Integrated Data Sets as a Source of Information on Refugees

Dr. Gunter Brückner, Jan Eberle – German Statistical Office
Introduction

Administrative data on residence status of foreigners
- Central Register of Foreigners (CRF)

Survey data on socio-economic situation of foreigners
- Labor Force Survey (LFS)

Combine data sources to fill data gaps on refugees
Data Integration Methods

- Record Linkage
  - UPIs
  - Probabilistic

- Statistical Matching
  - Parametric
  - Hot-Deck
    - Random
## Statistical Matching

<table>
<thead>
<tr>
<th>Donor</th>
<th>Recipient</th>
<th>Synthetic Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>$C$</td>
<td>$\hat{R}$</td>
</tr>
<tr>
<td>$S$</td>
<td>$C$</td>
<td>$C$</td>
</tr>
<tr>
<td>$S$</td>
<td>$S$</td>
<td>$S$</td>
</tr>
</tbody>
</table>

Donor: $R$, $S$  
Recipient: $C$  
Synthetic Dataset: $\hat{R}$, $C$, $S$
Random Hot Deck Matching

<table>
<thead>
<tr>
<th>Donor (CRF)</th>
<th>Recipient (LFS)</th>
<th>Synthetic Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>C</td>
<td>♂</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>♂</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>♂</td>
</tr>
<tr>
<td>♂</td>
<td>C</td>
<td>♂</td>
</tr>
<tr>
<td>♂</td>
<td>S</td>
<td>♂</td>
</tr>
</tbody>
</table>

© German Statistical Office, Migration and Integration Statistics
Case-Study

Integrating 2016 LFS and Central Register of Foreigners

Focus on arrivals between 2013 and 2015
  - Allows for comparison with benchmark survey

Random hot deck imputation algorithm
  - Donation classes: Country of origin and year of first arrival
Case-Study

Assumption of conditional independence

\[ P(S, R \mid C) = P(S \mid C) \times P(R \mid C) \]

Explanatory power using country of origin and year of arrival

- 70% of refugees correctly predicted
- 85% of non-refugees correctly predicted

© German Statistical Office, Migration and Integration Statistics
Results

**Education**

<table>
<thead>
<tr>
<th>Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low ISCED 0-2</td>
<td>50±5</td>
</tr>
<tr>
<td>Medium ISCED 3-4</td>
<td>30±3</td>
</tr>
<tr>
<td>High ISCED 5-8</td>
<td>10±1</td>
</tr>
</tbody>
</table>

Note: The graph depicts the percentage of people seeking humanitarian protection between age 18 and 65 in Germany between 2013 and 2015 with high, medium and low educational attainments according to ISCED. Error bars indicate minimum and maximum outcomes after one hundred imputations, the colored bar represents the average outcome after one hundred imputations.

**Employment**

<table>
<thead>
<tr>
<th>Year after arrival</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st year</td>
<td>10±5</td>
</tr>
<tr>
<td>2nd year</td>
<td>20±5</td>
</tr>
<tr>
<td>3rd year</td>
<td>30±5</td>
</tr>
</tbody>
</table>

Note: The graph depicts the percentage of people seeking humanitarian protection between age 18 and 65 in Germany between 2013 and 2015 who are employed according to the ILO-concept of employment. Error bars indicate minimum and maximum outcomes after one hundred imputations, the colored bar represents the average outcome after one hundred imputations.

Random hot deck imputation
Benchmark: German refugee survey
Results

- **Percentage with medium or high education**
  - Refugees: 33%
  - Benchmark: 66%

- **Percentage below age 30**
  - Refugees: 100%
  - Benchmark: 0%

- **Percentage in employment**
  - Refugees: 100%

- **Average monthly net income**
  - Refugees: €1000, €1500, €2000, €2500
  - Benchmark: €1500, €1000, €2000, €2500

© German Statistical Office, Migration and Integration Statistics
Conclusion and outlook

Preliminary results are promising

Open questions

- Adjusting LFS survey weights
- Improving explanatory power

Provide case study for IRRS compilers manual
Thank you for your attention