Composition Component or Conditional Component? A New Explanation on the Change of the Income Gap among Chinese Residents from the Perspective of Counterfactual

Jiqing Ma 1,2, Rundong Luo 1,*

1 School of Business, Shandong University, No. 180 West Culture Road, Weihai, Shandong, China
2 School of Business, Linyi University, Middle Section of Shuangling Road, Linyi, Shandong, China

Abstract: The income gap among residents is one of the important factors that restrict the development of economy and society. The emphasis of this study lies in investigating the reasons for the change of the income gap among Chinese residents. The Gini coefficient, a set of generalized entropy index and the relative polarization index are obtained on the data of CFPS. The results show that the income gap in China has been increasing year by year since 2012. Different from some ideas that components are the main factors affecting the change of income gap, the decomposition results show that the conditional components make the significant variation of income gap. This provides a new perspective to explain the causes of income gap changes. It may be necessary to increase the income of low-income groups and improve their status with dynamic assistance by stages and categories.

1 Introduction

As the old saying goes, one should not worry about scarcity but inequality. Marx also believed that distribution has an important influence on consumption and production[1]. According to "China Statistical Yearbook 2019", there were still about 610 million people in the low and middle income groups. These reflect the fact that the income gap among Chinese residents is large at present. Therefore, it is necessary to clarify the issues: First, what is the actual situation of the income gap among Chinese residents? Second, what cause the change of it?

Since 1912, the Gini coefficient has become the main index to measure income gap[2]. However, First, in view of the robustness and sensitivity, it is preferable to use a combination of various statistical indicators of inequality [3,4]. Second, the relative distribution method is a supplement to absolute measurement method[5]. In addition, based on the Oaxaca decomposition method and DiNardo decomposition method[6,7], some scholars decomposed the change of the income gap from the perspective of component and conditional component[8].

Therefore, the paper will focus on the following issues: First, based on the data of CFPS project, the income gap and its variation trend will be measured and compared. Second, the dynamic method is used to decompose the change of income gap into composition components and conditional components. Third, the conclusion and a brief discussion are obtained.

2 Materials and Methods

2.1. Data

The analysis in this paper are carried based on the data of China Family Panel Studies(CFPS) conducted by the Institute of Social Science Survey (ISSS) of Peking University, China. The project was launched in 2010 and will be surveyed every other year thereafter. The latest survey was conducted in 2018. As the full data for 2018 have not yet been officially released, this paper only uses the data of the first four periods for analysis. In order to make the sample more representative, the national resampling sample will be selected as the new population. In order to make the income comparable, this paper will exclude the samples with annual income less than 100 RMB and take the year of 2010 as the basis for analysis. The variables used in the analysis mainly include age, age squared, region, gender, household registration, ethnic group, education, party affiliation, administrative position, physical health, intelligence level, labor union status and marriage status, etc.

2.2. Indices

For the choice of income gap measurement indicators, we should mainly pursue the complementarity between these indicators. Therefore, when a single distribution is measured, the Gini coefficient and the generalized entropy is provided. The Gini coefficient is chosen because it is widely used and sensitive to income changes in the middle
of the income distribution. The reasons why we choose the four indexes with specified parameter values in the generalized entropy index family are as follows: Firstly, the generalized entropy family meets the axiomatic requirements of income gap measurement. Secondly, many other measurement indexes such as Atkinson index family and coefficient of variation can be transformed into corresponding measurement indexes of generalized entropy. Thirdly, in the generalized entropy index family, when the parameter \(\theta\) is -1, it is sensitive to the change of income at the bottom of income distribution. When the value of parameter \(\theta\) is 0 and 1, it is sensitive to the change of income in the middle of income distribution. While the value of parameter \(\theta\) is 2, it is sensitive to the change of income at the top of income distribution. Therefore, this combination of measurement indicators can meet the complementarity of single distribution measurement. When it comes to the comparison between distributions, different single distribution summary measurements are used. In addition, relative distributions are constructed by combining the information of different distributions, and the information of distribution changes is provided by using median relative polarization, lower relative polarization and upper relative polarization.

### 2.3. Methods

According to the information above, the selection of the income gap measurement indicators should mainly pursue the complementarity, therefore, when a single distribution is measured, the Gini coefficient and the generalized entropy index with parameter \(\theta\) of -1, 0, 1 and 2 are selected, when it comes to the comparison between distributions, supplementary information of distribution changes is provided by the relative polarization index. In addition, when it comes to the decomposition of the income gap variation, the decomposition formula of indices can be written as follows[4]:

\[
\Delta I = I_n - I_m = I(\hat{\beta}_n; C_n) - I(\hat{\beta}_m; C_m) \\
= \{I(\hat{\beta}_n; C_n) - I(\hat{\beta}_c; C_m)\} + \{I(\hat{\beta}_n; C_m) - I(\hat{\beta}_c; C_m)\} \tag{1}
\]

where \(\hat{\beta}_n\) and \(\hat{\beta}_m\) are the estimated value of covariate coefficients, \(C_n\) and \(C_m\) are the covariate composition of the nth period and the mth period, \(I_n\) and \(I_m\) are the income gap measurement indexes calculated by the actual marginal distribution, \(I(\hat{\beta}_n; C_m)\) is the income gap measurement index calculated by the counterfactual marginal distribution, \(I(\hat{\beta}_n; C_m) - I(\hat{\beta}_n; C_m)\) reflects the change of income gap caused by the change of covariate composition, while \(I(\hat{\beta}_n; C_m) - I(\hat{\beta}_n; C_m)\) reflects the change of the overall income gap caused by the change of coefficient. In addition, formula (1) only considers one counterfactual order, the other should be considered in practical application, therefore, the average of the two provides the final decomposition result.

### 3 Results and Discussion

#### 3.1 Descriptive Statistics Analysis

Before the detailed analysis, it is necessary to know something about the basic situation of the using data. Table 1 presents the sample size and the main statistics of income.

| Year | Obs  | Min | Mean | P50  | Std. Dev. | Max  |
|------|------|-----|------|------|-----------|------|
| 2010 | 20435| 100 | 10272| 5000 | 19210     | 80000|
| 2012 | 8731 | 90  | 20426| 16919| 27066     | 1084505|
| 2014 | 9970 | 88  | 18685| 14764| 20966     | 387719|
| 2016 | 8942 | 100 | 23212| 17391| 31314     | 1043478|

As shown in Table 1, it can be found that income has three obvious characteristics from an overall perspective: First, the average income and median income in 2016 are much higher than the corresponding items in 2010, which reflects the fact of economic growth. Second, their average income is higher than the median income in any given year, which shows that there is a certain income gap in both years. Third, compared to 2010, the gap between the average income and the median income in 2016 has increased in absolute terms, but decreased in relative terms, which reflects the changing trend of the income gap during this period.

#### 3.2. Measurement of Income Gap

According to the instructions above, complementary indicators such as the Gini coefficient, the generalized entropy indices and the relative polarization index are used to evaluate the income gap. The measurement results are shown in Table 2.

The results in Table 2 show that except for \(GE_{-1}\), the measurements of \(G, GE_0, GE_1\) and \(GE_2\) are all smaller than the corresponding values in 2010. It shows that except for the increase among low-income groups, the income gap among other groups has a trend of statistically significant decline. However, if 2010 is excluded, the values of \(G, GE_0\) and \(GE_1\) increase year by year, while the values of \(GE_{-1}\) first increase significantly and then decrease insignificantly, so it also shows an increasing trend on the whole. For \(GE_2\), its value is not significant when it decreases, but it is significant when it increases, so it shows an increasing...
Table 2. Measurement and changing trend of income gap

| Items | $G$ | $GE_{-1}$ | $GE_0$ | $GE_{1\text{ (T)}}$ | $GE_2$ | MRP | LRP | URP |
|-------|-----|-----------|--------|---------------------|--------|-----|-----|-----|
| Year  |     |           |        |                     |        |     |     |     |
| 2010  | 0.622 | 2.107 | 0.830 | 0.749               | 1.749   | —   | —   | —   |
| 2012  | 0.450 | 1.490 | 0.463 | 0.401               | 0.878   | —   | —   | —   |
| 2014  | 0.524 | 4.426 | 0.770 | 0.494               | 0.629   | —   | —   | —   |
| 2016  | 0.540 | 4.220 | 0.786 | 0.542               | 0.910   | —   | —   | —   |
| Variation |     |           |        |                     |        |     |     |     |
| 2010:2016 | -0.082*** | 2.114*** | -0.045*** | -0.208*** | -0.839*** | 0.541*** | 0.768*** | 0.314*** |
| 2012:2016 | 0.090*** | 2.730*** | 0.323*** | 0.141*** | 0.032 | 0.309*** | 0.347*** | 0.270*** |
| 2012:2014 | 0.074*** | 2.936*** | 0.307*** | 0.093*** | -0.248 | 0.160*** | 0.133*** | 0.187*** |
| 2014:2016 | 0.016** | -0.205 | 0.015 | 0.048** | 0.280* | 0.224*** | 0.331*** | 0.115*** |

Note: ***, ** and * represent significance at the levels of 1%, 5%, and 10%, respectively. Although the statistical significance of all indicators in each period was not marked, they were all significant at the significance level of 1%. In Table 2, MRP is the median relative polarization, LRP is the lower-portion relative polarization, and URP is the upper-portion relative polarization.

In addition, the MRP, LRP and URP also show a greater dispersion of income in 2016. Especially the LRP value between the 2010 and 2016 is as high as 0.768, indicating that the lower end of the 2016 income distribution is more dispersed than that of the 2010. The similar information is shown in Figure 1 and Figure 2.

According to Figure 1, the income distribution of 2012 Lorenz dominates the income distribution of 2010 and 2016, so the income gap in 2010 and 2016 is more serious than that in 2012. It also can be seen that the Lorenz curve of 2010 and 2016 intersect at the position of the 35th percentile. These show that the income gap of low-income people in 2016 is larger than that in 2010 and the income gap of middle and high-income people in 2016 has improved. However, in the case of crossover, the income gap between 2010 and 2016 cannot be clearly compared according to the Lorenz curve alone. For this reason, based on the data obtained by multiplying the cumulative share of income by the average income of each period, the generalized Lorenz curve shown in Figure 2 can be drawn. The information in Figure 2 clearly shows that the income gap in 2016 is generally smaller than that in 2010. In addition, the information in Figure 2 also shows that the generalized Lorenz curves for 2012 and 2016 cross each other around the 75th percentile, which is consistent with the information in Figure 1.

Table 3. Model-Based Decomposing Income Variation to Compositional and Conditional Components

| Decomposition | $G$ | $GE_{1\text{ (T)}}$ | $GE_2$ | $GE_{1\text{ (T)}}$ | $GE_2$ |
|---------------|-----|---------------------|--------|---------------------|--------|
| Simulative 2012 | $\tilde{F}(y_1; C_1)$ | 0.437 | 1.346 | 0.428 | 0.345 | 0.529 |
| Simulative 2016 | $\tilde{F}(y_2; C_2)$ | 0.462 | 1.764 | 0.497 | 0.369 | 0.456 |
| Overall change | 0.025*** | 0.418*** | 0.069*** | 0.024** | -0.073 |
| Counterfactual 1 | $\tilde{F}(y_2; C_1)$ | Compositional | 0.004 | 0.096 | 0.011 | -0.004 | -0.063 |
| | Conditional | 0.021*** | 0.322*** | 0.058*** | 0.028** | -0.010 |
revealed by the previous summary measurements. According to the previous summary measurement data, the indicators that are sensitive to the bottom and the middle of the income distribution all show that the income gap has widened over time, although the change of $GE_2$, which is sensitive to the top of the income gap, is positive from 2012 to 2016, it is not statistically significant.

### 3.3. Deconstruction of income gap variation

So far, we have measured and compared the income gap and its changing trend. What are the reasons for the change? Here we will use the method of quantile regression combined with counterfactual to decompose the variation of the income gap. The results are shown in Table 3.

Similar to those in Table 2, the measurements in Table 3 using the simulated marginal distribution also show a higher income gap in 2016 than in 2012. The average values of the two counterfactuals show that none of the indices listed are significant in terms of components, and most of them are significant in terms of conditional components at the 1% level of significance. These results suggest that changes of the resource classification mechanism are the main causes of the changes in income disparities.

### 4 Conclusions

In terms of the results of the measurement and the decomposition on the income gap among Chinese residents above, the conclusions are as follows: First, overall, the income gap is the greatest in 2010 and the smallest in 2012 than that in other years. The income distribution of 2012 Lorenz dominates the income distribution in 2010 and 2016. The dispersion of income distribution measured by relative polarization has gradually increased over time, and the income gap in China has been gradually expanding from 2012 to 2016. Second, the decomposition results of quantile regression combined with counterfactual methods show that the variation between 2012 and 2016 is mainly due to the income classification mechanism determined by factors such as family status, social relationships and luck, in other words, the conditional component, and the decomposition results also show that the compositional components have no significant impact on the income gap, which proves the conclusion that the variance of income gap mainly depends on the mechanism of income classification from a dynamic perspective.

In summary, from the results of measurement and dynamic decomposition, for the increasing income gap mainly results from different income classification mechanisms, it is necessary to focus on improving the income level of specific groups. The principles are as follows: First, according to the maximin criterion of Rawls, we should focus on improving the income gap of low-income groups. Second, given the complexity of optimizing the income classification mechanism, we should help low-income groups from two aspects: short run and long run. In the short run, we should help the low-income groups in employment and entrepreneurship through the development of industry and inclusive financial support, and in the long run, we should mainly improve intergenerational mobility through education, taxation and other means. Third, we should enrich the income sources of rural residents by means of macro-planning such as urbanization and rural revitalization, so as to promote the accumulation and the status of low-income groups. Fourth, although economic development is not the subject of this paper, we should always keep in mind that development is the last word, because both theory and practice have proved that rising inequality can be offset to some extent by an increase in average income under certain conditions[9]. Obviously, most of the suggestions above are summary. It is difficult to accomplish the whole task at one stroke, but it is just the remaining problems that propose direction for the follow-up research.

### Acknowledgments

This paper is one of the achievements of the National Natural Science Foundation of China project “Research on the structural substitution and reallocation effect of heterogeneous artificial intelligence model on human capital” (72073082).

### References

1. Karl Marx, Frederick Engels. (2016) Karl Marx and Frederick Engels. People's Publishing House, Beijing.
2. Gini, C. (1912) Variabilità e Mutabilità. Tipografia di Paolo Cuppini, Bologna.
3. Atkinson, A.B. (1970) On the Measurement of Inequality. J. Econ. Theory, 2: 244-263.

| Counterfactual 2 | $\tilde{F}(y_i; C_2)$ |
|------------------|----------------------|
| Compositional    | -0.011***           | -0.121*** | -0.026*** | -0.001 | 0.123 |
| Conditional      | 0.037***            | 0.539***  | 0.095***  | 0.025* | -0.196** |

| Average Counterfactual 1 and 2 |
|-------------------------------|
| Compositional                 | -0.003           | -0.012    | -0.007    | -0.003 | 0.030 |
| Conditional                   | 0.029***         | 0.431***  | 0.077***  | 0.027*** | -0.103* |

Note: All the index values in the top two rows of the table are the income gap measurement values of a certain marginal distribution. ***, ** and * represent significance at the levels of 1%, 5%, and 10%, respectively.
4. Hao, L.X., Naiman, D.Q. (2010) Assessing Inequality. SAGE Publications, Inc., London.
5. Handcock, M.S.; Morris, M. (1998) Relative Distribution Methods. Sociol Methodol, 28: 53-97.
6. Oaxaca, R. (1973) Male-Female Wage Differentials in Urban Labor Markets. Int. Econ. Rev, 14: 693-709.
7. DiNardo, J.; Fortin, N.M.; Lemieux, T. (1996) Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica, 64: 1001-1044.
8. Machado, J.A.; Mata, J. (2005) Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. J. Appl. Econom, 20: 445-465.
9. Cowell, F.A. (2011) Measuring Inequality. Oxford University Press, New York.