Geography and Social Space: Evidence from Watersheds and Dialects

Kalle Kappner (HU Berlin)
Nikolaus Wolf (HU Berlin)
with Sascha O. Becker (Monash)
and Marvin Suesse (Trinity)

DH Coffee Talks

December 2021
Motivation

- How does Physical Geography shape Social Space?
- Economics focus on
  - endowments
  - disease environments
  - cost of mobility
This Project

- Can watersheds (drainage divides) affect the intensity of social interaction across space? How large and how persistent are these effects?
This Project

Figure 1: Example map
Data and empirical strategy

- Use data on German dialects to capture persistent features of social space: Digitaler Wenker-Atlas (DiWA) based on Wenker (1887)

- Use data on European watersheds with Pfaffstetter Coding: Catchment Characterisation and Modelling (CCM2)

- Map these two spatially continuous measures to discrete grid cells

- Merge and test for causal effects of watersheds on dialect space
Data and empirical strategy

- Correlate similarity measures on the cell-pair level:

\[ \mathbf{I}^i = \{ l^i_1, l^i_2, \ldots, l^i_K \} \text{ with } l^i_k \in \{0, 1\} \text{ for } K \text{ realizations} \]

\[ L_{ij} = \left( \mathbf{I}^i \times \mathbf{I}^j \right) / M \in [0, 1] \]

- “Pfaffstetter similarity”:

\[ P_{ij}^p = \begin{cases} 1 & \text{if } i, j \text{ in same watershed at Pfaffstetter level } p \\ 0 & \text{else} \end{cases} \]

\[ P_{ij} = \frac{\sum_{p=1}^{9} P_{ij}^p}{9} / 9 \in [0, 1] \]

- Control for cell fixed effects and elevation-corrected distance (Tobler, 1993):

\[ L_{ij} = \alpha_i + \alpha_j + \beta P_{ij} + \gamma \text{dist}_{ij} + \epsilon_{ij} \]
Linguistic similarity: Wenker’s original maps

Georg Wenker's original map no. 5
Shading shows variation in the pronunciation of 'flieg-en'.
Linguistic similarity: Wenker’s original maps

Georg Wenker's original map no. 5

Shading shows variation in the pronunciation of 'flieg-en'.
Linguistic similarity: Our measure

Linguistic similarity with Berlin
13 Categories. Darker shading indicates higher similarity.
Pfaffstetter similarity: European watersheds

CCM2 watersheds
Map shows 201 distinct watersheds at the 5th Pfaffstetter level.
Pfaffstetter similarity: Our measure

Pfaffstetter similarity with Berlin
9 Categories. Darker shading indicates higher similarity.
Some preliminary correlations

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>OLS</th>
<th>OLS</th>
<th>PPML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep. var.: Linguistic similarity</td>
<td>.52***</td>
<td>.51***</td>
<td>.40***</td>
<td>.34***</td>
</tr>
<tr>
<td>Pfaffstetter similarity</td>
<td>(.01)</td>
<td>(.00)</td>
<td>(.01)</td>
<td>(.01)</td>
</tr>
<tr>
<td>ln(distance in km)</td>
<td>.018</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
</tr>
<tr>
<td>Cell FEs no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>N</td>
<td>3176460</td>
<td>3176460</td>
<td>3176458</td>
<td>3176458</td>
</tr>
<tr>
<td>R2</td>
<td>.14</td>
<td>.14</td>
<td>.35</td>
<td>.04</td>
</tr>
</tbody>
</table>

Table 1: Regression results. Note: Marginal effects. Both similarity measures run from 0 to 1. Robust SE clustered at the grid level in parantheses.
Methodological Challenges

- Measuring dialect similarity
- Measuring topological watersheds
- Definition of observations (and N)
- Omitted variables, spatial correlation
- Modern Outcome Variables at fine-level of disaggregation, direction-specific data?
Outlook

- Watershed topology with potentially large effect on social spaces
- Need more and better data, tackle methodological challenges