/j.telpol.2017.05.009

Blinder A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources* (8): 436–455. https://doi.org/10.2307/144855

Braesemann, F., Lehdonvirta, V., & Kässi, O. (2022). ICTs and the urban-rural divide: can online labour platforms bridge the gap? *Information, Communication & Society*, 25(1), 34-54. https://doi.org/10.1080/1369118X.2020.1761857

Canh, N. P., Schinckus, C., Thanh, S. D., & Ling, F. C. H. (2020). Effects of the internet, mobile, and land phones on income inequality and The Kuznets curve: Cross country analysis. *Telecommunications Policy*, 44(10), 102041. https://doi.org/10.1016/j.telpol.2020.102041

Caroline L Freund, & Diana Weinhold. (2004). The effect of the Internet on international trade. *Journal of International Economics*, 62(1). https://doi.org/10.1016/S0022-1996(03)00059-X

Cheng, M., & Zhang, J. (2019). Internet popularization and urban-rural income gap: a theoretical and empirical analysis. China Rural Economy, (02):19-41. (In Chinese).

China Internet Network Information Center. (2022). The 50th Statistical Report on China's Internet Development. http://www3.cnnic.cn/n4/2022/0914/c88-10226.html.

Colas, M., & Ge, S. (2019). Transformations in China’s internal labor migration and hukou system. *Journal of Labor Research*, 40(3), 296-331. https://doi.org/10.1007/s12122-019-9283-5

Donzuso, N. N. (2015). ‘Equality of opportunities’ in education for migrant children in China. *Global Social Welfare*, 2(1), 9-13. https://doi.org/10.1007/s40609-014-0012-y

Fang, X. D., & Chen, S. (2019). 25 years of the Internet in China. Modern Communication, 41(04): 1-10. (In Chinese).

Fernando, & Rios-Avila. (2020). Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition. *The Stata Journal*, 20(1), 51-94. https://doi.org/10.1177/1536867X20909690

Firpo, S. P., Fortin, N. M., & Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, 6(2), 28. https://doi.org/10.3390/econometrics6020028

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3), 953-973. https://doi.org/10.3982/ECTA6822

Fong, M. W. (2009). Digital divide between urban and rural regions in China. *The Electronic Journal of Information Systems in Developing Countries*, 36(1), 1-12. https://doi.org/10.1002/j.1681-4835.2009.tb00253.x

Glasmeier, A. K. (2018). Income inequality and growing disparity: Spatial patterns of inequality and the case of the USA. *The new Oxford handbook of economic geography*, 63-77. https://doi.org/10.1093/oxfordhb/9780198755609.013.3

Gong, X., Zhang, J., Zhang, H., Cheng, M., Wang, F., & Yu, N. (2020). Internet use encourages pro-environmental behavior: Evidence from China. Journal of Cleaner Production, 256, 120725. https://doi.org/10.1016/j.jclepro.2020.120725

GSMA intelligence. (2021). The Mobile Economy China 2021. https://www.gsma.com.

Han, L., Shi, L., Lu, L., & Ling, L. (2014). Work ability of Chinese migrant workers: the influence of migration characteristics. *BMC Public Health*, 14(1), 1-8. https://doi.org/10.1186/1471-2458-14-353

Krueger, A. B. (1993). How computers have changed the wage structure: evidence from microdata, 1984–1989. *The Quarterly Journal of Economics*, 108(1), 33-60. https://doi.org/10.2307/2118494

Lakhanpal, P., Mukherjee, J., Nag, B., & Tuteja, D. (2021). Trade, Investment and Economic Growth. Springer, Singapore.

Liu, S. L., Zhang, X. M., & Yang, Z. S. (2021). Impact of internet usage on rural residents' income. Journal of Quantitative and Technical Economics, 38 (04):103-119. (In Chinese).

Liu, Z. (2005). Institution and inequality: the hukou system in China. *Journal of comparative economics*, 33(1), 133-157. https://doi.org/10.1016/j.jce.2004.11.001

Loges, W. E., & Jung, J. Y. (2001). Exploring the digital divide: Internet connectedness and age. *Communication research*, 28(4), 536-562. https://doi.org/10.1177/009365001028004007

Luo, L. X., & Lin, Z. W. (2020) Development Tendency of Double-Layer Digital Divide in Mainland China: A Longitudinal Study Based on Perspective of Group Segmentation. In International Development Informatics Association Conference (pp.170-185). Springer, Cham. https://doi.org/10.1007/978-3-030-52014-4_12

Ma, X. (2022). Internet usage and gender wage gap: evidence from China. *Journal for Labour Market Research*, 56(1), 1-17. https://doi.org/10.1186/s12651022003209

Ma, X. (2022). Internet use and income gaps between rural and urban residents in China. *Journal of the Asia Pacific Economy*, 1-21. https://doi.org/10.1080/13547860.2022.2054133

Maddala, G. S. (1986). Limited-dependent and qualitative variables in econometrics (No. 3). Cambridge university press.

Maiti, D., Castellacci, F., & Melchior, A. (2020). Digitalisation and development: issues for India and beyond. In Digitalisation and Development (pp. 3-29). *Springer, Singapore*. https://doi.org/10.1007/978-981-13-9996-1_1

Mao, Y. F.., Hu, W. X., & Zeng, X. Q. (2021). Expanding or Narrowing: The Impact of internet usage on Household Wage Gap-Empirical Evidence Based on CGSS Data. Collected Essays on Finance and Economics, 269(2), 3. (In Chinese).

Matthess, M., & Kunkel, S. (2020). Structural change and digitalization in developing countries: Conceptually linking the two transformations. *Technology in society*, 63, 101428. https://doi.org/10.1016/j.techsoc.2020.101428

Mincer, J. (1974) Schooling, experience and earnings. Columbia University Press, New York.

Mincer, J. (1997) Changes in wage inequality: 1970-1990. Research in Labor Economics. 16(1), 1-18. https://doi.org/10.7916/D8F76M3J

National Bureau of Statistics of China. (2022). Income and consumer spending of the population in 2021. http://www.stats.gov.cn/xxgk/sjfb/zxfb2020/202201/t20220117_1826442.html

Ningsih, C., & Choi, Y. J. (2018). An effect of internet penetration on income inequality in Southeast Asian countries. https://doi.org/10.52558/ism.2019.12.20.3.121

Oaxaca, R. (1973). Male–female wage differentials in urban labor markets. *International Economic Review*, 14: 693–709. https://doi.org/10.2307/2525981

Official Website of the Central Government of China. (2022). http://www.gov.cn/xinwen/2022-08/22/content_5706298.

Rosenbaum, P. R., Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*,70(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

Shipman, J. E., Swanquist, Q. T., & Whited, R. L. (2017). Propensity score matching in accounting research. *The Accounting Review*, 92(1), 213-244. https://doi.org/10.2308/accr-51449

Song, Q., & Smith, J. P. (2019). Hukou system, mechanisms, and health stratification across the life course in rural and urban China. *Health & place*, 58, 102150. https://doi.org/10.1016/j.healthplace.2019.102150

Song, Y. (2014). What should economists know about the current Chinese hukou system? *China Economic Review*, 29, 200-212. https://doi.org/10.1016/j.chieco.2014.04.012

Van Deursen, A. J., Helsper, E., Eynon, R., & Van Dijk, J. A. (2017). The compoundness and sequentiality of digital inequality. *International Journal of Communication*, 11, 452-473. https://doi.org/1932-8036/20170005

Várallyai, L., Herdon, M., Botos, S. (2015). Statistical analyses of digital divide factors. Procedia Economics and Finance, 19, 364-372. https://doi.org/10.1016/S2212-5671(15)00037-4

Yang, J., & Mukhopadhaya, P. (2022) Monetary and non-monetary poverty within urban locals and rural migrants in China using inequality-sensitive poverty measures. The B. E. *Journal of Economic Analysis & Policy*. 22(1), 29-66. https://doi.org/10.1515/bejeap-2021-0193

Zhang, J. P., Wang, R., & Lu, C. (2018). A quantitative analysis of Hukou reform in Chinese cities: 2000–2016. *Growth and Change*, 2018; 00: 1–21. https://doi.org/10.1111/grow.12284