A Comprehensive Analysis of the Factors Influencing Regional Wage Level Differences in China

Yuming Li*

Linnan College, Sun Yat-sen University, Guangzhou, China, 510000

* Corresponding Author Email: liym235@mail2.sysu.edu.cn

Abstract. Based on the data of China Urban Statistical Yearbook (2019-2021), this paper analyzes the effects of economic development, science and technology innovation, higher education, openness, and the role of government on regional wage differences in China based on the theoretical basis of urban economics. The paper uses instrumental variables estimation, Zou test, Gini coefficient decomposition, factor analysis, Markov chain and other methods to draw a preliminary conclusion that all factors except openness have a significant positive effect on wage level, while openness does not have a significant effect. This paper gives suggestions based on the mechanism of influence, pointing out the important role of local government support in raising the wage level.

Keywords: Wage level; Gini coefficient decomposition; Markov chain analysis.

1. Introduction

Since the reform and opening up in 1978, China's economy has developed at a rapid pace, and the per capita disposable income and wage levels have been increasing, but there are significant regional differences, with residents' income growing fast in the eastern region, slightly slower in the central region, and slowest in the western region, leading to a gradual widening of the disparity in wage levels between regions. At the same time, China's economic development policies for regions are also undergoing a shift from an unbalanced and coordinated development strategy to a coordinated regional economic development strategy. Since the early stage of reform and opening up, the policy direction has gone through the "two big picture" concept, the "special economic zone" experiment in the east, the "western development" and the "rise of the central region" strategy. The "Rise of the Central Region" strategy. As socialism enters a new era, China places more emphasis on the strategy of coordinated regional economic development.

The report of the 19th Party Congress points out that the main contradiction in China has changed to the contradiction between the people's growing need for a better life and unbalanced and insufficient development. An important manifestation of unbalanced and insufficient development is the imbalance of regional development and the imbalance of income distribution. Gini coefficient is a common indicator used to measure the income disparity of residents in a country or region. For a long time, the Gini coefficient of per capita disposable income in China is higher than the warning line of 0.4. While the Gini coefficient of per capita disposable income of residents remains high, the Gini coefficient of per capita wages in China also shows a rising trend, and the unbalanced development of wages has an expanding trend. It is worth paying attention to the fact that although regional wage levels and regional per capita disposable income have a large correlation, their specific changes and trends are still significantly different. There have been a lot of studies on the unbalanced development of per capita disposable income among regions. However, regional differences in wages, an important basis for studying income distribution, have rarely been studied. For achieving high-quality economic development, the balanced development of workers' wage income in different regions is an aspect that cannot be ignored. And to achieve the balanced development of regional wages, the factors affecting the average wages of workers in different regions need to be explored.

This paper uses regression analysis, Gini coefficient decomposition, and factor analysis to analyze the different factors that influence regional differences in workers' wage levels based on national data for each prefecture-level city from 2018 to 2020. This paper focuses on answering the following questions: (1) What are the main factors that affect regional wage level differences? (2) Among these factors, which are the factors that have a greater impact on regional wage level differences? (3) Are the effects of these factors on regional wage levels significant? (4) Does the new crown epidemic have an impact on the role of these factors?

1.1. Impact of economic development level on regional wage differences

The Kuznets inverted U-curve hypothesis suggests that at the stage of low income level, economic growth is accompanied by the widening of income distribution gap; however, when the income level reaches a certain level, economic growth helps to alleviate income distribution inequality (Kuznets, 1955) [1]. Therefore, this paper proposes the hypothesis that

The level of economic development has a significant effect on regional wages, and a large part of regional wage differences depend on the differences in the level of economic development.

1.2. Impact of higher education level on regional wage differences

Higher education provides an important basis for the development of a large number of modern-oriented, knowledgeable, competent and qualified personnel. A sufficient pool of human resources is necessary for the emergence of innovative entrepreneurs, good managers and high-quality workers. However, differences in educational resources and differences in wage levels tend to converge. Therefore, this paper makes the following assumptions.

There is a positive relationship between the level of higher education and regional wages, and the imbalance in education contributes to regional wage differences

1.3. The influence of the degree of openness to the outside world on regional wage differences

According to the theory of international trade, the freedom of trade can give full play to comparative advantages, bring about greater economies of scale, improve efficiency, acquire technology, etc., thus promoting regional economic development and regional income increase. The lagging regions can give full play to their location advantages and help the reasonable flow of economic factors in the region, thus reducing the regional differences. Therefore, this paper makes the hypothesis that

The degree of opening to the outside world has a significant effect on regional wages, and opening to the outside world is conducive to the reduction of regional wage differences

2. Data sources and processing and variable definitions

2.1. Data source and processing

The data source used for the study is the China Urban Statistical Yearbook, which is an informative annual publication sponsored by the Department of Urban Socio-Economic Survey of the National Bureau of Statistics that comprehensively reflects the economic and social development of Chinese cities. The main data set used in the study is extracted from the original material, and the required explanatory variables are selected and the null values are eliminated; the data set used in the second part of the empirical analysis is a new data set formed by selecting some of the data needed in the main data set and then reformatting and integrating them. Its content is consistent with the original data.

2.2. Notations

Important notations used in this paper are listed in Table 1.

Table 1. Notations

Variable Name	Meaning	Unit
wage(_xy)	(20xy) Average wage of employees on the job, xy is two digits, the same below	Yuan
gdp_per(_xy)	(20xy) Gross regional product per capita	Yuan
patent(_xy)	Number of patents granted (20xy)	individual
school(_xy)	(20xy) Number of colleges and universities	individual
expend(_xy)	(20xy) General public budget expenditure divided by regional GDP	-
trade(_xy)	(20xy) Total trade divided by regional GDP	-
exp_sci(_xy)	(20xy) Public expenditure on science and technology	million yuan
exp_edu(_xy)	(20xy) Public expenditure on education	million yuan
por23(_xy)	(20xy) Share of secondary and tertiary industries in GDP	-
deposit(_xy)	(20xy) Balance of deposits and loans of financial institutions at the end of (20xy)	million yuan

3. The effect of economic development level on regional wage differences

3.1. Model Setting

Based on the existing studies and considering the availability of data, this section uses the average wage of employees (wage_19) as the explanatory variable, which represents the wage level of a city. The first five explanatory variables represent a city's overall economic development level, science and technology innovation level, education level, openness level and the government's development in the region, respectively. The first five explanatory variables represent the overall economic development level of a city, the level of science and technology innovation, the level of education, the degree of openness and the role of government in the economic activities of the city [2].

$$\text{wage}_1 9=\beta_0+\beta_1\text{gdp}_\text{p}\text{er}_1 9+\beta_2\text{patent}_1 9+\beta_3\text{school}_1 9+\beta_4\text{trade}_1 9+\beta_5\text{expend}_1 9+\beta_6\ \text{location}+u \tag{1}$$

The first five explanatory variables of the model are likely to be endogenous due to possible measurement error, omission of important explanatory variables, or bi-directional causality. For example, an increase in wage levels in a city is likely to attract more highly qualified and innovative talent to the city, which in turn increases the city's STI capacity and leads to an increase in the number of patents granted, and there is a two-way causal relationship between STI levels and wage levels, with patent_19 being an endogenous explanatory variable. Therefore, the results of direct OLS estimation are likely to be biased and inconsistent [3].

In order to address the endogeneity issue, the instrumental variables approach is adopted by selecting the first five explanatory variables with a one-period lag as the instrumental variables. The explanatory variables and the explanatory variables of the model are used in 2019, and the six instrumental variables are gdp_per_18, patent_18, school_18, trade_18, and general public budget_18 in 2018. expend_18, and gdp_per_17 in 2017.

3.2. Model estimation

IV estimation with Stata yielded the following results in Table 2.

Table 2. IV Estimated results

	Coefficient	t
gdp_per_19 0.236***	(7.58)	patent_19
0.138**	(2.46)	school_19
294.3***	(4.88)	trade_19
0.300	(0.99)	expend_19
5.611***	(5.92)	location
7522.1***	(4.71)	_cons
42725.0***	(12.08)	N
225	R-sq	0.674

Underidentificationtest obtained a p-value of 0.0000, which allows rejecting the original hypothesis that instrumental variables are unrelated to endogenous variables at 1% significance level. From the first stage regression, the values of the corresponding R^2 and F statistics are large and it can be assumed that there is no weak instrumental variable problem.

To test the exogeneity of the instrumental variables, the over-identification constraint (Sargan) test is performed and the obtained p-value is 0.1274, which cannot reject the original hypothesis that the instrumental variables all satisfy the exogeneity at the 10% significance level [4].

To test the endogeneity of the explanatory variables, Wu_Hausman test was conducted and the obtained P-value was 0.0043[5], which rejects the hypothesis that H0, the explanatory variable, is exogenous at 5% level of significance, indicating that the explanatory variables are endogenous, OLS estimation is inconsistent and IV estimation is better than OLS estimation.

According to the estimation results, there is a significant positive effect of gross regional product per capita, level of science, technology and innovation, level of education and the role played by the government in economic activities on the wage level, while the degree of openness measured by the ratio of total imports and exports to gross regional product has no significant effect on the wage level. The dummy variable location is significant, implying that on average, holding other things constant, the average wage of employees in cities in eastern provinces is about $7,522.1 higher than the average wage of employees in cities in non-eastern provinces. The intercept term is not significant because at least two explanatory variables, patent_19 and gdp_per_19, take values much greater than 0. It is not surprising that the regression equation does not perform well when the explanatory variables take values close to 0.

4. Open Inquiry - Gini Coefficient Score

4.1. Theoretical Foundations of Gini Coefficient Decomposition

If the variable X_i becomes a significant influence on income Y, the within-group sample differences in Y can be decomposed into a combination of differences in individual variables X_i. The decomposition model under the analytical framework of Wan, Guanghua (2004) is [6].

$$G_y = CO_a + \sum_{k=1}^{r} \frac{E\left(Y^k\right)}{E(Y)} C_k \Bigg|^{\text{rank by } Y} + CO_\varepsilon \tag{2}$$

That is, the Gini coefficient of income Y is decomposed into three components, namely the intercept term, the regression variable, and the contribution of the error term (CO). For the case of individual data, Ping-Sheng Dai (2013) points out and proves based on Shorrocks' (1999) natural decomposition rule [7,8], by the income share method this paper has.

$$G = \sum_{k=1}^{r} \frac{S_k}{S} G_k + \sum_{k=1}^{r} \sum_{i=1}^{n} \frac{q_i y_i^k}{S} \left(\omega_i - \omega_i^k\right) \tag{3}$$

$$S_k = \sum_{i=1}^{n} s_i^k \qquad S = \sum_{k=1}^{r} S_k \tag{4}$$

This paper considers the essence of the Gini coefficient as a measure of within-group variation. In this paper, we mainly draw on the idea of Gini coefficient decomposition in our study, and use the term Gini coefficient even though the study is not the Gini coefficient in a narrow sense. The differences in the explanatory variables are decomposed into linear combinations of the differences in each explanatory variable in order to explore the composition of the sources of differences in per capita wages in the years under study.

4.2. Gini coefficient scores based on sample regression models

Using Stata's ineqdeco, glcurve and descogin commands, this paper first calculates the Gini coefficients of per capita wages in 2019 and 2020 as 0.1116 and 0.1076, respectively, and plots the Lorenz curves. Further decompose the Gini coefficients of per capita wages in 2019 and 2020 respectively by the explanatory variables after considering endogeneity, and the decomposition results are shown in Table 3.

Table 3. Results of variable difference decomposition using Gini coefficient idea

Share of variance in each explanatory variable in wage variance (%)		
	2019	2020
gdp_per	36.04	29.52
patent	9.70	-2.88
school	13.19	6.72
trade	1.88	5.98
expend	-8.02	13.09
location	9.08	7.9

From the decomposition results, the explanatory contribution of GDP per capita variance to per capita wage variance in 2019 is 36.04%, which is much higher than other explanatory variables; the number of patents granted in the previous year, the number of colleges and universities and geographical variance have 9.70%, 13.19% and 9.08% respectively, which have certain contribution; the contribution of total trade and general public budget expenditure is smaller. Compared with 2019, the contributions of GDP per capita, number of patents granted in the previous year, number of colleges and universities, and geographical differences to per capita wage differences in 2020 decrease to 29.52%, 6.72%, 7.96%, and 6.61%, respectively, which is seen to be a large decrease, presumably due to the weakening influence of the above factors in causing income differences among urban residents during the New Crown epidemic; while the contribution of general public budget differences The contribution of the general public budget variance to the per capita wage variance rose significantly to 13.09%, indicating that the per capita wage variance was sharply influenced by the fiscal budget expenditure.

5. Open Inquiry - Markov Chain Analysis

5.1. Theoretical Foundations of Markov Chains

Markov chains are mainly used to analyze the internal dynamics of variables and their evolutionary processes (Quah, 1996) [9]. Let {ξ_n, n=1,2,...} be a random sequence and the state space *E* be a finite or columnar set, for any positive integers m,n, if $i,j,i_k \in$E(k=1,L,n-1), we have.

$$P\left\{\xi_{n+m}=j \middle| \ \xi_n=i,\xi_{n-1}=i_{n-1},\cdots,\xi_1=i_1\right\}=P\left\{\xi_{n+m}=j \middle| \ \xi_n=i\right\} \tag{5}$$

If the conditional probability on the right-hand side of the equation is independent of n that is.

$$P\left\{\xi_{n+m}=j \middle| \ \xi_n=i\right\}=p_{ij}(m) \tag{6}$$

It is the probability of transferring the system from state i to state j after m time intervals (or m steps).

5.2. Markov chain score of per capita wages

Using Markov chain analysis, this paper attempts to explore the trends of inter-year relative per capita wage level changes in different regions. By dividing the regional per capita wage levels studied in this paper into N types, an N × N dimensional state transfer probability matrix P of labor market development levels can be obtained through Markov chains, as shown in the following equation.

$$P = p_{ij} = \begin{vmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots \\ \vdots & \vdots & & \vdots & \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix} \tag{7}$$

$$p_{ij} \geqslant 0, ij \in N \tag{8}$$

$$\sum_{j \in N} p_{ij} = 1, ij \in N \tag{9}$$

Drawing on Quah (1996), Pu Yingxia (2005), Ye Changhua et al. (2018) and others[9-10], urban per capita wage levels are divided into four intervals, with those whose average wage levels are in the overall average wage level ranking 25% and before being category I; those whose average wage levels are in the overall average wage level ranking 25%-50% being category II; those whose average wage levels are in the overall average wage level ranking 50%-75% being category III; and those whose average wage levels are after the overall average wage level ranking 75% being category IV.

Table 4 gives the results of the statistical estimation of the probability of shifting the category of the average wage level in Chinese cities in 2019-2020, and it can be seen that the probability of being located on the diagonal line of the table is much higher than the probability of being located on the non-diagonal line, indicating that there is relatively stable between the category states of the average wage level in Chinese cities in 2019-2020, with the characteristic of low mobility. The data in row 2 of the table illustrate that in 2020 compared to 2019, 91.67% of the cities have their per capita wages still in category I, while 0.05% of the cities have dropped one place, and it is of interest that 3.33% of the cities have dropped three places in a row from category I to category IV. The results in row 3 show that in 2020, compared to 2019, 81.67% of cities still have their per capita wages in category II, 10.00% and 6.67% of cities have dropped one place and increased one place respectively, and 1.67% of cities have dropped 2 places to become category IV. The data in row 4 shows that in 2020, compared to 2019, 75.00% of cities will still be in category III, while 16.67% and 6.67% of cities will be down one place and up one place respectively, and 1.67% of cities will be up two places to become category I. The data in row 5 shows that 73.47% of cities still have their per capita wage level in category IV, while 18.37%, 8.16%.

Table 4. Probability Shift Moments for Interannual Markov Chain Analysis 2019-2020

2019/2020	Class I	Class II	Class III	Class IV
Class I	0.9167	0.0500	0.0000	0.0333
Class II	0.0667	0.8167	0.1000	0.0167
Class III	0.0167	0.0667	0.7500	0.1667
Class IV	0.0000	0.0816	0.1837	0.7347

The 2019-2020 interannual Markov chain analysis shows that the mobility between the states of per capita wage levels in urban agglomerations is generally low, but the relative positions of various types of areas in the distribution of per capita wage levels have shown signs of instability, especially in Class III and Class IV areas, where the relative levels of per capita wages are more variable. Under the impact of the "black swan" of the new epidemic, different cities will have certain changes in per capita wage levels due to different industries, population structures, economic patterns and geographical locations. Localities should strengthen their ability to cope with contingent events, maintain the resilience of economic development and people's income growth, and minimize the impact of the epidemic shock on personal material life.

6. Conclusion

Per capita gross regional product, education level, and the role played by the government in economic activities have significant positive effects on wage levels, while openness, as measured by the ratio of total imports and exports to gross regional product, has a non-significant effect on wage levels, and the difference in wage levels between eastern and western regions is significant.

Based on the effect of the epidemic on time variation, the results of the grouped regressions show that the explanatory strength of the level of openness in the regional average wage difference again decreases under the influence of the epidemic and appears to be insignificant, while the explanatory strength of the other four basic variables is good and their coefficients are significant. Meanwhile, according to the time-varying Zou test, it is known that the impact of the epidemic on each coefficient of the explanatory variable of the regional average wage level is more reflected in the intercept term and the slope coefficient of the economic development variable, i.e. GDP per capita. The impact of the epidemic mainly changes the structure of these two coefficients, while the slope coefficients of the other four explanatory variables do not show significant changes.

The difference in per capita wages between cities is mainly influenced by the difference in the level of technological creation and economic strength of cities, but the dependence on fiscal tilt deepens after the outbreak. According to the Gini coefficient decomposition results, GDP per capita differences contribute prominently to per capita wage differences in 2019, and there are also sizable contributions from the number of patents granted in the previous year, the number of universities and geographical differences; however, the contributions of all the above four indicators drop more significantly in 2020, followed by a sharp rise in the impact of general public budget expenditures. In general, GDP per capita, number of patents granted in the previous year, number of colleges and universities, and geography are representations of inherent regional characteristics and should have some stability and inheritance in the short term.

The level of personal income is influenced by both the growth factor regarding public expenditure and investment in science and innovation and the structural factor reflecting the structure of the economy, and the growth factor has a stronger impact. This finding provides a good paradigm for local governments to promote the growth of local residents' personal income: to give full play to the guiding and supporting role of public finance, to support the development of science and innovation education with fiscal inclination, to use fiscal funds to "protect people's livelihood and promote happiness" during the risk period, to fully mobilize all positive factors, and finally to raise This will help to mobilize all the positive factors, and ultimately improve the income and sense of access of the residents.

The new crown epidemic has caused large relative changes in the average regional wage levels, and the risk contains opportunities. 2019-2020 inter-year Markov chain analysis shows that the relative positions of various types of regions in the distribution of per capita wage levels are relatively unstable, and the degree of inter-year changes cannot be ignored, especially in regions with relatively high income levels, the relative level of per capita wage changes are large. Under the impact of the "black swan" of the new epidemic, different cities will experience fluctuations in per capita wage levels due to different industries, demographic structures, economic patterns and geographical locations. Local governments should actively strengthen their ability to respond to contingent events, maintain the resilience of economic development and people's income growth, and minimize the impact of the epidemic on individuals' material lives.

References

[1] William F. Ogburn. National Income and Its Composition, 1919-1938. Simon Kuznets [J]. American Journal of Sociology, 1942, 48(2).

[2] Zhang Jianhong, J. Paul Elhorst, Arjen van Witteloostuijn. Analysis of the factors influencing regional wage differences in China [J]. Economic Research, 2006(10): 62-71.

[3] Zhong Xiaohan. Labor Mobility and Wage Differences [J]. China Social Science, 2006(1): 34-46.

[4] Cheng, Zhonghua, Yu, Binbin. Industrial Agglomeration and Regional Wage Gap: A Spatial Econometric Analysis Based on Chinese Urban Data [J]. Contemporary Economic Science, 2014, 36(06): 86-94+125.

[5] Wang Jingjing, Yue Zhonggang. Financial agglomeration, local demand size and regional wage differences: an empirical analysis based on panel data of prefecture-level and above cities in China [J]. Contemporary Finance and Economics, 2016(03): 87-95.

[6] Sang, R. C., Wang, H. L.. Local market demand, industrial agglomeration and regional wage differences [J]. DOI: 10.13269/j.cnki.ier.2011.06.003.

[7] Liu Xiuyan, He Xiaohai, Yin Shengmin. Market potential and regional wage gap: An empirical study based on prefecture-level panel data in China [J]. Management World, 2007(09): 48-55.

[8] Kuznet S. 1955 "Economic Growth and Income Inequality". American Economic Review45: 1-28.

[9] Dai Pingsheng. Gini coefficient decomposition based on regression equation [J]. Quantitative Economic and Technical Economics Research, 2013, 30(01): 150-160.

[10] Xiong Yuyi. Factor analysis and investment strategy construction of new energy generation enterprises [D]. Zhejiang University, 2020.