Residential Segregation in Chinese Cities

Junbiao Chen (NYU), Xiaoruo Feng (PKU), Yizhen Gu (PKU),
Changcheng Kan (Baidu), Rongxi Quan (Tsinghua), Ben Zou (Purdue)

DSE HKU, Dec 2025

Motivation

- Residential segregation by income, race, birthplace, etc., has been widely observed in the developed world.
- A lot of attention given its negative consequences on
  - educational attainment (e.g., Cutler et al., 1999; Bohlmark & Willen, 2020)
  - labor market outcomes (e.g., Chetty & Hendren, 2018; Bohlmark & Willen, 2020)
  - intergenerational mobility (Borjas, 1995)
  - health outcomes (Collins & Williams, 1999; Almond et al., 2006)
  - access to public goods (Asher et al., 2024; Harari, 2024)
- No systematic evidence on segregation levels in China due to the lack of nationwide data at fine spatial levels.
  - Administrative data: not publicly available
  - Survey data: small sample size

This Paper

- The first detailed examination of residential segregation in Chinese cities.
  - China's economic transition provides a unique setting
  - how housing market and agglomeration economy has contributed to segregation in a short time period.
- Documents the pattern and the trend of segregation using
  - a novel source of grid-level population by education/income data derived from frequent information captured by smartphones
  - home mortgage data from one large city during 2005-2014
- Explores predictors of residential segregation: housing market, agglomeration, etc.
  - using cross-city variation
  - estimating a dynamic sorting model within a city

Summary of Findings

- Chinese cities display modest level of residential segregation.
  - The average dissimilarity index is 0.174 by education and 0.139 by income, respectively.
  - much lower than international counterparts
- Sorting on the housing market led to an increase in segregation.
- Housing supply, service industry GDP ratio and agglomeration level, and other attributes explain almost half of cross-city segregation variation.
- Estimation results of the dynamic model:
  - Low-edu people have a higher housing price elasticity.
  - Both high- and low-edu people value access to job opportunities.
  - Consistent with the steepening housing and education gradients

Related Literature

- Patterns of residential segregation in cities globally
  - The literature focuses on developed countries: data availability
  - China: case studies with census data at the township level (~240 km^2) or the village/residents' committee level (~96 km^2)
  - This paper: measures segregation using granular smartphone data in a developing country
- Causes of segregation
  - Historical changes, racial preferences, income and education levels, etc.
  - This paper: a transitional economy featuring increasing income inequality, rising agglomeration, and housing market sorting
- Dynamic Spatial Sorting
  - The literature focuses on various residential amenities
  - This paper: adds commuting and models access to job opportunities as amenities

Background: Pre-1978 Reform

- low degree of segregation
  - high level of income parity
  - public housing system (*Danwei*): $3.6m^2$ per capita living space in 1978
  - low degree of agglomeration
    - only 1/4 GDP from state-controlled service industries
    - basic services were integrated with the *Danwei* system

Background: Post-Reform

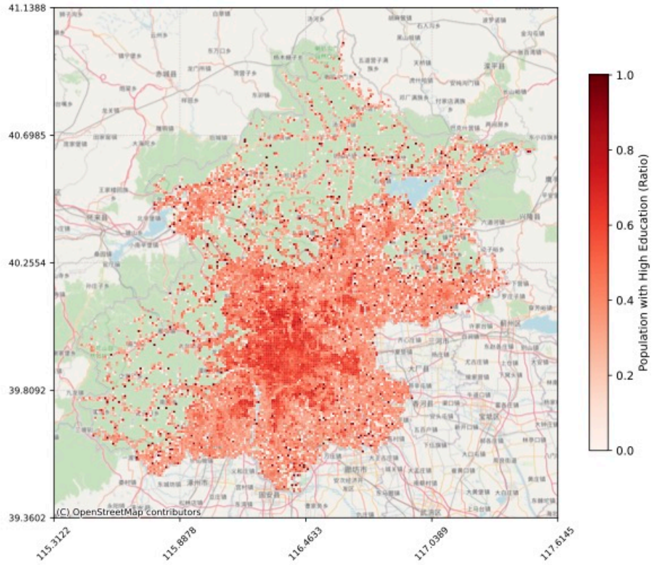
- Income inequality has risen
  - The Gini coefficient almost doubled from ~0.30 in 1980 to 0.55 in 2012 (Xie & Zhou, 2014).
  - The top 10 percent income share rose from 27 percent to 41 percent between 1978 and 2015 (Piketty et al., 2019).
- Tiebout sorting enabled by market-oriented housing and land reforms
  - residential building stock increased by nearly 150% during 1980-2020
- Agglomeration economy
  - rapid expansion in the service sector: GDP ratio from 24.7% in 1978 to 56.7% in 2024
  - rising spatial concentration of service jobs: ratio of newly registered service firms around city centers from 67.3% in 1988 to 86.1% in 2018

A Corner in Shanghai

Grid-Level Population Data

- Baidu Maps is a leading provider of location services.
  - embedded in many mobile apps, beyond Baidu products
  - capture locations when using the app or in background (with permission)
- For each mobile device, Baidu infer
  - home: typical location at night
  - gender, education, income, etc., applying a multi-model approach featuring the fusion algorithm and data on users' search engines usage features, travel routes and app usage habits
- Aggregate to 1km-by-1km grids with population by education/income in Nov 2019
  - education: high school and below, some college and above
  - monthly income: <2500, $>=2500$
  - limited access: codes sent to an engineer at Baidu Maps

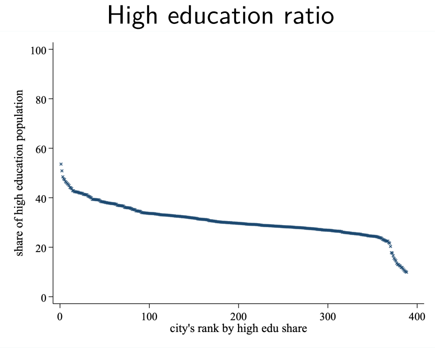
Example Data: Grid Level High Education Ratio in Beijing

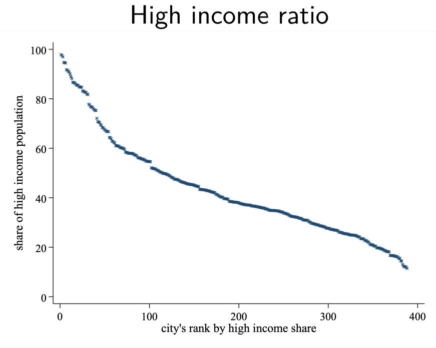


(1km-by-1km Grid)

Education Segregation as the Baseline

Prefecture level average

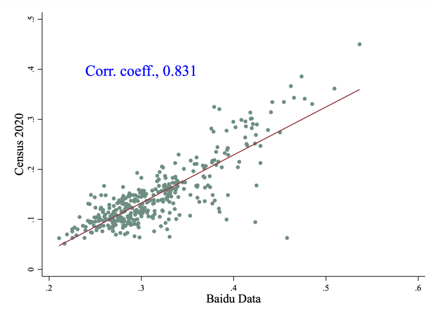




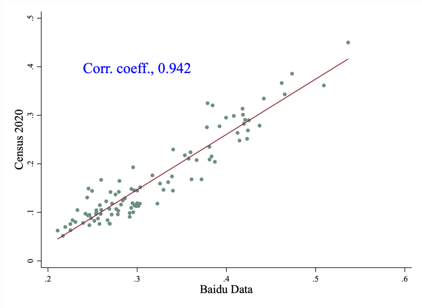
- relatively consistent cutoffs across prefectures
- more validation can be done

Data Validation: Prefecture-Level High Education Ratio

All cities



Top 100 largest cities



(Validations)

Home Mortgage Data

- Administrative records of home mortgages in a large city issued by an anonymous mortgage lender between 2000 and 2014.
- For each mortgage record, detailed information on
  - buyer's demographic characteristics (education level, gender, marital status),
  - monthly salary,
  - addresses of their current and new homes,
  - the total price and the size of the new home.

	N	Mean	SD
Total housing price (thousand yuan)	332,049	748.54	587.88
Individual income (yuan)	332,049	6,557.22	194,581.52
Share with college degree	332,049	0.72	0.45

City Features

- Housing market: GHS building volume data
- Agglomeration economy: firm level data and statistical yearbooks
- Demographic features: 2020 Census
- Topography features: OSM data (Topography)

	N	Mean	SD	Min.	Max.
Log change of residential building vol.	336	16.36	1.04	11.41	18.43
GDP ratio of service industry	336	0.49	0.08	0.28	0.84
Spatial Gini of service industry	336	0.85	0.04	0.49	0.99
Log population	337	14.88	0.97	7.75	17.28
Grid-like measure	337	0.12	0.05	0.06	0.40
Latitude	337	33.05	6.91	16.85	51.99

Measure of Residential Segregation

- The index of dissimilarity: most widely used
  - $D_j = \frac{1}{2} * \sum_{i=1}^{n} |(x_{ij}/X_j) - (y_{ij}/Y_j)|$
  - x_{ij} and y_{ij}, the size of groups x and y in unit i of city j,
  - X_j and Y_j, the size of groups x and y in city j.
- Other indices are highly correlated with the dissimilarity index. (More indices)

	Education Segregation			Income Segregation		
	D	D_{dc}	A	D	D_{dc}	A
D	1			1		
D_{dc}	0.932***	1		0.993***	1	
A	0.906***	0.848***	1	0.954***	0.932***	1

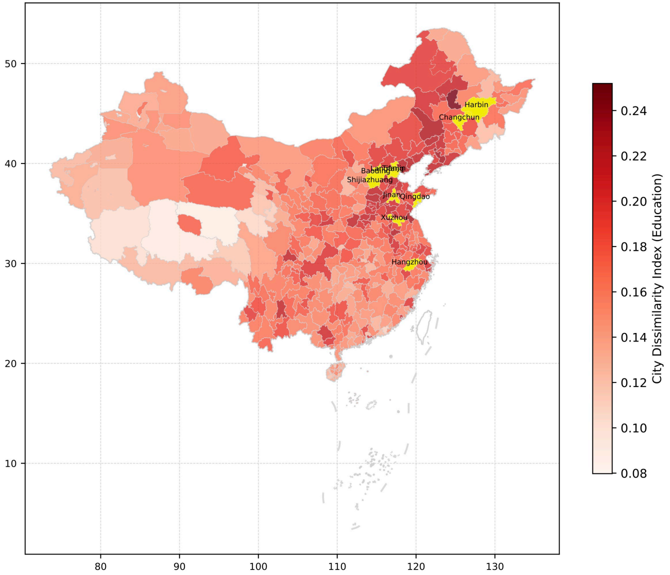
Patterns of Residential Segregation in 2019

	Education-based Segregation				Income-based Segregation			
City Definition	Mean	Std	Min	Max	Mean	Std	Min	Max
Prefecture	0.174	0.028	0.087	0.252	0.139	0.040	0.075	0.376
Districts	0.162	0.030	0.065	0.245	0.423	0.045	0.280	0.566
MA	0.162	0.033	0.061	0.250	0.418	0.043	0.242	0.582

(Alternative Data) (Commuting MA) (Spatial Inequality)

- International comparison
  - U.S. Cities
    - education segregation: [0.139, 0.441], avg. 0.288 (Florida & Mellander, 2018)
    - income segregation: [0.278, 0.463] (Cutler et al., 1999)
  - European Cities, income segregation
    - [0.22, 0.27] in Northern Europe (Helsinki, Oslo),
    - [0.39, 0.46] in Southern Europe (Madrid, Milan) (Tammaru et al., 2020)
    - 0.463 in U.K. (Nottingham) (Cauvain & Rae, 2022)

Prefecture-Level Education Segregation Map



(Income Segregation)

Most Segregated Cities

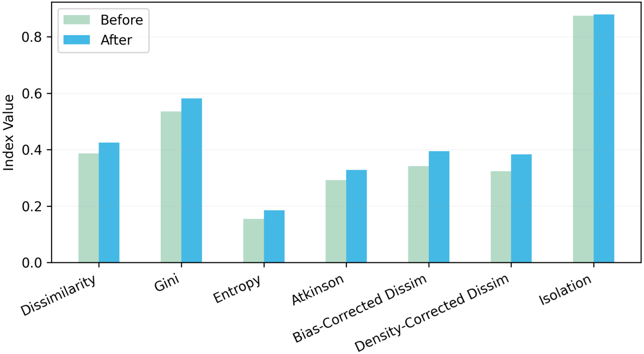
Table: Segregation Ranking of the 100 Largest Cities

Rank	City	Index	Rank (full sample)
1	Shijiazhuang	0.252	1
2	Harbin	0.248	2
3	Hangzhou	0.242	4
4	Xuzhou	0.239	5
5	Jinan	0.238	6
6	Langfang	0.235	7
7	Baoding	0.233	8
8	Tianjin	0.233	9
9	Qingdao	0.229	12
10	Changchun	0.227	14

Note: We rank the most segregated cities based on the dissimilarity index. The fourth column reports the rank of the index among all Chinese cities.

Trend of Segregation: Example of One City

- Calculate segregation indices based on buyers' current and new location, respectively: about 10% increase
  - assuming each individual has only one residential change during the period.
  - using mortgage records in which new house was bought after 2010 and the borrower's income is available (~100,000 purchases). (Validations)
  - aggregating to 1*km* grids, and keeping only grids that have residents in both data.

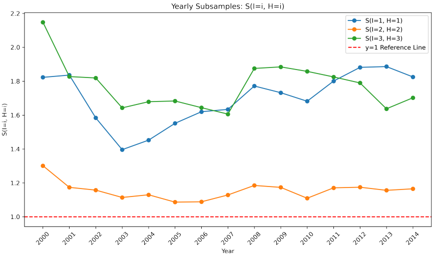


Sorting Contributes to Segregation

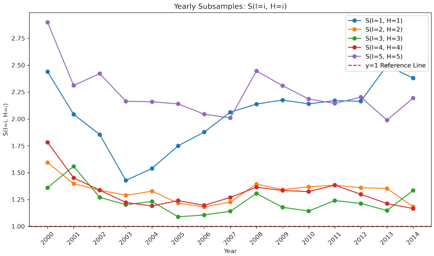
- Calculate buyer-housing sorting parameter for each year: $s(I_i, H_j) = \frac{P(I=i, H=j)}{P(I=i)P(H=j)}$
  - I, the income level of the buyer,
  - H, the price level of housing the buyer purchased,
  - $i, j = 1, ..., k$, the category.
- Show the values of sorting parameters on the diagonal.
  - buyers and housings at the same level, that is, $i = j$.
- Positive assortative matching means that
  - Buyers with level i purchase a house with level j more frequently than what would be expected under a random purchasing pattern.

Sorting Parameter Values By Year

Three groups



Five groups

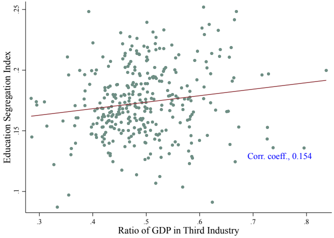


(Bi-year and Tri-year)

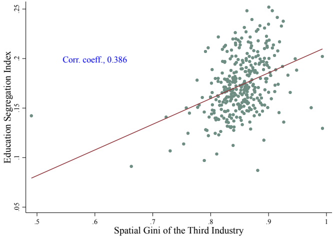
- The values are always greater than one, in particular for high-income and low-income buyers.

Predictors of Segregation Using Cross-City Variation

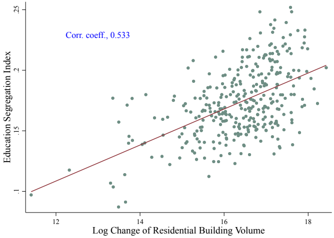
Service Ind GDP Ratio



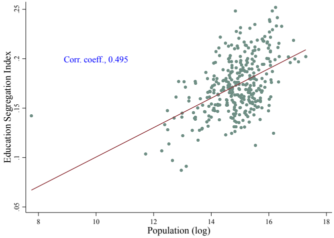
Spatial Gini Index



Residential Volume Change



Population



Put Together

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Education Dissimilarity Index				
Log population	0.015***						0.011***	0.010
	(0.002)						(0.003)	(0.006)
Grid-like measure		0.201***					0.158***	0.142***
		(0.044)					(0.029)	(0.038)
Latitude			0.001***				0.001***	0.003***
			(0.000)				(0.000)	(0.001)
Ratio of GDP in third industry				0.054**			0.018	0.055
				(0.022)			(0.017)	(0.040)
Spatial gini of third industry					0.259***		0.147***	0.217***
					(0.051)		(0.043)	(0.062)
Log change of residential building volume						0.015***	0.004*	-0.006
						(0.001)	(0.002)	(0.004)
Observations	337	337	337	336	336	336	335	99
R-squared	0.245	0.108	0.045	0.024	0.149	0.284	0.455	0.642

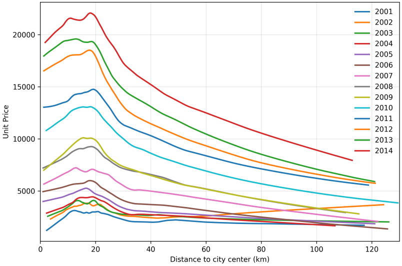
Standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

- Newly added residential floor area, service industry GDP ratio, agglomeration level, population and other city characteristics explain almost half of cross-prefecture segregation variation.

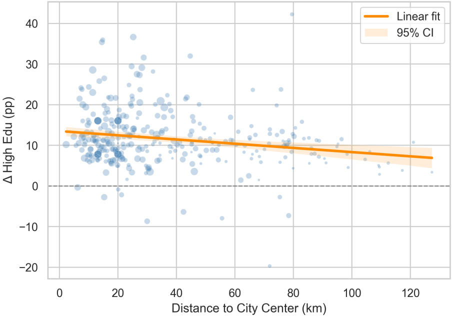
Dynamic Sorting Model: Motivating Evidence

- Steepening housing price gradients and education gradient

Housing Price Gradients



Change in High-Edu Ratio



Dynamic Sorting Model: Setup -1

- Environment:
  - dynamic residential location choice: k individual's education type
  - j discrete residential locations
  - workplace locations w are simplified into "downtown" and "suburban area"
  - each location endowed with exogenous amenities which change over time
- Preference:

$$u_t^k(j \mid j_{i,t-1}, w_i) = \theta_{jt} + \gamma^k \log r_{jt} + \varepsilon^k \log \kappa_{j,w_i,t} - \mathfrak{C}^k(j, j_{i,t-1}) \tag{1}$$

  - $j_{i,t-1}$ is the previous residential location, w_i is the fixed workplace
  - θ_{jt} is the residential place amenity
  - r_{jt} is the housing price
  - $\kappa_{j,w_i,t}$ is the commuting cost
  - $\mathfrak{C}^k(j, j_{i,t-1})$ is the moving cost of leaving the previous location

Dynamic Sorting Model: Setup -2

- Value Function:

$$V_t^k(\mathbf{x}_{it}, \boldsymbol{\omega}_t, \boldsymbol{\xi}_{it}) = \max_{j \in \{0,1,\dots,J\}} \Big\{ u_t^k(j \mid \mathbf{x}_{it}, \boldsymbol{\omega}_t) + \xi_{ijt} + \beta\, \mathbb{E}\Big[V_{t+1}^k(\mathbf{x}_{i,t+1}, \boldsymbol{\omega}_{t+1}, \boldsymbol{\xi}_{i,t+1}) \,\Big|\, j, \mathbf{x}_{it}, \boldsymbol{\omega}_t \Big] \Big\}. \tag{2}$$

  - $x_{it} = [j_{t-1}, w_i]$ denotes the individual state
  - $\boldsymbol{\omega}_t = \left\{ (r_{jt})_{j=1}^J, (\theta_{jt})_{j=1}^J, (\kappa_{j,w,t})_{j=1,\dots,J;\ w=1,\dots,J} \right\}$ denotes the vector of aggregate states (housing prices, amenities and commuting costs)
  - $\xi_{it} = (\xi_{1it}, \xi_{2it}, \dots, \xi_{Jit})$ is a vector of idiosyncratic preference shocks across locations
  - β is the discount factor.

- Conditional Choice Probability:

$$P_t^k(j \mid j_{t-1}, w_i) = \frac{\exp\left(\tilde{V}_t^k(j \mid j_{t-1}, w_i) \right)}{\sum_{j'} \exp\left(\tilde{V}_t^k(j' \mid j_{t-1}, w_i) \right)}. \tag{3}$$

Dynamic Sorting Model: Setup -3

- Housing developer chooses floor space h_j (in log) subject to the FAR constraint in each location j

$$\begin{aligned} \max_{h_j} \quad & r_j h_j - C_j h_j - \kappa h_j^2 \\ \text{subject to:} \quad & h_j \leq \bar{h}_j + \nu_{j|R} \end{aligned} \tag{4}$$

  - $\nu_{j|R}$: an idiosyncratic shock to the constraint in ring-road R cluster j
  - C_j is composed of a deterministic component based on observables (L_j, A_j) and an idiosyncratic shock ξ_j:

$$C_j = \beta_0 + \beta_1 L_j + \beta_2 A_j + \xi_j \tag{5}$$

  - FAR shocks and costs shocks follow:

$$\xi_j \overset{iid}{\sim} \mathcal{N}(0, \sigma_\xi^2) \tag{6}$$

$$\nu_{j|R} \overset{iid}{\sim} \mathcal{N}(0, \sigma_R^2) \tag{7}$$

Dynamic Sorting Model: Estimation -1

- Following Davis et al. (2024), we estimate a stationary dynamic location choice model with a Hotz-Miller type demand system.

- Under **stationary equilibrium** we treat all aggregate states as time-invariant and denote them by $(r_j, \theta_j, \kappa_{j,w})$.
  - Flow utility:

$$u_{j,w}^k = \underbrace{\theta_j + \gamma^k \log r_j + \varepsilon^k \log(\kappa_{j,w})}_{\equiv \delta_{j,w}^k} - \mathfrak{C}_{j,j_{t-1}} \tag{8}$$

  - Moving Cost:

$$\mathfrak{C}_{j,j_{t-1}} = \begin{cases} b^k \log\left[\text{dist}(j, j_{t-1})\right] & j \neq j_{t-1}, \\ 0 & j = j_{t-1}. \end{cases} \tag{9}$$

Dynamic Sorting Model: Estimation -2

- For each type k, we construct empirical CCPs by pooling all adjacent years:

$$\widehat{P}^k(j \mid j_{t-1}, w) = \frac{\sum_{i,t} 1\{j_{i,t-1} = j_{t-1},\, w_i = w,\, j_{it} = j\}}{\sum_{i,t} 1\{j_{i,t-1} = j_{t-1},\, w_i = w\}}. \tag{10}$$

- Define markets
  - given type k, each unique pair $\text{mkt} := (j_{t-1}, w)$ defines a market
  - within that market, all J possible destinations j are products
  - market shares for each product j is :

$$s^k_{j,\text{mkt}} = \widehat{P}^k(j \mid j_{t-1}, w).$$

- These stationary transition probabilities summarize the dynamic behavior of individuals without explicitly solving a Bellman equation.

Dynamic Sorting Model: Estimation -3

- GMM Estimation of Moving Costs via Hotz-Miller
  - Under logit structure, the CCP satisfy:

$$P^k_{j,\text{mkt}} = \frac{\exp\left(\tilde{V}^k(j \mid j_{t-1}, w)\right)}{\sum_l \exp\left(\tilde{V}^k(l \mid j_{t-1}, w)\right)}, \tag{11}$$

  - Apply Hotz-Miller inversion, and use flow utility specification

$$\delta^k_{j,w} - \mathfrak{C}^k(j, j_{t-1}) - \delta^k_{j',w} + \mathfrak{C}^k(j', j_{t-1}) = \log \frac{s^k_{j,\text{mkt}}}{s^k_{j',\text{mkt}}}, \tag{12}$$

- Given the moving-cost function $\mathfrak{C}^k(j, j_{t-1})$, we estimate the moving-cost parameter b^k and the set of mean utilities $\{\delta^k_{j,w}\}$ by minimizing a GMM objective function:

$$\min_{b^k, \{\delta^k_{j,w}\}} \quad \frac{1}{M} \frac{1}{J(J-1)} \tag{13}$$

$$\sum_{\text{mkt}} \sum_{j \neq j'} \left(\log \frac{s^k_{j,\text{mkt}}}{s^k_{j',\text{mkt}}} - \left[\delta^k_{j,w} - \mathfrak{C}^k(j, j_{t-1}) - \delta^k_{j',w} + \mathfrak{C}^k(j', j_{t-1}) \right] \right)^2, \tag{14}$$

Dynamic Sorting Model: Estimation -4

- Preference Parameters of Demand
  - Estimate equation

$$\widehat{\delta}_{j,w}^{k} = \alpha^{k} + \gamma^{k} \log r_j + \varepsilon^{k} \log \kappa_{j,w} + \mathbf{A}_j' \eta^{k} + \xi_{j,w}^{k}, \tag{15}$$

    - $\mathbf{A}_j$: amenity characteristics including the number of local employment, restaurants, public facilities and schools
  - Gandhi-Houde Instrumental Variables

$$Z_j^{(\ell)} = \frac{1}{|\mathcal{N}(j)|} \sum_{j' \in \mathcal{N}(j)} \left(O_{j'}^{(\ell)} - O_j^{(\ell)} \right)^2 .$$

    - $\mathcal{N}(j)$: all other town-cluster locations within the city
    - $O_j^{(\ell)}$: housing characteristics (average building age and plot ratio) in location j
    - captures the local differences in housing characteristics around location j

Dynamic Sorting Model: Estimation -5

- Supply Estimation (Simplified Version)
  - Estimate the log-linear relationship between quantity and price

$$\log q_j = \alpha + \lambda \log p_j + \mathbf{X}_j' \beta + \varepsilon_j, \tag{16}$$

- λ : supply elasticity, which is $\frac{1}{\kappa}$ in the marginal cost function.
- $\mathbf{X}_j$: controls, include average allowed floor-area ratio (FAR), distance to the city center, and the composition of parcel types (e.g., affordable housing share)

(MLE Supply Estimation)

Dynamic Sorting Model: Data Preparation

- Demand Estimation Setup
  - Data source: home mortgage records of a large city, 2006–2014.
  - Choice set: 64 town-cluster units
  - Workplace: downtown vs. suburban
  - Housing prices: average unit housing price from mortgage record during 2009–2011.
  - Amenities: 2010 Point-of-Interest (POI) data, 2008 Economic Census
  - Commuting cost: 2015 household travel survey.

- Supply Estimation Setup
  - Data source: land transactions from 2000 to 2019.
  - Aggregation Units: 64 town clusters
  - Variables: For each cluster j, we compute:
    - total residential land supplied (hectares), q_j
    - average and median price per hectare, p_j^{avg} and p_j^{med}
    - average planned FAR, FAR_j; average distance to the CBD, dist_j
    - parcel-type composition ratios: residential share, affordable-housing share, etc.

Dynamic Sorting Model: Results - 1

- Demand Estimation Results

Table: GMM Estimation of Moving Costs

	High Education	Low Education
Moving-cost b^k	2.389	2.265

	High Education		Low Education	
	OLS	IV (robust)	OLS	IV (robust)
Housing price (log)	-1.280***	-3.252***	-1.983***	-4.314***
Commuting time (log)	-0.731	-2.170***	-0.361	-2.060***
Employees (log)	0.189	0.376**	0.017	0.238*
Retail POI	-0.001	0.000	-0.001	0.000
Public facilities POI	-0.003***	-0.002**	-0.001*	0.000
Schools POI	0.033***	0.028***	0.020***	0.014***
Constant	12.158***	34.387***	19.014***	45.279***
Observations	128	128	128	128
First-stage F	—	12.108	—	12.998

Notes: All regressions use robust standard errors. The observations are at the year-township-cluster level. Instruments follow Gandhi–Houde (2019). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Dynamic Sorting Model: Illustrative Example -1

- An Illustrative Example:
  - 3 residential locations, 2 worker types (high vs. low education).
  - **Housing demand (logit share):**

$$s_{jt}^k = \frac{\exp(\beta^k r_{jt})}{\sum_{j'=1}^J \exp(\beta^k r_{j't})},$$

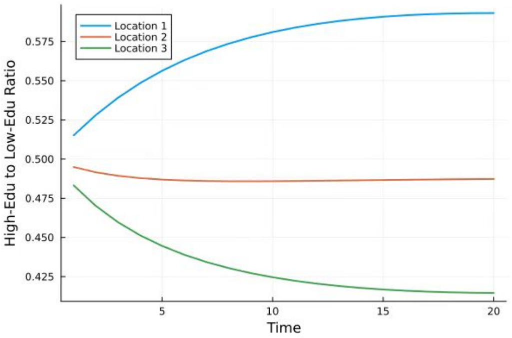
  - s_{jt}^k: share of group k in location j at t;
  - β^k: elasticity (low vs. high education);
  - r_{jt}: housing price in location j.
  - **Price dynamics:**

$$\log(r_{j,t+1}) = \gamma \log(r_{jt}) + \eta \left(\frac{N_{jt}^{\text{low}} + N_{jt}^{\text{high}}}{T_j} \right),$$

  - γ: persistence; η: sensitivity to demand
  - $N_{jt}^{\text{low}}, N_{jt}^{\text{high}}$: populations by education
  - T_j: land supply

Dynamic Sorting Model: Illustrative Example -2

- An Illustrative Example:
  - 3 residential locations, 2 worker types (high vs low education)
  - measure segregation by the ratio of high- to low-education residents
  - $\beta^{\text{low}} = -0.01$, $\beta^{\text{high}} = -0.002$, $\gamma = 0.9$, $\eta = 0.05$
  - Initial conditions: prices $r_0 = [10, 5, 2]$, populations $N^{\text{low}} = [1000, 800, 600]$, $N^{\text{high}} = [500, 400, 300]$, land supplies $T = [1000, 800, 600]$.



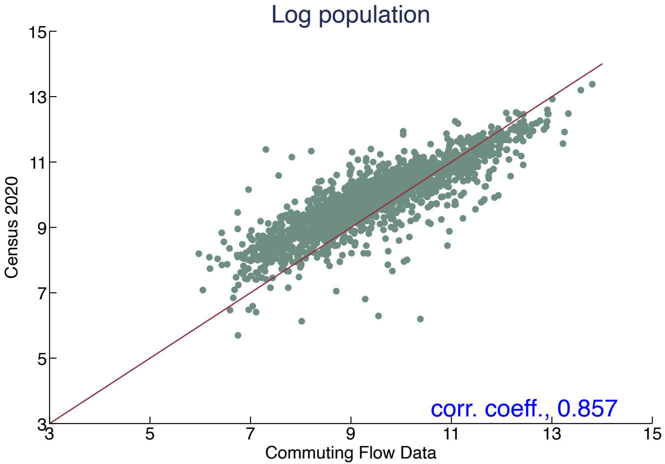
Conclusions and Next Step

- Conclusions
  - modest segregation levels in Chinese cities, but on the increase
  - sorting on the housing market and the spatial concentration of service jobs are the major contributors to segregation
- Next step
  - construct city-level income (online consumption) inequality
  - Complete the dynamic sorting model (e.g., supply side)
  - Counterfactual outcomes under alternative land supply, and transportation infrastructure development

Examples of the $1km$-by-$1km$ Grid

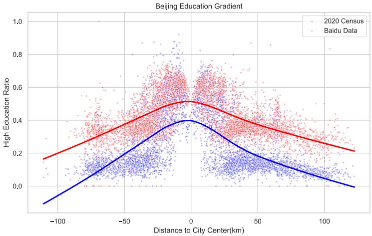
(Back)

Data Validation: Township-Level Population Count

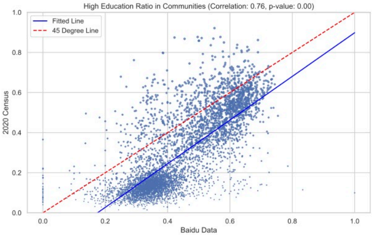


Data Validation: Residents' Committee-Level, Beijing

Education Gradient

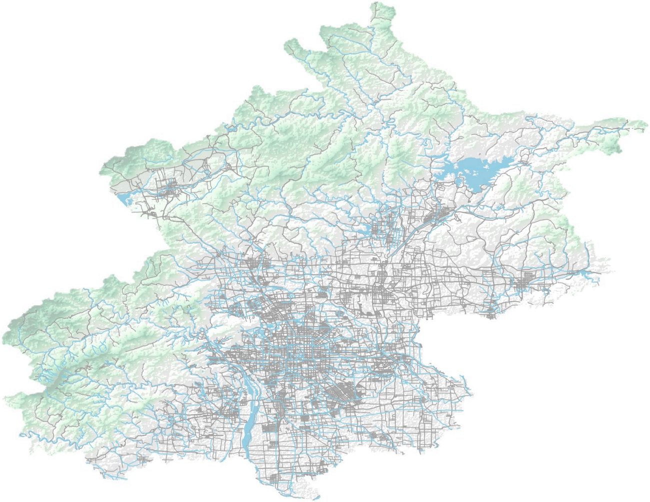


Scatterplot

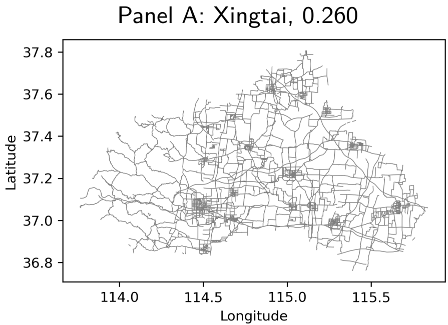


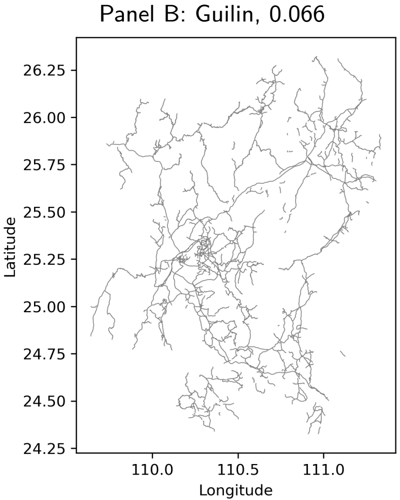
(Back)

Topography



Grid-like Road Network





(Back)

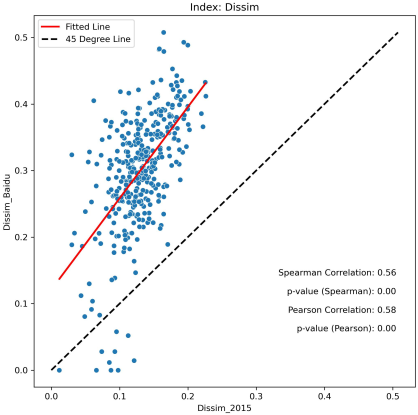
More Segregation Indices

Dimension	Indices	Notes	Range
Unevenness	Dissimilarity	· The most commonly used segregation index. · Problems: fail to satisfy the transfer principle; value is biased when the unit size is small.	[0,1]
	Gini	· Satisfy all 4 principles discussed in James & Taeuber(1985)	[0,1]
	Entropy	· Problem: fail to satisfy the compositional invariance principle.	[0,1]
	Atkinson	· Satisfy all 4 principles and is better than Gini because it forces the researcher to confront the problem of crossing Loronz curve · Problem: value depends on parameter b.	[0,1]
	Bias Corrected Dissimilarity	· Adjust the bias in dissimilarity index when unit sizes and/or minority proportions are small. · Not work well for small unit sizes combined with small values of Dpop	[0,1]
	Density Corrected Dissimilarity	· Compared with BCD, it reduces enough of the bias, provided unit sizes are not too small.	[0,1]
	Spatial Dissimilarity	· Take spatial factors into account, adding the effects arising from the interaction between units into the index.	/
Exposure	Isolation	· Negatively related to the exposure.	[0,1]
Concentration	Relative Concentration	· Consider both the concentration of majority group and minority group.	[-1,1]
Centralization	Absolute Centralization		[-1,1]
Clustering	Spatial Proximity	· Euclidean distance in every 2 units are calculated	[1,+finitely)

(Back)

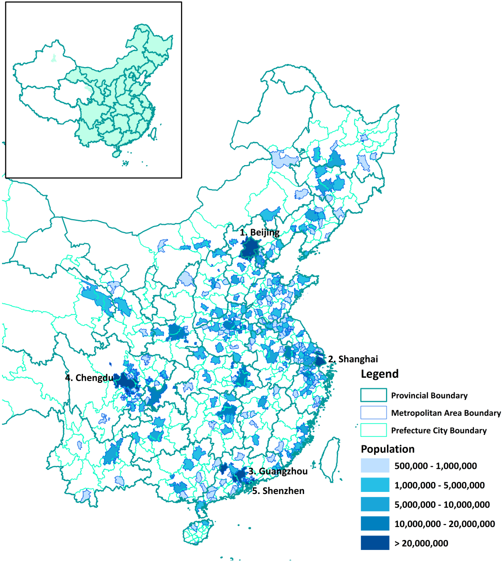
Result Validation

- Prefecture-level Segregation using township level data
  - Baidu 2019 vs. mini-Census 2015



(Back)

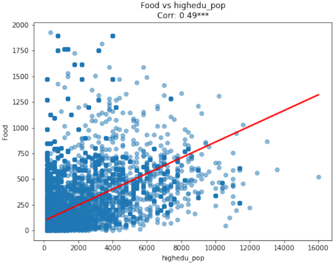
Commuting-Based MA



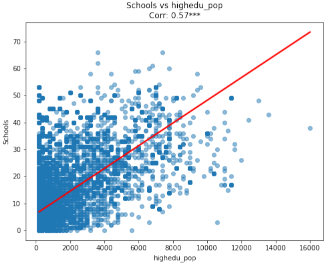
(Back)

Spatial Inequality

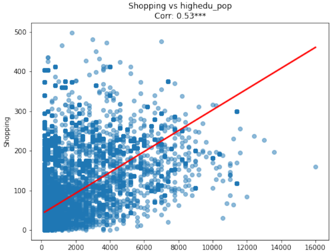
Dining



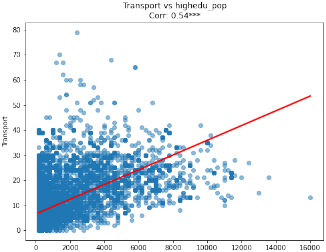
School



Shopping



Transit

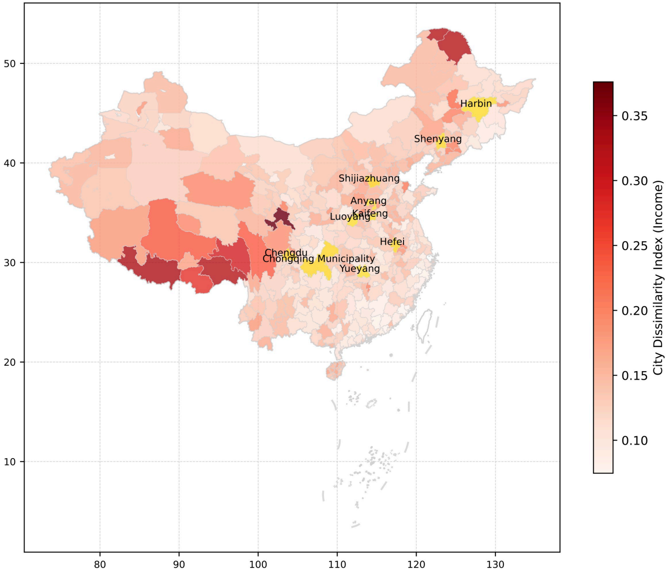


Spatial Inequality

Number of POI in Category	(1) Shopping	(2) Schools	(3) Healthcare	(4) Transport	(5) Life Services	(6) Sports	(7) Leisure	(8) Food
Share of High Education Population	178.447***	28.973***	62.797***	21.456***	2.992***	67.724***	2.765***	514.915***
	(2.510)	(0.395)	(0.862)	(0.297)	(0.053)	(1.020)	(0.063)	(7.604)
Observations	30,057	30,057	30,057	30,057	30,057	30,057	30,057	30,057
R-squared	0.177	0.212	0.206	0.206	0.112	0.162	0.050	0.151
Mean dep.var	59.113	9.164	18.581	8.499	0.846	18.127	0.711	150.697

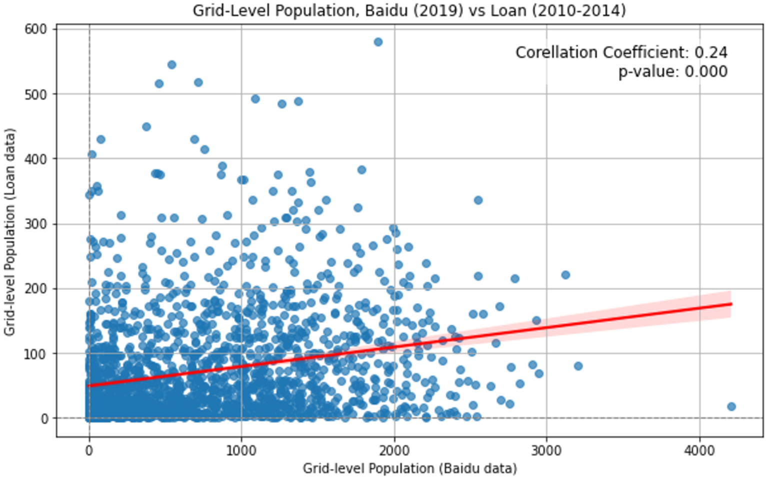
(Back)

Prefecture-Level Income Segregation Map



(Back)

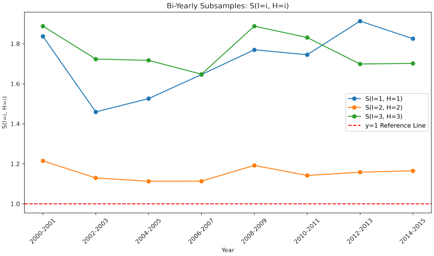
Mortgage Data Validation



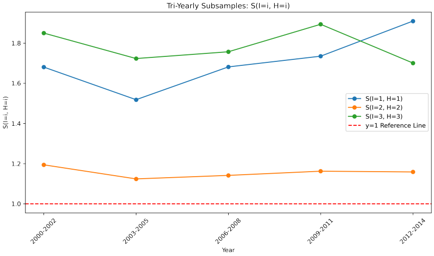
(Back)

Sorting Parameter Values: Bi-Year and Triple-Year

Bi-year



Triple-year



(Back)

Dynamic Sorting Model: Supply Estimation - 1

- Two unobserved supply measures:
  - Unconstrained supply

$$h_n^* = \frac{r_n - (\beta_0 + \beta_1 L_n + \beta_2 A_n) - \xi_n}{2k}$$

    where $\xi_n \sim N(0, \sigma_\xi^2)$ is an economic shock. We can rewrite this as $h_n^* \sim \mathcal{N}(\mu_{1n}, \sigma_1^2)$, where $\mu_{1n} = \frac{r_n - \beta_0 - \beta_1 L_n - \beta_2 A_n}{2k}$.
  - Constrained supply (regulation-limited)

$$h_n^{con} \sim \mathcal{N}(S_n \bar{h}_n, \sigma_\nu^2)$$

    where ν_n is a regulatory shock.
- Observed Supply:
  - Binding constraint determines observed floorspace:

$$h_n = \min(h_n^*, h_n^{con})$$

Dynamic Sorting Model: Supply Estimation - 2

- MLE Parameters: $\theta = \{\beta_0, \beta_1, \beta_2, k, \sigma_\xi, \{\sigma_R\}_{R=1}^{6}\}$
  - Unconstrained:

$$P(h_n \mid \text{Unconstrained}) = \phi\left(\frac{h_n - \mu_{1n}}{\sigma_1}\right)\frac{1}{\sigma_1} \times \left[1 - \Phi\left(\frac{h_n - S_n \bar{h}_n}{\sigma_v}\right)\right]$$

  - Constrained:

$$P(h_n \mid \text{Constrained}) = \phi\left(\frac{h_n - S_n \bar{h}_n}{\sigma_v}\right)\frac{1}{\sigma_v} \times \left[1 - \Phi\left(\frac{h_n - \mu_{1n}}{\sigma_1}\right)\right]$$

- Log-likelihood Function: (Back)

$$\mathscr{L}(\theta) = \sum_{n=1}^{N} \ln \left[\underbrace{\frac{1}{\sigma_1}\phi\left(\frac{h_n - \mu_{1n}}{\sigma_1}\right)\bar{\Phi}\left(\frac{h_n - \mu_{2n}}{\sigma_2}\right)}_{\text{Unconstrained Probability}} + \underbrace{\frac{1}{\sigma_2}\phi\left(\frac{h_n - \mu_{2n}}{\sigma_2}\right)\bar{\Phi}\left(\frac{h_n - \mu_{1n}}{\sigma_1}\right)}_{\text{Constrained Probability}} \right] \tag{17}$$