

Race in motion: sorting, segregation, colocation in Brazil

Hyoungchul Kim (Wharton, UPenn)

April 30, 2025

Motivation

- Socioeconomic inequality is closely linked to residential choice, both across and within cities (Moretti, 2013).
- As local amenities and productivities endogenously respond to the local population, this feedback loop could amplify inequality.
- One important socio-demographic when discussing location choice: **Race**.
- Important issue in both developed and developing country: racial segregation.
- It is important to characterize factors affecting people's location preferences by race to fully understand the dynamics of residential patterns.

Research question (tentative)

Some potential research questions

1. How do preference heterogeneity (by race) over amenities and productivity shape location sorting and inequality?
2. Quantify push- and pull- factors of people's location choices by race and its implications for local amenities and productivities in developing country setting.
3. Develop a unified framework for across-city location choices w.r.t. productivity and amenities.

(Brief) Literature review

1. Sorting across/within cities: Ahlfeldt et al. (2015); Diamond (2016); Sebastian et al. (2015); Ferreira and Wong (2023); Almagro and Dominguez (2025);.
⇒ Understand factors leading to sorting and colocation by race.
2. Residential segregation: Wong (2013); Asher et al. (2024); Harari (2024).
⇒ Focus more on fully characterizing push- pull- **factors** leading to residential patterns by **race across** cities.
3. Method-wise: Hsieh et al. (2019); Bryan and Morten (2019); Morten and Oliveira (2024)

1. **Residential** Census microdata of Brazil (1970, 1980, 1991, 2000, 2010)
 - Variable: **Race, Past locations**, socio-demographics.
2. **Workplace** RAIS data (formal sector employment; 1980 - 2023)
 - Admin data on all formally employed workers across jobs in different industries and locations.
3. **Workplace** ECINF data (informal sector employment; 2003)
 - Sample survey for self-employed workers in informal sector.

Research design: descriptive patterns about migration by race

Goal

Use location choice model by race to estimate relative migration attractiveness of Brazil municipalities

⇒ Show evidence of both colocation and separating force in play.

Motivation

- Goal: Show variations and overlap of preferences of regions by race.
- It is hard to consider all factors that affect people's relative preferences for cities from scratch (Making restrictive assumptions).
- Reveal preference approach: observing where people move implicitly shows us their preferences without the need to fully characterize such factors upfront.
- It also has additional benefits (e.g. Compared to surveys and population growth).

Location choice setup

Following Lee, Lee and Lin (2021); Kim, Lee and Park (2023):

Environment setup

- J number of regions with initial populations.
- Each region offers multiple opportunities (k) to a person with number increasing with population (# of houses, jobs, etc).

- (Individual) $\max_k U_{odk} = \overbrace{u_d}^{\text{Destination utility}} - \underbrace{c_{od}}_{\text{Migration cost}} + X + \varepsilon_{odk}.$

- **Note 1:** # of opportunity generally increase with size of region.
- **Note 2:** Stayers are assumed to have chosen the region they are currently in.

Choice probability of choosing region d

$$\pi_{od} = \frac{N_d^\gamma \exp(u_d - c_{od})}{\sum_j N_j^\gamma \exp(u_j - c_{oj})} \Rightarrow \frac{\exp(\delta_d - \log(d_{od})\beta)}{\sum_j \exp(\delta_j - \log(d_{oj})\beta)} \quad (1)$$

- N_d^γ is the total # of opportunities offered by each city.
- γ is a parameter that sets the relation between city size and number of opportunities.

City characteristics

- Parametric assumption: $\delta_d = \overbrace{Z_d' \alpha}^{\text{City characteristics}} + \gamma N_d \Rightarrow u_d = \delta_d - \gamma N_d$.
- $Z_d' \alpha$ is linear unbiased predictor of city's welfare levels.

Example results

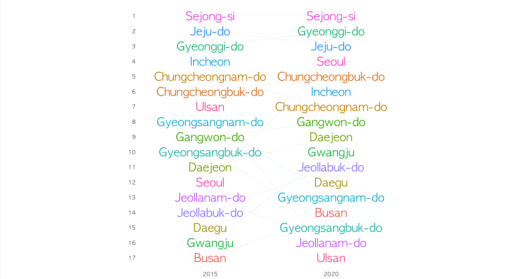


Figure 2: Changes in Weighted Average Ranking of Attractiveness by Population

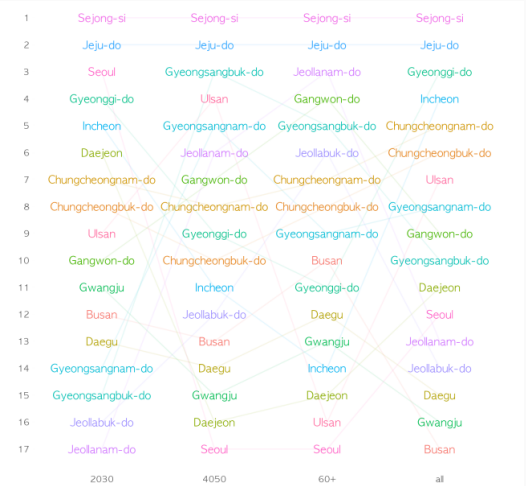


Figure 4: Average Ranking of City Attractiveness by Age (Metropolitan Areas and Provinces)

Possible results (hopefully)

- Hopefully, there will be some variations and some correlations in people's ranking of cities by race.
- **Variations:** implies that there can be some homophily and/or segregation effects happening between race.
- **Correlations:** implies that there is also some preferences for certain types of people with different ethical backgrounds to be close to each other.
- Nice way to visualize some possible effects happening and a starting point for some full model to characterize possible mechanisms.

Why do we need a model?

- GE effects: our previous results assume partial eqm.
- Integration of differential and overlapping patterns of people by race.
- Dynamics!
- Deep parameters of interest: e.g. (dis)-utility from density.
- Setup a full unified framework that captures this migration sorting and eqm effects.

(Preliminary) Economic model: key building blocks

Environment setup

- Individuals i indexed by their race type $r \in R$. N_r is total number of individuals of race r .
- Individuals consume goods and housing accord with a C-D utility function.
- Individuals also enjoy amenity B_{dt} of living in location d .
- Given this indirect utility, individual living in origin o will choose which destination d to live in (migration cost c_{odt}). iid shock for each destination.
- Each location can use labor to produce goods (productivity draw)...

I am looking into literature to think more about the model setup.

Conclusion & Next steps

1. Apply to get access to necessary data.
2. First do some descriptive analysis I mentioned in the slide to see overall patterns.
3. If previous steps look promising, start focusing bit more about the model.