Intro
○○

Method
○○○○

Data
○

Results
○

Assessment
○

High-speed Rail and Inter-provincial Inequality in Payoffs to Human Capital

Yiqing Zheng

MACSS Project Proposal
April 15, 2020

Motivation

- Research question: How does high-speed rail development influence regional inequality in payoffs to human capital?

- A large wave of migration is a significant phenomenon in China, which indicates that payoffs to human capital are not identical across provinces.

- High-speed rail network witnesses unprecedented development in China in the past few years.

- It might be interesting to investigate whether the development in infrastructure influences regional inequality.

Intro
○●
Method
○○○○
Data
○
Results
○
Assessment
○

Introduction

Introduction

- Transportation development benefits (regional) economic growth: Chong, Chen, & Qin, 2019; Fujita & Thisse, 2013; Duranton & Turner, 2012; Lin, 2017

- Transportation development reduces regional economic disparities: Chen & Haynes, 2017
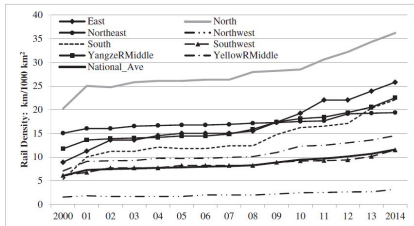
Figure: Change of Chinese railway network density by region (from Lin, 2017)

Intro
○○

Method
●○○○

Data
○

Results
○

Assessment
○

Measurement of inequality in payoffs to human capital

- Zax (2019) introduces a counter-factual strategy to measure regional inequalities in valuations in human capital.

Location cost – first regression

$$y_{ji} = \beta_{j0} + \beta_{j1}\,age_{ji} + \beta_{j2}\,education_{ji} + \epsilon_{ji}$$

$$\hat{y}_i = \hat{\beta_{j0}} + \hat{\beta_{j1}}\,age_{ji} + \hat{\beta_{j2}}\,education_{ji}$$

$$y_i^* = \max_{k \in K} \hat{\beta_{k0}} + \hat{\beta_{k1}}\,age_{ji} + \hat{\beta_{k2}}\,education_{ji}$$

$$LocationCost_i = y_i^* - \hat{y}_i \in (-\infty, 0]$$

where j indexes province, i indexes individual, K means the set of all provinces.

Intro
○○

Method
○●○○

Data
○

Results
○

Assessment
○

Measurement of inequality in payoffs to human capital

- Run regression inside each province and obtain individual predicted earnings inside her home province.
- Put individual characteristics into regression results of other provinces and get maximum predicted earnings.
- Compare the maximum predicted earnings and home province predicted earnings to get location cost.
- Location cost reflects the differences in payoffs to human capital.

Intro
○○

Method
○○●○

Data
○

Results
○

Assessment
○

Measurement of inequality in payoffs to human capital

- Zax (2019) just uses OLS to conduct counter-factual strategy. The accuracy to predict individual earning level is critical to the conclusion.

Computational Part

I plan to use multiple computational regression models to get most accurate estimation strategy according to their MSE. Potential models:

- Lasso regression
- Ridge regression
- Random forest

High-speed rail and location cost

- Run regression of individual location cost and whether the home province has HSR travelling to the province where she gets maximum earnings.

second regression

$$LocationCost_i = \beta_0 + \beta_1 D_i + \sum_{s=1}^{S} \beta_s x_{si} + \epsilon_i$$

where D_i is a dummy indicating whether she can travels to the province where she obtains maximum predicted earnings by HSR and x_{si} are covariates.

Still use multiple strategies to get most accurate estimation.

Intro
oo

Method
oooo

Data
●

Results
o

Assessment
o

Data

- The China Household Income Project (CHIP) provides detailed information on individual address, earning level, and other characteristics.

- HSR data: Chinese Research Data Services Platform provides detailed data on HSR from 2002. The data include the opening time, passing stations, and departure time of each HSR line.

Intro
oo

Method
oooo

Data
o

Results
•

Assessment
o

Results

- As for the location cost, I expect average location cost in Guangdong province to be lowest according to existing literature while location cost for less developed region such as Gansu would be higher.

- As for the relationship between location cost and HSR, I expect the existence of HSR to reduce location cost.

- The logic is that for those location cost is not zero and have suitable HSR to travel to the province where they can earn maximum income, the reason for them to stay is that the location cost is not that high. On the other hand, for those who lack access to the right province, the location cost is higher for them.

Intro
oo

Method
oooo

Data
o

Results
o

Assessment
•

Contribution and limits

- Contribution:
  (1) Use computational methods to get more accurate estimation of location cost.
  (2) Construct a relationship between HSR and location cost.

- Limits:
  (1) There might be endogeneity in the second regression that needs more computational method to address.
  (2) I plan to use IV or spatial model to further investigate this issue.