IMPACT OF RURAL-URBAN MIGRATION ON CHILDREN'S ACADEMIC PERFORMANCE IN CHINA

By

ZHANG, Yanze

Thesis

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MDP

2019

IMPACT OF RURAL-URBAN MIGRATION ON CHILDREN'S ACADEMIC PERFORMANCE IN CHINA

By

ZHANG, Yanze

Thesis

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MDP

2019

Professor Shun Wang

IMPACT OF RURAL-URBAN MIGRATION ON CHILDREN'S ACADEMIC PERFORMANCE IN CHINA

By

ZHANG, Yanze

Thesis

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MDP

Committee in charge:

Professor Shun Wang, Supervisor

Professor Kim, TaeJong

Approval as of August, 2019

ABSTRACT

IMPACT OF RURAL-URBAN MIGRATION ON CHILDREN'S ACADEMIC PERFORMANCE IN CHINA

By

Yanze Zhang

Left-behind children and rural-urban migrant children have become two large vulnerable groups in China. What is the well-being status of them? In this paper, I tested their academic performance by utilizing the data from CFPS. After solving endogeneity issue with instrumental variables, I obtained empirical results suggesting that urban-urban children' academic performance do not improve after migration, while left-behind children suffer from low academic performance due to the separation from their parents. This study highlights the significant difference between rural and urban education quality, and implies the importance of parenting in children's education. It also provides policy recommendations to Chinese government to deal with this issue.

ACKNOWLEDGEMENTS

First of all, I sincerely thank my thesis supervisors, Professor Shun Wang and Professor Taejong Kim. Without their kind guidance and insightful feedback, I would not be able to actually conduct this research that had been interesting to me from a long time ago. I am also very grateful to Professor Jongil You and Professor Jisun Baek for encouraging me to pursue an academic career. The time I spent at the KDI School of Public Policy and Management was truly meaningful for my whole life. It rewarded me a deeper understanding and sincere respect to different cultures, and I will keep it deep in my mind forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

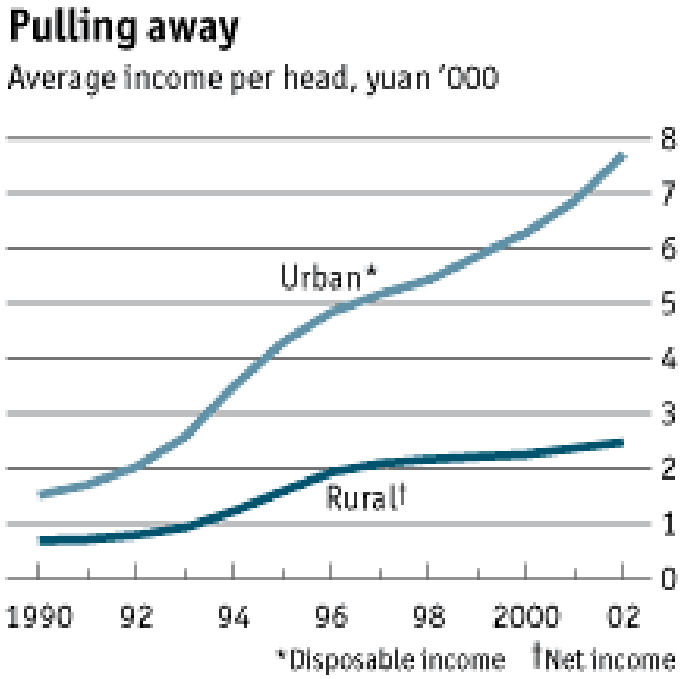
1.1 Background

China has achieved dramatic economic growth in the past several decades. It became the second largest economy in the world in 2010. During the same period, we also witness large waves of labor migration from rural to urban areas. This phenomenon emerged after *the reform and opening up* along with the process of industrialization and urbanization. The total number of migrant workers was in an increasing trend and reached 285.52 million in 2017, according to the National Bureau of Statistics of China (NBS, 2018).

The practice of many developed countries showed that rapid economic growth is accompanied by the transfer of agricultural labor to non-agricultural industries. This pattern is also what is going on in China. Scholars argue that labor migration is one vital driving force of China's economic development (Jia, Du, & Wang 2017).

In the individual level, rural works make the migrant decisions primarily because of the significant income gaps between rural and urban areas (Zhu, 2002). Many of them, especially the young and middle-aged generations, choose to work far away from their hometowns for the sake of maximizing family income and support their old parents and young children (Hugo, 1998). Figure 1 (the Economist, 2003) shows the vast income gap between rural and urban areas.

Figure 1: the Income Gap between Rural and Urban Areas

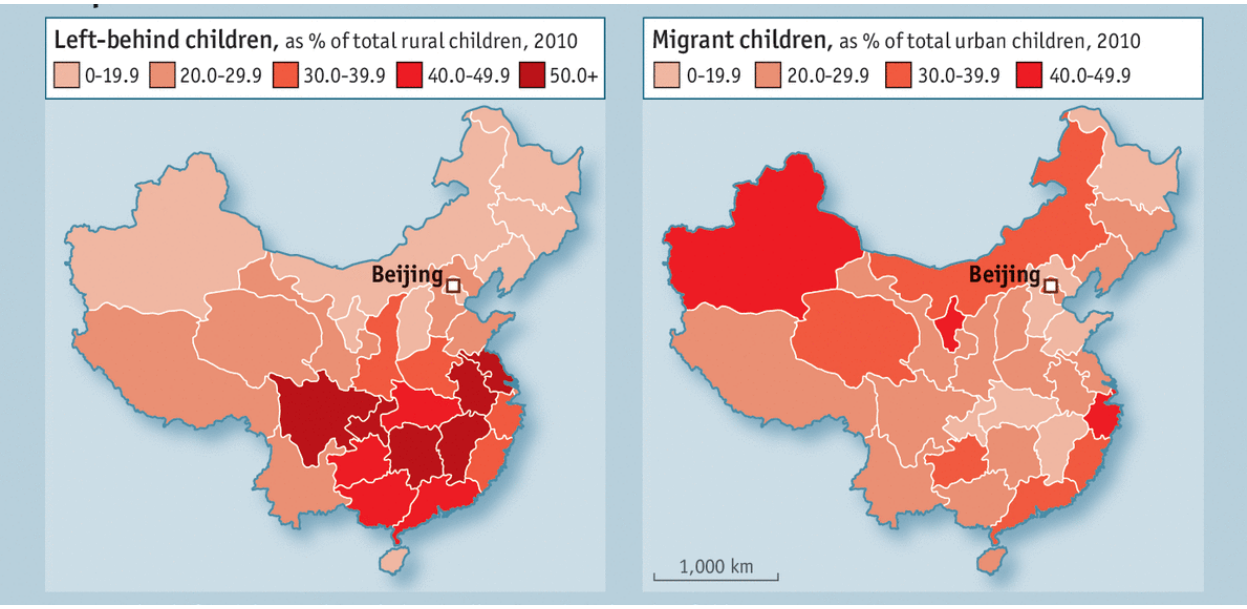


However, this choice is definitely not costless. Migrant workers suffer from various family issues, for example, being separated from their children and parents who need to be taken care of. These children who are called *left-behind children* lack parental care and, therefore, have worse wellbeing states. In terms of mental health, some of them are introverted and suffer from low self-esteem. In terms of the relationship between children and parents, some of them do not understand or even resent their parents. In terms of education, some of them lose interest in studying due to lack of supervision and support.

The left-behind children issue is prevalent and cannot be neglected anymore. According to a national survey, there were 61.03 million left-behind children in 2013, accounting for 21.88% of the whole population of children in China (Wu & Zhang, 2015). Figure 2 (the Economist, 2015) shows the proportion of left-behind children and migrant children in the total rural children

population. We can see that in some southwestern provinces, the proportion of left-behind children is even over 50%.

Figure 2: the Proportion of Left-behind Children and Migrant Children in the Total Rural Children Population by Province



Why are children left behind? The fundamental reason is the Hukou system, which impedes parents to bring their children together to urban areas because it is hard for the children to attend local the public schools and enjoy local welfare (Liang, Guo & Duan, 2008). Attending private schools is almost impossible because the migrant parents typically have too low income to afford high tuition fees for private schools.

1.2 Hypothesis Development

Previous research on the left-behind children has covered different aspects of wellbeing (Gao, Li, Kim, Congdon, Lau, & Griffiths, 2010; Jia & Tian, 2010; Wen & Lin, 2012; Zhao & Yu, 2016). The majority of them suggest negative effects of being left behind on the children's health, mental status, and education outcome.

Many of the previous studies which examined the education of left-behind children focus more on years of schooling and school attendance rate, instead of academic performance (Lu, 2012; Morooka & Liang, 2009). The comparison was commonly conducted between only two groups of migrant children, left-behind children and rural children without parents' migration. However, it is also important to consider children who migrate from rural to urban areas.

Moreover, these studies failed to effectively take care of the endogenous nature of migration decisions, making the results unreliable.

Based on the literature review, I decided to epically test the relationship between children migration status and academic performance with survey data collected in China. I categorized four types of children migration status, including rural-rural children, left-behind children, rural-urban children, and urban-urban children, in order to provide a more comprehensive picture. I also proposed two types of instrumental variables, time spent on the way to school and age of grandparents, as treatments of endogeneity issue.

More specifically, I tested the following three hypotheses:

Hypothesis 1: The academic performance of children who are originally from urban areas are better than that of their peers who are originally from rural areas.

Hypothesis 2: The academic performance of migrant children who are originally from urban areas are better than that of their peers who stay in rural areas with their non-migrant parents.

Hypothesis 3: The academic performance of left-behind children is worse than that of their peers who stay in rural areas with their non-migrant parents.

I found that urban children's academic performance is better than that of rural children. Urban-urban children's academic performance does not benefit from the migration, while the left-behind children suffer from the separation from parents.

The remaining contents of this paper were structured as follows. In chapter 2, I introduced the features of the survey data I used and explained the sampling process. Then I listed the dependent variables, independent variables, instrument variables, and control variables, and explained how they were measured in detail. At the end of chapter 2, I illustrated why 2SLS is a suitable regression method in my study. In chapter 3, I provided the regression results by tables and explanations. Finally, in chapter 4 I draw conclusions based on the empirical results and propose policy recommendations to the Chinese government. I also stated the limitation of this research and showed future research directions.

Chapter 2: Methodology

2.1 Data and Sampling

The sample of this paper was collected from the survey data China Family Panel Studies (CFPS). CFPS is a biennial social survey project implemented by Peking University, China. I summarized the important facts and characters of CFPS as follows:

1) The baseline survey was launched in 2010. By 2019, the survey data of four periods are available (2010, 2012, 2014 and 2016).
2) The CFPS sample initially covered 25 provinces and the target sample size was 16000 households. Since 2015, it has covered all 31 provinces of the mainland of China and the sample size is greater than 32500 households.

3) The sampling unit is each household. If one family is chosen, then each family members whose age is larger than nine will receive the questionnaire.
4) Once being included in the sample, one becomes a permanent member of the database and will receive follow-up survey ever since.
5) Since CFPS designed several different questionnaires, the data collected are organized into subsets, such as child subset, adult subset, and family subset.

To sum up, CFPS has a wide coverage of the population and measures a variety of social aspects on individual, family and community levels. Moreover, it is a reliable data source for academic research. It uses computer-aided survey technology during the interviews to improve access efficiency and ensure data quality. Thus, it has attracted attention from scholars working in different research topics.

CFPS is a rich source of data. For example, the questionnaire for children consists of 734 variables. I first excluded the variables that are irrelevant to my study in each subset, and then merged them into one dataset. I selected the children whose age is between 10 and 15. The younger children's questionnaires were not answered by themselves but by their guardians. To avoid any potential influence of the difference in respondents, I excluded the children under 10 from the sample. After dealing with missing values, the final sample consisted of 2964 observations, which is sufficient in terms of the size.

2.2 Measurement

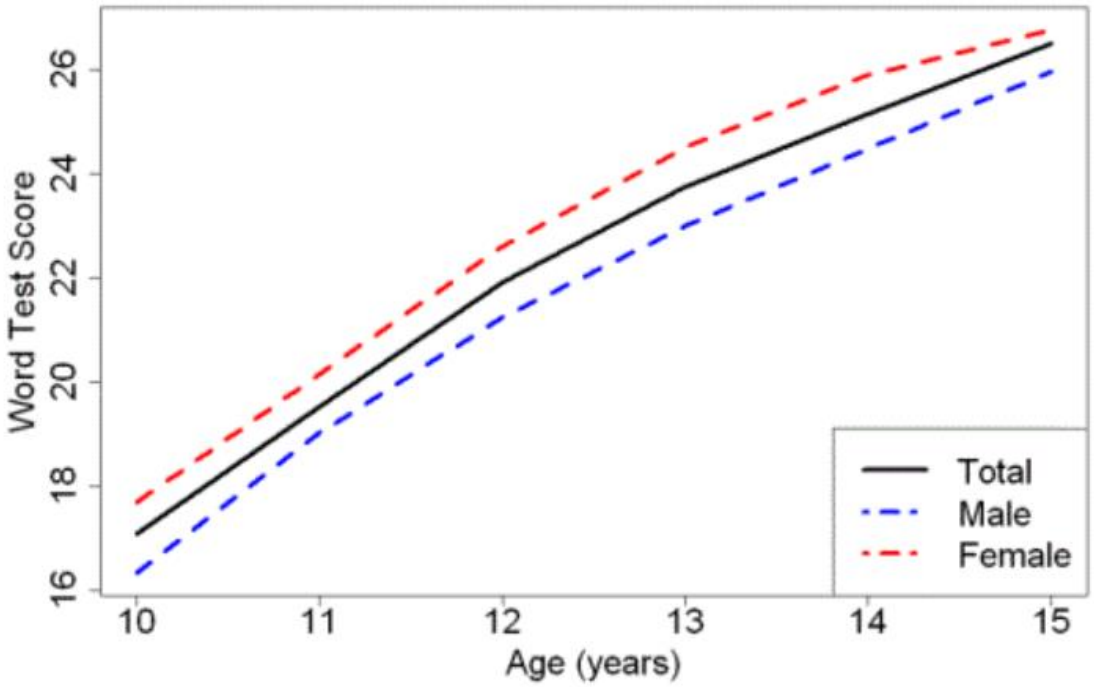
Dependent variables

Academic performance is an important aspect of education. However, it is empirically hard to measure in previous research in the Chinese context due to the lack of a universal grading

system. Scores gained from different schools are normally not comparable. Thus, the majority of research has been using years of schooling and school attendance rate as independent variables, which is not a good measure for the quality of education. One of the advantages of CFPS is that it access the academic ability of each respondent in the same way. The researchers conducted two tests, word test and math test. These two scores can be viewed as valid and reliable measures for academic performance.

The first variable is the verbal score, which is originally the word test score in CFPS. Respondents were asked to answer a series of questions starting from the easiest one to the most difficult one. There are 34 questions in total. They would not be asked the remaining questions if they failed to give the correct answers for three times continuously. The value of the word test score is the number of the most difficult question that the respondent answered correctly. The starting point of the questions varied by the respondent's education level. Figure 3 (China Family Panel Studies User's Manual, 2017) shows the relationship between age and verbal score of all children in the whole survey. The respondents attained higher score along with the increase in their age. Additionally, female children have better performance than male children in the verbal score at every age.

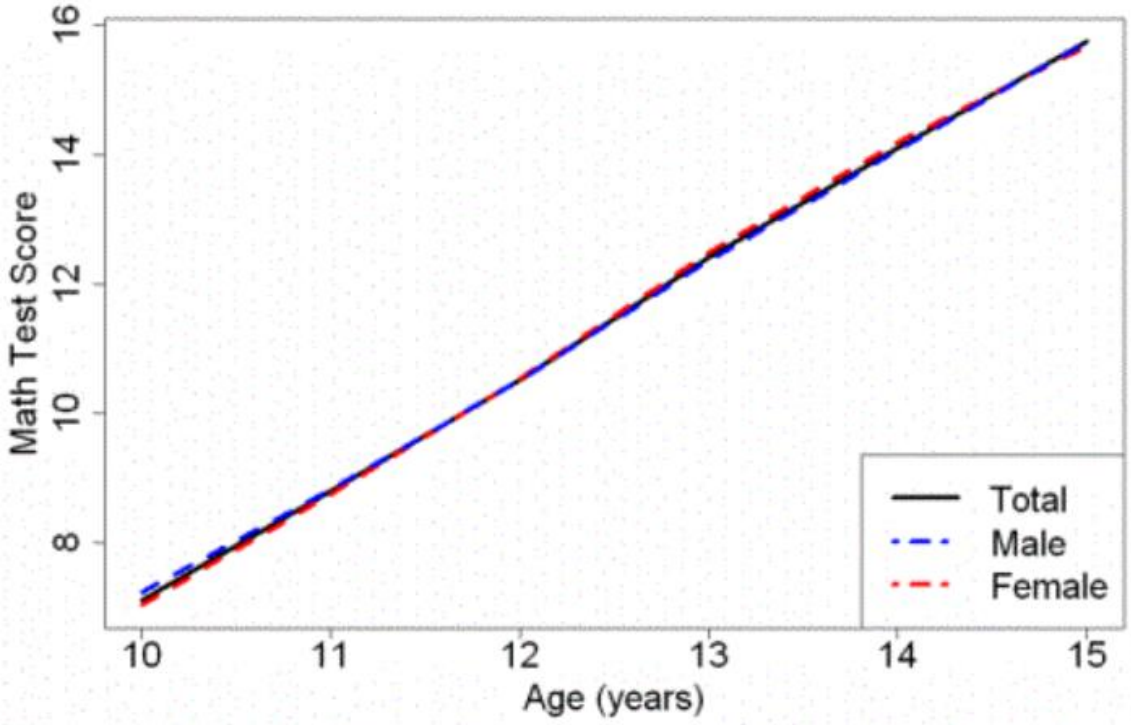
Figure 3: Verbal Score by Gender and Age



Source: CFPS (2017)

The math score was measured in the same way. As Figure 4 (China Family Panel Studies User's Manual, 2017) shows, the respondents tended to have higher math as age increases. Unlike the verbal score, female and male children have very similar performance in terms of math.

Figure 4: Math Score by Gender and Age



Source: CFPS (2017)

Independent variables

I defined four types of migration status and listed them in Table 1 where rural-rural children serve as the baseline for comparison.

Table 1: Types of Children Migration Status

Urban-urban (uu)	the child is originally from urban to rural areas and neither of two parents are migrant workers
Rural-rural (rr)	the child is originally from rural areas and neither of two parents are migrant workers
Rural-urban (ru)	the child is originally from rural areas, but has moved to urban areas since at least one parent is a migrant worker

Left-behind (lb)	the child is originally from rural areas but is separated from at least one migrant parent

As we can infer from the table, only Urban-urban children are originally from urban areas. Note that there are two scenarios for left-behind children: both parents are migrant workers, or only one parent is a migrant worker. Moreover, this typology does not cover all the possible situations, but only the ones that are relevant to my research questions.

Instrumental variables

The migration status of children is endogenous. Assume there is a migrant father who pays much attention to children academic performance. He would probably prefer to bring his children to the city where he works so that he can oversee children's education by himself. Here, the attention paid to children education is one potential omitted variable.

Thus, I decided to use instrumental variables, time spent on the way to school and grandparents age, to get reliable results. Both of them have no direct impact on academic performance. However, if time spent on the way is a determent whether to migrate or not. Intuitionally, parents would prefer schools where reduce the time spent on the way, either in rural or urban areas.

Grandparents are the people who can take care of children instead of their parents to a certain degree, especially in the context of Chinese tradition. But if they are older than 65, this role will be weakened given their degreasing health status. Thus, I expect that having grandparents whose age is under 65 increases the chance of being left behind in rural areas.

After running the correlation test, I found that the maternal grandmother age is highly correlated with the maternal grandfather age, and the paternal grandmother age is highly correlated with the

paternal grandfather age. Thus, I only included maternal grandmother age and paternal grandmother age in my analysis.

Control variables

To be consistent with the stream of research on children education, I included gender and age of children, father education level and household income as control variables. I also included the province dummies. As we saw in Figure 2, the proportion of left-behind varies across provinces. As a matter of fact, people in different provinces have different preference over migration, so its impact has to be controlled.

2.3 Regression Method

In the next chapter, I first ran OLS regression on the relationship between origins and academic performance of children to test whether there is an urban-rural difference in terms of education quality. In this case, endogeneity is not a concern because the origins of children were decided at birth as given. Thus, I believe OLS regression is sufficient to get reliable effect sizes.

Then I compare the three groups of children whose origin is rural areas, i.e., rural-rural, rural-urban and left-behind children. I ran 2SLS regression due to endogeneity issue as I have already mentioned.

Chapter 3: Results

3.1 Descriptive Statistics

Table 2 shows the summary statistics of the variables. The average verbal score is 21.723, and the average math score is 11.201. About 45% of the whole sample is rural-rural children. Both urban-urban and rural-urban children account for around 20% of the sample. The proportion of

left-behind children is the least, roughly around 16%. In terms of gender, the proportion of male children (53%) is higher than that of female children. The numbers of children of different ages are very similar. Around 41% of the fathers did not hold middle school degree, while only 20% of the fathers have high school degree or higher degrees. In the unreported correlation test, I found no evidence that multicollinearity would be a concern in this paper.

Table 2: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
verbal	21.723	7.152	0	34
math	11.201	4.455	0	24
uu	0.192	0.394	0	1
ru	0.199	0.400	0	1
lb	0.159	0.366	0	1
rr	0.449	0.497	0	1
male	0.529	0.499	0	1
lninc	8.249	1.255	-1.791	12.941
child_11	0.165	0.371	0	1
child_12	0.158	0.365	0	1
child_13	0.159	0.366	0	1
child_14	0.168	0.374	0	1
child_15	0.186	0.389	0	1
f_edu_2	0.262	0.440	0	1
f_edu_3	0.389	0.488	0	1
f_edu_4	0.128	0.334	0	1
f_edu_5	0.074	0.262	0	1

3.2 Regression Results

Table 3 shows the OLS regression results for the urban-rural difference. The coefficients of urban children dummy is positive and significant with p-value smaller than 0.05 for both the verbal score and the math score. Thus, Hypothesis 1 that urban children have better academic performance is supported.

Additionally, female students have significantly stronger performance in verbal, but similar performance in math. Both of them are in line with what was shown in Figure 3 and Figure 4. The coefficient of log household income is positive and significant at 95% confidence level. The coefficients of age dummies being positive and significant shows consistency with the trends shown in Figure 3 and Figure 4. Lastly, we can see that the higher the father education level, the larger its positive impact on children academic performance.

Table 3: OLS Regression Results for Urban-rural Difference

	Verbal	**Math**
Urban children	1.635***	0.938***
	(0.348)	(0.185)
Male	-1.135***	-0.011
	(0.261)	(0.140)
Log household income	0.435***	0.204***
	(0.141)	(0.079)
Child age:11	3.096***	1.319***
	(0.477)	(0.227)
Child age:12	5.207***	3.101***
	(0.479)	(0.246)
Child age:13	7.256***	5.244***

	(0.467)	(0.262)
Child age:14	8.500***	7.113***
	(0.455)	(0.230)
Child age:15	9.544***	8.336***
	(0.475)	(0.235)
Father education: primary school	2.270***	0.673***
	(0.463)	(0.248)
Father education: middle school	3.077***	1.160***
	(0.434)	(0.241)
Father education: high school	3.651***	1.421***
	(0.528)	(0.289)
Father education: post-secondary	4.981***	2.062***
	(0.680)	(0.330)
Province dummy	Y	Y
Obs.	2964	2964
Adj. R-sq	0.368	0.529

Robust standard errors in parentheses;

* p<0.1, ** p<0.05, *** p<0.01

As I have discussed, the instrument variable approach was applied to solve the endogeneity issue. From Table 4, we can see that both models have significant F statistics, meaning that the instruments I chose have enough explanatory power. We can also see that the magnitude of paternal mother's age is the largest when it comes to the left-behind decision making. This is identical to our institution. It is more common that paternal mothers take care of children instead of maternal grandmothers in China, especially in rural areas. If the paternal mother is able to take care of the children, then the parents would prefer leaving them in the rural area to avoid the high

cost of children migration. While time spent on the way to school has an insignificant influence on the left-behind decision, it has a significantly negative influence on children migration.

Table 4: The First Stage of the 2SLS Regression

	ru	**lb**
Time spent on the way to school	-0.002***	0.0001
	(0.0003)	(0.0003)
Paternal mother's age greater than 65	-0.055	0.159***
	(0.229)	(0.015)
Maternal grandmother's age greater than 65	0.07***	0. 041**
	(0.026)	(0.019)
F(3, 2317)	20.92	49.74
Prob > F	0.000	0.0000

Robust standard errors in parentheses;

* p<0.1, ** p<0.05, *** p<0.01

In Table 5, I present the results for OLS regression and 2SLS regression for comparison. In both Model 3 and Model 4, the rural-urban migrant has no significant coefficients, while left-behind has significantly negative coefficients. Thus, Hypothesis 2 is rejected, and Hypothesis 3 is supported. The directions the control variables are the same as in Table 5.

Table 5: OLS Regression results and the second stage regression results for rural students

	OLS Regression		2SLS Regression	
	Verbal	**Math**	**Verbal**	**Math**
	Model 1	**Model 2**	**Model 3**	**Model 4**
Rural-urban	0.831**	0.084	0.550	-0.094
	(0.357)	(0.189)	(2.322)	(1.495)
Left-behind	-0.572	-0.317	-4.087**	-2.148**
	(0.490)	(0.245)	(1.957)	(1.010)
Male	-1.263***	-0.049	-1.351***	-0.071
	(0.303)	(0.165)	(0.308)	(0.165)
Log household income	0.356**	0.206**	0.370*	0.220*
	(0.159)	(0.091)	(0.201)	(0.123)
Child age:11	3.211***	1.399***	3.241***	1.423***
	(0.544)	(0.256)	(0.560)	(0.270)
Child age:12	4.942***	2.897***	4.929***	2.863***
	(0.564)	(0.289)	(0.578)	(0.305)
Child age:13	7.038***	5.008***	7.171***	5.010***
	(0.546)	(0.309)	(0.584)	(0.332)
Child age:14	8.878***	7.120***	8.781***	7.081***
	(0.528)	(0.266)	(0.541)	(0.281)
Child age:15	9.638***	8.328***	9.757***	8.398***
	(0.557)	(0.278)	(0.568)	(0.290)
Father education	2.291***	0.570**	2.444***	0.652**
	(0.476)	(0.254)	(0.487)	(0.262)
Father education	2.921***	0.972***	3.197***	1.145***
	(0.463)	(0.255)	(0.543)	(0.296)
Father education	4.234***	1.810***	4.566***	1.826***
	(0.607)	(0.343)	(0.688)	(0.352)

Father education	5.422***	1.916***	5.488***	1.932***
	(1.394)	(0.518)	(1.407)	(0.506)
Province dummy	Y	Y	Y	Y
Obs.	2375	2375	2356	2356
Adj. R-sq	0.347	0.499	0.320	0.481

Robust standard errors in parentheses;

* p<0.1, ** p<0.05, *** p<0.01

Chapter 4: Discussion

4.1 Conclusion

The findings can be summarized as follows:

- Urban children's academic performance is better than that of rural children after controlling for family income and father education level. The reasons are multi-dimensioned. 1) Urban schools on average have better facilities and teachers, enable children to receive education of higher quality. 2) Parents in urban areas, especially in big cities, are more eager to improve children's academic performance. They are more likely to send their children to private education institutes even though it can be very costly.
- There is no significant improvement of rural-urban migrant children's academic performance compared with that of rural-rural children. It is difficult for migrant children to get enrolled in good schools in rural areas. Due to limited resources that their parents have, they hardly get access to high-quality education as the urban children do. At the same time, it takes time for them to adjust to urban life and build a new friendship with their peers because of discrimination. Moreover, their migrant parents are busy with work and generally fail to spend enough time caring for the children.
- Left-behind children's academic performance is worse than that of normal rural children. This shows that living with parents is vital for children academic performance. Even though there are other relatives who can take care of the left-behind children, such as grandparents, the role of is parenting is seriously missed since long-distance migration and long-lasting separation is involved in most cases.

4.2 Policy Implications

This paper shows that left-behind children have the most serious issue in terms of education compared to other groups of children. To solve it, there are multiple policies that the government can implement:

- The Chinese government should create a good employment environment to encourage migrant workers to return to their hometowns and start their own businesses. By doing so, the total number of left-behind children will be significantly reduced. Moreover, this policy can help develop the rural economy and reduce urban-rural income inequality.
- The Chinese government should help migrant workers who have motivations to bring their children to urban areas. Currently, the largest barrier is the difficulty for migrant children to enroll in good public school in urban areas due to the Hukou system. Thus, the government can either release the strict limitation on school enrollment or build schools with good felicities for the migrant children.
- For the children who have already moved to rural areas, it is import to make sure they are well adjusted to the new life. The government can motivate the residential committee to take responsibility by providing additional subsidies. NGOs can also play a role during the process.
- If the separation is unavoidable in some cases, the village committee, school teachers, parents, and relatives should coordinate with each other to ensure that left-behind children get enough care and support.

4.3 Limitations

Although I have applied the instrumental variable approach to deal with the endogeneity issue, there are still some other methods that I can take to improve the consistency of estimators. It would be more pervasive if I run an additional test to check the heteroscedasticity assumption. Another way method could be conducting robustness check with more control variables. For example, the city where the children migrate to can serve as a control variable. Previous research shows that it is less difficult to get enrolled in a local public school in small cities than in the big cities. Given the time limitation, I could not collect information about the additional control variables.

4.4 Future Research

One interesting question that this paper does not answer is whether the time of separation matters. It is possible to be empirically tested thanks to the panel study structure of this survey. Presumably, the earlier separation happened, and the longer the separation lasts, the stronger should be the negative impacts.

As the Chinese government has recognized the seriousness of left-behind children issues, some policies have been or are going to be implemented. This also provides chances for us to evaluate policies aiming at which factor are most effective empirically.

BIBLIOGRAPHY

China Family Panel Studies User's Manual. (2017) http://www.isss.pku.edu.cn/cfps/docs/20180928161838082277.pdf

Gao, Y., Li, L. P., Kim, J. H., Congdon, N., Lau, J., & Griffiths, S. (2010). The impact of parental migration on health status and health behaviours among left behind adolescent school children in China. *BMC public health*, *10*(1), 56.

Hugo, G. J. (1998). Migration as a survival strategy: the family dimension of migration.

Liang, Z., Guo, L., & Duan, C. C. (2008). Migration and the well-being of children in China. Yale China Health Journal, 5, 25–46.

Lu, Y. (2012). Education of children left behind in rural China. *Journal of Marriage and Family*, *74*(2), 328-341.

Jia, P., Du, Y., & Wang, M. (2017). Rural labor migration and poverty reduction in China. *China & World Economy*, *25*(6), 45-64.

Jia, Z., & Tian, W. (2010). Loneliness of left-behind children: a cross-sectional survey in a sample of rural China. *Child: care, health and development*, *36*(6), 812-817.

Wen, M., & Lin, D. (2012). Child development in rural China: Children left behind by their migrant parents and children of nonmigrant families. *Child development*, *83*(1), 120-136.

Morooka, H., & Liang, Z. (2009). International migration and the education of left-behind children in Fujian, China. *Asian and Pacific migration journal*, *18*(3), 345-370.

National Bureau of Statistics of China (NBS) (2018). *China Yearbook of Household Survey*, Beijing: China Statistics Press [in Chinese].

The Economist (2003). *Rich Man, Poor Man.* https://www.economist.com/asia/2003/09/25/rich-man-poor-man

The Economist (2015). *Little match children.* https://www.economist.com/briefing/2015/10/17/little-match-children

Wu, X., & Zhang, Z. (2015). Population migration and children's school enrollments in China, 1990–2005. *Social science research*, *53*, 177-190.

Zhao, F., & Yu, G. (2016). Parental migration and rural left-behind children's mental health in China: A meta-analysis based on mental health test. *Journal of Child and Family Studies*, *25*(12), 3462-3472.

Zhu, N. (2002). The impacts of income gaps on migration decisions in China. *China Economic Review*, *13*(2-3), 213-230.

Zhu, Y. (2003). The floating population's household strategies and the role of migration in China's regional development and integration. *International Journal of Population Geography*, *9*(6), 485-502.