Hetergeneity in Human Cpital Accumulation and intergenerational mobility in Urban and Rural Chinaz; Data Method Section

Luxi Han 10449918

May 7th, 2017

1 Methodology

1.1 Theoretical Model

In general the theotetical model that is used to generate the econometric specification is overlapping generations model.

This paper uses a overlapping generations model with three periods and two generations.

Following the traditional human capital literature, we use a additively separable three period utitlity function in logrithm form:

$$U(c_1, c_2, c_3^c) = log(c_1) + \beta log(c_2) + \beta^2 log(c_3) + \beta^2 \alpha log(c_3^c)$$

In the first period, we assume the individual's consumption is entirely determined by their paretns, then we can reduce the utility function in the following form:

$$U(c_2, c_3^c) = log(c_2) + \beta log(c_3) + \beta \alpha log(c_3^c)$$

In the utility function, c_t denotes the consumption at period t. c_3^c denotes the consumption of the individual's child. In particular, we follows the specification of Qin, et.al (2014), where the consumption is another additive terms that is added into the utility function. Another variation can be made to consider transfer from one's offspring. In which we would consider the substitution between saving and investment in human capital.

Now we have the following budget constraint:

$$\begin{aligned} c_2 + s_2 + i_t &\leq y_2 \\ c_3 &\leq (1 + r_t) s_2 \end{aligned}$$

Additionally, we assume that one's income is a function of their human capital and other types of capital (for example, one's family's wealth, network effect) and the law of human capital is determined by the investment of human capital and parent's human cpaital.

$$y_t = (h_t)^{\gamma}(k_t)^{1-\gamma}$$
$$h^c_{t+1} = A_t(i^p_t)^{1-\sigma}(h^p_t)^{\sigma}$$

Notice, for different regions of China, γ, which represents the contribution of human capital to one's income, may differ. In general, the urban labor market is more complete, which means that human capital has higher returns. For the law of motion of capital, A_t stands for the productivity of investment and parents' human capital. In Qin's(2014) paper, they assume that A_t is independently and identically distributed. But in general, we will consider A_t as a function of one's innnate ability and environment effect, for example quality of education at different schools.

To sum up, we have the following theoretical model:

$$\begin{aligned} U(c_2, c^c_3) &= log(c_2) + \beta log(c_3) + \beta\alpha log(c^c_3) \\ s.t. & c_2 + s_2 + i_t \leq y_2 \\ & c_3 \leq (1 + r_t)s_2 \\ & y_t = (h_t)^{\gamma}(k_t)^{1-\gamma} \\ & h^c_{t+1} = A_t(i^p_t)^{1-\sigma}(h^p_t)^{\sigma} \end{aligned}$$

We omit the solution here and proceed the the basic econometric model.

1.2 Econometric Specification

We form the initial econometric specification. This model is used in Qin(2104). But this model suffers from endogeneity and homogneity problem. Specifically, the model ignores the effect of ones' parents' human capital and wealth on ones' income. And the model also ignores the heterogeneity by assuming common discount factor and altruistic parameter. Also ignoring regional differences excludes instituitional effects. But we first explore the basci model and the proceed to the full model in the final version of the paper.

The following equation demonstrates the econometric specification.

FIrst Stage Regression:

$$Educ^c = \beta_1 Euc^p + \beta_2 Income^p + \beta_3 X + \epsilon_1$$

Second Stage Regression

$$Income^c = \beta_1 Educ^c + \beta_2 Income^p + + beta_3 X + \epsilon_2$$

We estimate a two stage least square model. We also assume that parents' education level only affect one's income through one's educational attainment. The varaible X contains several control variables: occupation, age, gender.

2 Data Section

This paper will use household survey data from China Family Panel Studies (CFPS). CFPS is a nationally representative, annual longitudinal survey of Chinese communities, families and individuals. This survey was launched in 2010 and now has 3 batches of data released. The survey primairly focuses on economic aspect of the household. The CFPS survey is funded by Peking University.

For this paper, we will be using 2014 survey data which is the latest batch of data release.

2.1 Summary Statistics

In the following two tables, we show the summary statistics for urban and rural China separatly. We excule all of the interviewees with age under 9.

We can see that the education level in the urban area is significantly higher than that of the rural area. Average education year in urban area is almost twice as that of the rural area. The same situation is true for their parents' educational attainment. For family income, we can see that uran houeholds earn as much as three times of that of rural households. At the same time, the rural household size is larger than urban houeholds. Then average income per capita for rural household is significantly lower than that for urban households.

Table 1: Summary Statistics - Rural Housedhodls

Statistic	N	Mean	St. Dev.	Min	Max
Gender	28,037	0.498	0.500	0	1
Education Year	28,023	5.998	4.640	0	18
Father Education Year	21,966	4.232	4.287	0	21
Mother Education Year	22,164	2.170	3.522	0	21
Personal Annual Income	7,809	25,161.650	19,750.020	0	300,000
Household Identity	28,037	1.000	0.000	1	1
Agricultrual Income	20,048	12,652.830	31,696.180	0	856,000
Self-Employment Income	2,483	38,536.480	47,350.870	0	1,200,000
Government Transfer	18,633	1,344.802	3,220.604	0	100,000
Retirement Stipend	8,709	4,700.195	9,533.636	0	100,000
Annual Wage	5,203	42,480.370	38,860.860	0	600,000
Annual Family Income	27,770	23,676.620	41,942.930	0	1,404,200
Job Sector	8,793	4.936	6.590	1	77

Table 2: Summary Statistics - Urban Household

Statistic	N	Mean	St. Dev.	Min	Max
Gender	9,991	0.505	0.500	0	1
Education Year	9,990	9.757	4.850	0	21
Father Education Year	7,753	6.160	4.918	0	21
Mother Education Year	7,915	4.169	4.778	0	21
Personal Annual Income	4,118	33,259.910	32,740.520	0	800,000
Household Identity	9,991	3.000	0.000	3	3
Agricultrual Income	1,383	11,723.590	36,868.880	0	410,000
Self-Employment Income	968	51,498.180	96,859.000	0	1,200,000
Government Transfer	2,201	3,154.818	5,652.985	1	100,000
Retirement Stipend	5,176	32,077.240	31,359.410	1	1,000,000
Annual Wage	5,617	62,366.540	117,943.200	0	4,000,000
Annual Family Income	9,909	60,572.920	104,408.200	0	4,043,000
Job Sector	4,803	4.672	9.181	1	77

3 Initial Regression Result

3.1 First Stage

In table 3, we present the first stage regression. We can see that all of the variables that we choose are significant. Since the sample size for CFPS is relatively large, the comparison of significant level may not be much intuitive. Instead, we can focus on the difference between the urban and rural houehold model. For intergenerational human capital transfer, we can see that father's education is more important than mother's education for both urban and rural household. The regression model for rural houesehold has slightly higher estimated value than urban household. But in comparison, the effect of income on education is much higher for rural household than urban household. Specifically, increase of 10 thousand yuan in annual family income will contribute to an increase of 0.096 year in education, while the estimate is 0.037 years for urban households. The effect of income is almost three times for rural households than for urban households. Then this result in a certain degree can subtantiate our hypothesis that financial constraint is more important for rural households.

3.2 Second Stage

Now we show the second stage regression for the earning equation. This is a simplified version of the result. Since the working experience variable will need further cleaning of the data, we are going to use age as a proxy for working experience. The variable of interest is education and its effect on personal annual income. We can see that the effect of education on earnings is larger for urban households than for rural households. On one hand, this result shows that the labor market in urban area is more complete than rural area, on the

Table 3: First Stage Regression

	Dependent variable:	
	Educational Attainment	
	Urban	Rural
	(1)	(2)
Father Education	0.168***	0.182***
	(0.012)	(0.007)
Mother Education	0.108***	0.182***
	(0.013)	(0.009)
Family Income	0.037***	0.096***
	(0.004)	(0.007)
Age	-0.089***	-0.099***
	(0.003)	(0.002)
Gender	0.808***	1.364***
	(0.092)	(0.055)
Constant	12.137***	8.472***
	(0.216)	(0.107)
Observations	6,973	18,883
R^2	0.298	0.330
Adjusted R^2	0.298	0.329
Residual Std. Error	3.839 (df = 6967)	3.752 (df = 18877)
F Statistic	592.152*** (df = 5; 6967)	1,855.720*** (df = 5; 18877)
Note:		*p<0.1; **p<0.05; ***p<0.01

other hand, this shows that the expected returns of education is lower in rural area, with urban area has a 5.7 percent increase of annual income corresponding to one year increase in edcuation and 2.7 percent for rural area. Then this indicates that the human capital investment incentive is lower for rural area.

Table 4: Second Stage Regression

	Dependent variable:	
	Personal Annual Income	
	Urban	Rural
	(1)	(2)
Education	0.057***	0.027***
	(0.004)	(0.003)
Log(Family Income)	0.174***	0.051***
	(0.012)	(0.006)
Age	0.003**	-0.001
	(0.001)	(0.001)
Age^2	0.312***	0.446***
	(0.029)	(0.023)
Gender	8.882***	9.371***
	(0.088)	(0.051)
Observations	3,766	6,586
R^2	0.145	0.079
Adjusted R^2	0.144	0.078
Residual Std. Error	0.877 (df = 3761)	0.905 (df = 6581)
F Statistic	159.053*** (df = 4; 3761)	140.280*** (df = 4; 6581)
Note:		*p<0.1; **p<0.05; ***p<0.01

The above results are initial results from the raw dataset. There are several changes to be made in the future: 1) The theoretical model needs to be solved; 2) the econometric specification is going to be more sophisticated and more variables are going to be added in response to the specification of theoreticfal model; 3) the data set needs further cleaning and variables such as education investment, parental income, working experience, need to be added into the model; 4) more econometric specification should be added for robustness test.