Measuring spatial polarization and inequality in cities using housing transaction big data

Jinhyung Lee¹,³, Harvey J. Miller¹,³, Nicholas B. Irwin² & Elena G. Irwin²,³

¹) Department of Geography
²) Department of Agricultural, Environmental, and Development Economics
³) Center for Urban and Regional Analysis (CURA)
The Ohio State University
Great Recession in the U.S.
• Economic earthquake
  o The U.S. labor market lost 8.4 million jobs

Recovery from recession
• Positive signs
  o Increase in GDP percent change
  o Urban revitalization
• However, unequal distribution of benefits
  : concentration of wealth and economic growth in small parts of city

Possible outcomes
1. Deepening spatial polarization (clusters of the rich vs. poor)
2. Increasing inequality (e.g., income)

New Urban Crisis! (Florida, 2017)
Geography of “New Urban Crisis”

How have polarization and inequality in cities evolved?
Research objectives

1. Investigate the evolution of spatial polarization in the U.S. cities
2. Investigate the evolution of inequality in the U.S cities

• Study area
  - State of Ohio: as a pilot study
  - Cities with different economic growth trajectories:
    ▪ Columbus (high), Cincinnati (medium), Cleveland (low)

• Methods
  - Big data analysis using high-resolution housing transaction data
  - Geovisualization of polarization over time
  - Measuring spatial polarization (Moran’s I) and inequality (Gini coefficient)
Data

Housing transaction data

- High spatial resolution: parcel-level
- Time range: 2000-2015
- Big data: # of observations in Columbus: 482,478
- City boundary: Urbanized Areas (UAs)

Data pre-processing

1. Removing housing transactions missing key characteristics
2. Removing outliers
   - Housing transactions less than $10,000 and more than $2.5 million are dropped
Housing price hedonic model

- Model specification
  - Estimate hedonic price models to measure location premium
    - Disaggregate price into attributes of house
  
  \[ \ln P_{it} = \beta_0 + \beta_X X_i + \beta_Q Quarter_q + \epsilon_{it} \]

- Natural log of price for house \( i \) in period \( t \)
- Function of a vector of house characteristics \( (X_i) \)
  and a quarter-by-year fixed effect \( (Quarter_q) \)
- Remaining unexplained variation → Location premium
<table>
<thead>
<tr>
<th>Variables</th>
<th>Cincinnati</th>
<th>Cleveland</th>
<th>Columbus</th>
</tr>
</thead>
<tbody>
<tr>
<td>House size (sqft)</td>
<td>0.000564</td>
<td>0.000660</td>
<td>0.000573</td>
</tr>
<tr>
<td>House size squared</td>
<td>-4.36e-08</td>
<td>-4.33e-08</td>
<td>-3.28e-08</td>
</tr>
<tr>
<td>Parcel size (acres)</td>
<td>0.173</td>
<td>0.0948</td>
<td>0.170</td>
</tr>
<tr>
<td>Parcel size squared</td>
<td>-0.0189</td>
<td>-0.0126</td>
<td>-0.0178</td>
</tr>
<tr>
<td>Age</td>
<td>-0.000318</td>
<td>0.00276</td>
<td>-0.0109</td>
</tr>
<tr>
<td>Age squared</td>
<td>-5.04e-05</td>
<td>-0.000100</td>
<td>5.94e-05</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>-0.0563</td>
<td>-0.0511</td>
<td>-0.152</td>
</tr>
<tr>
<td>Full bathrooms</td>
<td>0.135</td>
<td>0.0347</td>
<td>0.145</td>
</tr>
<tr>
<td>Stories</td>
<td>-0.0149</td>
<td>-0.106</td>
<td>-</td>
</tr>
<tr>
<td>Pool</td>
<td>0.0420</td>
<td>-0.162</td>
<td>0.0904</td>
</tr>
<tr>
<td>Finished basement</td>
<td>-0.0517</td>
<td>0.0763</td>
<td>-0.150</td>
</tr>
<tr>
<td>Fireplace</td>
<td>0.192</td>
<td>0.239</td>
<td>0.116</td>
</tr>
<tr>
<td>Central air conditioning</td>
<td>0.130</td>
<td>0.194</td>
<td>0.581</td>
</tr>
<tr>
<td>Garage</td>
<td>0.138</td>
<td>0.0700</td>
<td>-0.0417</td>
</tr>
<tr>
<td>Constant</td>
<td>10.38</td>
<td>10.70</td>
<td>10.60</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.475</td>
<td>0.526</td>
<td>0.447</td>
</tr>
<tr>
<td>Observations</td>
<td>414,665</td>
<td>512,844</td>
<td>482,477</td>
</tr>
<tr>
<td>Quarter-year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Note:** LogPrice is the dependent variable. All variables are significant at 0.01 levels. Stories is not reported for Columbus due to a lack of observations.
Geovisualization of polarization

• Spatial interpolation: Kriging
  - Creating surfaces of location premium over time
  - Ordinary Kriging / Exponential semivariogram model (lowest RMSE)

Measuring polarization and inequality

• Moran’s I values
  - Measure spatial polarization of location premium
  - Unit of analysis: Census block group

• Gini coefficients
  - Measure inequality of location premium
Location premium surfaces

Columbus

- Top 5%
- Bottom 5%
Location premium surfaces

Cleveland

2000-2003

2004-2007

2008-2011

2012-2015

Top 5%

Bottom 5%
Measuring spatial polarization: Moran’s I

- Increasing spatial polarization trends
- Peak right before GR (2007), sharp drop after GR
- Growing spatial polarization trends during recovery (after 2011)
Measuring inequality: Gini coefficient

- Similar to the Moran’s I values, general increasing trends of inequality
Conclusion

• Three biggest Ohio cities are becoming more polarized!
  - Location premium surfaces: clustering of high (or low) values → Divided cities
  - Spatial polarization: increasing trends in Moran’s I values
  - Inequality: growing trends in Gini coefficients

Implications

1. Methodological: unprecedented high-resolution data
2. Empirical: evidence of growing polarization / inequality trends
3. Policy / planning: data-driven, map-based decision making

Next steps

• Scale up to the national level
• GWR against location premium → Why does an area have high LP values?
• Predicting future polarization using GWR-TS (time series)
Thank you!

Any Questions?

Contact

• Email: lee.7738@osu.edu (Jinhyung Lee)
miller.81@osu.edu (Harvey J. Miller)
irwin.114@osu.edu (Nicholas B. Irwin)
irwin.78@osu.edu (Elena G. Irwin)