Academic Journal of Business & Management
ISSN 2616-5902 Vol. 5, Issue 21: 96-102, DOI: 10.25236/AJBM.2023.052114

A Study of the Gender Wage Gap Based on Big Data Regression Analysis of the Urban Employed Population and Wages: A Technological Progress Perspective

Boxin Hao

Chengdu Jinjiang Jiaxiang Foreign Languags High School, Chengdu, China

Abstract: *The concept of wage propensity coefficient is introduced, a mathematical model is established to accurately measure the gender wage propensity index, and a regression model is further established to analyze the impact of technological progress on gender wage gap by using the big data of China's urban employed population and wages. The results found that technological progress and the proportion of urban female employed persons to all urban employed persons are important factors affecting the gender wage gap in urban areas, and the female wage propensity index increases with the increase of technological progress, and decreases with the increase of the proportion of urban female employed persons to all urban employed persons. Among them, technological progress has a greater impact, for every one-point increase in technological progress, the wage propensity index of women increases by 5.15%. Technological progress has helped to increase the wage propensity index of employed women, thereby contributing to the reduction of the gender wage gap. It is suggested that, firstly, the influence of women's own characteristics such as education and years of working experience should be improved; secondly, women's employment should be promoted, and the proportion of urban women employed in the total number of urban employed persons should be expanded; and thirdly, the simultaneous improvement of women's employment and wages should be ensured, and technological advancement should be strengthened even more.*

Keywords: *Population; Wage; Gender wage gap; Technological progress; Big data*

1. Introduction

At present, a certain gender wage gap still exists in China's labor market. Li Shi et al. (2014) analyzed and found that the gender wage gap in China rose significantly from 2002 to 2007 [1]. The results of the data analysis by Ma Chao et al. (2013) showed that men's wage level is higher than women's [2]. Several scholars, including Zhu Jinxia et al. (2014), Wei Wei and Yang Heqing (2015), found through empirical analysis that there is a general trend of gradual increase in the gender wage gap between employed men and women [3-4].

Existing research on the gender wage gap usually examines special groups such as urban, regional, rural-urban, or rural migrant workers as a whole labor market, and uses the difference decomposition method to attribute the gender wage gap to two major factors: a component that can be explained by individual characteristics such as education and years of work (experience), and a component that cannot be explained: labor market discrimination. As for the impact of technological progress on the gender wage gap, which is an important driving force to promote economic development and form the income distribution gap, there are relatively few studies in China. Xing Chunbing, Jia Shuyan and Li Shi (2014) conducted a study about it earlier, and due to the difficulty of measuring technological progress and the specific skills of laborers, the authors chose the returns to education as a proxy variable to indirectly study the impact of scientific and technological progress on the gender wage gap [5]. Hao Cuihong and Li Jianmin (2018) et al. measured technological progress with the provincial technological progress variable to test the impact of technological progress on gender wage gap [6]. Suqin Ge and Yu Zhou (2020), Wang Beibei (2022), and Huacong Liu et al. (2022) used the use of robots and computers as a variable to study the impact of technological progress on gender wage gap [7-9]. While Qi Yudong and Liu Cuihua (2020), Feng Xiliang et al. (2021) and others in recent years have studied the impact of internet use on gender wage gap [10-11]. Based on this, this article empirically analyzes the impact of total factor productivity as a measure of technological progress and its effect on

ISSN 2616-5902 Vol. 5, Issue 21: 96-102, DOI: 10.25236/AJBM.2023.052114

gender wage gap as a variable.

2. Gender Wages Difference

At this stage, many scholars in China have studied the gender wage gap problem in China's labor market from theoretical research and empirical analysis, in which empirical research mostly uses the difference decomposition method, due to the difficulty of directly obtaining the average wages of women and men in various industries or occupations, while fuzzy data analysis and measurement based on the assumptions of the model is used to obtain approximate data, making the results deviate from the actual, based on this, this paper utilizes the wage propensity coefficient and its estimation results to calculate the gender wage propensity index, which reflects the relationship between the gender wage gap and technological progress (total factor productivity).

Xu Linqing (2004) proposed a wage propensity index based on the distribution of female occupations [12], which can be used to make a judgment on the gender wage-earning gap:

$$P_F = \frac{\sum F_i \overline{W}_{fi}}{F\overline{W}} \tag{1}$$

Where, P_F denotes the wage propensity index of female occupational distribution, F denotes the total number of female employed persons in all industries, and $\overline{W}$ denotes the total average wage of all industries, F_i denotes the number of female employed persons in industry i, and $\overline{W}_{fi}$ denotes the average wage of female employed persons in industry i. If the female labor force is not discriminated against, then its occupational distribution is completely random, then $P_F = 1$, if $P_F < 1$, then it indicates that there is a phenomenon that females are employed in industries with lower wages due to discrimination. The smaller the P_F value, the more females are employed in lower wage industries.

In practice, it is difficult to directly obtain the average wages of female employed persons in various industries $\overline{W}_i$. The available data are mainly obtained through the following methods: first, questionnaires, and second, modeling and econometric estimation, the second of which mostly uses the Blinder-Oaxaca wage differential decomposition method:

$$\ln \overline{W}_m - \ln \overline{W}_f = \beta_0^m - \beta_0^f + \sum \beta_i^m \left(\bar{X}_i^m - \bar{X}_i^f\right) + \sum \bar{X}_i^f \left(\beta_i^m - \beta_i^f\right) \tag{2}$$

Existing calculation methods can approximate the gender wage propensity index, but due to the questionnaire is not only a large amount of work, but also often a large error due to the inconsistency of the standard of a certain concept (such as the wage, not sure whether it refers to the basic wage or all the income from the labor) or the respondent's subjective consciousness problems (such as deliberately overstating the amount of wages or understatement of the amount of wages), etc., and the model-based estimation is both complex and generates errors.

In order to eliminate errors in the process of measuring the gender wage propensity index and to make the data calculation more objective and simpler, Liu Renbao and Liu Guanjun (2017) proposed the concept of gender wage propensity coefficient, that is to say, when calculating the gender wage propensity index, if the average wage of female (or male) employed persons is directly substituted by the average wage of the industry, a difference exists, which is called the gender wage propensity coefficient [13].

For example to calculate the female wage propensity index, let the average wage of industry $\overline{W}_i$, the female wage propensity index replaced by the average wage of the industry is $P_F{}'$, then:

$$P_F{}' = \frac{\sum F_i \overline{W}_i}{F\overline{W}} \tag{3}$$

The gender wage propensity coefficient d can be expressed as:

$$d = P_F' - 1 \tag{4}$$

To summarize, the female wage propensity index P_F can also be expressed as:

$$P_F = 1 - d \tag{5}$$

Based on the data from the China Labor Statistics Yearbook for each year from 2004-2019, this paper uses this method to calculate the urban female wage propensity index based on occupational distribution for each year from 2003-2018 (as shown in Table 1), and the results indicate that the phenomenon of gender wage gap currently exists in the China's labor market.

ISSN 2616-5902 Vol. 5, Issue 21: 96-102, DOI: 10.25236/AJBM.2023.052114

Table 1: Wage Propensity Index for Chinese Urban Women Based on Occupational Distribution (2003-2018)

particular year	Number of employed persons at the end of the year (10,000 persons)	Number of women in industry (10,000)	Number of men in industry (10,000)	Percentage of women in industry	Percentage of men in industry	Average wage (dollars)	Female Wage Propensity Index
2003	10969.7	4156.1	6813.6	37.89%	62.11%	13969	0.9979
2004	11098.9	4227.3	6871.6	38.09%	61.91%	15920	0.9981
2005	11404	4324.6	7079.4	37.92%	62.08%	18200	0.9991
2006	11713.2	4445.7	7267.5	37.95%	62.05%	20856	0.9980
2007	12024.4	4540.3	7484.1	37.76%	62.24%	24721	0.9928
2008	12192.5	4579.6	7612.9	37.56%	62.44%	28898	0.9899
2009	12573	4678.5	7894.5	37.21%	62.79%	32244	0.9872
2010	13051.5	4861.5	8190	37.25%	62.75%	36539	0.9846
2011	14413.3	5227.7	9185.6	36.27%	63.73%	41799	0.9778
2012	15236.4	5458.9	9777.5	35.83%	64.17%	46769	0.9559
2013	18108.4	6338.3	11770.1	35.00%	65.00%	51483	0.9750
2014	18277.8	6546.2	311.0	35.82%	64.18%	56360	0.9735
2015	18062.5	6527.0	435.9	36.14%	63.86%	62029	0.9646
2016	17888.1	6517.6	297.0	36.44%	63.56%	67569	0.9597
2017	17643.8	6545.3	148.5	37.10%	62.90%	74318	0.9571
2018	17258.2	6427.6	41.6	37.24%	62.76%	82413	0.9542

Data source: China Labor Statistics Yearbook (2004-2019).

3. Measurement analysis

3.1. Model

Drawing on Solow's exogenous economic growth model, combined with China's socio-economic reality as well as data sources, here we take the female wage propensity index as the object of examination (independent variable), adopt the rate of technological progress (total factor productivity, TFP), the proportion of female employed persons, and fixed capital stock as the dependent variables, and construct the following econometric model to test the impact of technological progress on the gender wage gap:

$$S_t = \beta_0 + \beta_1 T_t + \beta_2 F_t + \beta_3 K_t + \mu \tag{6}$$

Among them:

S denotes the female wage propensity index, using the results calculated in Table 1 (2003-2018).

T denotes technological progress, total factor productivity (TFP) growth rate as a proxy variable, reference to Battese and Coelli's model, using SFA method for calculation, data sources China's statistical yearbooks of all years, CNRDS, CSMAR database, the compilation of domestic historical data, historical data, and foreign databases of China's relevant data, the results are shown in Table 2.

Table 2: Total Factor Productivity (TFP) Growth Rate (2003-2018)

Particular year	2003	2004	2005	2006	2007	2008	2009	2010
TFP growth rate	0.2276	0.2973	0.3411	0.3779	0.4177	0.4337	0.4175	0.4165
particular year	2011	2012	2013	2014	2015	2016	2017	2018
TFP growth rate	0.4014	0.3741	0.3348	0.3012	0.2757	0.2539	0.2507	0.2537

F represents the ratio of urban female employed persons to the total number of all urban employed persons, and is calculated using the data in Table 3.

K denotes fixed capital stock. Investment in fixed assets by major industry and deflated by price index, based on data from the China Statistical Yearbook (2004-2019).

ISSN 2616-5902 Vol. 5, Issue 21: 96-102, DOI: 10.25236/AJBM.2023.052114

Table 3: Summary of Data

Year	S	T	F	K
2003	0.9979	0.2276	0.3789	381472.33
2004	0.9981	0.2973	0.3809	408528.06
2005	0.9991	0.3411	0.3792	447732.42
2006	0.9980	0.3779	0.3795	497560.07
2007	0.9928	0.4177	0.3776	558527.60
2008	0.9899	0.4337	0.3756	627499.79
2009	0.9872	0.4175	0.3721	718780.25
2010	0.9846	0.4165	0.3725	826639.50
2011	0.9778	0.4014	0.3627	946925.42
2012	0.9559	0.3741	0.3583	1079764.38
2013	0.9750	0.3348	0.3500	1223141.01
2014	0.9735	0.3012	0.3582	1370035.91
2015	0.9646	0.2757	0.3614	1516757.42
2016	0.9597	0.2539	0.3644	1667833.28
2017	0.9571	0.2507	0.3710	1803806.07
2018	0.9542	0.2537	0.3724	1937066.14

Source: China Statistical Yearbook (2004-2019).

3.2. Tests of significance

Stata 15.0 software was used to enter and calculate the relevant data and to do significance test of correlation, ADF test and regression analysis.

Table 4 shows that the correlation between the female occupational wage propensity index S and the ratio of urban female employment to total urban employment F (positive correlation) and the logarithm of gross fixed capital investment lnK (negative correlation) is significant at the 10 percent confidence level.

Table 4: Significance Test Results

Variables	(1)	(2)	(3)	(4)
(1) S	1.0000			
(2) T	0.4679	1.0000		
	0.3429			
(3) F	0.5978*	0.0806	1.0000	
	0.0836	0.9998		
(4) lnK	-0.9813*	-0.3605	-0.6905*	1.0000
	0.0000	0.6734	0.0183	

3.3. ADF test

In order to further test whether the data are smooth or not, the unit root test was next performed on the variables in question, and the results of the test are shown in Table 5:

Table 5: ADF Test Results

		Test statistic	1% Critical Value	5% Critical Value	10% Critical Value	MacKinnon Approximate p-value for Z(t)	Number of obs
S*	Z(t)	-3.168	-4.380	-3.600	-3.240	0.0909	15
T**	Z(t)	-3.659	-4.380	-3.600	-3.240	0.0252	15
F	Z(t)	-0.891	-4.380	-3.600	-3.240	0.9571	15
d.F*	Z(t)	-2.859	-3.750	-3.000	-2.630	0.0504	14
lnK	Z(t)	-1.518	-4.380	-3.600	-3.240	0.8227	15
d.lnK	Z(t)	-1.541	-3.750	-3.000	-2.630	0.5132	14
d2.lnK	Z(t)	-2.243	-3.750	-3.000	-2.630	0.1911	13
d3.lnK***	Z(t)	-6.569	-3.750	-3.000	-2.630	0.0000	12

Note: *** indicates that the original hypothesis is rejected at the 1%, 5%, and 10% confidence levels; ** indicates that the original hypothesis is rejected at the 5% and 10% confidence levels; * indicates that the original hypothesis is rejected at the 10% confidence level.

ISSN 2616-5902 Vol. 5, Issue 21: 96-102, DOI: 10.25236/AJBM.2023.052114

Through the ADF test, it was found that the relevant variables F, lnK could not reject the original hypothesis at 1%, 5% and 10% confidence level and accepted the original hypothesis that there is a unit root, and the relevant variables have a unit root, therefore, it is necessary to test the relevant variables further by doing first-order differencing or second-order differencing.

After the first order difference test, the hypothesis of having a unit root was rejected with a p-value of 0.0504 for the F-variable, which can also be obtained by observing the value of -2.859 for the Z(t), which rejects the original hypothesis at the 10% confidence level (-2.630), so there is no unit root in the first order difference data for the F-variable. However, the P-value of the first-order difference of the lnK variable is 0.5132, which accepts the assumption that there is a unit root, which can also be obtained by observing the value of -1.541 for the first-order difference Z(t) of the lnK variable, which fails to reject the original assumption at the 1% confidence level (-3.750), the 5% confidence level (-3.000), and the 10% confidence level (-2.630), and therefore the first-order difference of the lnK variable still has a unit root in the data and it needs to be second-order differenced before proceeding with the test. lnK variable with a p-value of 0.1911 accepts the hypothesis that there is a unit root, which can also be obtained by observing that the value of Z(t) (-2.243) rejects the original hypothesis at the 1% confidence level (-3.750), 5% confidence level (-3.000), and 10% confidence level (-2.630), thus there is a unit root in the second-order differenced data of lnK variable, which needs to be third-order differenced before proceeding with the test.

After the third order difference test, the p-value of the lnK variable is 0.0000, which rejects the hypothesis that there is a unit root, which can also be obtained by observing the value of Z(t) (-6.569), which rejects the original hypothesis at the 1% confidence level (-3.750), 5% confidence level (-3.000), and 10% confidence level (-2.630), so there is no unit root in the third order difference data of the lnK variable .

3.4. Regression analysis

After ADF test and difference treatment, the initial model (6) becomes:

$$S_t = \beta_0 + \beta_1 T_t + \beta_2 d.F_t + \beta_3 \,\mathrm{d3}.\ln K_t + \mu \tag{7}$$

Regression analysis was performed using Stata 15.0 software and the results are shown in Table 6:

Table 6: Results of Regression Analysis

Source	SS	df	MS	Number of obs	=	13
Model	0.0020	3	0.0007	F(3, 9)	=	15.49
Residual	0.0004	9	0.0000	Prob > F	=	0.0007
Total	0.0024	12	0.0002	R-squared	=	0.8378
				Adj R-squared	=	0.7837
				Root MSE	=	0.0066

S	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
T	0.2032	0.0354	5.73	0.000	0.1230	0.2833
D. F	0.5906	0.4692	1.26	0.240	-0.4708	1.6519
D3. lnK	-0.0426	0.2219	-0.19	0.852	-0.5448	0.4594
_cons	0.9058	0.0121	74.79	0.000	0.8784	0.9332

As can be seen in Table 6, the coefficients on variable T are highly significant at the 99% confidence interval, and the coefficients on variables D.F, D3.lnK are highly insignificant.

The regression equation for the model is:

$$S_t = 0.2032T_t + 0.5906d.F_t - 0.0426 \,\mathrm{d3}.\ln K_t + 0.9058 \tag{8}$$

From the above analysis it can be seen that the urban female wage propensity index (S) based on occupational distribution is positively and significantly affected by technological progress (T).

In model (8), since the coefficients of the variables d.F and d3.lnK are insignificant, the regression analysis is re-run after excluding the variable and the results are obtained as shown in Table 7.

Table 7: New Results of Regression Analysis

Source	SS	df	MS				
Model	0.0008	1	0.0008	Number of obs	=	16	
Residual	0.0028	14	0.0002	F(1,14)	=	3.92	
				Prob > F	=	0.0676	
Total	0.0036	15	0.0002	R-squared	=	0.2189	
				Adj R-squared	=	0.1631	
				Root MSE	=	0.0142	
S	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]	
T	0.1022	0.0516	1.98	0.068	-0.0085	0.2129	
_cons	0.9459	0.0177	53.45	0.000	0.9080	0.9839	

From Table 7, it can be seen that the model is still very significant at the 90% confidence interval and has a high explanatory power, therefore, the results of this regression can be used as the final regression model:

$$S_t = 0.1022T_t + 0.9459 \tag{9}$$

4. Results

From the analysis of the model (9), it can be seen that the urban women's wage propensity index (S) based on occupational distribution is positively and significantly affected by technological progress (T), and the women's wage propensity index increases with technological progress, and for every one point increase in the rate of technological progress measured by total factor productivity (TFP), the urban women's wage propensity index based on occupational distribution increases by 10.22 percent.

5. Conclusion

Based on the previous methods of calculating the wage propensity index based on women's occupational distribution, this paper utilizes the urban women's wage propensity index based on occupational distribution calculated by the gender wage propensity coefficient for each year of 2003-2018, and constructs an econometric model to test the impact of technological progress, measured by total factor productivity (TFP) on the gender wage gap. From the urban female wage propensity index based on occupational distribution, the phenomenon of gender wage gap still exists in China's labor market at present, and the average wage and salary of urban female employed persons is lower than that of male employed persons in general. In terms of the sources of income distribution and the factors that influence it, the role is played by technological progress, which has helped to raise the wage propensity index of female employees; for every 1-point increase in the rate of technological progress, as measured by total factor productivity (TFP), the wage propensity index of urban females based on the distribution of occupations has risen by 10.22%.

Therefore, in order to raise the wage level of urban female employees and reduce the gender wage gap, it is necessary, on the one hand, to ameliorate the effects of women's own characteristics, such as education and years of work experience, on the other hand, to eliminate discrimination in the labor market by improving the relevant laws and regulations and strengthening governmental and social control, so as to truly realize equal pay for men and women for work of equal value. As women's wages rise, it will also promote women's employment. To ensure that women's employment and wages rise in tandem, technological progress should be stepped up. To this end, the state and relevant government departments should formulate and improve laws, regulations and policies to support and promote technological progress, such as guiding innovative elements, tax incentives, project establishment, financial support, and rewards for technological progress, etc., so as to encourage and guide enterprises and scientific research institutes, as well as technological researchers, to actively carry out scientific research and technological innovation, and to realize the transformation of technological achievements, and to promote technological progress. Enterprises should strengthen the awareness of technological innovation, change the way of research and development and introduction of technological progress, and strengthen the independent innovation of enterprises; technological personnel of colleges and universities and scientific research institutes should carry out scientific research and technological innovation in-depth, actively transform technological achievements, and strive to improve the level of technological research and development; and laborers should receive education and training on various fronts and at various levels to improve their own qualities, cultivate the spirit of innovation, and improve the ability of innovation.

ISSN 2616-5902 Vol. 5, Issue 21: 96-102, DOI: 10.25236/AJBM.2023.052114

References

[1] LI Shi, SONG Jin, LIU Xiaochuan. The Evolution of the Gender Wage Gap of the Staff of China's Cities and Towns [J]. Management World, 2014(3):53-65,187.
[2] MA Chao, GU Hai, LI Jiajia. A Study of Changes in Gender Wage Differentials in the Chinese Labor Market--Evidence from Panel Quantile Regression Decomposition Methods [J]. World Economic Papers, 2013(2):96-108.
[3] ZHU Jinxia, LU Kangyin, ZHANG Li. Gender Income Gap and the Changes of Gender Discrimination [J]. Research on Technical Economy and Management, 2014(6):83-89.
[4] Wei Wei, Yang Heqing. An Empirical Study of Labor Market Discrimination and the Problem in China [J]. China Labor, 2015(24):69-74.
[5] XING Chunbing, JIA Shuyan, LI Shi. Technological Change, Return to Education, and Gender Wage Differentials in Urban China [J]. Labor Economics Research, 2014 (3):42-62.
[6] Hao Cuihong, Li Jianmin. Technological Progress, R&D Input and Gender Wage Gap - Based on the Empirical Analysis of CGSS Data [J]. Journal of Guizhou University of Finance and Economics, 2018 (05).
[7] Suqin Ge, Yu Zhou. Robots, Computers, and the Gender Wage Gap [J]. Journal of Economic Behavior and Organization, 178(2020):194-222.
[8] Wang Beibei. The Impact of Robot Imports on the Gender Wage Gap in Firms: An Empirical Analysis Based on the Transportation Equipment Manufacturing Industry [D]. Supervisor: Yi Miao. East China Normal University, 2022.
[9] Huacong Liu , Frank Fernandez , Gregor Dutz . Educational Attainment, Use of Numeracy at Work, and Gender Wage Gaps: Evidence from 12 Middle-income Countries [J]. International Journal of Educational Development, 2022 (92):1-9.
[10] Qi Yudong, Liu Cuihua. Does the Use of Internet Reduce the Difference in Gender Pay in the Context of Digital Economy – An Empirical Analysis Based on the Comprehensive Social Survey in China [J]. Economic Theory and Economic Management, 2020 (9):70-87.
[11] FENG Xiliang, GAO Panpan, LUO Rongbo. The Influence of Internet Use on the Gender Wage Income Gap of Migrant Workers [J]. Population and Economics, 2021 (5):111-124.
[12] Xu, Linqing. The Trade-Wage Preference in Women's Employment and Sexual Discrimination [J]. Collection of Women's Studies, 2004(2):34-38.
[13] LIU Renbao, LIU Guanjun. The Influence of Scientific and Technological Progress on the Gender Wage Differentials of Urban Workers [J]. Research on Economics and Management, 2017, 38(11):50-57.