

ORIGINAL ARTICLE

Open Access



# Internet use and gender wage gap: evidence from China

Xinxin Ma\*

## Abstract

This study explores the influence of Internet use on the gender wage gap in China by using national longitudinal survey data. A fixed effects and instrumental variable method were employed to address individual heterogeneity and other endogeneity problems. The study contributes in the form of four key conclusions. First, considering the endogeneity problems, the return to Internet use is higher for men than for women, unlike the results derived using the ordinary least squares model, which indicates the opposite. The individual heterogeneity problem considerably affects the results, thus suggesting a bias in earlier studies. The results of robustness checks based on the Internet use frequency for different purposes confirm the conclusions. Second, the gender difference in return to Internet use is higher in the low-education group and older age cohorts. Third, both the components—the gender disparity in Internet access and gender difference in return to Internet use—widen the gender wage gap, with the gender difference in return to Internet use having a higher impact. Fourth, the effects of these two components on the gender wage gap vary with the educational attainment and age cohorts.

**Keywords:** Gender wage gap, Internet use, Return to Internet use, Internet access, China

**JEL Classification:** J16, J24, J31, O33

## 1 Introduction

With the progress of information and communication technology (ICT) since the 1970s, Internet usage has expanded worldwide (OECD 2018). The gender digital gap in Internet access rose in developed countries in the early stages of ICT development (Bimber 2000; DiMaggio et al. 2001; Fatehikia et al. 2018) but reduced with the increasing diffusion of digital technologies (Haight et al. 2014; Ono and Zavodny 2007; Rice and Katz 2003). Women in developing countries have a significantly lower likelihood of Internet access than men, and this gender disparity in Internet use can enlarge the overall socioeconomic gender gap (Alozie and Akpan-Obong 2017; Broadband Commission 2013; Hafkin and Huyer 2007; OECD 2018). Although research has established that Internet use can affect wages (Krueger 1993; Liu et al.

2021; Miller and Mulvey 1997; Pabilonia and Zoghi 2005, etc.) and employment (Alam and Mamun 2017; Atasoy 2013; Deylyng 2017; Mao and Zeng 2017, etc.), empirical studies on the impact of Internet use on the gender wage gap are scarce. This study attempts to bridge this gap by providing new evidence from China—a country that has seen rapid Internet diffusion and gender wage gap growth in the last two decades.

The China's gender disparity in Internet use can be highlighted through the *Statistical Report on the Development of the Internet in China* No. 45 (CNNIC 2020), which reveals that the number of Internet users in China reached 904 million in April 2020, of which 48.1% were women (30.4% in 2000).<sup>1</sup> The statistics suggests the existence of a gender disparity in Internet access in China.

Additionally, the gender wage gap in China has expanded since the 1980s (Gustafsson and Li 2000;

\*Correspondence: xxma@hosei.ac.jp

Faculty of Economics, Hosei University, 4342 Machita-shi Aiharamachi, Tokyo 194-0298, Japan

<sup>1</sup> The proportion of women in the total population in China was 48.71% in 2020 (World Bank 2022).

Iwasaki and Ma 2020; Ma 2018). Several empirical studies reveal that the main determinants are gender differences in human capital and workplace discrimination against women (Gustafsson and Li 2000; Li and Yang 2010; Li and Yang 2010; Ma et al. 2013). Others indicate that occupational segregation and industry/ownership sector segmentation also contribute to the creation of the gender wage gap (Ge 2007; Li and Ma 2006; Liu et al. 2000; Ma 2018; Meng and Zhang 2001; Wang and Cai 2008; Wang 2005). However, empirical studies on the effects of Internet use on the gender wage gap are scarce (Qi and Liu 2020; Wu 2021; Zhuang et al. 2016) and limited to cross-sectional data analyses, which can result in endogeneity problems.

This study contributes to the literature in four ways: First, in contrast to earlier studies (Qi and Liu 2020; Wu 2021; Zhuang et al. 2016), we examine the influence of Internet use on the gender wage gap in China using the panel data analysis method, such as the fixed effects (FE) model, random effects (RE) model, and the lagged variable (LV) model to address endogeneity problems. This study is the first to investigate the association between Internet use and the gender wage gap in China based on the panel data analysis method. Second, besides investigating the extensive margin effect of Internet use (whether used Internet), we also ascertain the intensive margin effects (the frequency of using the Internet) for different purposes (work and study, communication, leisure) on the gender wage gap, which has been not considered in previous studies. Third, our study is the first to decompose the determinants of the gender wage gap into two components—the gender disparity in Internet access (e.g., gender disparity in the proportion of workers using the Internet) and gender difference in return to Internet use (e.g., the gender difference in the rising range of wage level in terms of Internet use)—based on the IV method. Since policy implications differ across these two components, it would be meaningful to investigate how these two components contribute to the effects of Internet use on the gender wage gap. For example, if the gender disparity in Internet access is found to be the main contributor, a policy promoting Internet accessibility among women is expected to reduce the wage gap. In contrast, if the main contributor is the gender difference in the return to Internet use, reducing the discrimination against women in the workplace may reduce the wage gap. Fourth, although Eastin et al. (2015), OECD (2018), Riggins and Dewan (2005), and Scheerder et al. (2017) have argued the Internet use divisions in education, age, and gender groups, they have not investigated the effect of Internet use on gender wage gap among education

and age cohort groups. The gender differences in Internet access and return to Internet use may differ among different education and age cohort groups; therefore, the effects of Internet use on the gender wage gap may differ per these groups. This study compares the effects of Internet use among educational attainment and age cohort groups to provide more evidence to understand the digital division issues in depth. Additionally, we use the three latest wave national longitudinal survey data of 2014–2018, which can provide new information.

The study resulted in four main findings. First, when addressing endogeneity problems, the return to Internet use is higher for men than for women, unlike the results using the ordinary least squares (OLS) model, which indicate the opposite. In this case, the individual heterogeneity problem considerably affects the results, thus suggesting a bias in earlier studies. Second, the results indicate that the gender disparities in return to Internet use are higher in the low-education and older age cohort than in others. Third, both the components—the gender disparity in Internet access and the gender difference in return to Internet use—widen the gender wage gap, with the gender difference in return to Internet use having a higher impact. Fourth, the effects of both components on the gender wage gap vary with the educational attainment and age cohorts.

The remainder of this paper is organized as follows. Section 2 discusses the channels whereby Internet use influences wages and summarizes the empirical literature on the issue. Section 3 introduces the data and the methodology. Section 4 presents and discusses the empirical results, and Sect. 5 concludes the study.

## 2 Literature review

### 2.1 Channels of influence of Internet use on the gender wage gap

Several economic theories can explain the gender wage gap. First, gender differences in the endowment of factors such as human capital can contribute to a gender wage gap. Based on the human capital theory (Becker 1964; Mincer 1974), the individual wage level in a perfectly competitive labor market is determined by the workers' labor productivity. Labor productivity is related to a worker's human capital (e.g., education and years of work experience). When men have higher educational attainment than women, they may earn a higher wage.

Second, according to the employer discrimination hypothesis (Becker 1957), whenever employers, customers, and colleagues discriminate against women, it can create a gender wage gap. During the economic transition period (after 1978), with the enforcement

of the Opening-up Policy, the private sector (i.e., the privately-owned enterprises, the Township and Village Enterprises,<sup>2</sup> etc.) has developed since the 1990s, and the proportion of workers in the private sector to total workers in urban areas has increased from 27.2% in 1985 to 67.4% in 2020 (NBS 2021).<sup>3</sup> In the planned economy period (1949–1977), women's employment promotion policies, such as the equality employment policy, were enforced in the public sector (e.g., the state-owned enterprises: SOEs; the government organizations) by the Chinese government, which led to less discrimination against women (Gustafsson and Li 2000; Ma 2018, 2021a, b). With the progress of SOE reforms, most SOEs have changed to privately-owned enterprises. Although the influences of the equality employment policies are still greater in the SOEs even in the current period, when discrimination against women persists in the private sector, a gender wage gap can arise. Iwasaki and Ma (2020) report that the gender wage gap is higher in privately-owned enterprises than in SOEs, and the gender wage gap expanded during the economic transition period from the 1990s to the 2010s, suggesting that the discrimination against women has become severe with the progress of the market-oriented economic reform in China.

Third, the statistical discrimination hypothesis (Arrow 1973; Phelps 1972) suggests that because of information asymmetry, employers must make employment and wage decisions for employees, both men, and women, based on the average values of some unobservable factors (i.e., work effort, probability of turnover, etc.). If the employer perceives that the probability of doing housework (i.e., childcare, geriatric care, domestic cleaning, cooking, etc.) is higher for women than for men, they may set a lower wage level for women, thus generating a gender wage gap.

In China, factors such as the deregulation of the one-child policy, lack of formal public childcare institutions (e.g., kindergarten), population aging, and insufficiency of institutional care for the elderly have increased the responsibilities of childcare and geriatric care for women (Connelly et al. 2018; Cook and Dong 2011; Ma 2021a, b). Family care has decreased the female labor force participation and reduced women's work efforts in China (Chen and Fan 2016; Chen et al. 2016; Chen 2019; Connelly et al. 2018; Ma 2021a, b). Therefore, statistical

discrimination may contribute to widening the gender wage gap in China.

Fourth, the crowding hypothesis (Bergmann 1974) stated that there remains gender occupational segregation in the labor market, which means women concentrate on female-dominated occupations (e.g., staff, service job), whereas men on male-dominated (e.g., manager, technician). When the wage levels of male-dominated occupations are higher than those of female-dominated occupations, the gender wage gap arises. The evidence from the empirical studies for the developed countries has supported the crowding hypothesis (Brown et al. 1980; Kidd and Shannon 1996; Kidd 1993; Miller 1987). Meng (1998) and Li and Ma (2006) also reported that in China, there exists gender occupational segregation, which contributes to the formation of the gender wage gap.

Finally, the monopsony power hypothesis suggests that imperfect competition may lead to a gender wage gap. When a firm has the monopsony power in the labor market and sets a lower wage level for women, the gender wage gap arises (Hirsch 2010). Vick (2017) reports that the monopsony power hypothesis is supported in Brazil.

Based on these theories or hypotheses, in this study, we considered that Internet use could lead to the formation of gender wage gap in China through two channels: (i) the explained component (e.g., gender disparity in Internet access), and (ii) the unexplained component—workplace discrimination against women (e.g., gender difference in return to Internet use as reflected in the wages).

Regarding the first component—the effect of gender disparity in Internet access on the gender wage gap, (i) based on human capital theory, the wage level in a perfectly competitive labor market is determined by a worker's labor productivity; the labor productivity is related to a worker's human capital. Internet use can be considered an element of human capital: Internet users are usually considered to be those with higher skills (or higher productivity) than non-users. If the percentage of Internet users in the women's group is lower than that in the men's, the gender wage gap arises. (ii) According to the crowding hypothesis, men may occupy male-dominated occupations (e.g., manager, technician), which have a higher likelihood of using the Internet for work, which may lead to the gender divisions in Internet access.

To consider the second component—the effect of gender differences in return to Internet use on the gender wage gap, (i) according to the employer discrimination and statistical discrimination hypotheses, when workplace discrimination exists, wage levels tend to be set lower for women than men despite similar endowments (e.g., Internet use skills). (ii) Based on the crowding hypothesis, if Internet use is evaluated higher for

<sup>2</sup> In China, the Township and Village Enterprises (TVEs) were collectively owned enterprises and developed by the end of the 1980s. Since the 1990s, most of them have transformed into privately owned enterprises, wherein operations and human resource management are similar to those in privately-owned enterprises.

<sup>3</sup> In the official data (NBS 2021), the private sector includes the privately-owned enterprises, collectively owned enterprises, foreign investment enterprises, and other privately-owned enterprises (e.g., the joint-stock company, joint company, limited liability company).

male-dominated occupations than female-dominated occupations by a firm, the gender wage gap arises. (iii) According to the monopsony power hypothesis, if a firm has monopsony power and values Internet use more highly for men than women, the gender wage gap widens.

Although it is assumed that the two components may contribute to the formation of the gender wage gap, there is no empirical study to investigate the issue. In this study, we attempt to estimate the contribution rates of these two components in the subsequent decomposition analysis and provide new evidence to understand the causes of the gender wage gap in China in depth.

## 2.2 Empirical studies on the gender wage gap in China

There are numerous empirical studies on the gender wage gap in developed countries and China (Iwasaki and Ma 2020).<sup>4</sup> We summarize only the main ones in China as follows.

Gustafsson and Li (2000), Liu et al. (2000), Li and Yang (2010), Li et al. (2014), and Ma et al. (2013) use the Blinder–Oaxaca model (Blinder 1973; Oaxaca 1973) for the decomposition analysis, thereby demonstrating that both the explained (e.g., gender differences in educational attainment) and unexplained (e.g., returns to education) components affect the gender wage gap. Most studies show that the contribution of the unexplained component is higher than that of the explained component, which suggests that workplace discrimination against women is the primary cause of the gender wage gap in China. The contribution rates of the unexplained component in the gender wage gap for the local urban resident group were reported to be 52.49% in 1988 and 63.20% in 1995 (Gustafsson and Li 2000); and 52.0% in 1995, 69.0% in 2002, and 77.7% in 2007 (Li et al. 2014). The rate for the migrant group was 74.32%–84.38% in 2008 (Li and Yang 2010), and that for all residents was 49.18% in 1996 (Meng and Zhang 2001). In addition, the values by wage percentiles were 86.08–101.80% in 2006 and 45.31–91.73% in 2009 (Ma et al. 2013). A few empirical studies also focus on the effect of segmentation by sector on the gender wage gap in China. Li and Ma (2006) and Wang (2005) analyze the influence of occupational segregation on the gender wage gap. Ge (2007), Ma (2018), and Wang and Cai (2008) explore the impact of segmentation by industry or enterprise ownership sector type on the gender wage gap in China and report that the unexplained component in intra-sector differentials drives the gender wage gap in China.

<sup>4</sup> Iwasaki and Ma (2020) conducted a meta-analysis based on the results of the inclusion of a gender dummy variable in the wage functions of 199 studies for China. For the current empirical studies on the gender wage gap in other countries, please refer to Ge and Zhou (2020) for the U.S., Biewen et al. (2020) for German, Masso et al. (2022) for Estonia.

## 2.3 Empirical studies on Internet use and gender wage gap

Empirical studies on Internet use and the gender wage gap are scarce in developed countries and China. Borghans et al. (2014) and Ge and Zhou (2020) find that Internet skills and their return affect the changes in the gender wage gap in developed countries. Beaudry and Lewis (2014) hold that changes in return to computer skills are an important factor that explains the decline in the gender wage gap. Using industrial data, Ge and Zhou (2020) report that robot use reduces the gender wage gap, while an increase in computer capital raises the gender wage gap in the US. However, these studies are not based on the decomposition method. Therefore, the contribution rates of the gender disparity in Internet access and gender difference in return to Internet use in forming the gender wage gap are unclear.

Regarding China, only three studies focus on the issue. Zhuang et al. (2016) use data from the Third Survey on Chinese Women's Social Status and apply the propensity score matching method to estimate the wage function. They find that the Internet use wage premium for women is 90.6% of that for men, suggesting that the return to Internet usage is lower for women than men. Using data from 2010, 2013, and 2015 Chinese General Social Surveys (CGSS) and the Blinder–Oaxaca decomposition method, Qi and Liu (2020) report that both Internet access and the return to Internet usage reduced the gender wage gap in 2013 and 2015. Wu (2021) uses data from the 2017 CGSS, the OLS model, and the Blinder–Oaxaca decomposition method to analyze the effect of Internet use on the gender wage gap and reveals that the contribution rate of gender disparity in Internet access is -4.34%, which reduces the gender wage gap.

Although these studies provide some evidence on the effects of Internet use on the gender wage gap in China, they have not addressed the individual heterogeneity problem owing to the use of cross-sectional analysis methods. Additionally, the intensive margin effects of Internet use (e.g., the frequency of Internet use) for different purposes (e.g., work and study, leisure) were not considered. Moreover, the differences in the effects of Internet use by the educational attainment and age cohorts were not considered. This study can address these neglects.

## 3 Methodology

### 3.1 Model

To estimate the gender disparity in return to Internet use, we estimate the wage function. The OLS model is expressed by Eq. (1).



$$\begin{aligned} \ln W_i = & a_0 + \beta_1 \text{Int}_i + \beta_2 \text{Female}_i \\ & + \beta_3 \text{Female} * \text{Int}_i + \gamma \sum X_i + \nu_i \end{aligned} \quad (1)$$

In Eq. (1),  $i$  represents the individual.  $\text{Int}$  is the Internet use variable,  $\text{Female}$  is the female dummy;  $\text{Female} * \text{Int}$  is the interaction term of Internet use and female dummy variable;  $\ln W_i$  denotes the logarithmic value of wages;  $\sum X$  represents a set of the factors (i.e., education, years of work experience, occupation, etc.) that affect the wages;  $\beta$  and  $\gamma$  are the estimated coefficients; and  $\nu$  denotes the error term. The total of  $\beta_1$  and  $\beta_3$  are the return to Internet use,  $\beta_3$  is the gender differences in return to Internet use.

However, the endogeneity problem may exist in Eq. (1) for three reasons: first, the omitted variable may influence the likelihood of using the Internet and the wage level. We constructed the variables to control factors that may affect wage level as much as possible. However, some unobservable variables may also affect the results. We used the IV method to address the issue expressed by Eqs. (2–4).

$$\begin{aligned} \text{Pr}(\text{Int}_i = 1) = & a_1 + \beta_Z Z_i + \beta_1 \text{Int}_i + \beta_2 \text{Female}_i \\ & + \beta_3 \text{Female} * \text{Int}_i + \gamma \sum X_i + u_i \end{aligned} \quad (2)$$

$$\begin{aligned} \ln W_i = & a_0 + \beta_1 \widehat{\text{Int}}_i + \beta_2 \text{Female}_i \\ & + \beta_3 \text{Female} * \text{Int}_i + \gamma \sum X_i + \delta_i \end{aligned} \quad (3)$$

$$\text{corr}(Z, \delta) = 0 \text{ and } \text{corr}(Z, u) \neq 0 \quad (4)$$

In Eqs. (2–4),  $u$ , and  $\delta$  denote the error terms, respectively.  $Z$  represents the IV. The internet penetration rate at the regional (provincial or local) level and the importance of using the Internet attitude were generally used in previous studies (e.g., Cao and Jiang 2020; Zhao and Li 2020). We performed several tests for these IVs, but the results rejected the hypothesis that these IVs are exogenous, suggesting that they were not valid for this study. We used two variables—(i) the provincial optical cable circuit in 1999 and (ii) the provincial long-distance cable line length in 1999 as IVs in this study. Both are the oldest data that we could obtain from the government's official dataset.<sup>5</sup> It can be assumed that Internet installations in recent survey years (2014, 2016, and 2018) are closely related to the regional telecommunication capability in the past (such as the optical cable circuits or the long-distance cable line length 15 to 19 years back), while

the influence of the regional telecommunication capability in the past on an individual's recent income levels is small, which may fit the conditions of the IV from the econometric perspective. We performed several tests, including the Durbin-Wu-Hausman test, the exclusion test (Hansen's J statistic), and the weak identification test (Cragg–Donald Wald F statistic). The results indicated that the two IVs are appropriate (see Tables 2, 3, 4 and 5 and discussions in Sect. 4).

Second, as  $\nu$  in Eq. (1) includes the errors related to individual-specific and time-invariant factors ( $\rho_i$ ) and the idiosyncratic error ( $\varepsilon$ ), an individual heterogeneity problem may arise in Eq. (1). We use the FE or RE model to address this heterogeneity problem. We perform the Hausman specification test to judge the validity of the FE and RE models. In Eq. (5),  $t$  denotes the longitudinal survey year.

$$\ln W_{it} = \alpha + \beta_1 \text{Int}_{it} + \beta_2 \text{Female} * \text{Int}_{it} + \gamma \sum X_{it} + \rho_i + \varepsilon_{it} \quad (5)$$

Third, the endogeneity issue may arise due to reverse causality. It is assumed that Internet use may affect wages (explored in this study). However, wages may also affect the likelihood of Internet use. For instance, the likelihood of using the Internet for work and study may be greater for high-wage workers than low-wage workers. Since there is a two-way relationship between Internet use and wage, we use a one-period lagged variable (LV) model to address the potential reverse causality. We assume that Internet use in period  $t - 1$  may affect the wage level in period  $t$ . However, wages in period  $t$  cannot influence the likelihood of using the Internet in period  $t - 1$ . In Eq. (6) below,  $t$  represents the recent period (e.g., 2018),  $t - 1$  represents the prior period (e.g., 2016), and  $\text{Int}_{t-1}$  expresses the Internet use in the prior period ( $t - 1$ ).

$$\begin{aligned} \ln W_{it} = & \alpha + \beta_1 \text{Int}_{it-1} + \beta_2 \text{Female} * \text{Int}_{it-1} \\ & + \gamma \sum X_i + \rho_i + \varepsilon_{it} \end{aligned} \quad (6)$$

Then, to investigate the effect of Internet use on the gender wage gap, we use the Blinder–Oaxaca decomposition method (Blinder 1973; Oaxaca 1973) to decompose the determinants into two components: (i) the explained component, comprising the gender endowment differences (i.e., the gender disparity in Internet access, etc.), and (ii) the unexplained component, composed of the gender differences in the evaluated price of each factor (i.e., the gender difference in the return to Internet use, etc.). Equations (7, 8) describe the model<sup>6</sup>:

<sup>5</sup> The information on these two IVs were obtained from the data published in *China Statistical Yearbook 1999* (NBS 1999).

<sup>6</sup> The constant is omitted for descriptive convenience. We conduct two decomposition analyses based on Eqs. (7) and (8); only one has been listed because the results are very similar.

$$\overline{LnW_m} - \overline{LnW_f} = \beta_m \left( \sum \bar{H}_m - \sum \bar{H}_f \right) + (\beta_m - \beta_f) \sum \bar{H}_f \quad (7)$$

$$\overline{LnW_m} - \overline{LnW_f} = \beta_f \left( \sum \bar{H}_f - \sum \bar{H}_m \right) + (\beta_f - \beta_m) \sum \bar{H}_m \quad (8)$$

In Eqs. (7) and (8),  $\overline{LnW_m} - \overline{LnW_f}$  is the gender wage gap;  $\beta_m$ ,  $\beta_f$  is coefficient of each factor calculated based on the male and female wage functions, respectively;  $\bar{H}$  denotes the mean value of each factor, including Internet use.  $\beta_m(\sum \bar{H}_m - \sum \bar{H}_f)$  or  $\beta_f(\sum \bar{H}_f - \sum \bar{H}_m)$  is the explained component, and  $(\beta_m - \beta_f)\sum \bar{H}_f$  or  $(\beta_f - \beta_m)\sum \bar{H}_m$  is the unexplained component which includes the discrimination against women in the workplace.

To compare the effects of Internet use by groups, we also calculate the estimations per educational attainment and age cohort group.

### 3.2 Data

Three waves of data (2014, 2016, and 2018) from the China Family Panel Studies (CFPS 2020) dataset are used in this study. The reasons for using the CFPS are considered as follows: first, the CFPS is a nationally representative longitudinal survey of Chinese communities, families, and individuals launched in 2010 by the Institute of Social Science Survey, Peking University, China. Although Chinese General Social Survey (CGSS) data was used in the previous studies (e.g., Qi and Liu 2020), since the CGSS is the cross-sectional survey data, the individual heterogeneity problem could not be addressed. On the contrary, this study performs the analysis using the panel data analysis method (e.g., the FE and RE models, the LV model) to deal with the endogeneity problems, this study, based on the CFPS, can provide the robustness results on the issue. Second, although the China Health and Retirement Longitudinal Survey (CHARLS) is a longitudinal survey having information on Internet use, the survey targets of CHARLS are individuals aged 45 and older, whereas the CFPS covers all age generations. Therefore, we can use the data from the CFPS to compare the differences in the effect of Internet use among the younger, middle-aged, and older generations in this study. Third, we can obtain the rich information on Internet use, such as the used Internet and the frequency of using the Internet for different purposes (e.g., working and study, communication, and leisure); the latter is the unique question item in the CFPS which is firstly utilized in this study on the issue.

The CFPS is designed for individual-, family-, and community-level longitudinal data collection in contemporary China and provides information on Internet use, wages, and other factors (education, years of work experience, sex, occupation, industry sector, etc.). The

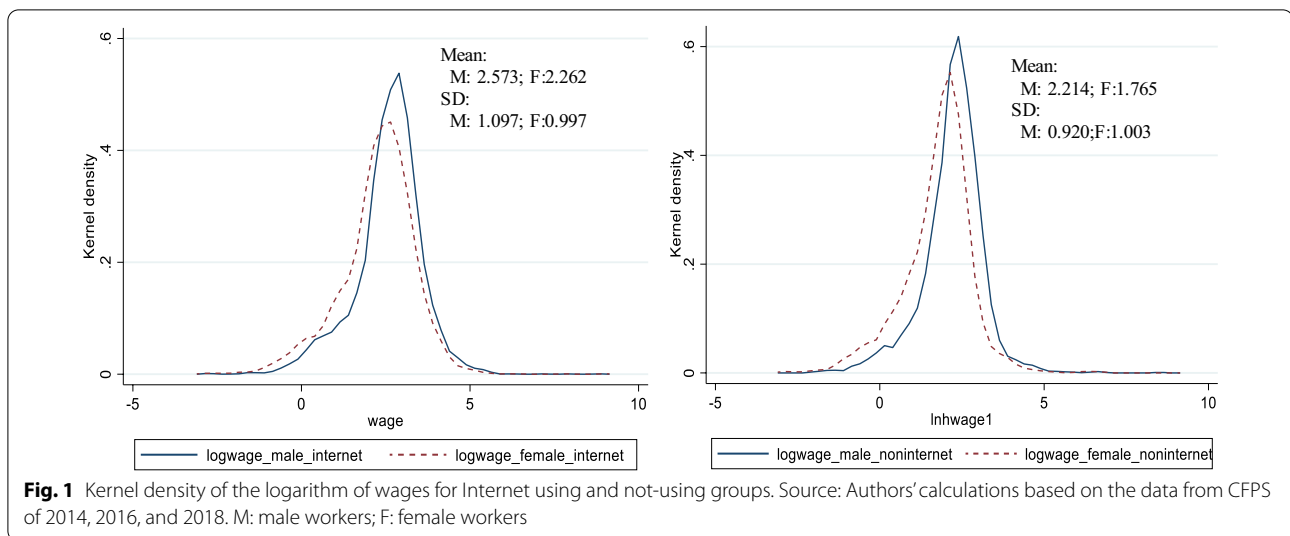
2010 CFPS baseline survey data was obtained through multi-stage probability sampling with implicit stratification. Multi-stage sampling reduces the operational cost of the survey and permits the analysis of the social context. In the 2010 baseline survey, the CFPS successfully interviewed nearly 15,000 families and 30,000 individuals within these families, with an approximate response rate of 79%. The respondents were tracked through annual follow-up surveys. The CFPS 2010 covers 25 provinces and municipalities. Only the latest three waves (2014, 2016, and 2018) of the CFPS, which include the survey item on Internet use, have been used in this study. The CFPS sample sizes are 37,147 (2014), 36,892 (2016), and 37,354 (2018).

The logarithmic value of the hourly wage is used as the dependent variable. The wages for 2014, 2016, and 2018 have been adjusted using the regional Consumption Price Index (CPI) in the rural and urban areas published by China's National Bureau of Statistics (NBS 1999) to account for inflation, using the 2014 CPI as the baseline.

We calculate the hourly wage as a dependent variable based on the annual earned income and corresponding working hours.

The key independent variable is an Internet use dummy variable based on the questionnaire item: "Did you use the Internet in the past year?" (1=has used the Internet in the past year, 0=otherwise); we primarily use the variable to estimate the extensive margin effect of Internet use, which was estimated in the literature. Based on the questionnaire items on the frequency of Internet use by purpose (work, study, communication, shopping, entertainment), we originally constructed three indicators to investigate the intensive margin effect of Internet use in this study: (a) frequency of using the Internet for work, including work and study, (b) frequency of using the Internet for communication, and (c) frequency of using the Internet for leisure, including shopping and entertainment. Based on the five questions items in the CFPS as "please answer the frequency of using the Internet for study, work, communication, entertainment, commercial activity (e.g., Internet payment, shopping): (i)almost every day; (ii)3–4 times a week; (iii)1–2 times a week; (iv) 2–3 times a month; (v) once a month; (vi) once few months; (vii) never use"; we re-coded each frequency as "7=almost every day; 6=3–4 times a week; 5=3–4 times a week; 4=1–2 times a week; 3=once a month; 2=once in few months; 1=never use". We calculated the total values for (a) and (c) and used their arithmetic mean in the analysis.

We constructed an interaction term of Internet use and a female dummy variable to investigate the gender difference in return to Internet use in wage functions.



Based on the economic theories and existing studies, we identified a set of variables that may affect wages as specified by the wage functions, such as years of schooling, years of work experience and its squared term, ethnicity (1=han, 0=minority), party membership (1=member of Communist Party of China, 0=non-member), urban residents (1=urban residents, 0=rural residents), occupation (1=manager and technician, 0=otherwise), industry sector (1=manufacturing industry sector, 0=otherwise), workplace ownership (1=state-owned sector, 0=otherwise), region (west, central, east), and survey year.

As described in Sect. 3.1, we used two variables: the provincial optical cable circuit and the provincial long-distance cable line length in 1999, as IVs in this study.

This analysis is limited to respondents aged 16–60 years and excludes missing values; the longitudinal sample size is 18,381.<sup>7</sup>

## 4 Empirical results and discussion

### 4.1 Descriptive statistics

Figure 1 shows the kernel density of the logarithmic value of wages by Internet-using and non-using groups. The wage level is higher for men than women in both groups, and the gender wage gap is higher in the Internet-using group than that in the not using one. The results indicate

that there exists a gender wage gap, and the gender wage gap differs among Internet using and non-using groups.

The proportions of individuals using the Internet by gender are summarized in Table 1. In general, an increase is observed from 2014 to 2018 for both men and women. However, the percentage is approximately 5% higher for men than women in each year of the three years, suggesting the existence of gender disparity in Internet access.

In terms of the disparity by educational attainment group,<sup>8</sup> the percentage of individuals using the Internet is higher for men than women in the low- and high-education groups, whereas the opposite holds in the middle-education group. Regarding the disparity by age cohort, the percentage of individuals using the Internet is higher for men than women in each age cohort. The gender disparity in Internet access is the largest in the middle-aged generation born during 1970–1989 and the least in the younger generation born after 1990.<sup>9</sup> The results indicate that the gender disparities in Internet access differ by educational background and age cohort. Therefore, the heterogeneous group should be considered in the analysis.

### 4.2 Basic results

Table 2 presents the results of the wage function analysis using five models—the OLS (Model 1), IV (Model 2),

<sup>7</sup> According to the *Interim Measures of the State Council on the Placement of Elderly, Weak, Sick and Disabled Cadres* (Article 4) and *Interim Measures of the State Council on Retirement and Retirement of Workers* (Article 1) published in 1978, the minimum employment age in China is 16 for both men and women. However, the mandatory retirement age differs by gender: it is 60 years for both blue- and white-collar male workers, while it is 50 years for blue-collar female workers and 55 years for white-collar female workers (e.g., women working as a civil servant/office executive, or working as a manager, technician, or a professor in the public sector).

<sup>8</sup> The questionnaire item on education attainment has eight categories: illiterate/semiliterate, elementary school, junior high school, senior high school, college, university, master, and doctor. We divided the educational groups into three categories: the low- (elementary school and lower), middle- (junior and senior high school), and high- (college and higher) educational groups.

<sup>9</sup> We divide the age cohorts considering the Internet diffusion situation. We define the younger generation as the group born in the year when the Internet was used in society, the middle-aged generation who used the Internet in their adolescence, and the older generation who used it in their middle- or older age.

**Table 1** The proportion of using the Internet by sex

	Unit: %			
	Total	Males	Females	Gap (M-F)
Total				
2014	34.84	37.82	32.08	5.74
2016	49.88	52.93	46.95	5.98
2018	62.41	65.19	59.74	5.46
Low education				
2014	6.61	8.37	5.61	2.77
2016	15.72	17.46	14.72	2.75
2018	34.04	38.21	31.18	7.03
Middle education				
2014	41.21	40.42	42.14	-1.71
2016	56.15	55.10	57.39	-2.29
2018	69.52	68.60	70.60	-2.00
High education				
2014	65.98	68.89	63.20	5.68
2016	74.28	75.63	72.81	2.82
2018	94.73	94.64	94.82	-0.19
Born before 1979				
2014	6.03	8.14	3.80	4.34
2016	11.32	12.75	9.73	3.02
2018	19.45	21.65	17.01	4.64
Born in 1979–1989				
2014	32.44	36.46	28.81	7.65
2016	47.36	51.41	43.56	7.85
2018	58.55	61.95	55.33	6.62
Born after 1990				
2014	81.28	81.94	80.69	1.25
2016	89.31	90.24	88.41	1.83
2018	93.14	93.45	92.83	0.62

Source: Authors' calculations based on the data from CFPS of 2014, 2016, and 2018

Gap = percentage of male Internet user - percentage of female Internet user

Age range: 16–60 years

Low education: elementary school and lower; Middle education: junior and senior high school; High education: college and higher

LV (Model 3), RE (Model 4), and FE (Model 5).<sup>10</sup> Regarding the appropriateness of the IVs, the endogeneity test result (Durbin–Wu–Hausman test) is statistically significant at the 1% level; therefore, the null hypothesis considering that all the variables are exogenous is rejected. The Hansen J statistic is not significant, thus revealing that the IV is exogenous in the second stage estimation. Furthermore, the Cragg–Donald Wald F statistic is 26.393, which is larger than 10, suggesting that the weak identification problem can be neglected. Therefore, it can

be considered that the IVs are valid. The results of the F test and the Breusch–Pagan Lagrange multiplier test indicate that both the FE and RE models are appropriate compared to the OLS model. The Hausman test results (2308.67,  $p=0.000$ ) suggest that the FE model is more appropriate than the RE model. The main findings are summarized as follows.

First, the coefficients of Internet use are significantly positive in Models 1–5, suggesting that Internet use may increase the wage levels when addressing individual heterogeneity and other endogeneity problems. The results can be explained based on human capital theory considering Internet use skill as a kind of human capital, which can potentially increase labor productivity. Additionally, Internet use may have a signaling effect whereby an employer may evaluate that an Internet-using worker possesses higher skills than a non-Internet-using worker.

Second, in terms of the gender differences in return to Internet use, the results of interaction item of Internet use and female dummy variable in Model 1 (OLS), Model 3 (LV), and Model 4 (RE) indicate that the return to Internet use is significantly greater for women than men. However, considering the endogeneity problem, the wage premium is significantly greater for men than women in Model 2 (IV). Additionally, although the result is non-significant in Model 5 (FE), it is a negative value. The results suggest a bias in the results derived from the OLS model. When using appropriate models (e.g., the IV or FE model) to address the endogeneity problem, the return to Internet use is higher for men than women.

### 4.3 Estimations on the intensive margin effect of Internet use

As shown in Table 2, although we analyzed the extensive margin effects of Internet use on the gender wage gap using a binary variable (whether used Internet), the intensive margin effect of Internet use (e.g., the frequency of using the Internet) on the wage gap should also be considered. Additionally, the effects of Internet use on the wage gap may differ for different purposes of Internet use. For example, when comparing the group using the Internet frequently for leisure (e.g., entertainment, shopping) to the group using it frequently for work or study, it is assumed that the group using the Internet frequently for work or study is likely to obtain higher earned income. To investigate the intensive margin effects of Internet use, we used the three indices: (i) the frequency of using the Internet for work and study, (ii) the frequency of using the Internet for communication, and (iii) the frequency of using the Internet for leisure including shopping and entertainment, to replace the binary variable of Internet use in Table 2 and re-estimate the models. The results are presented in Table 3.

<sup>10</sup> Due to space limit constraints, results, including all control variables, were reported in Additional file 1: Table S1.



**Table 2** The gender differences in return to Internet use

	(1) OLS		(2) IV		(3) LV		(4) RE		(5) FE	
	Coef	t	Coef	t	Coef	t	Coef	t	Coef	t
Internet use	0.603***	6.68	14.900***	5.59	0.107	0.98	0.606***	6.65	0.405**	2.26
Female	-0.849***	-11.29	4.054***	4.40	-0.697***	-5.76	-0.844***	-10.90	-	-
Internet use x Female	0.540***	5.19	-8.881***	-5.04	0.360**	2.51	0.531***	5.02	-0.106	-0.42
Control variables	Yes		Yes		Yes		Yes		Yes	
First stage estimation										
IV1			0.083***	3.47						
IV2			-0.243***	-11.58						
Observations	18,381		18,381		7777		18,381		18,381	
R-squared	0.476				0.367		12,876		12,876	
R-sq. Between							0.190		0.197	
Within							0.574		0.017	
Overall							0.476		0.008	
F-test (Prob > F)							66.51 (p=0.000)		0.000	
BP test (Prob > chi2)										
Hausman test (Prob > chi2)										
Endogeneity test (DWH)			69.216 (p=0.000)							
Hansen J statistic			p=0.224							
Cragg-Donald Wald F statistic			26.393							

Source: Authors' Calculations Based on the data from CFPS of 2014, 2016 and 2018

Control variables, including years of schooling, years of work experience and its squared term, party membership, urban, occupation, industry sector, workplace ownership, region, and survey year variables, have been calculated, but the results are not listed in the table owing to space limit constraints

OLS: ordinary least squares; IV: instrumental variable method; LV: lagged variable model; RE: random effects model; BP test: Breusch-Pagan Lagrange multiplier test; DWH: Durbin-Wu-Hausman test; IV1: the provincial optical cable circuit in 1999; IV2: the provincial long-distance cable line length in 1999

\*\*\*p < 0.01; \*\*p < 0.05; \*p < 0.1

**Table 3** Frequency of Internet use for different purposes

	(1) Work and study		(2) Communication		(3) Living	
	IV	FE	IV	FE	IV	FE
Internet use	2.020*** (4.96)	− 0.003 (− 0.14)	2.540*** (5.59)	0.056* (1.94)	1.385*** (6.58)	0.010 (0.45)
Female	2.709*** (3.77)		2.987*** (4.21)		1.922*** (4.44)	
Internet use × Female	− 1.317*** (− 4.66)	− 0.008 (− 0.29)	− 1.540*** (− 5.11)	− 0.016 (− 0.39)	− 0.797*** (− 5.93)	− 0.015 (− 0.54)
First stage estimation						
IV1	0.204*** (4.66)		0.151*** (4.91)		0.411*** (9.78)	
IV2	− 0.351*** (− 9.55)		− 0.249*** (− 9.63)		− 0.612*** (− 17.32)	
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,605	18,605	18,605	18,605	18,605	18,605
Groups		12,920		12,920		12,920
R-sq. Between		0.192		0.198		0.192
Within		0.015		0.030		0.015
Overall		0.007		0.033		0.006
F test (Prob > F)		0.000		0.000		0.000
BP test (Prob > chibar2)		73.70 ( $p = 0.000$ )		77.77 ( $p = 0.000$ )		77.22 ( $p = 0.100$ )
Hausman test		2406.81 ( $p = 0.000$ )		2355.36 ( $p = 0.000$ )		2380.03 ( $p = 0.000$ )
Endogeneity test (DWH)	$p = 0.000$		$p = 0.000$		$p = 0.000$	
Hansen J statistic	$p = 0.917$		$p = 0.082$		$p = 0.482$	
Cragg-Donald Wald F statistic	18.682		26.134		47.343	

Source: Authors' Calculations Based on the data from CFPS of 2014, 2016 and 2018

Control variables, including years of schooling, years of work experience and its squared term, party membership, urban, occupation, industry sector, workplace ownership, region, and year variables have been calculated but are not listed in the table owing to space limit constraints

IV: Instrumental variable method; FE: fixed effects model; BP test: Breusch–Pagan Lagrange multiplier test; DWH: Durbin–Wu–Hausman test

\*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ .  $t$ -values are shown in parentheses

The results from the Hausman specification test, F-test, and the Breusch–Pagan Lagrange multiplier test indicate that the FE model is more appropriate than the OLS and RE models. Regarding the validity of the IV method, the results in the first stage estimations indicate that both IVs significantly affect the likelihood of using the Internet at the 1% level; the results from the Durbin–Wu–Hausman test, Hansen J statistic, and Cragg–Donald Wald F statistic values indicate that the IV method is valid. Therefore, we report the results using the FE and IV models in Table 3.<sup>11</sup>

First, the results from the IV method indicate that the coefficients of Internet use are positive and significant at the 1% level in Models 1–3. It is also positive and significant at the 10% level in Model 2 from the FE model. The conclusions are consistent with the results in Table 2.

Second, the results from the IV methods show that the coefficients of the interaction term of Internet use and

female dummy variables are negative values and significant at the 1% level in Models 1–3, but they are not significant in the results from the FE model. These results are consistent with those in Table 2.

#### 4.4 Estimations considering heterogeneous group

The results for the heterogeneous group based on educational attainment and age cohort are summarized in Tables 4, 5.<sup>12</sup> The results from the Hausman specification test, F-test, and the Breusch–Pagan Lagrange multiplier test indicates that the FE model is more appropriate than the OLS and RE models. Regarding the validity of the IV method, the results in the first stage estimations indicate that both IVs significantly affect the likelihood of using the Internet at the 1% level; the results from the Durbin–Wu–Hausman test, Hansen J statistic, and Cragg–Donald Wald F statistic values indicate that the IV method is

<sup>11</sup> Due to space limit constraints, results, including all control variables, were reported in Additional file 1: Table S2.

<sup>12</sup> Due to space limit constraints, results, including all control variables, were reported in Additional file 1: Tables S3 and S4.

**Table 4** The gender differences in return to Internet use by educational group

	(1) Low		(2) Middle		(3) High	
	IV	FE	IV	FE	IV	FE
Internet use	- 11.427 (-0.94)	0.496 (0.93)	10.277*** (4.60)	0.475** (2.22)	19.852** (2.37)	- 0.083 (- 0.14)
Female	- 2.598 (- 1.41)		3.403*** (3.50)	16.978** (2.25)		
Internet use × Female	9.648 (- 0.97)	0.139 (0.21)	- 6.856*** (- 4.40)	- 0.370 (- 1.21)	- 18.215 (- 2.28)	- 0.033 (- 0.04)
First stage estimation						
IV1	- 0.167*** (- 3.71)		0.165*** (5.40)		0.466*** (4.58)	
IV2	- 0.060 (- 1.23)		- 0.277*** (- 10.83)		- 0.472*** (- 5.70)	
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4652	4652	10,284	10,284	3445	3445
Groups		3289		7070		1905
R-sq. Between		0.017		0.200		0.200
Within		0.008		0.004		0.004
Overall		0.000		0.001		0.001
F test (Prob > F)		0.000		0.000		0.000
BP test (Prob > chibar2)		0.000 ( $p=0.987$ )		17.67 ( $p=0.000$ )		19.58 ( $p=0.000$ )
Hausman test		248.75 ( $p=0.000$ )		874.72 ( $p=0.000$ )		874.72 ( $p=0.000$ )
Endogeneity test (DWH)	$p=0.131$		$p=0.000$		$p=0.000$	
Hansen J statistic	$p=0.001$		$p=0.652$		$p=0.127$	
Cragg-Donald Wald F statistic	1.199		21.914		19.93	

Source: Authors' Calculations Based on the data from CFPS of 2014, 2016 and 2018

Low: elementary school and lower; middle: junior and senior high school; high: college and higher

Control variables, including years of schooling, years of work experience and its squared term, party membership, urban, occupation, industry sector, workplace ownership, region, and survey year variables, have been calculated, but the results are not listed in the table owing to space limit constraints

IV: Instrumental variable method; FE: Fixed effects model; BP: test Breusch-Pagan Lagrange multiplier test; DWH: Durbin-Wu-Hausman test

\*\*\*:  $p < 0.01$ ; \*\*:  $p < 0.05$ ; \*:  $p < 0.1$ .  $t$ -values are shown in parentheses

valid in most cases. Therefore, we report the results using the FE and IV models in Tables 4, 5.

Table 4 presents the low-, middle- and high-educational group results. To secure enough samples for the analysis and consider the distribution of workers by education attainment levels,<sup>13</sup> we distinguished the samples into three groups (i) the low-education group (elementary school and lower); (ii) the middle-education group (junior and senior high school); and (iii) the high-education group (college and higher). The main findings are as follows.

First, in terms of the return to Internet use by the educational group, the results from the IV method show that the coefficients of Internet use are positive and significant at the 1% level in the middle- and high-educational groups, whereas it is not significant for the low-education group. The results from the FE model reveal that the Internet use coefficient is significant only for the middle education group at the 5% level. They suggest the effects of the return to Internet use on wages are much more significant for the middle- and high-educational groups. Additionally, the individual heterogeneity problem considerably affects the results.

Second, in terms of the gender difference in return to Internet use by the educational group, the results from the IV method show that the coefficient of the interaction of Internet use and female dummy is negative and significant at the 5% level for the middle- and high-educational

<sup>13</sup> The distribution proportion of workers by the educational group in the samples is 25.3% for the elementary school and lower group, 55.9% for the junior and senior high school group, and 18.8% for the college and above group.

**Table 5** The gender disparity in return to Internet use by age cohort

	(1) Born before 1969		(2) Born from 1970–1989		(3) Born after 1990	
	IV	FE	IV	FE	IV	FE
Internet use	28.375*** (3.41)	− 0.822 (− 1.32)	12.618*** (4.88)	0.341* (1.75)	− 24.696 (− 0.90)	1.322 (2.62)
Female	0.077 (0.28)		3.510*** (3.78)		− 22.009 (− 0.97)	
Internet use × Female	− 23.690*** (− 3.25)	1.404 (1.14)	− 8.217*** (− 4.45)	0.011 (0.04)	24.273 (0.96)	− 1.207 (− 1.25)
First stage estimation						
IV1	0.290*** (5.09)		0.088*** (3.32)		− 0.032 (− 0.44)	
IV2	− 0.383*** (− 11.13)		− 0.236*** (− 9.99)		− 0.155* (− 1.91)	
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,512	3,512	12,701	12,701	2,168	2,168
Groups		2,256		9,229		1,676
R-sq. Between		0.230		0.198		0.200
Within		0.002		0.030		0.004
Overall		0.000		0.033		0.001
F test (Prob > F)		0.999		0.000		0.000
BP test (Prob > chibar2)		0.000 ( $p = 1.000$ )		49.00 ( $p = 0.000$ )		1.63 ( $p = 0.100$ )
Hausman test		67.74 ( $p = 0.000$ )		2736.48 ( $p = 0.000$ )		57.86 ( $p = 0.000$ )
Endogeneity test (DWH)	$p = 0.000$		$p = 0.000$		$p = 0.083$	
Hansen J statistic	$p = 0.554$		$p = 0.604$		$p = 0.625$	
Cragg-Donald Wald F statistic	11.590		20.405		11.201	

Source: Author's calculations based on the data from CFPS of 2014, 2016 and 2018

Control variables, including years of schooling, years of work experience and its squared term, party membership, urban, occupation, industry sector, workplace ownership, region, and survey year variables, have been calculated, but are not listed in the table owing to space limit constraints

IV: Instrumental variable method; FE: Fixed effects model; BP test: Breusch–Pagan Lagrange multiplier test; DWH: Durbin–Wu–Hausman test

\*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ .  $t$ -values are shown in parentheses

groups, while it is insignificant for the low-education group, thus indicating when the other factors are held consistent, the return to Internet use is lower for female workers than male workers, and the gender difference in return to Internet use is higher in the middle- and high-educational groups than the low-education group. However, the results from the FE model are not significant in three educational groups.

Third, comparing the magnitude of the coefficients from the IV method show that the gender difference in return to Internet use is larger in the high-education group (− 18.215) than in the middle education group (− 6.856). The possible reasons can be considered as follows: the discrimination against high-educated women in the workplace, which may be caused by the glass-ceiling problem, or due to the gender occupational segregation

among the high-education group is serious much more than that in the low- and middle-educational groups.

Table 5 summarizes the results by three age cohorts: the younger (born after 1990), middle-aged (born in 1970–1989), and older (born before 1969) generations. First, the results from the IV method show that the coefficients of Internet use are positive and significant at 1% levels in both the middle-aged and older cohort groups, while they are not significant for the younger cohort. The FE model results reveal that the coefficients of Internet use are significant for both the younger and middle-aged cohort groups at 5 and 10% levels, while it is not significant for the older cohort. The results suggest that, in general, there is a positive effect of Internet use on wages in each age cohort group, and the influence of individual heterogeneity on the return to Internet use is greater for the older cohort.



**Table 6** Decomposition results of Internet use and the gender wage gap

	Value		Percentage	
	Explained	Unexplained	Explained (%)	Unexplained (%)
Total	0.471	0.449	51.2	48.8
Internet use	0.461	0.665	50.1	72.4
Education	-0.052	0.005	-5.7	0.5
Experience	-0.044	1.087	-4.8	118.2
Ethnicity	-0.002	-0.107	-0.2	-11.6
Party	-0.027	-0.029	-3.0	-3.2
Occupation	-0.018	-0.005	-2.0	-0.5
Industry	0.021	-0.160	2.3	-17.4
State-owned	0.000	-0.188	0.0	-20.4
Urban	0.067	0.037	7.3	4.0
Region	0.001	0.219	0.1	23.8
Year	0.064	0.002	7.0	0.2
Constant	0.000	-1.077	0.0	-117.1

Source: Authors' Calculations Based on the data from CFPS of 2014, 2016 and 2018

The decomposition is based on the results from the IV method

Second, in terms of the gender difference in return to Internet use by age cohorts, the results from the IV method show that the interaction coefficients of Internet use and female dummy are negative values and significant at 1% or 5% level for both middle-aged and older cohorts, and not significant for the younger cohort. The results from the FE model are not significant in the three age cohorts.

Third, the gender difference in return to Internet use was found higher in the older cohort than the other age cohorts when compared using the magnitude of the coefficients based on the IV method. It may be caused by that the discrimination against female workers is serious much more among older generations than those among the younger and middle-aged generations.

**4.5 Decomposition results**

The results in Table 1 indicate that Internet access differs by gender, and those in Tables 2, 3 suggest that the return to Internet use is different for men and women. However, how the two components affect the formation of the gender wage gap is unclear. Therefore, we conduct a decomposition analysis to calculate the contribution rates of these two components (Table 6). We also perform the decomposition analyses based on the educational attainment and age cohort (Tables 7, 8).

Table 6 reports the decomposition results for the total sample. First, the explained and unexplained components contribute to the formation of the gender wage gap. The influence is slightly less for the former (46.3%) than the latter (53.7%).

Second, in terms of the effects of Internet use, it is shown that both the gender disparity in Internet access and gender difference in the return to Internet use contribute to widening the gender wage gap, and the contribution rates of both are higher than the other factors (e.g., education, occupation). Additionally, the contribution rate is larger for the return to Internet use

**Table 7** Decomposition results of Internet use and the gender wage gap by educational group

	Value		Percentage	
	Explained	Unexplained	Explained (%)	Unexplained (%)
(a) Low (N = 4652)				
Total	0.579	1.019	36.2	63.8
Internet use	-0.842	-2.097	-52.7	-131.3
Other variables	1.421	3.116	88.9	195.1
(b) Middle (N = 10,284)				
Total	0.051	0.357	12.4	87.6
Internet use	-0.224	-0.484	-54.9	-118.7
Other variables	0.275	0.841	67.3	206.3
(c) High (N = 3445)				
Total	-0.113	0.229	-97.3	197.3
Internet use	-0.255	3.480	-219.6	2996.8
Other variables	0.142	-3.251	122.3	-2799.5

Source: Authors' Calculations Based on the data from CFPS of 2014, 2016 and 2018

The decomposition is based on the results from the IV method

Low: elementary school and lower; Middle: junior and senior high school; High: college and higher

"Other variables" include years of schooling, years of work experience, party membership, urban, occupation, industry sector, workplace ownership, region, and years

**Table 8** Decomposition results of Internet use and the gender wage gap by age cohort

	Value		Percentage	
	Explained	Unexplained	Explained (%)	Unexplained (%)
(a) Born before 1969 (N = 3512)				
Total	1.000	1.162	46.3	53.7
Internet use	1.023	- 2.401	47.3	- 111.0
Other variables	- 0.023	3.563	- 1.0	168.3
(b) Born between 1970–1989 (N = 12,701)				
Total	0.538	0.445	54.7	45.3
Internet use	0.536	- 0.004	54.6	- 0.4
Other variables	0.002	0.449	0.1	45.7
(c) Born after 1990 (N = 2168)				
Total	0.132	0.136	49.1	50.9
Internet use	0.044	- 3.022	16.3	- 1129.4
Other variables	0.088	3.158	32.8	1180.3

Source: Author's Calculations Based on the data from CFPS of 2014, 2016 and 2018

The decomposition is based on the results from the IV method

"Other variables" include years of schooling, years of work experience, party membership, urban, occupation, industry sector, workplace ownership, region, and years

(72.4%) than the Internet access (50.1%). The results indicate that although both components drive the gender wage gap in China, the effect of the gender difference in return to Internet use is greater, suggesting that workplace discrimination against women in terms of Internet use is a serious issue in China.

Table 7 shows the decomposition results based on the educational attainment group. The results indicate that the effects of gender differences in both the Internet access and return to Internet use differ across different educational groups.

First, the gender disparity in Internet access reduces the wage gap in three educational groups; its effect is greater in the high-education group than those in the low- and middle-education groups.

Second, the gender difference in return to Internet use widens the gender wage gap in the high-education group, whereas they reduce the wage gap in the low- and middle-education groups.

Hence, these results suggest that the workplace discrimination against female workers is greater for the high education group in terms of Internet use than the other group, which widens the gender wage gap in China.

Table 8 presents the results of the decomposition analysis for three age cohorts: (i) the older cohort (born before 1969), (ii) the middle-aged cohort (born between 1979 and 1989), and (iii) the younger cohort (born after 1990).

First, the gender disparity in Internet access widens the gender wage gap in three age cohorts; the effects are greater for the middle-aged (54.6%) and older age cohorts (47.3%) than for the younger age cohorts (16.3%).

Second, the gender difference in return to Internet use reduces the gender wage gap in three cohorts; the effects are greater for the younger age cohorts.

The results of Table 8 suggest that the gender disparity in Internet access in the middle/older age cohorts and the workplace discrimination against the younger women in terms of Internet use widens the gender wage gap in China.

These results contradict those of Qi and Liu (2020), who report that both components reduce the gender wage gap in the younger, middle-aged, and older groups. There are two reasons: first, the method of analysis differs. This study conducts the decomposition analysis based on the IV method, whereas Qi and Liu (2020) use the OLS model; it can be concluded that an endogeneity problem such as the unobservable omitted variable issue might exist in the earlier studies. Second, the period of analysis in this study is 2014–2018, whereas it is 2010–2015 in Qi and Liu (2020). With Internet diffusion and Internet technology progressing, the effects of both components may change over time. The results in this study suggest that the gender divisions in Internet accessibility might become severe in the middle-aged and older age cohorts in the recent period (2014–2018).

#### 4.6 Further discussions on the limitations of this study

It should be noted that this study has several limitations. First, although we used the IV, LV, RE, and FE models to attempt to address the endogeneity problem, future research should also explore the causal association between Internet use and the gender wage gap.

Second, the gender gap in the return to Internet usage may also result from the gender disparity in

Internet-using abilities (or skills). Ge and Zhou (2020) report that the firm attributes (e.g., capital, trade exposure) and firm technology level (e.g., robot or computer use situations) affect the gender wage gap. Future studies can conduct a detailed survey on individuals' Internet use skills and workplace technology levels.

Third, although the occupation, industry sector, and enterprise ownership type were used to control the influence of the workplace on the gender wage gap, other factors (e.g., wage and employment systems, enterprise attributes) may also affect the wage gap. We should conduct an employer-employee survey on the issue in the future.

Fourth, some studies have found that the work preferences differ by gender (e.g., Beblo and Görge 2018). This gender preference disparity in using new technology can also affect Internet access and return to Internet use. Further research should consider the influence of personality factors and self-selection on the issue.

Finally, due to the limitation of the survey period, we only investigated the issue in the current period (2014, 2016 and 2018), and the longer-term analysis has become a new challenge in the future.

## 5 Conclusions

Using national longitudinal data from CFPS of 2014, 2016, and 2018, this study empirically analyzed the influence of Internet use on the gender wage gap in China, considering the endogeneity problems. It yields the following four main conclusions.

First, according to the results derived from the OLS model, the return to Internet usage is higher for women than men, meaning the results are similar to those of earlier studies. However, when longitudinal survey data is used to address heterogeneity and other endogeneity issues based on the IV and FE models, the results show that the return to Internet usage is higher for men than women. The individual heterogeneity problem considerably affects the estimations, thus suggesting an estimation bias in the existing literature. The results based on the frequency of Internet use for different purposes confirm the conclusions.

Second, the gender difference in return to Internet use differs by heterogeneous groups: it is higher in the middle-/high- education groups and middle-aged/older age cohorts than those in low education and younger age cohorts.

Third, the decomposition results indicate that, in general, the two components (the gender disparity in Internet access and the gender difference in return to Internet use) drive the gender wage gap in China; the effect of gender difference in the return to Internet use is greater.

Fourth, the influence of Internet usage on the gender wage gap varies with the educational attainment and age cohort groups. For example, the gender difference in the return to Internet use widens the gender wage gap in the high-education group while they reduce the wage gap in the low- and middle-education groups; the influences of gender disparity in Internet access on the formation of gender wage gap are greater for the middle-aged and older generations than the younger generation.

The study highlights two policy implications. First, there exists a large gender difference in return to Internet use even when the other factors (e.g., education, occupation) are held constant, which highlights the prevalence of workplace discrimination against women in terms of Internet use, especially for the higher-educated women. The enforcement of the implementation of equality policies, such as the equality employment policy and "equal pay for equal work" policy, can be expected to reduce the gender wage gap. Moreover, the discrimination against women may be caused by the more family responsibilities for women than men (Connelly et al. 2018; Ma 2021a, b). Thus, policies that reduce the responsibilities of childcare and geriatric care, such as the one to promote the establishment of public kindergartens and long-term care insurance, are also expected to close the gender wage gap in the long run. Second, the results suggest that the policies aimed at the reduction of the Internet access disparities among various groups, such as women in the middle-aged and older generations, may contribute to reducing the gender wage gap.

Despite these limitations mentioned above, this study investigated the influence of Internet use on the gender wage gap and provided new evidence on the determinants of the gender wage gap in the era of the digital economy from China. We anticipate that the insights about the gender disparities in Internet access and gender difference in return to Internet use (including the discrimination against women in terms of Internet use in the workplace), that contribute to the formation of the gender wage gap in China, can provide valuable lessons for other countries as well.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12651-022-00320-9>.

**Additional file 1: Table S1.** The gender differences in return to Internet use. **Table S2.** Results using frequency of Internet use for different purposes. **Table S3.** The gender differences in return to Internet use by educational group. **Table S4.** The gender differences in return to Internet use by age cohort.

### Acknowledgements

The authors are grateful to professor Boris Hirsch, and anonymous reviewers for their helpful comments and suggestions.

### Author contributions

X. MA designed the study, collected the data, performed the formal analyses, and wrote the original manuscript; the author read and approved the final manuscript.

### Author information

Dr. X. MA is a professor at the Faculty of Economics, Hosei University, Japan. She was the editor of the *Japanese Journal of Comparative Economics*, *Asian Studies*, and the *Journal of Chinese Economics*. Her articles have been published in peer-reviewed journals such as *China Economic Review*, *Journal of Asian Economics*, *Economic Systems*, *Education Economics* and *Journal for Labor Market Research*. She is the author of *Economic Transition and Labor Market Reform in China* (Palgrave Macmillan 2018), *Female Employment and Gender Gap in China* (Springer 2021), and the single editor of *Employment, Retirement and Lifestyle in Aging East Asia* (Palgrave Macmillan 2021).

### Funding

This research was supported by JSPS (Japan Society for the Promotion Science) Grant-in-Aid for Scientific Research (Grant Numbers: 20H01512 and 20H01489).

### Data availability

The data can be obtained through the website <https://opendata.pku.edu.cn/dataset/CFPS?language=en>. The datasets constructed by the author are available upon request.

### Declarations

#### Ethics approval and consent to participate

The dataset used in this study, the China Family Panel Studies (CFPS), is publicly available (<http://opendata.pku.edu.cn/en>). The medical ethics committee approved the CFPS study, and all interviewees were required to sign informed consent. Ethics approval for the data collection in CFPS was obtained from the Biomedical Ethics Review Committee of Peking University (IRB00001052-14010). Ethics approval for the use of CFPS data was obtained from the University of Newcastle Human Research Ethics Committee. And all the participants gave their written informed consent before any study procedures began.

#### Consent for publication

The author of this study declares that this manuscript does not infringe copyright, moral rights, or other intellectual property rights of any other person. The author of this study testifies that this paper is original and has not incurred any type of plagiarism. The author of this study declares that the article is original, has not already been published in a journal, and is not currently under consideration by another journal. The author of this study agrees to the terms of the Springer Open Copyright and License Agreement.

#### Competing interests

The author declares no conflicts of interest associated with this study.

Received: 19 March 2022 Accepted: 10 August 2022

Published online: 18 September 2022

### References

- Alam, K., Mamun, S.A.K.: Access to broadband Internet and labour force outcomes: a case study of the Western Downs Region. Queensland. *Telemat. Inform.* **34**, 73–84 (2017)
- Alozie, N.O., Akpan-Obong, P.: The digital gender divide: confronting obstacles to women's development in Africa. *Dev. Policy Rev.* **35**, 137–160 (2017)
- Arrow, K.J.: The theory of discrimination. In: Asehnenfelter, O., Rees, A. (eds.) *Discrimination in Labor Market*. Princeton University Press, Princeton (1973)
- Atasoy, H.: The effects of broadband Internet expansion on labor market outcomes. *ILR Rev.* **66**, 315–345 (2013)
- Beaudry, P., Lewis, E.: Do male–female wage differentials reflect differences in the return to skill? Cross-city evidence from 1980–2000. *Am. Econ. J. Appl. Econ.* **6**, 178–194 (2014)
- Beblo, M., Gorges, L.: On the nature of nurture: The malleability of gender differences in work preferences. *J. Econ. Behav. Organ.* **151**, 19–41 (2018)
- Becker, G.S.: *The economics of discrimination*. University of Chicago Press, Chicago (1957)
- Becker, G.S.: Human capital: a theoretical and empirical analysis, with special reference to education. Columbia University Press, New York (1964)
- Bergmann, B.R.: Occupational segregation, wages and profits when employers discriminate by race and sex. *East. Econ. J.* **1**, 103–110 (1974)
- Biewen, M., Fitzenberger, B., Seckler, M.: Counterfactual quantile decompositions with selection correction taking into account Huber/Melly (2015): an application to the German gender wage gap. *Labour Econ.* **67**, 101927 (2020)
- Bimber, B.: Measuring the gender gap on the Internet. *Soc. Sci. Q.* **81**, 868–876 (2000)
- Blinder, A.: Wage discrimination: reduced form and structural estimation. *Hum. Resour. Manag. Res.* **8**, 436–455 (1973)
- Borghans, L., Ter Weel, B.T., Weinberg, B.A.: People skills and the labor-market outcomes of underrepresented groups. *ILR Rev.* **67**, 287–334 (2014)
- Broadband commission: Doubling digital opportunities: enhancing the inclusion of women and girls in the information society. UNESCO ITU (2013).
- Brown, R.S., Moon, M., Zoloth, B.S.: Incorporating occupational attainment in studies of male/female earnings differentials. *J. Hum. Resour.* **15**, 3–28 (1980)
- Cao, J., Jiang, T.: The impact of Internet usage on women's income: An analysis based on CFPS data. *J. Tianjin Univ. Fin. Econ.* **371**, 79–95 (2020). (曹景林, 姜甜: 互联网使用对女性收入的影响-基于CFPS数据的经验证据. 《天津财经大学学报》, **371(12)**, 79–95 (2020))
- Chen, X.: The evolution of female labour force participation in urban China. *Econ. Transit.* **27**, 267–299 (2019)
- Chen, L., Fan, H.: Does family elderly care reduce female employment? Two-Stage Residual Inclusion Approach. *Popul. Res.* **40**, 71–81 (2016). (陈璐, 范红丽: 家庭老年照料会降低女性劳动参与率吗?—基于两阶段残差介入法的实证分析. 《人口研究》, **40(3)**, 71–81 (2016))
- Chen, L., Fan, H., Zhao, N., Chu, L.: The impact of informal care on employment for women in China. *Econ. Res. J.* **51**, 176–189 (2016). (陈璐, 范红丽, 赵娜, 褚兰兰: 家庭照料对女性就业的影响. 《经济研究》, **51(3)**, 176–189 (2016))
- China Family Panel Studies (CFPS): <https://opendata.pku.edu.cn/dataset/CFPS?language=en>. Accessed 20 May 2020.
- Connelly, R., Dong, X., Jacobsen, J., Zhao, Y.: The care economy in post-reform China: feminist research on unpaid and paid work and well-being. *Fem. Econ.* **24**, 1–30 (2018)
- Cook, S., Dong, X.Y.: Harsh choices: Chinese women's paid work and unpaid care responsibilities under economic reform. *Dev. Change* **42**, 947–965 (2011)
- Deyyiling, L.J.: Broadband in the labor market: the impact of residential high-speed internet on married women's labor force participation. *Ind. Lab. Relat. Rev.* **70**, 451–482 (2017)
- DiMaggio, P., Hargittai, E., Neuman, W.R., Robinson, J.P.: Social implications of the internet. *Annu. Rev. Sociol.* **27**, 307–336 (2001)
- Eastin, M.S., Cicchirillo, V., Mabry, A.: Extending the digital divide conversation: examining the knowledge gap through media expectancies. *J. Broadcast. Electron. Media* **59**, 416–437 (2015)
- Fatehkia, M., Kashyap, R., Weber, I.: Using Facebook ad data to track the global digital gender gap. *World Dev.* **107**, 189–209 (2018)
- Ge, Y.: Sectoral choice and the gender wage gap: 1988–2001. *China Econ. Q.* **2**, 607–628 (2007). (葛玉好: 部门选择对工资性别差距的影响: 1988–2001年. 《经济季刊》, **2**, 607–628 (2007))
- Ge, S., Zhou, Y.: Robots, computers, and the gender wage gap. *J. Econ. Behav. Organ.* **178**, 194–222 (2020)
- Gustafsson, B., Li, S.: Economic transformation and the gender earnings gap in urban China. *J. Popul. Econ.* **13**, 305–329 (2000)
- Hafkin, N.J., Huyer, S.: Women and gender in ICT statistics and indicators for development. *Inf. Technol. Dev.* **4**, 25–41 (2007)
- Haight, M., Quan-Haase, A., Corbett, B.A.: Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet,



- level of online activity, and social networking site usage. *Inf. Commun. Soc.* **17**, 503–519 (2014)
- Hirsch, B.: *Spatial Monopsony and the Gender Pay Gap*. Springer, New York (2010)
- Iwasaki, I., Ma, X.: Gender wage gap in China: a large meta-analysis. *J. Lab. Mark. Res.* **54**, 17 (2020)
- Kidd, M.P.: Sex discrimination and occupational segregation in the Australian labour market. *Econ. Rec.* **69**, 44–55 (1993)
- Kidd, M.P., Shannon, M.: Does the level of occupational aggregation affect estimates of the gender wage gap. *Ind. Lab. Relat. Rev.* **49**, 317–329 (1996)
- Krueger, A.B.: How computers have changed the wage structure: evidence from microdata, 1984–1989. *Q. J. Econ.* **108**, 33–60 (1993)
- Li, S., Ma, X.: Occupational segregation and gender wage differentials in urban China. *Chin. J. Popul. Sci.* **5**, 2–13 (2006). (李实, 马欣欣: 中国城镇职工的性别工资差异与职业分割的经验分析. 《中国人口科学》, **5**, 2–13 (2006))
- Li, S., Yang, X.: The determinants of gender wage gaps in migrants. *Comp. Econ. Soc. Syst.* **5**, 82–89 (2010). (李实, 杨修娜: 农民工工资的性别差异及其影响因素. 《经济社会体制比较》, **5**, 82–89 (2010))
- Li, S., Song, J., Liu, X.: Evolution of the change of the gender wage gap among China's urban employees. *Manag. World* **3**, 53–65 (2014). (李实, 宋锦, 刘小川: 中国城镇职工性别工资差距的演变. 《管理世界》, **3**, 53–65 (2014))
- Liu, P., Meng, X., Zhang, J.: Sectoral gender wage differentials and discrimination in the transitional Chinese economy. *J. Popul. Econ.* **13**, 331–352 (2000)
- Liu, S., Zhang, X., Yang, Z.: Impact of Internet use on rural residents' income. *Quant. Econ. Tech. Econ. Res.* **38**, 103–119 (2021). (刘生龙, 张晓明, 杨竺松: 互联网使用对农村居民收入的影响. 《数量经济技术经济研究》, **38**(4), 103–119 (2021))
- Ma, X.: Ownership sector segmentation and the gender wage gap in urban China during the 2000s. *Post-Communist Econ.* **30**, 775–804 (2018)
- Ma, X.: *Female employment and gender gap in China*. Springer, Singapore (2021a)
- Ma, X.: Grandchildren care and labor supply of middle-aged grandmothers: evidence from China. *J. Fam. Econ. Issues* (2021b). <https://doi.org/10.1007/s10834-021-09794-0>. (Early published online)
- Ma, C., Gu, H., Li, J.: Change of gender wage gaps in Chinese labor market: evidence from quantile decomposition analysis of panel data. *World Econ. Pap.* **2**, 96–108 (2013). (马超, 顾海, 李佳佳: 中国劳动力市场上的性别工资差异变化研究: 来自面板分位数回归分解方法的证据. 《世界经济文汇》, **2**, 96–108 (2013))
- Mao, Y., Zeng, X.: Does Internet use promote female employment? An empirical study based on the data from the CGSS. *Econ. Dyn.* **6**, 21–131 (2017). (毛宇飞, 曾湘泉: 互联网使用是否促进了女性就业-基于CGSS数据的经验分析. 《经济学动态》, **6**, 21–131 (2017))
- Masso, J., Meriküll, J., Vahter, P.: The role of firms in the gender wage gap. *J. Comp. Econ.* **50**, 454–473 (2022)
- Meng, X.: Gender occupational segregation and its impact on the gender wage differential among rural-urban migrants: a Chinese case study. *Appl. Econ.* **30**, 741–752 (1998)
- Meng, X., Zhang, J.: Two-tier labor market in urban China: occupational segregation and wage differentials between urban residents and rural migrants in Shanghai. *J. Comp. Econ.* **29**, 485–504 (2001)
- Miller, P.W.: The wage effect of the occupational segregation of women in Britain. *Econ. J.* **97**, 885–896 (1987)
- Miller, P., Mulvey, C.: Computer skills and wages. *Aust. Econ. Pap.* **36**, 106–113 (1997)
- Mincer, J.: *Schooling, experience and earning*. Columbia University Press, New York (1974)
- National Bureau of Statistics NBS: *China statistical yearbook 1999*. China Statistics Press, Beijing (1999). (国家统计局: 中国统计年鉴1999. 中国统计出版社, 北京 (1999))
- National Bureau of Statistics NBS: *China statistical yearbook 2021*. China Statistics Press, Beijing (2021). (国家统计局: 中国统计年鉴2021. 中国统计出版社, 北京 (2021))
- Oaxaca, R.: Male–female wage differentials in urban labor markets. *Int. Econ. Rev.* **14**, 693–709 (1973)
- Ono, H., Zavodny, M.: Digital inequality: a five country comparison using microdata. *Soc. Sci. Res.* **36**, 1135–1155 (2007)
- Organization for Economic Co-operation and Development (OECD): *Bridging the digital gender divide: include, upskill, innovate*. OECD, Paris (2018)
- Pabilonia, S.W., Zoghi, C.: Returning to the returns to computer use. *Am. Econ. Rev.* **95**, 314–317 (2005)
- Phelps, E.S.: The statistical theory of racism and sexism. *Am. Econ. Rev.* **62**, 659–661 (1972)
- Qi, Y., Liu, C.: Does the use of internet reduce the difference in gender pay in the context of digital economy? *Econ. Theory and Econ. Admi.* **9**, 70–87 (2020). (戚聿东, 刘翠花: 数字经济背景下互联网使用是否缩小了性别工资差异. 《经济理论与经济管理》, **9**, 70–87 (2020))
- Rice, R.E., Katz, J.E.: Comparing internet and mobile phone usage: Digital divides of usage, adoption, and dropouts. *Telecommun. Policy* **27**, 597–623 (2003)
- Riggins, F.J., Dewan, S.: The digital divide: Current and future research directions. *J. Assoc. Inf. Syst.* **6**, 298–337 (2005)
- Scheerder, A., van Deursen, A., van Dijk, J.: Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. *Telemat. Inform.* **34**, 1607–1624 (2017)
- Vick, B.: Measuring links between labor monopsony and the gender pay gap in Brazil. *IZA J. Dev. Migr.* **7**, 10 (2017)
- Wang, M.: Work chance and wage differentials in the urban labor market: Work and wage of migrants. *Chin. Soc. Sci.* **5**, 36–46 (2005). (王美艳: 城市劳动力市场上的就业机会与工资差异-外来劳动力就业与报酬研究. 《中国社会科学》, **5**, 36–46 (2005))
- Wang, M., Cai, F.: Gender earnings differential in urban China. *Rev. Dev. Econ.* **12**, 442–454 (2008)
- World Bank, <http://data.worldbank.org/indicator>. Accessed 30 Jun 2022.
- Wu, E.: Analysis of Internet usage and gender wage difference. *Coop. Econ. Technol.* **16**, 138–140 (2021). (吴恩喆: 互联网使用与性别工资差异分析. 《合作经济与科技》, **6**, 138–140 (2021))
- Zhao, Y., Li, X.: The impact of internet use on individual health: Based on the evidence analysis of the Chinese Family Panel. *Stud. Sociol.* **5**, 139–144 (2020). (赵颖智, 李星颖: 互联网使用对个体健康的影响-基于中国家庭追踪调查数据的实证分析. 《江汉论坛》, **5**, 139–144 (2020))
- Zhuang, J., Liu, A., Sun, C.: Does Internet reproduce gender inequality? A study on Internet use wage premium. *Social* **36**, 88–106 (2016). (庄家焜, 刘爱, 玉孙超: 网络空间性别不平等的再生产: 互联网工资溢价效应的性别差异以第三期妇女地位调查为例. 《社会》, **36**(5), 88–106 (2016))

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)