Ethnic Cleavages in EPR–ETH
A new dataset

Nils-Christian Bormann,
Manuel Vogt, Seraina Rüegger
ETH Zürich

Prepared for the ENCoRe Meeting in Oslo from 21st -22nd September 2012
Goals

- Map the content of ethnicity
The Data

• For each group in EPR we coded
  – the three largest language segments
  – the three largest religious segments

  – usually only if segment > 10% of group
  – no bilingulaism, no syncretism
The Origins

• Language
  – Ethnologue

• Religion
  – Joshua Project
  – Library of Congress
  – Regional experts
  – Additional case studies
### What does it look like?

<table>
<thead>
<tr>
<th>Group</th>
<th>State</th>
<th>Religion1</th>
<th>Rsize1</th>
<th>Religion2</th>
<th>Rsize2</th>
<th>Religion3</th>
<th>Rsize3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites</td>
<td>USA</td>
<td>Protestants</td>
<td>.56</td>
<td>Atheists</td>
<td>.2</td>
<td>Catholics</td>
<td>.17</td>
</tr>
<tr>
<td>Latino</td>
<td>USA</td>
<td>Catholics</td>
<td>.7</td>
<td>Protestants</td>
<td>.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blacks</td>
<td>USA</td>
<td>Protestants</td>
<td>.78</td>
<td>Atheists</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Language

<table>
<thead>
<tr>
<th>Group</th>
<th>State</th>
<th>Language1</th>
<th>Lsize1</th>
<th>Language2</th>
<th>Lsize2</th>
<th>Language3</th>
<th>Lsize3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheduled Castes &amp; Tribes</td>
<td>India</td>
<td>Hindi</td>
<td>.4</td>
<td>Marathi</td>
<td>.06</td>
<td>Telugu</td>
<td>.06</td>
</tr>
<tr>
<td>Other Muslims</td>
<td>India</td>
<td>Urdu</td>
<td>.75</td>
<td>Bengali</td>
<td>.17</td>
<td>Malayalam</td>
<td>.056</td>
</tr>
<tr>
<td>Kashmiri Muslims</td>
<td>India</td>
<td>Kashmiri</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Descriptive Statistics

Unique number of categories

- Count
  - Languages
  - Religions

Average number of categories by group

- Mean
  - Languages
  - Religions

Average segment size (all)

- Mean
  - Language
  - Religion

Average segment size

- Mean
  - Languages
  - Religions
Limitations

- Mostly time-invariant
- Data from the 2000s
- Validity of Joshua Project as a non-academic source
What can we do with it?

• Investigate
  – ethnic identity

• Compute
  – ethnic differences
  – ethnic distances
  – ethnic group fragmentation
Dyadic Arguments

Dyadic Arguments

- Who is the (in) government/polity?
  1. All included groups
  2. Most powerful (and largest) ethnic group
Application: Clash of Civilizations?
Hypotheses to test

• Clash of Civilizations (Huntington)

• Cleavages (Chandra, Stein & Rokkan)

• Primordial vs constructivist arguments (Laitin, Toft, Fearon, etc.)

• Conflict leads to more cohesion (Kalyvas)
Database connections

- Socio-economic country-year
  - EPR-ETH
  - Docking
  - NSA
  - ACD
  - EC
    - Ethnologue
    - Religious Hierarchies
• **Ethnologue Definition based on ISO 639-3**
  – Two related varieties are normally considered varieties of the same language if speakers of each variety have inherent understanding of the other variety at a functional level (that is, can understand based on knowledge of their own variety without needing to learn the other variety).
  – Where spoken intelligibility between varieties is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central variety that both understand can be a strong indicator that they should nevertheless be considered varieties of the same language.
  – Where there is enough intelligibility between varieties to enable communication, the existence of well-established distinct ethnolinguistic identities can be a strong indicator that they should nevertheless be considered to be different languages.

  http://www.ethnologue.com/ethno_docs/introduction.asp
Definitions: Language

- Indo-European
  - Germanic
  - Celtic
  - Indo-Iranian
  - Italic
  - ...  

- North
  - East Scandinavian
    - Danish-Swedish
      - Danish-Bokmal
      - Danish-Riksmal
      - Norwegian
      - Swedish
  - High German
    - German
      - Middle German
      - East Middle German
      - German, Standard
      - German, Swiss
      - Swabian
      - Upper German
      - Allemanic
Definitions: Religion

• Geertz: “[A] system of symbols which acts to establish powerful, pervasive, and long-lasting moods and motivations in men by formulating conceptions of a general order of existence and clothing these conceptions with such an aura of factuality that the moods and motivations seem uniquely realistic.”

• We define religious adherence as confession towards a specific religion.
Definitions: Religion

Christianity

- Catholic
- Eastern Orthodox
  - Greek
  - Romanian
  - Russian
  - ...
- Oriental Orthodox
  - Coptic
  - Armenian Apostolic
  - Ethiopian
  - ...
- Protestant
  - Baptism
    - Lutheranism
    - Methodism
  - ...
- Anglican
- Others