Socio-Economic Aspects of Urban Development

Unit 10: Measuring Socio-Spatial Segregation

Prof. Melanie Krause, Ph.D.

Winter Semester 26/27

Overview of Unit 10

- Measuring socio-spatial segregration and inequality in cities
  - An index of segregration between groups: The Dissimilarity Index (DI)
    - ★ Practical exercise and comparison with empirical results
  - An index of neighborhood diversity: The Entropy Index (EI)
    - ★ Practical exercise
  - Measuring income inequality: The Gini coefficient
    - ★ Correlation of segregation with income inequality
  - Policy Responses to high levels of socio-spatial segregration

Measuring Socio-Spatial Segregation

- In Unit 9, we have looked at various reasons why rich residents might tend to cluster in certain parts of the city, and the (relatively) poor in others. It is typically the rich who, thanks to their financial means, can be selective in choosing their neighborhoods, while poorer households have to stick with what is left and what they can afford.
- But how can we measure the mixture - or separation - of people in neighborhoods?
- Literature we are going to use here:
  - White (1986) gives an overview over widely-used indices of segregation.
  - Forest (2005) provides calculation examples.
  - The case studies in the book by van Ham (2021) compute these indices for 24 cities around the world and analyze patterns.

The Challenge of Grouping People

- Indices of socio-spatial segregation rely on a certain **classification of people into groups**
- This can be based on **income**
  - Two groups: High-income, low-income. Or rather three groups? Most European cities have a huge middle class
  - Two groups: People working in high-skilled jobs, low skilled jobs? Based on university education?
- Or based on **ethnic characteristics**
  - Several groups: White, Black, Asian, Hispanic, other/mixed (often used in the US)? Or white, non-white?
  - Two groups: With and without background of migration? (how to define that?)
- Where to make the cut-off points between income classes? Into how many groups should you split the city population? Based on individual or household-level data? $\Longrightarrow$ All of this can influence the resulting index numbers, keep it in mind when comparing different cities!

The Dissimilarity Index (DI)

- The **Dissimilarity Index (DI)** is widely used as a measure of residential segregation between socio-economic groups
- It reflects their relative distribution across neighborhoods within the city.
- It always refers to the distribution of two groups, although there might be more groups in the city.
- Its values range between 0 and 1:
  - 0: Both groups are distributed equally across the neighborhoods of the city. $\Longrightarrow$ Perfect mixture
  - 1: Both groups are located in different neighborhoods of the city. $\Longrightarrow$ Total segregation
- The DI value can also be interpreted as the proportion of a group that would have to move in order to create an equal distribution of the population across neighborhoods.

The Dissimilarity Index (DI)

- Calculation of the DI for the distribution of two groups A and B across N neighborhoods of the city :

$$DI = \frac{1}{2} \cdot \sum_{i=1}^{N} \left| \frac{A_i}{A_{Tot}} - \frac{B_i}{B_{Tot}} \right| \tag{1}$$

  where

  - A_i and B_i are the populations numbers of, respectively, group A and B in neighborhood i
  - A_{Tot} and B_{Tot} are the total population numbers of, respectively, group A and B in the city

- Note that if one group is more represented than the other group in a certain neighborhood (given the group's overall population number), then the absolute difference between the fractions is positive. It adds up towards the DI, which is the sum of all the absolute differences over all neighborhoods.

The Dissimilarity Index (DI): Exercise

- Let us implement the illustrative example by Forest (2005) and compute DIs from an (imaginary) American city with 705 residents from 3 different groups (White, Black, Asian). There are 6 neighborhoods.
- Take a look at the table in the excel sheet *Unit10_UrbanSegregation*. What do you note about the distribution across neighborhoods?
- Then compute the DI for (i) White-Black, (ii) White-Asian, (iii) Black-Asian.

Neighborhood	White	Black	Asian
1	100	0	10
2	120	10	20
3	40	0	5
4	50	20	50
5	100	20	20
6	30	90	20
Total	440	140	125

The Dissimilarity Index (DI): Exercise

- First step: For each group and each neighborhood, compute the relative population share of this group in the neighborhood, e.g. $\frac{W_1}{W_{Tot}}$ for the Whites in neighborhood 1.
  - Note that for each group (column), these fractions add up to 1.

Neighborhood	White	Black	Asian	w_i/w_Tot	b_i/b_Tot	a_i/a_Tot
1	100	0	10	22.73%	0.00%	8.00%
2	120	10	20	27.27%	7.14%	16.00%
3	40	0	5	9.09%	0.00%	4.00%
4	50	20	50	11.36%	14.29%	40.00%
5	100	20	20	22.73%	14.29%	16.00%
6	30	90	20	6.82%	64.29%	16.00%
Total	440	140	125	100.00%	100.00%	100.00%

The Dissimilarity Index (DI): Exercise

- Second step: For each DI (White-Black, White-Asian, Black-Asian), compute the absolute value of the difference of the shares of the respective groups for each neighborhood. Finally, sum up these differences and multiply by 1/2 to obtain the DIs. Interpret your results.

Neighborhood	White	Black	Asian	w_i/w_Tot	b_i/b_Tot	a_i/a_Tot	Abs (w_i/w_Tot-b_i/b_Tot)	Abs (w_i/w_Tot-a_i/a_Tot)	Abs (w_i/w_Tot-a_i/a_Tot)
1	100	0	10	0.2273	0.0000	0.0800	0.2273	0.1473	0.0800
2	120	10	20	0.2727	0.0714	0.1600	0.2013	0.1127	0.0886
3	40	0	5	0.0909	0.0000	0.0400	0.0909	0.0509	0.0400
4	50	20	50	0.1136	0.1429	0.4000	0.0292	0.2864	0.2571
5	100	20	20	0.2273	0.1429	0.1600	0.0844	0.0673	0.0171
6	30	90	20	0.0682	0.6429	0.1600	0.5747	0.0918	0.4829
Total	440	140	125	1.0000	1.0000	1.0000			
						DI	0.6039	0.3782	0.4829

- 60% of the Whites (or the Blacks) would hypothetically have to move in order to achieve an equal distribution of these two groups across the city. The DIs between Whites and Asian as well as between Blacks and Asian are lower, but still sizeable. There is a notable concentration of groups in certain neighborhoods, which is captured by these indices.

The Dissimilarity Index (DI): Empirical Results

- Which DI value from 0 to 1 can be considered high? van Ham et al. (2021) point out that it depends on the application and the countries considered:
  - For ethnic and racial segregation, based on US values:
    - ★ Values below 0.3: Low segregation
    - ★ Values between 0.3 - 0.6: Moderate segregation
    - ★ Values above 0.6: High segregation
  - For segregation based on income or occupation
    - ★ Values tend to vary with the countries and cities considered, because of the overall socioeconomic conditions and income inequality within a country
    - ★ Computing DI between top- and bottom occupation groups, a value of 0.5 would be high in Europe, but only moderate in Latin America.

The Dissimilarity Index (DI): Empirical Results

van Ham et al. (2021) compare DI case study analyses from 24 cities on five continents, all using Functional Urban Areas and three groups (top, medium, bottom).

Limitations: Mostly DI between the top- and bottom occupational groups, while some studies (Mexico City) use top- and bottom educational groups (higher within-group inequality within educational than occupational groups?)

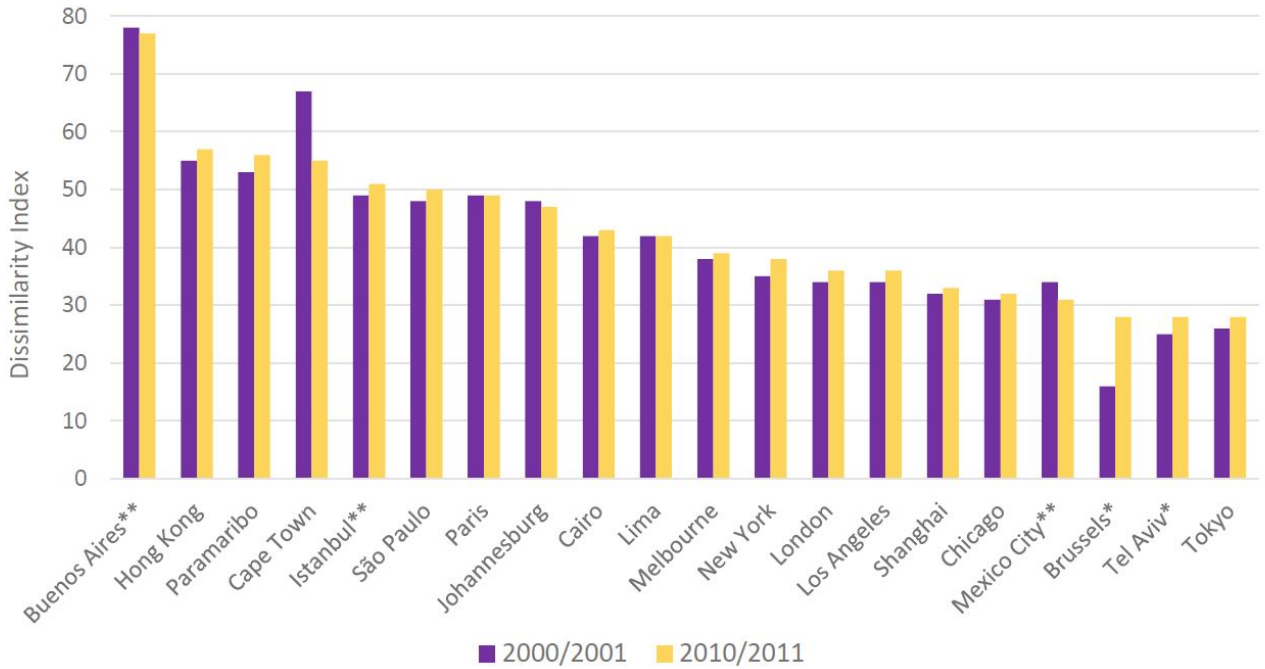


Fig. 1.3 Residential segregation between top and bottom socio-economic groups, 2000/2001 and 2010/2011. *Source* Individual chapters in this book, see Appendix 4 for more details). *Notes* *Top and bottom groups based on income; **Top and bottom groups based on educational attainment. Data for Paramaribo 2004 and 2012, Paris 1999 and 2015, Cairo 1996 and 2016, Lima 1993 and 2007, New York 2000 and 2013-2017, Mexico City 1990 and 2010, Tel Aviv 1995 and 2008

The Dissimilarity Index (DI): Empirical Results

- European cities do not necessarily have low segregation levels: Paris has as high a DI as Johannesburg and Sao Paulo.
  - Most high-income cities show an increase in segregation in the 10 years considered. In middle-income cities, the picture is mixed.
  - Note the high segregation level in Hong Kong. This is atypical for Asia, but consider the colonial past.
- The Anglo-American cities in the studies show a modest level of socio-economic segregation.
- Very low level of segregation in Tokyo
  - European cities have an overlap between ethnic and socio-economic segregation due to the on average low incomes of migrants compared to natives. Tokyo has very few international migrants.
  - Tokyo has low income inequality, strong public sector involvement in the economy, the housing market, and urban planning.
  - A densely populated compact city, Tokyo provides few opportunities for residential segregation. Maybe vertically? Not measured!

The Entropy Index (EI)

- The Dissimilarity Index can only measure the segregation of two groups compared to each other. By contrast, the **Entropy Index (EI)** captures the spatial distribution of multiple groups simultaneously.
- The EI is calculated at the neighborhood level. It tells you how diverse the neighborhood is in terms of group composition: Dominated by one group or does its population belong to all groups in equal proportions?
- The EI in neighborhood j is calculated as

$$EI_j = -\sum_{j=1}^{k} p_{ij} \cdot ln(p_{ij}) = -[p_{1j} \cdot ln(p_{1j}) + p_{2j} \cdot ln(p_{2j}) + ... + p_{Kj} \cdot ln(p_{Kj})] \quad (2)$$

  - where k are the number of groups (from group 1 to group K),
  - p_{ij} is the proportion of the population in neighborhood j that is accounted for by group i

The Entropy Index (EI)

- Calculating and interpreting

$$EI_j = -\sum_{j=1}^{k} p_{ij} \cdot ln(p_{ij}) = -[p_{1j} \cdot ln(p_{1j}) + p_{2j} \cdot ln(p_{2j}) + ... + p_{kj} \cdot ln(p_{kj})] \quad (3)$$

  - EI ranges between 0 and $ln(K)$ (the natural logarithm of the number of groups). For example, with three groups, $ln(3) \approx 1.10$.
  - $0 \cdot ln(0)$ is defined as 0.
  - Neighborhoods with higher values of EI are more diverse.
    - ★ A neighborhood with $EI = 0$ contains only a single group: $K - 1$ groups have $p_{ij} = 0$, so they contribute 0. For the single group K, it holds $p_{Kj} \cdot ln(p_{Kj}) = 1 \cdot ln(1) = 0$, so it also contributes 0.
    - ★ A neighborhood with $EI = ln(k)$, e.g. $ln(3)$ houses all groups in equal proportion:
      $-[1/3 \cdot ln(1/3) + 1/3 \cdot ln(1/3) + 1/3 \cdot ln(1/3)] = -ln(1/3) = ln(3)$.

The Entropy Index (EI): Exercise

- Calculate the diversity, as measured by the Entropy Index, of the six neighborhoods in the imaginary American city.
- Start by copying the population data from *Unit10_UrbanSegregation* into a new spreadsheet. Calculate the total population P_j in each neighborhood, as well as the share p_{ij} that is accounted for by each group i

Neighborhood	White(1)	Black(2)	Asian(3)	Total	White: p_(1j)	Black: p_(2j)	Asian: p_(3j)
1	100	0	10	110	0.9091	0.0000	0.0909
2	120	10	20	150	0.8000	0.0667	0.1333
3	40	0	5	45	0.8889	0.0000	0.1111
4	50	20	50	120	0.4167	0.1667	0.4167
5	100	20	20	140	0.7143	0.1429	0.1429
6	30	90	20	140	0.2143	0.6429	0.1429
Total	440	140	125	705	0.6241	0.1986	0.1773

The Entropy Index (EI): Exercise

- For each neighborhood, calculate

$$EI_j = -\sum_{j=1}^{3} p_{ij} \cdot ln(p_{ij}) = -[p_{1j} \cdot ln(p_{1j}) + p_{2j} \cdot ln(p_{2j}) + p_{3j} \cdot ln(p_{3j})] \quad (4)$$

  - Manually correct the error that the calculation runs into with $0 \cdot ln(0)$.
- The most diverse/mixed neighborhood is no.4, which has an EI of 1.03, close to the maximum of 1.10 for $K = 3$. The least diverse is neighborhood no.1. You can also see which neighborhoods lie above and below the city's overall EI of 0.92.

Neighborhood	White(1)	Black(2)	Asian(3)	Total	White: p_(1j)	Black: p_(2j)	Asian: p_(3j)	EI
1	100	0	10	110	0.9091	0.0000	0.0909	0.3046
2	120	10	20	150	0.8000	0.0667	0.1333	0.6277
3	40	0	5	45	0.8889	0.0000	0.1111	0.3488
4	50	20	50	120	0.4167	0.1667	0.4167	1.0282
5	100	20	20	140	0.7143	0.1429	0.1429	0.7963
6	30	90	20	140	0.2143	0.6429	0.1429	0.8921
Total	440	140	125	705	0.6241	0.1986	0.1773	0.9220

Discussion of DI and EI

- Note the differences: *DI* as a measure of segregation between two groups looks at the spatial distribution of these groups across neighborhoods. *EI* as a measure of diversity refers to the neighborhood level, not the group level.
- The *EI* has an upper bound that depends on the number of groups K. Note that cities with more groups K can achieve a higher *EI*, because $ln(K+1) > ln(K)$. This fits its interpretation of a diversity index, because cities with more groups are more diverse.
- For both *DI* and *EI*, we do not need individual-level data. We just need the total population numbers of the groups in the city, as well as the population numbers for each group in each city part/district/neighborhood i.
  - Many cities provide socio-economic data at the district level which allows to group people and which can be used to calculate certain *DI* or *EI* measures:
    - ★ Number of residents with and without background of migration?
    - ★ Number of residents with and without a job?
    - ★ Number of residents with income below or above a certain threshold?

Measuring Inequality

- The most widely-used index of income inequality is the **Gini coefficient**. In order to calculate it, we need individual-level income data. Note that there is no spatial component to it.
- The Gini coefficient can be calculated and interpreted as half of the expected relative income difference of two random people i and j:
  - Take the absolute value of the income differences between all individuals and sum them up. Adjust for double-counting, mean income as well as population size by dividing by $2 \cdot \bar{\mu} \cdot N^2$.

$$Gini = \frac{E|y_i - y_j|}{2\bar{\mu}} = \frac{1}{2\bar{\mu}} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |y_i - y_j| \tag{5}$$

- It ranges between 0 (everybody has the same income) and 1 (one person has all the income, the rest have zero). For a detailed discussion on the properties of the Gini coefficient, see the textbook by Cowell (2011).

Gini Coefficient: Example

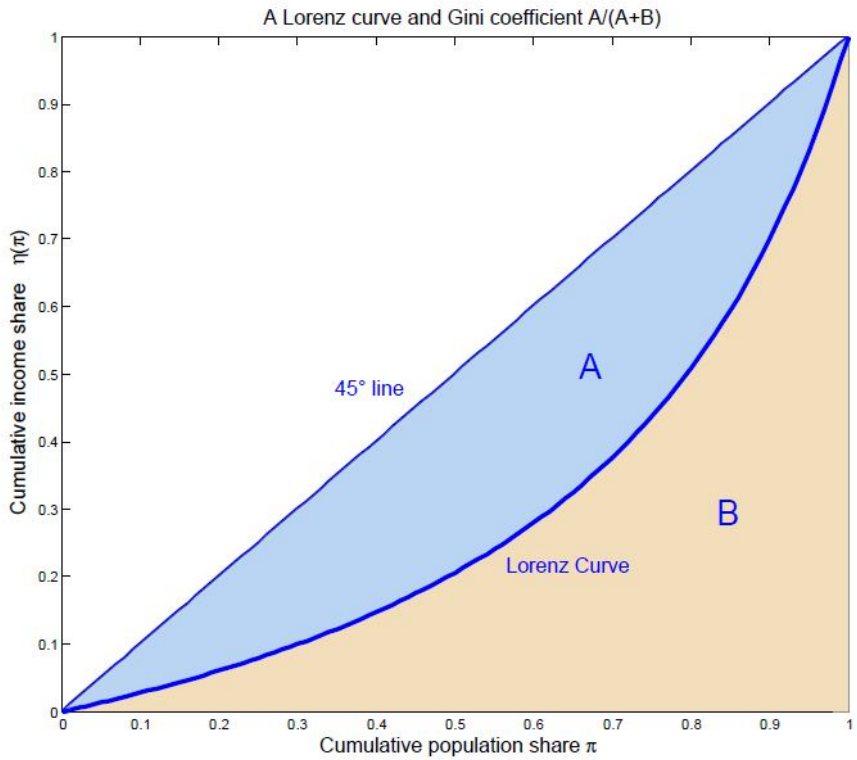
50	Individuals have	EUR 100 each
50	Individuals have	EUR 200 each
75	Individuals have	EUR 300 each
25	Individuals have	EUR 400 each
200		47500 Euro

$$Gini = \frac{1}{2\bar{\mu}}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|y_i - y_j| = \frac{1}{2 \cdot 237.5} \cdot \frac{1}{200^2} \cdot SUM = \underline{\underline{0.2303}} \quad (6)$$

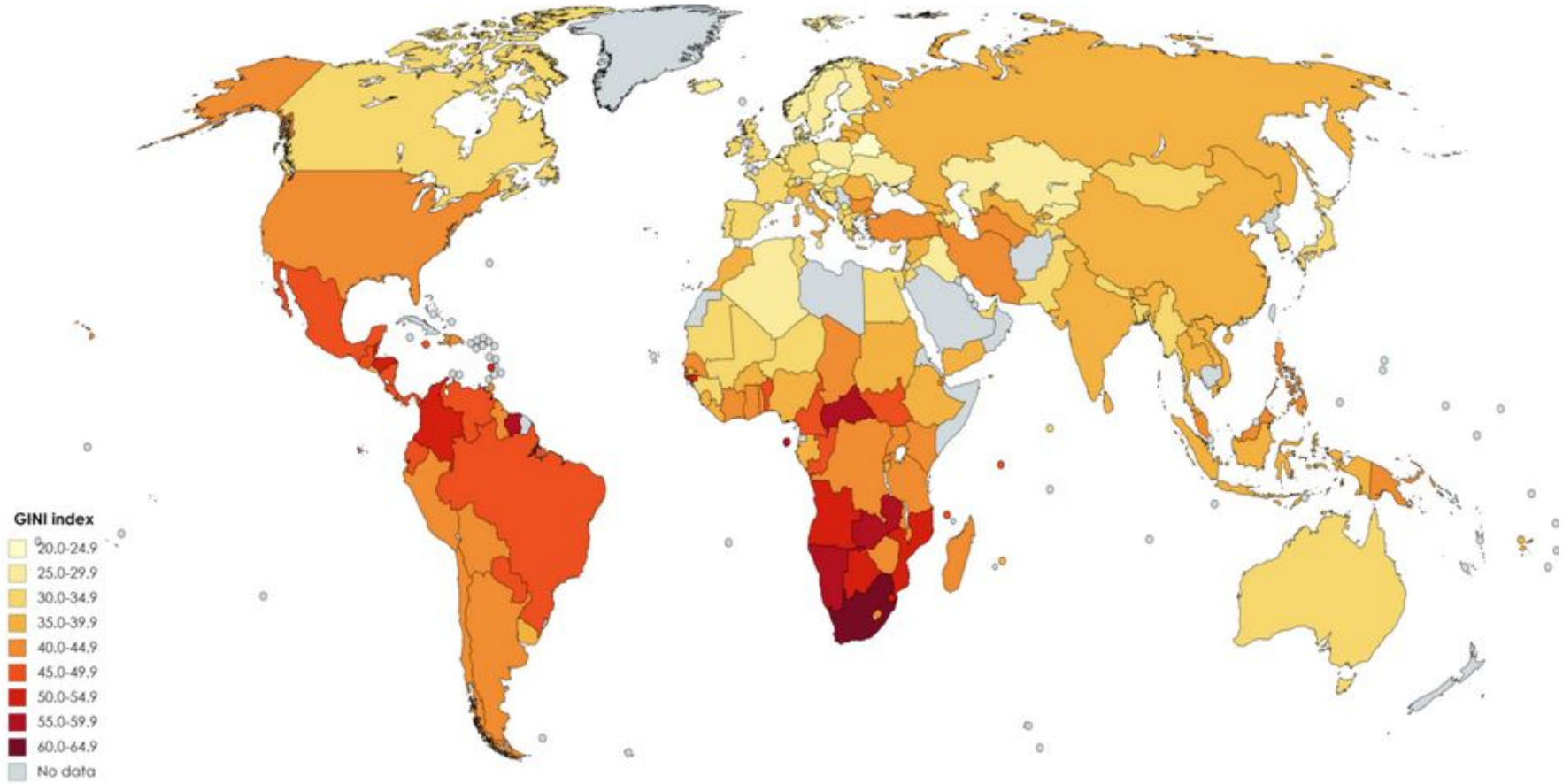
$$\begin{aligned} SUM &= \sum_{i=1}^{50}[50 \cdot |100 - 200| + 75 \cdot |100 - 300| + 25 \cdot |100 - 400|] \\ &+ \sum_{i=1}^{50}[50 \cdot |200 - 100| + 75 \cdot |200 - 300| + 25 \cdot |200 - 400|] \\ &+ \sum_{i=1}^{75}[50 \cdot |300 - 100| + 50 \cdot |300 - 200| + 25 \cdot |300 - 400|] \\ &+ \sum_{i=1}^{25}[50 \cdot |400 - 100| + 50 \cdot |400 - 200| + 75 \cdot |400 - 300|] \\ &= 1.375.000 + 875.000 + 1.312.500 + 812.000 = 4.375.000 \end{aligned}$$

Gini Coefficient: Relation to Lorenz Curves

- The **Lorenz Curve** plots the proportion of overall income held by the bottom p% of the population. Hence, the x-axis shows the cumulative share of individuals (ordered by income); the y-axis shows their corresponding cumulative share of income.
- $Gini = 1 - 2 \cdot$ Area below the Lorenz Curve $= 1 - 2 \cdot B$
- Total Equality: Lorenz Curve equals 45 degree line $\Longrightarrow Gini = 1 - 2 \cdot 0.5 = 0$
- Total Inequality: Lorenz Curve equals 0 for all but the richest individual, then it jumps to 1. $\Longrightarrow Gini = 1 - 2 \cdot 0 = 1$



Gini Coefficents of Net Income (World Bank 2018)



Typical values for net income inequality at the country level range between 0.20-0.35 (Europe and Central Asia), 0.30-0.45 (South/East Asia), 0.40-0.50 (USA), and 0.45-0.60 (Latin America/parts of Africa)

Inequality and Segregation: Empirical Results

van Ham et al. (2021) show scatter plots between segregation (DI) in the cities and income inequality (Gini) at the country level.

Note the positive correlation - and keep in mind that the Gini coefficient at the country level is typically lower than at the city level.

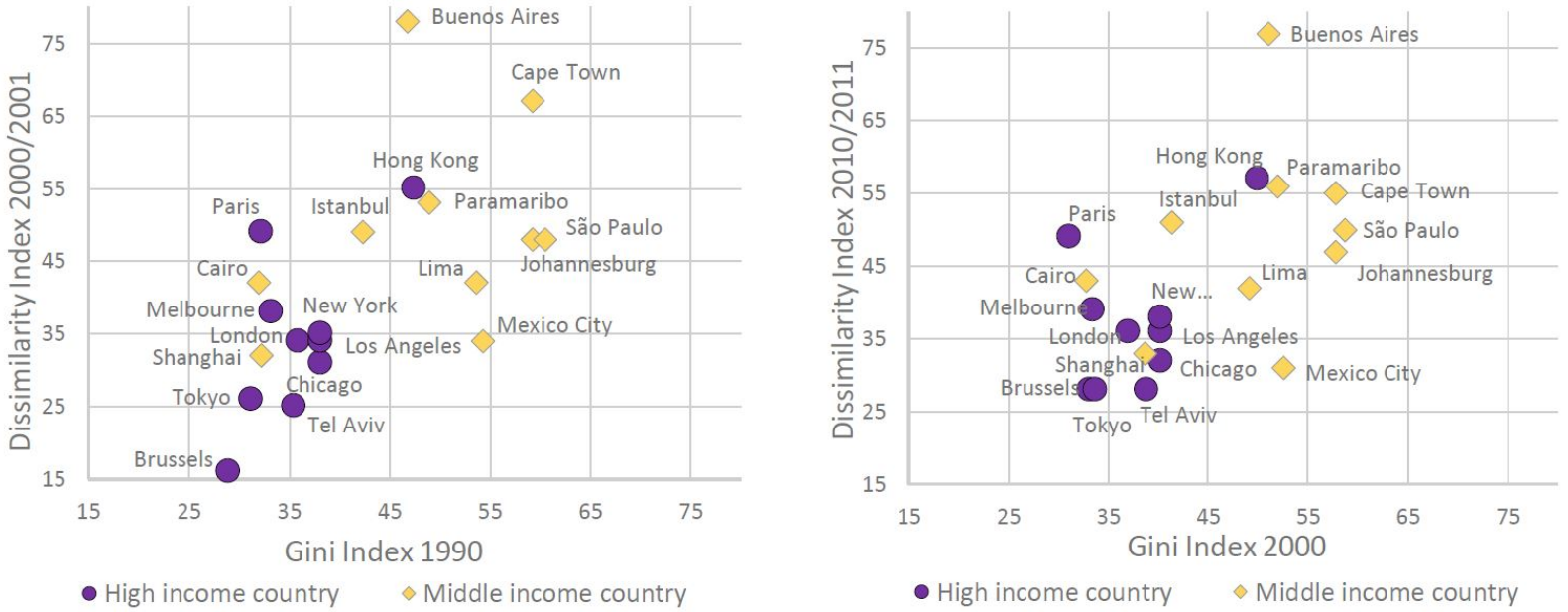


Fig. 1.4 Relationship between income inequality (lagged 10 years and) socio-economic residential segregation between Top and Bottom socio-economic groups. *Source* Individual chapters in this book and the World Bank

Inequality and Segregation: Empirical Results

- The authors work with lagged income inequality values: It takes time for changes in income inequality to materialize in terms of residential segregation patterns.
  - In a market-dominated housing system with little public interventions in housing, changes in income inequality may translate rather quickly into changes in residential sorting based on income.
  - If there is a high share of social or public housing - and strong policy interventions - changes in income inequality and residential segregation become more de-coupled.
- The two scatter plots show a positive correlation between the (lagged) Gini coefficient and the DI.
  - Correlation coefficient: 0.529 in 2010/11, lower than 0.583 in 2000/01.
  - Outliers: Paris has a higher segregation than its income inequality level would suggest. Lower-income migrant households in housing estates, market forces sorting households with different financial means into different neighborhoods?

Socio-Economic Problem of High Segregation

- A high level of segregration might go in line with the concentration of socio-economic problems in some particular low-income neighborhoods:
  - Low income and high unemployment rates
  - Bad schools
  - Bad infrastructure connections to other areas
  - Potentially a concentration of minority groups and immigrants
  - Low opportunities and bad perspectives of social mobility
  - Potentially (organized) crime
  - $\Longrightarrow$ Bad reputation of the neighborhood, perpetuating its problems, because it draws only those new residents that cannot afford to live elsewhere

Policy Response to High Segregation

- What can policymakers do when faced with high segregation levels? How can they help disadvantaged neighborhoods?
  - **Public investment in disadvantaged neighborhoods**
    - ★ Investment in schools in poor areas, sending good teachers
    - ★ Investment in higher-quality housing stock
  - **Encouraging more mixing with richer residents**
    - ★ School vouchers to send poor kids to good schools in other neighborhoods
    - ★ Better transportation links between neighborhoods
  - **Facilitating the move into richer areas**
    - ★ In richer areas: Ensuring affordable housing, revisiting zoning laws that restrict further development (against resistance of rich voters?)
  - **Forced moves as an extreme case**
    - ★ Denmark's controversial 'social experiment' to tear down social housing estates in certain low-income neighborhoods, offering residents a place to live elsewhere and instead rebuilding higher-quality houses to draw in wealthier residents

Wrap-Up of Unit 10

- The **Dissimilarity Index (DI)** is an index of segregation between two groups. Ranging between 0 and 1, it can be interpreted as the share of a group that would have to be moved to ensure an equal distribution of the groups across the city.
- The **Entropy Index (EI)** indicates how diverse a neighborhood is. It has a value of 0, if only one group populates the neighborhood and can obtain a maximum value of $ln(K)$, if all k groups are equally represented in the neighborhood.
- Segregation in a city correlates with **income inequality** at the country level, as measured by the Gini coefficient. But the relation is far from perfect and urban planning as well as public policy matter, as the case of low-inequality, high-segregation Paris shows.
- **Policymakers** can decrease segregation by ensuring the mixture of households from different background, to which infrastructure connections, educational and job opportunities as well as affordable housing thoughout the city can contribute.

Bibliography of Unit 10



Cowell, F. (2011).

Measuring Inequality.

Oxford University Press.



Forest, B. (2005).

Measures of Segregation and Isolation.

Technical Report Dartmouth College.



van Ham, M., T. Tammaru, R. Ubarevičienė, and H. Janssen (2021).

Urban Socio-Economic Segregation and Income Inequality - A Global Perspective.

Springer - The Urban Book Series.



White, M. (1986).

Segregation and Diversity Measures in Population Distribution.

Population Index 52, 198–221.