Sorting based on amenities and income

Mark van Duijn  m.van.duijn@vu.nl
Jan Rouwendal

Department of Spatial Economics

(Work in progress)

Seminar
Utrecht School of Economics
25 September 2013
Projects

- **Economic valuation of cultural heritage (NICIS)**
  - Maintenance costs vs. Benefits
  - Does cultural heritage attract specific households?
  - What is the willingness to pay for cultural heritage?
  - Location choice models (revealed preference)
    - Tiebout (1956): *Households sort or ‘vote with their feet’ to choose their most preferred community*

- **Highly educated location preferences (NWO-HELP)**
  - How to attract or retain high educated households?
  - Work location vs. Amenities
  - (Focus on extending the location choice models)
  - (Focus on gaining additional insights from the location choice models)
Today

- Motivation
- The model: Equilibrium sorting model
- Some extensions
- Econometric issues
- Data and study area
- Estimation results
- Discussion
Motivation

- **Urban amenities are becoming more important for the location choice of households**
  - Traditional focus on cities as center of employment (Alonso-Muth-Mills models)
  - But other (consumption) needs are growing in relative importance
    - School quality
    - Restaurants
    - Theatres
    - Cultural status
    - Demographic composition
  - Producer city $\rightarrow$ Consumer city
Motivation

- **Literature**
  - *Brueckner, Thisse and Zenou (1999)*
    - Why is central Paris rich and downtown Detroit poor?
    - An amenity based theory
    - Amenities are luxuries that affect the location choices of households
  
  - This has likely consequences for economic growth
  
  - *Glaeser, Kolko, Saiz (2001); Carlino and Saiz (2008)*
Mechanisms

- **Amenities attract high incomes**
  - Example: cultural heritage
    - Canals of Amsterdam

- **High incomes attract each other**
  - Households are attracted to similar households
  - Social interaction effect
  - ‘multiplier’ on the impact of neighbourhood characteristics

- **The presence of high incomes may attract endogenous amenities**
  - Like shops, musea, theatre performances,...
Location equilibrium models

- **Sorting models**
  - Households choose their residential location
  - Between a fixed choiceset (municipalities / neighbourhoods / specific houses)
  - These locations differ in quality...
    - Distance to city center
    - Cultural heritage
    - Natural amenities
    - Demographic composition
    - ...

Sorting based on amenities and income
Location equilibrium models

Why are sorting models used?

- Sorting models are structural models that explain house prices
  - More advanced than simple hedonic methods

- Sorting models are able to account for heterogenous households
  - Marginal willingness to pay for various types of households (education, income, age, household size, etc.)

- Equilibrium property can be exploited to study (exogenous) shocks
  - Counterfactual simulations
  - Segregation, gentrification, demographic composition

- Sorting models are able to account for unobserved location characteristics
- There is room for extensions (more later)
Location equilibrium models

- **Basically: the sorting model is a logit model for location choice**
  - Choice alternatives are neighbourhoods
  - Decision makers are heterogeneous
    - Heterogeneity is related to household characteristics
    - Like education and income

- **We take into account unobserved neighbourhood characteristics**
  - Using the methodology of Berry, Levinsohn and Pakes (1995)
  - Two-step estimation procedure
  - Alternative-specific constants are further analysed in second step
The model (1)

Choice probabilities:

\[ Pr_{i,n} = \frac{\exp(w_{i,n})}{\sum_{m=1}^{M} \exp(w_{i,m})} \]

- Prob that household \( i \) chooses neighbourhood \( n \)
  - \( w_{in} \): deterministic part of the utility of neighbourhood \( n \) for household \( i \)
  - \( \varepsilon_{in} \): stochastic part of the utility of neighbourhood \( n \) for household \( i \)

- Total utility: \( u_{in} = w_{in} + \varepsilon_{in} \)

- Households maximize utility based on preferences and budget constraint

Sorting based on amenities and income
The model (2)

Utility: 
\[ u_{i,n} = \sum_{k=1}^{K} \alpha_{i,k} X_{k,n} + \xi_n + \epsilon_{i,n} \]

- Note that the coefficients are \( i \)-specific
  There is heterogeneity in tastes

- The \( \xi \) denotes unobserved neighborhood attributes
  Observed by the household, but not by the researcher
  Not \( i \)-specific
Further specification of coefficients:

\[ \alpha_{i,k} = \beta_{0,k} + \sum_{l=1}^{L} \beta_{k,l} (Z_{i,l} - \bar{Z}_l) \]

- Linear function of household characteristics \( Z \)
- Household characteristics are de-meaned
- \( \beta_{0,k} \) is the average value of the coefficient for characteristic \( k \)
The model (4)

Substitute into the utility function:

\[ u_{i,n} = \sum_{k=1}^{K} \beta_{0,k} X_{k,n} + \xi_n + \sum_{k=1}^{K} \left( \sum_{l=1}^{L} \beta_{k,l} (Z_{i,l} - \bar{Z}_l) \right) X_{k,n} + \varepsilon_{i,n} \]

And rewrite:

\[ u_{i,n} = \delta_n + \sum_{k=1}^{K} \left( \sum_{l=1}^{L} \beta_{k,l} (Z_{i,l} - \bar{Z}_l) \right) X_{k,n} + \varepsilon_{i,n} \]

\[ \sum_{i=1}^{I} P_{r_{i,n}} = S_n \]

- \( \delta \)'s are alternative-specific constants (mean indirect utility)
  - \( \delta \)'s and \( \beta_{k,l} \) are estimated in the first step
  - \( \delta \)'s are then further analyzed in the second step
The model (5)

After estimating the logit model we write again:

\[ \delta_n = \sum_{k=1}^{K} \beta_{0,k} X_{k,n} + \xi_n \]

- And use techniques for linear models to estimate the coefficients
- The unobserved heterogeneity now appears as an error term
Extensions

- **Social interactions → Multiplier effects**
  - Include demographic composition of the neighbourhood
  - Share of high income households
  - Share of high income households attract other (endogenous) amenities?

- **Characteristics of surrounding neighbourhoods**
  - Spatial lags of exogenous neighbourhood characteristics

- **(Movement costs)**
- **(Extending supply side)**
Econometric issues

- **Why not estimate a simple logit model?**
  - Unobserved characteristics are not taken into account
  - They may have an impact on observed neighborhood characteristics

- **Example: housing price**
  - If $\xi$ is high, a neighborhood is attractive
  - Housing price will be relatively high there
  - But we do not observe the reason and will run the risk of interpreting this as low price sensitivity
Econometric issues

- How to deal with this issue?
- Recall that in the second step we have a linear equation:

\[ \delta_n = \sum_{k=1}^{K} \beta_{0,k} X_{k,n} + \xi_n \]

- The price is one of the X-s
  - We have an endogeneity problem

- We can use 2SLS instead of OLS to deal with the endogeneity
Instruments?

- **We can create an instrument exploiting the equilibrium property:**
  - Use the model to predict the prices that would be observed if all the $\xi$-s are equal to zero
    - These prices are uncorrelated with the $\xi$-s
    - And (probably) highly correlated with the observed prices
    - And should not be included in the estimation equation

- **Since we do not yet know the true coefficients, an iterative procedure is used**
Social interactions

- **We want to include the possibility of preferences for the demographic composition of the neighborhood**
  - Especially of the share of high income households

- **This gives rise to a second endogeneity issue**

- **Which can be solved similarly**
  - We can compute the counterfactual share of high income households that would be observed if there were no unobserved heterogeneity
We study household location in the Amsterdam area

Household data
- Microdata Statistics Netherlands (GBA + IHI + SEC)

Neighbourhood data
- Price of a standard house (based on a simple hedonic model with neighbourhood fixed effects – NVM)
- Percentage of high income households (Top 25% – CBS)
- Conservation areas in km2 (RCE)
- Distance to the nearest 100,000 jobs (PBL)
Historic city centre
Historic city centre
Historic city centre
Maps

- **Percentages of high income households are higher around Amsterdam**
  - Rental sector is around 60-70% in Amsterdam with a large amount of social housing

- **Percentages of high income homeowners are more equally distributed in the Amsterdam area**
  - In the sorting model we focus mostly on homeowners (real choices)
  - Interpreting the results for renters is difficult (not always a choice)
## Descriptives

<table>
<thead>
<tr>
<th>Variables</th>
<th>Data source</th>
<th>Mean</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross primary household income</td>
<td>CBS (2008)</td>
<td>42,835</td>
<td>55,740</td>
<td>0</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Household with children (-18)</td>
<td>CBS (2008)</td>
<td>0.240</td>
<td>0.427</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Age of oldest household member</td>
<td>CBS (2008)</td>
<td>48.730</td>
<td>17.461</td>
<td>16</td>
<td>107</td>
</tr>
<tr>
<td><strong>Social Economic Category</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>CBS (2008)</td>
<td>0.053</td>
<td>0.223</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(Self-)Employed</td>
<td>CBS (2008)</td>
<td>0.559</td>
<td>0.496</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Unemployed (Social assistance benefits)</td>
<td>CBS (2008)</td>
<td>0.176</td>
<td>0.381</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Retired</td>
<td>CBS (2008)</td>
<td>0.212</td>
<td>0.409</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Neighborhood characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historic city center (km2)</td>
<td>RCE (2012)</td>
<td>0.027</td>
<td>0.134</td>
<td>0.000</td>
<td>1.029</td>
</tr>
<tr>
<td>Distance to the nearest 100,000 jobs (km)</td>
<td>PBL (2005)</td>
<td>8.287</td>
<td>3.355</td>
<td>0.637</td>
<td>18.407</td>
</tr>
<tr>
<td>Percentage rich households (%)</td>
<td>CBS (2008)</td>
<td>33.325</td>
<td>14.433</td>
<td>0.000</td>
<td>77.707</td>
</tr>
<tr>
<td>Price of standard house (in euros)</td>
<td>NVM (2009)</td>
<td>209,858</td>
<td>49,587</td>
<td>112,877</td>
<td>390,691</td>
</tr>
</tbody>
</table>
First step results

- **Deviations from the alternative specific constant for homeowners**

<table>
<thead>
<tr>
<th>Neighborhood characteristics</th>
<th>Household characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Income</td>
</tr>
<tr>
<td>Standardized house price (in euros)</td>
<td>0.01259</td>
</tr>
<tr>
<td></td>
<td>(0.0006)***</td>
</tr>
<tr>
<td>Historical city center (km²)</td>
<td>0.00313</td>
</tr>
<tr>
<td></td>
<td>(0.0004)***</td>
</tr>
<tr>
<td>Historical city center in surrounding neighborhoods</td>
<td>-0.00031</td>
</tr>
<tr>
<td></td>
<td>(0.00005)***</td>
</tr>
<tr>
<td>High income households (%)</td>
<td>0.00027</td>
</tr>
<tr>
<td></td>
<td>(0.00001)***</td>
</tr>
<tr>
<td>Distance to the nearest 100,000 jobs (km)</td>
<td>0.00001</td>
</tr>
<tr>
<td></td>
<td>(0.00004)</td>
</tr>
</tbody>
</table>
## Second step results

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) OLS (se)</th>
<th>(2) 2SLS (se)</th>
<th>(3) 2SLS (se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardized house price (in euros)</td>
<td><strong>-1.2582</strong></td>
<td><strong>-26.6315</strong></td>
<td><strong>-37.9354</strong></td>
</tr>
<tr>
<td></td>
<td>(0.5621) **</td>
<td>(7.976) ***</td>
<td>(10.434) ***</td>
</tr>
<tr>
<td>Historical city center (km²)</td>
<td><strong>1.3146</strong></td>
<td><strong>5.7193</strong></td>
<td><strong>7.5236</strong></td>
</tr>
<tr>
<td></td>
<td>(0.3482) ***</td>
<td>(1.9397) ***</td>
<td>(3.327) **</td>
</tr>
<tr>
<td>Historical city center in surrounding neighborhoods</td>
<td><strong>0.0521</strong></td>
<td><strong>1.2362</strong></td>
<td><strong>1.7907</strong></td>
</tr>
<tr>
<td></td>
<td>(0.0435)</td>
<td>(0.3828) ***</td>
<td>(0.517) ***</td>
</tr>
<tr>
<td>High income households (%)</td>
<td><strong>-0.0079</strong></td>
<td><strong>0.1634</strong></td>
<td><strong>0.2618</strong></td>
</tr>
<tr>
<td></td>
<td>(0.0087)</td>
<td>(0.0577) ***</td>
<td>(0.0812) ***</td>
</tr>
<tr>
<td>Distance to the nearest 100,000 jobs (km)</td>
<td><strong>-0.1323</strong></td>
<td><strong>-0.1383</strong></td>
<td><strong>-0.1692</strong></td>
</tr>
<tr>
<td></td>
<td>(0.0285) ***</td>
<td>(0.0922)</td>
<td>(0.1393)</td>
</tr>
<tr>
<td>Constant</td>
<td><strong>15.5797</strong></td>
<td><strong>317.8204</strong></td>
<td><strong>451.915</strong></td>
</tr>
<tr>
<td></td>
<td>(6.661) **</td>
<td>(95.043) ***</td>
<td>(124.15) ***</td>
</tr>
<tr>
<td>Price instrumented</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>High income households instrumented</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>F-statistic</td>
<td>11.427</td>
<td>6.598</td>
<td></td>
</tr>
</tbody>
</table>
## Marginal willingness to pay

### Marginal willingness to pay – in terms of house prices – for homeowners

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Income (+10,000)</td>
<td>Employed</td>
<td>Retired</td>
</tr>
<tr>
<td>Historic city center (+km2)</td>
<td>€40,274</td>
<td>€2,175</td>
<td>€-1,300</td>
</tr>
<tr>
<td>Historic city center in surrounding n'hoods (+km2)</td>
<td>€9,842</td>
<td>€91</td>
<td>€-84</td>
</tr>
<tr>
<td>High income households (+%)</td>
<td>€1,414</td>
<td>€137</td>
<td>€55</td>
</tr>
<tr>
<td>Distance to nearest 100,000 jobs (-km)</td>
<td>€644</td>
<td>(ns)</td>
<td>€10</td>
</tr>
</tbody>
</table>

- High income households prefer to live in neighbourhoods with a high concentration of high income households (Social interaction effect)
- High income households prefer to live in or around the historic city centre (+ its endogenous amenities)
Simulation if we...

- **Eliminate the historical center of Amsterdam**
  - Exploit the equilibrium property of the sorting model

<table>
<thead>
<tr>
<th>Neighborhoods</th>
<th>Standardized house price (in euros)</th>
<th>Predicted house price (in euros)</th>
<th>Difference</th>
<th>Percentage</th>
<th>Percentage rich households</th>
<th>Predicted percentage rich households</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amstel III en Bullewijk</td>
<td>119,581</td>
<td>191,581</td>
<td>+72,000</td>
<td>+60%</td>
<td>11.3%</td>
<td>12.6%</td>
</tr>
<tr>
<td>Bijlmer-Oost E, G en K</td>
<td>144,981</td>
<td>180,890</td>
<td>+35,909</td>
<td>+25%</td>
<td>19.1%</td>
<td>16.7%</td>
</tr>
<tr>
<td>Bijlmer-Centrum D, F en H</td>
<td>146,714</td>
<td>181,313</td>
<td>+34,599</td>
<td>+24%</td>
<td>17.4%</td>
<td>15.3%</td>
</tr>
<tr>
<td>Grachtengordel-Zuid</td>
<td>359,220</td>
<td>204,869</td>
<td>-154,351</td>
<td>-43%</td>
<td>31.6%</td>
<td>32.6%</td>
</tr>
<tr>
<td>Grachtengordel-West</td>
<td>359,694</td>
<td>204,790</td>
<td>-154,904</td>
<td>-43%</td>
<td>32.4%</td>
<td>33.2%</td>
</tr>
<tr>
<td>Museumkwartier</td>
<td>380,141</td>
<td>210,465</td>
<td>-169,676</td>
<td>-45%</td>
<td>37.0%</td>
<td>41.1%</td>
</tr>
</tbody>
</table>

- How to explain the shift in share of rich households (work in progress)
Social interactions

- **Strong impact of the share of high income households on the attractiveness of neighbourhoods**

- **What is behind this results?**
  - People want to meet high income households
    - e.g. want their children to go to school with children from high income households
  - High income households attract shops, restaurants, ... to neighborhoods
    - That are also appreciated by others → Multiplier effect

- **Simple regressions**
  - Does the concentration of high income households explain the number of shops, musea, theatre performances?
### Shops and high income households

Simple regressions of different type of shops

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grocery shops (#)</td>
<td>Fashion &amp; Luxury shops (#)</td>
<td>Leisure &amp; culture shops (#)</td>
<td>Musea (#)</td>
<td>Theatre performances (#)</td>
</tr>
<tr>
<td>OLS (se)</td>
<td>17.0435 **</td>
<td>59.7460 **</td>
<td>134.6857 ***</td>
<td>5.1371 ***</td>
<td>1.5641 ***</td>
</tr>
<tr>
<td>(7.0643)</td>
<td>(24.6781)</td>
<td>(36.8945)</td>
<td>(1.1375)</td>
<td>(0.4568)</td>
<td></td>
</tr>
<tr>
<td>Population (#)</td>
<td>0.0035 ***</td>
<td>0.0024 **</td>
<td>0.0034 **</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(0.0004)</td>
<td>(0.0011)</td>
<td>(0.0014)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>High income households (%)</td>
<td>-0.0365</td>
<td>0.1241</td>
<td>-0.0026</td>
<td>-0.0054</td>
<td>-0.0022</td>
</tr>
<tr>
<td>(0.0352)</td>
<td>(0.1306)</td>
<td>(0.1224)</td>
<td>(0.0071)</td>
<td>(0.0040)</td>
<td></td>
</tr>
<tr>
<td>High income households in surrounding neighborhoods (%)</td>
<td>-0.1808 **</td>
<td>-0.5672 **</td>
<td>-1.2462 ***</td>
<td>-0.0199</td>
<td>-0.0150 **</td>
</tr>
<tr>
<td>(0.0727)</td>
<td>(0.2685)</td>
<td>(0.3374)</td>
<td>(0.0171)</td>
<td>(0.0080)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>7.6977 **</td>
<td>15.5943 *</td>
<td>45.0938 ***</td>
<td>1.4116 **</td>
<td>0.7394</td>
</tr>
<tr>
<td>(3.2179)</td>
<td>(9.2797)</td>
<td>(13.7106)</td>
<td>(0.5963)</td>
<td>(0.3460)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>290</td>
<td>290</td>
<td>290</td>
<td>231</td>
<td>231</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.6155</td>
<td>0.2946</td>
<td>0.5570</td>
<td>0.5051</td>
<td>0.3949</td>
</tr>
</tbody>
</table>
Conclusions

- **Strong impact of cultural heritage on attractiveness of neighbourhoods**
  - Especially, high income households are willing to pay more for living in or close to a historic city center

- **Social interactions**
  - Households prefer to live in neighbourhoods where high income households reside
  - Especially other high income households

- **Simulation**
  - Even without cultural heritage, high income households cluster in and around the city center

- **Endogenous amenities cannot explain this**
Thank you for attending!
Questions?

Mark van Duijn  m.van.duijn@vu.nl
Jan Rouwendal

Department of Spatial Economics

Seminar
Utrecht School of Economics
25 September 2013