To what extent does where people are born account for income and wealth inequality in China?

LIS Conference | Gdansk

Dongyuan Mu

Graduate School of Economics
University of Tokyo

September 28, 2023

Content

1. **Background**

2. **Methodology**

3. **Data and variables**

4. **Empirical results**

5. **Conclusion**

Background

Background

- China experienced a long and rapid development period since the Opening-up Reform, and inequality measured by GINI and Theil also launched from that time, there are several reputable discussing the causes of inequality in China. As representative studies, Xie and Zhou (2014) and Xie and Jin (2015) commonly regard the **residence** of households is the largest contributor to inequality of income and wealth, respectively.

- Besides the residential status, *hukou*, the unique identification system in China, is also regarded as an important driver for China's inequality.

Background: hukou system

Hukou system was carried out and prompted since 1950s. It experienced several revisions and help China's urbanization process in the past. Everyone's hukou consist of **registration location**[1] and **population type**[2]

There are two characteristics of this unique ID system in China:

1 Stringency in *hukou* conversion. One can converts their *hukou* into another province, but very difficult in specific case.

2 Valuable merits brought by owning a local *hukou*, especially in a developed city.

Given this background, one's ID can affect one's residence decision with no doubt, thus discuss the importance of residence situation without control of *hukou* can bias the results.

---

[1]Typically the born place.

[2]Population type contains "agriculture" and "non-agriculture", commonly called by "rural" and "urban" in short.

Purposes and framework

So, the purposes of current study are mainly two issues:

1 Does between-residence inequality and the importance of residence issue overestimated? And what is the reason?
2 Will the residence issue still be the largest contributor to explain the inequality when we take Chinese ID system into consider?

Methodology

Empirical strategies

To address these two questions above, I show evidence in two phases.

Phase 1   I decompose the residential inequality in this phase. Employing the Theil index, I
calculate the between-residence inequality and a **conditional**.

1. The former one can be a *gross effect* of residence because it contains effects caused by other factor.
2. The conditional between-residence inequality is hence regarded as a *direct effect*.
3. The difference between the former two is regarded as *indirect effect* caused by controlling these factors.

Then I use Shapley decomposition to allocate the *indirect effect* into factors being conditional on.

Phase 2   I make Shapley decomposition on total inequality in this phase. I show the results
contain and not contain the *hukou* factor and see what changes especially changes to
residence caused by adding *hukou*.

Intuition of conditional between-residence inequality
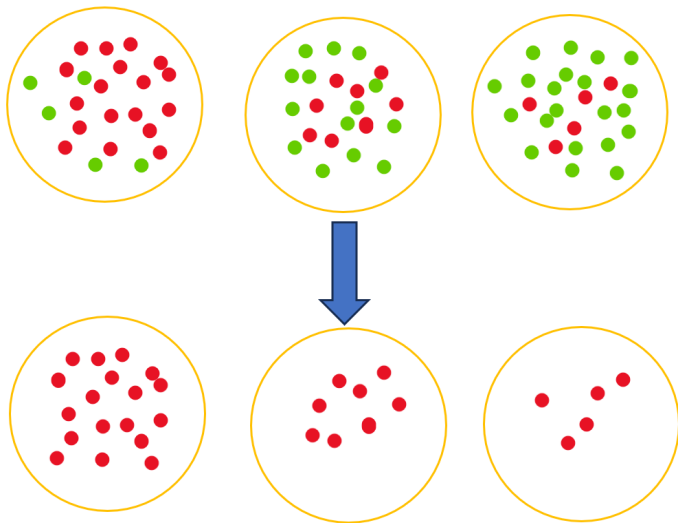
Data and variables

Data

To make the factors existing in every year and make sure the sample size, the present study employ the **CHFS** (Chinese household financial survey) data in 2017 and 2019.

- CHFS is conducted by China Southwestern University of Financial and Economics. This data contains a vast quantity of households' financial information compared with others. It also provide detailed including non-financial and financial information, and it is a survey that containing personal survey so we can know about a detailed situation of households.

I focus on income and wealth inequality, and the dependent variables are scale-modified with reference to the modified equivalent household scale standard provided by OECD.

Variable setting

Then I introduce other variables. Notice that conditional Theil index is based on dividing group, thus we need to transform variables into categorical form.

- *residence*: it is the combination of residential province and area type.
- *hukou*: it is the combination of *hukou* province and population type including agriculture, non-agriculture and uniform.
- *education*: it is a dummy variable that takes a value of 1 if household head graduates from a high school.
- *gender*: dummy variable to present household head's gender.
- *agegroup*: household head's age which is divided into 10-years interval from 17 to 75.
- *he*: dummy variable of whether this household has their own house.
- *business*: dummy variable for whether this household participant in business.

Summary statistics

2017	Mean	S.D.	2019	Mean	S.D.
income	48533.48	109078.52	income	45860.19	101829.29
gender	0.84	0.40	gender	0.76	0.43
education	0.37	0.48	education	0.35	0.48
age	52.73	12.30	age	54.04	11.99
he	0.88	0.32	he	0.91	0.29
business	0.15	0.36	business	0.09	0.29
Sample amount	34559		Sample amount	30417	

Empirical results

Empirical results of income: Phase 1

Table 2: Baseline results in 2017 and 2019

2017	Value	2019	Value
Residential inequality	0.1203	Residential inequality	0.1397
Direct effect	0.0597	Direct effect	0.0467
Indirect effect	0.0606	Indirect effect	0.0930
Decomposition on indirect effect			
gender	-0.0048	*gender*	-0.0038
education	0.0174	*education*	0.0240
agegroup	-0.0124	*agegroup*	-0.0072
hukou	**0.0695**	**hukou**	**0.0920**
he	-0.0050	*he*	-0.0063
business	-0.0041	*business*	-0.0058

Findings

- First of all, we can see that the majority of the gross residential inequality comes from indirect effect, and the direct effect only caused by residential status can only account for less than 50% in both year.

- Within the indirect effect, residential inequality mainly receive the indirect effects from **education** and **hukou**, and does not receive sizable effects from other factors. Meanwhile, *hukou* account for the majority of the indirect effect.

Empirical results of income: Phase 2

Table 3: Shapley decomposition of total inequality in 2017 and 2019

2017	value	2017	value	2019	value	2019	value
Total inequality		0.7413		Total inequality		0.7091	
explained	0.3337	explained	0.4308	explained	0.3232	explained	0.4103
residual	0.4075	*residual*	0.3104	*residual*	0.3860	*residual*	0.2988
Decomposition on *expained*				Decomposition on *expained*			
gender	0.0085	*gender*	0.0136	*gender*	0.0108	*gender*	0.0160
education	0.0797	*education*	0.0680	*education*	0.0836	*education*	0.0706
agegroup	0.0496	*agegroup*	0.0653	*agegroup*	0.0431	*agegroup*	0.0585
residence	0.1292	**residence**	0.0911	**residence**	0.1463	**residence**	0.0967
he	0.0109	*he*	0.0177	*he*	0.0103	*he*	0.0134
business	0.0557	*business*	0.0638	*business*	0.0290	*business*	0.0375
		hukou	0.1113			**hukou**	0.1176

Empirical results of wealth: Phase 1

I then make analysis on wealth issue in the same order to income's.

Table 4: decomposition on residential asset inequality

2017	Value	2019	Value
Residential inequality	0.37537	Residential inequality	0.39958
Direct effect	0.05062	Direct effect	0.07518
Indirect effect	0.32475	Indirect effect	0.32440
Decomposition on indirect effect			
gender	0.00132	*gender*	-0.00218
education	0.04515	*education*	0.06506
agegroup	-0.01011	*agegroup*	-0.03742
he	-0.01218	*he*	-0.01442
business	-0.00158	*business*	-0.00733
hukou	0.30078	**hukou**	0.32068

Empirical results of wealth: Phase 2

Table 5: Shapley decomposition of wealth inequality in 2017 and 2019

S	Value	S	Value	S	Value	S	Value
Total inequality		1.0288		Total inequality		1.2802	
explained	0.5824	*explained*	0.6762	*explained*	0.6995	*explained*	0.8536
residual	0.4464	*residual*	0.3526	*residual*	0.5807	*residual*	0.4265
Decomposition on *expained*				Decomposition on *expained*			
gender	0.0119	*gender*	0.0146	*gender*	0.0222	*gender*	0.0244
education	0.1031	*education*	0.0787	*education*	0.1292	*education*	0.1294
agegroup	0.0260	*agegroup*	0.0390	*agegroup*	0.0532	*agegroup*	0.0746
residence	0.3546	**residence**	0.2064	**residence**	0.4146	**residence**	0.2482
he	0.0524	*he*	0.0528	*he*	0.0591	*he*	0.0575
business	0.0344	*business*	0.0422	*business*	0.0213	*business*	0.0289
		hukou	0.2426			**hukou**	0.2907

Findings and compared with income

The results are similar but more significant in wealth case. That means:

- The residence's importance is **overestimated in a larger extent** in wealth than that in income.
- The residence status and *hukou* status have larger importance than that in income case.

This can be explained by the follows:

- Wealth can be stocked from generation to generation, by income and asset accumulation. Given the stringency of *hukou* conversion, the income gap can be accumulated and this will lead to a larger importance of *hukou* for wealth inequality.

Conclusion

Conclusions

We make decomposition on residential inequality and total inequality, and we have three main results:

- Current high residential inequality can be overestimated by ignoring *hukou* issues.
- *Hukou* is the largest contributor to total inequality in China.
- In term of wealth inequality, residential inequality is overestimated in a larger extent than that in income.

The relations between *hukou* and residence choice are widely discussed, while how to discuss them simultaneously in single framework is remained undone. Governors need to pay their attention to *hukou* with no doubt, the unbalanced regional development, welfare policies and treatments should be the true cause to these phenomena.

The End

Appendix: methodology summary

	Index decomposition	Blinder-Oaxaca decomposition	Machado-Mata decomposition
Merit	Non-parametric Decompose inequality index	Decompose average gap Counterfactual	Decompose distribution Counterfactual
Weakness	Only for categorical variables Only for hierarchy factors	Only decompose mean gap Parametric	Fail to capture the "imporatance" Parametric
Shapley value and Theil index (Present study)			
Merit	Non-parametric Decompose inequality Available for factors in any relations		
Weakness	Only for categorical variables Only available for positive objectives Fail to capture the causality		