International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume: 05, Issue: 03 "March 2020"

AN ANALYSIS OF THE INCOME GAP BETWEEN INSIDE AND OUTSIDE CHINA'S SYSTEM: BASED ON CFPS PANEL DATA

Ma Yiming

School of economics and management, Nanjing University of
Aeronautics and Astronautics, Jiangsu 211106, China

DOI: 10.46609/IJSSER.2020.v05i03.005 **URL:** https://doi.org/10.46609/IJSSER.2020.v05i03.005

ABSTRACT

The purpose of this paper is to assess the income gap between within and outside the system. To overcome the problem of self-selection of the sample, this paper focuses on the group that entered the system from outside between 2010 and 2012, using a combination of propensity score matching method and the double difference method to construct the counterfactual. In 2010 and 2012 CFPS data base, we control the observable factors as much as possible, while eliminating the influence of the variables which do not change or change evenly over time using the double difference method, in order to obtain the pure impact on revenue of entering the system as precisely as possible. The results showed that, entering the system can increase the average individual income by 17.35 percent. Further classifying discussion showed that coming into the state-owned enterprises make revenue increased by 34.95 percent, and into the party and government organs, institutions will make the income increased by 6.70 percent.

Keywords: System inside and outside, Income gap, Self-selection, Propensity score matching, Difference in difference

INTRODUCTION

In public economics, income redistribution is a major function of the government. Large income gap will lead to a series of economic and social problems, even social unrest. Therefore, when the income gap is too large, the government needs to spend more energy and resources to carry out income redistribution, thus affecting the speed of economic development. Therefore, many scholars began to decompose the factors that affect the income gap, hoping to predict the future trend of China's income gap. Many studies show that gender, education, family background and other factors have a very significant impact on income. However, there is a great controversy about the research on the income gap between the inside and outside of the system. Many researchers think that the income of the personnel in the system, including civil servants and employees of state-owned enterprises, is lower than the social average level, but many scholars think that the income of the personnel in the system is actually higher than that outside the system, especially the income of the employees of many state-owned enterprises which occupy a monopoly position in the market. The two views are at loggerheads. Therefore, the goal of this paper is to evaluate the income gap inside and outside the system, and decompose the pure impact on income generated by entering the system from outside the system.

MODEL

The problem of income gap has been studied in a wide range. Some launch research from a relative macro perspective, such as Wang & Fan (2005), who studied the problem by testing the hypothesis of existence of Kuznets curve. Chen et al. (2010), Wang & Ouyang (2007), Li & Luo (2007) studied the factors that have an impact on the income gap such as education and household register. Ren (2002) and Lu et al. (2012) studied the income gap between inside and outside system.

There is no doubt that inside and outside the system is an important factor affecting the income gap, but in addition, there are many other factors that will have a significant impact on the income, such as gender, education, family background mentioned earlier, so simple income comparison cannot get accurate results. The usual method is to bring all these variables into the model and make multiple linear regression, so as to control other variables and obtain the income gap inside and outside the system. However, the problem of this method is that it fails to solve the problem of self-selection of samples, because there are significant differences in the distribution of education, family background and other factors in different groups inside and outside the division system. These factors can explain to some extent whether different individuals enter the system or not. Generally speaking, the groups with higher incomes have advantages in these aspects. With high education level and better family background, it is more likely to enter the system, which leads to the variable we are concerned about is not randomly distributed in the sample. In the multiple linear regression model, the problem of self-selection of samples has not been solved, which will produce endogenous problems, so the result is biased.

In general, when meeting endogenous problems in multiple regression model, the more common solution is to use tool variables, but tool variable method is not suitable for solving endogenous problems caused by self-selection. A good tool variable requires both tool relevance and tool externality. Taking family background as an example, family background is obviously related to whether an individual enters the system and satisfies tool relevance. However, family background also has an impact on other factors of an individual, including education, and even directly affects income. It is difficult to ensure that family background and residual items are not correlated. So family background does not satisfy the tool exogenous, and it is not a good tool variable. Similarly, in the face of endogenous problems caused by non-random selection, it is difficult to find appropriate tool variables, and the most important point is that tool variable method still does not effectively solve the problem of sample self-selection.

To sum up, we must consider the interrelationship between whether individuals enter the system and the income level and peel the influence of the former upon other factors, in order to get unbiased results. In general, for this kind of non-random selection problem, we can use the Propensity Score Matching (PSM) to reduce the influence of the endogenous problem caused by self-selection by constructing the counterfactual, and when it is combined with the Differences-in-Differences (DID).

Therefore, this paper will combine PSM and DID to launch the research based on the panel data of China Family Panel Studies (CFPS) in 2010 and 2012, focusing on the people who are transferred from outside the system to inside the system, control as many variables as possible and analyze the income gap between individuals inside and outside the system, so as to obtain the net impact on income brought by entering the system.

Set 2010 data as early group, 2012 data as late group, in which all individuals in early group are outside the system, the group that is still outside the system in late group is set as control group, and the group that enters the system is set as treat group. The idea of estimating income gap between inside and outside the system based on DID is as follows:

$$ATT_{DID}=E(Y_1\text{-}Y_0|treat=1)\text{-}\ E(Y_1\text{-}Y_0|treat=0) \qquad (1)$$

Among them, treat is a dummy variable that distinguishes treat group from control group. "1" represents treat group and "0" represents control group. Y_1 represents late group's income and Y_0 represents early group's income. On the right side of the equation, through the first difference, we can get the income change of treat group and control group respectively, and then carry out the second difference, which can eliminate the influence of the common change trend of all individuals. The result is the net impact of the transfer from work outside the system to work inside the system on the income.

The advantage of DID is that it can eliminate the influence of a large number of unobservable factors on income through difference, including the factors that do not change with time and the factors that change at a constant speed with time. However, this method cannot solve the endogenous problem caused by self-selection, so simply using DID cannot effectively solve the problem raised in this paper.

In the analysis of this paper, the distribution of whether to work within the system is not random, but a self-selection process influenced by many factors such as personal ability, family background, etc. the causal relationship between whether to work within the system and income level is difficult to distinguish. The direct use of multiple linear regression will lead to biased estimation results. So we construct the counterfactual here. Suppose that a person who works outside the system, with all other conditions remaining the same, enters the system. The change of the person’s income is the income gap between inside and outside the system. However, in reality, it is impossible for a person to have two jobs and two different incomes inside and outside the system at the same time. Therefore, it is necessary to use the method of PSM and other covariates to score whether he or she works in the system. Generally, probit model or logit model is used to estimate the probability of working in the system, and then group and compare the individuals with similar probability but in different group (inside the system and outside the system) to get the income difference. The idea is as follows:

$$ATT_{PSM}= E_{P(X)|treat=1}\{E[Y|P(X),treat=1]-E[Y|P(X),treat=0]\} \qquad (2)$$

Among them, treat is a dummy variable that distinguishes treat group from control group. "1" represents treat group and "0" represents control group. Y is the income level, X is the covariates, and P (x) is the score obtained by the covariates, that is, the probability of an individual entering the system obtained by probit model or logit model.

PSM can well solve the endogenous problems caused by self-selection, but it also has some defects. The biggest problem is that tendency score can only be estimated by some observed variables, so it is likely to miss variables. When the fitting effect of probit model and logit model is not good enough, the estimated income gap may have large errors.

The above two methods have their own advantages and disadvantages. Therefore, in this paper, we can use panel data to combine the two models to fully integrate the advantages of PSM and DID, which is to control the influence of observable variables and non-observable variables that do not change with time or uniformly change with time, and to eliminate the endogenous problems caused by self-selection. The idea is as follows:

$$ATT_{PSM\text{-}DID}= E_{P(X)|treat=1}\{E[Y_1\text{-}Y_0|P(X),treat=1]-E[Y_1\text{-}Y_0|P(X),treat=0]\} \qquad (3)$$

Among them, treat is a dummy variable that distinguishes treat group from control group. "1" represents treat group and "0" represents control group. Y_1 represents late group's income and Y_0 represents early group's income. X is the covariates, and P (x) is the score obtained by the covariates. Compared to PSM model before, the difference is that the dependent variable has been changed from income levels in cross section data to income level after differential in panel data. This is the PSM-DID model, which was used in the studies of Huang & Liu (2013) and Wan & Li (2013).

The data we used in this paper is from China Family Panel Studies (CFPS).After merging the individual data in 2010, family data in 2010 and individual data in 2012, we exclude mismatched samples. Then we simplify variables such as the year of joining the party, remove the sample of the students, the unemployed and the missing data.

Then define the variables ins2010 and ins2012 as dummy variables of whether they are in the system in 2010 and 2012, which are defined by the type of workplace. Individuals working in "government departments / party and government organs / people's organizations / army", "state owned / collective institutions / institutes / scientific research institutes" and "state owned enterprises / state holding enterprises" are defined as work within the system, while individuals working in other departments are defined as work outside the system. After that, all samples within the system in 2010 (ins2010 = 1) will be removed, and the rest of the individuals all work outside the system in 2010 (ins2010 = 0). Then, the samples outside the system in 2012 (ins2012 = 0) will be set as control group (treat = 0), and the samples within the system in 2012 (ins2012 = 1) will be set as treat group (treat = 1).And then, calculate the difference between the income level of each individual in 2012 and that in 2010, and get the new variable income_cha, which is the change of income level. Finally, 1544 samples were obtained, including 107 in treat group and 1437 in control group. And we get the covariates in Table 1.

Table 1: Covariates of the PSM-DID model

gender
age
urban or not
minority nationalities or not
Party member or not
worked in a cadre school or not
joined the army or not
administrative and management position or not

depression score
health score
public praise score
expression score
number of brothers and sisters
Spouse has administrative / management position or not
Years of mother's education
living area

Number of working years in current position as of 2010	family deposit
social status score	household expenditure
happiness score	number of family cars
confidence score	family health score
math test score	current housing price
career reputation score	family net worth

In this paper, the entry into the system is a self selection process determined by many factors, so we need to estimate the probability of individuals entering the system by propensity score which can be fulfilled by logit or probit model. Here we build a logit model:

$$logit(treat = 1) = \beta_0 + \beta_1 X + \varepsilon \qquad (4)$$

Among them, treat, as a dummy variable, reflects whether individuals outside the system in 2010 entered the system in 2012. X is all the covariates listed above that may affect the entry into the system.

RESULT

According to logit model, the regression results of Table 2 are obtained:

Table 2: The regression results of the logit model

variables	Coef.	Std.
gender	0.098	0.228
age	0.030	0.017
urban or not	0.822	0.243
minority nationalities or not	0.982	0.402
Party member or not	0.903	0.348
worked in a cadre school or not	1.238	1.557
joined the army or not	-0.489	0.619
administrative and management position or not	-0.335	0.369
Number of working years in current position as of 2010	0.019	0.013

social status score	-0.225	0.131
happiness score	-0.131	0.129
confidence score	0.151	0.124
math test score	0.063	0.026
career reputation score	0.011	0.009
depression score	0.078	0.043
health score	-0.158	0.146
public praise score	0.097	0.150
expression score	0.105	0.141
number of brothers and sisters	0.096	0.073
Spouse has administrative / management position or not	-0.754	0.553
Years of mother's education	-0.011	0.030
living area	-0.001	0.001
family deposit	0.000	0.000
household expenditure	0.000	0.000
number of family cars	-0.221	0.281
family health score	-0.090	0.127
current housing price	0.000	0.000
family net worth	0.000	0.000
_cons	-7.195	1.516
Pseudo R2	0.906	
n	1544	

From the regression results, we can see that most of the variables have a significant impact on the dependent variables. Pseudor2 has 0.906, which means the model has a considerable degree of explanatory power.

After obtaining the probability of individuals entering the system, the next step is to test whether there is a significant difference between the treatment group and control group in these covariates and tendency scores, that is, the balance test. Under the hypothesis of conditional exogenous, we require that there is no significant difference in the distribution between the two groups. Generally speaking, there are two ways to compare: the first is to test the double t distribution of each covariate; the second is to evaluate the whole model, including comparing R2, P values, etc. The results are presented in Table 3 and Table 4.

Table 3: The balance test 1

variables	match	t	p
gender	not matched	1	0.319
	matched	-0.14	0.888
age	not matched	3.5	0
	matched	-0.33	0.739
urban or not	not matched	5.3	0
	matched	-0.28	0.783
minority nationalities or not	not matched	2.25	0.025
	matched	0.52	0.605
Party member or not	not matched	4.22	0
	matched	0.19	0.851
worked in a cadre school or not	not matched	2.4	0.016
	matched	0	1
joined the army or not	not matched	0.88	0.377
	matched	-0.93	0.355
administrative and management position or not	not matched	0.01	0.99
	matched	1.41	0.16
Number of working years in current position as of 2010	not matched	3.37	0.001
	matched	0.05	0.958
social status score	not matched	-1.4	0.163
	matched	0.22	0.826
happiness score	not matched	-0.87	0.383
	matched	0.57	0.571
confidence score	not matched	0.19	0.849
	matched	0.5	0.619
math test score	not matched	3.57	0
	matched	1.2	0.231
career reputation score	not matched	1.96	0.05
	matched	0.6	0.552
depression score	not matched	1.45	0.147

	matched	0.1	0.992
health score	not matched	-0.42	0.676
	matched	0.73	0.464
public praise score	not matched	0.6	0.55
	matched	0.14	0.891
expression score	not matched	1.01	0.313
	matched	0.98	0.326
number of brothers and sisters	not matched	2.19	0.029
	matched	-0.26	0.792
Spouse has administrative / management position or not	not matched	-1.33	0.185
	matched	1.36	0.176
Years of mother's education	not matched	-0.31	0.758
	matched	-0.18	0.854
living area	not matched	-1.35	0.176
	matched	-0.02	0.982
family deposit	not matched	-0.31	0.754
	matched	0.45	0.65
household expenditure	not matched	-0.19	0.845
	matched	0.35	0.728
number of family cars	not matched	-1.22	0.223
	matched	1.59	0.113
family health score	not matched	-0.04	0.965
	matched	-0.21	0.836
current housing price	not matched	1.38	0.166
	matched	-0.69	0.494
family net worth	not matched	0.23	0.817
	matched	-0.35	0.724

Table 4: The balance test 2

	Ps R2	LR chi2	p>chi2
matched	0.107	83.02	0.000
not matched	0.042	12.34	0.995

From the above table, we can see that the p value of most covariates in "matched" is relatively large, which shows that there is no significant difference between treat group and control group, and their covariate distribution is consistent. The PS R2 and P values of the joint test also show that the treat group and control group matched have similar distribution.

If there is a big difference between the propensity scores of all treat group individuals and control group individuals, then the PSM-DID model is invalid, because in this case, the matching quality of the two groups is poor. So before formally estimating the income gap, we need to test the common support hypothesis. The common support hypothesis is to remove the samples near the end of the scores in the two groups, so that the maximum and minimum scores in the two groups are the same, so as to reduce a certain sample size and improve the matching quality. Finally, the number of samples in treat group and control group is 107 and 1169 respectively.

After the completion of scoring, balance test and common support test, the next step is to estimate the income gap between inside and outside the system. In order to reduce the error, this paper estimates by several different matching methods and different parameters, and obtains the average value as the conclusion. The estimated results are presented in Table 5.

Table 5: The estimated results of the income gap

matching methods	parameters	average treatment effect
attnd	/	1779.683
atts	/	5392.146
attr	r=0.001	2972.843
	r=0.002	3669.861
	r=0.005	4202.542
attk	k:gaus bw:0.01	2987.614
	k:gaus bw:0.06	4059.664
	k:epan	3209.955

	bw:0.01	
	k:epan bw:0.06	3470.516
average	/	3527.203

It can be seen from the last row of the table that, on average, the income gap between within and outside the system, i.e. the average treatment effect, is about 3527 yuan. And the average income of 1276 selected samples is about 20324 yuan. Therefore, on average, the income will increase by 17.35% (3527 / 20324) simply because of the entry into the system.

Considering that the system is also divided into different departments, and the probability of entering different departments is not necessarily the same, simply classifying it into one category for calculation may generate new endogenous problems. Therefore, this paper will divide the sample of treat group into two categories: the group entering the state-owned enterprises and the group entering other institutions (mainly the party and government organs and public institutions). The number of samples is 60 and 47 respectively, and then discuss them by classification.

The steps are basically the same as the previous analysis, and the results are presented in Table 6.

Table 6: The estimated results of the income gap in different categories.

matching methods	parameters	average treatment effect	
		state-owned enterprises	other institutions
attnd	/	7564.167	1651.284
atts	/	7132.143	3217.028
	r=0.001	7207.509	1827.714
attr	r=0.002	7581.989	-82.098
	r=0.005	6893.189	873.832

attk	k:gaus bw:0.01	6669.199	2861.431
	k:gaus bw:0.06	7056.188	3069.565
	k:epan bw:0.01	7059.227	-600.075
	k:epan bw:0.06	6764.617	-575.957
average	/	7103.136	1360.636

It can be seen from the table that the income gap between the sample of state-owned enterprises and the sample outside the system is significantly larger than that between the sample inside other systems and the sample outside the system. The former is about 7103 yuan, and entering the state-owned enterprises from outside the system will increase the income of personnel outside the system by 34.95%; the latter is about 1361 yuan, and entering the party and government organs and institutions from outside the system will increase the income of personnel outside the system by 6.70%.

CONCLUSION

The goal of this paper is to estimate the income gap between the people inside and outside the system. In the analysis process, we pay special attention to the group who transferred from outside the system to inside the system from 2010 to 2012, and use the combination of PSM and DID. On the whole, no matter which matching method and parameters are used, there is a significant income gap between them. On average, under the same conditions, entry into the system will increase the income of personnel outside the system by 17.35% in the short term.

In addition, this paper also discusses the groups of different parts in the system. The results show that the income gap between inside and outside the system is mainly caused by the excessive income of the staff of state-owned enterprises. In contrast, the income gap between the party, government organs, public institutions and people outside the system is relatively small, which is only 6.70%.

REFERENCES

[1] Wang Xiaolu, Fan Gang. Income Inequality in China and Its Influential Factors [J]. Economic Research Journal, 2005(10):24-36.

[2] Chen Binkai, Zhang Pengfei, Yang Rudai. Government educational expenditure, human capital investment and urban-rural inequality in China [J]. Management World,

2010(01):36-43.

[3] Wang Shaoping and Ouyang Zhigang. The Rural-urban Income Disparity and Its Effects to Economic Growth in the Case of China [J]. Economic Research Journal, 2007, 42 (10):44-55.

[4] Li Shi, Luo Chuliang. Re-estimating the Income Gap between Urban and Rural Households in China [J]. Journal of Peking University (Philosophy and Social Sciences), 2007(02):111-120.

[5] Ren Taizeng. Is the Income Level of Chinese Civil Servants Low [J]. Economist, 2002(06):16-22.

[6] Lu Zhengfei, Wang Xiongyuan, Zhang Peng. Do Chinese State-Owned Enterprises Pay Higher Wage? [J]. Economic Research Journal, 2012, 47(03):28-39.

[7] Huang Jingbo, Liu Shulin. Higher Employees' Income Growth in Exporting Firms? With Propensity Score Matching Method [J]. Finance and Trade Research, 2013,24 (06):62-69.

[8] Wan Haiyuan, Li Shi. The Effects of Household Registration System Discrimination on Urban-rural Income Inequality in China [J]. Economic Research Journal, 2013,48(09):43-55.