Empirical
International and Regional Economics
Introduction

Jun-Prof. Dr. Wolfgang Dauth

Winter Term 2018/19
0.1 Administrative Issues
Goals of the course

Main goal: Students learn to carry out empirical studies in International and Regional Economics

Lecture (Room 226, Wednesday, 14:15-15:45)
  ▶ We will discuss various topics on the basis of current research studies (will be provided one week in advance)
    ▶ We cover the theoretical intuition of each topic,
    ▶ but leave more room for empirical applications
      “If applied econometrics was easy, theorists would do it.”
      (“Last Words” in Angrist/Pischke 2009)
Goals of the course

Lab exercise (CIP-Pool II (409), Thursday, 16:00-17:30)

▶ Hands-on exercise using Stata
  ▶ Familiarization with Software
  ▶ Learn to prepare *raw, real-world* datasets
  ▶ Get to know your data by obtaining descriptive statistics and making nice graphs
  ▶ Carry out regression analyses and report results

▶ Clean code will be provided after each lab
Literature

Textbook
▶ Wooldridge, J. (2006), Introductory Econometrics, Thomson South-Western, chapters 2-4, 6-7, 14.1, 15

Research Papers
▶ will be provided as we go along

Further Reading
▶ Angrist, J.D., J.-S. Pischke (2009), Mostly Harmless Econometrics, Princeton University Press  
  (Not “harmless” at all, but definitely an essential companion for any empirical Masters-/PhD-Thesis)
Assessment

Midterm Paper: During the semester, students will write a paper on one of four selected lab tutorials. The students will carry out their own empirical analysis. In the paper (max 10 pages), they discuss the relevance of the topic, provide an intuition on the underlying theory, explain each step of their empirical analysis, and report and interpret the results. In order to obtain a (very) good grade, students must go beyond a mere replication of the tutorial. The appendix of the paper must include a clean version of the code.

Final Paper: In the final week, Students will be given a current empirical paper. They will write a critical comment (max. 5 pages) on the empirical strategy of the paper and discuss its results.

Grade: The final grade is the weighted average (2/3 + 1/3) of the grades of the two papers.
Chapter outline

0 Introduction

1 Administrative issues
2 Motivation: Stylized facts about international trade and (local) labor markets
3 Causality
4 Data sources / Types of datasets

I Firm heterogeneity and the exporter wage premium

II The Gravity Model of Trade – theory and estimation

III International trade and local labor markets

IV Geographical concentration of economic activity
0.2
Motivation: Stylized Facts about International Trade and (Local) Labor Markets
Trade Volumes

Source: UN Comtrade

German World imports and exports
Whom we trade with

Source: UN Comtrade

German imports and exports by origin and destination country
What we Trade

Source: UN Comtrade

German World imports and exports by product category
What Determines Trade

Source: UN Comtrade, CEPII

Japan’s imports from the EU, 2005
What Determines Trade

Source: UN Comtrade, CEPII

Japan’s exports to the EU, 2005

W. Dauth

Empirical International and Regional Economics
What Determines Trade II

Source: UN Comtrade, CEPII

France’s imports, 2005
What Determines Trade II

Source: UN Comtrade, CEPII

France’s exports, 2005
The decline of the manufacturing sector

<table>
<thead>
<tr>
<th>Year</th>
<th>Western Germany</th>
<th>Eastern Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>7.019</td>
<td>0.877</td>
</tr>
<tr>
<td>1985</td>
<td>6.473</td>
<td>0.799</td>
</tr>
<tr>
<td>1990</td>
<td>7.003</td>
<td>0.739</td>
</tr>
<tr>
<td>1995</td>
<td>6.062</td>
<td>0.773</td>
</tr>
<tr>
<td>2000</td>
<td>5.717</td>
<td>0.816</td>
</tr>
<tr>
<td>2005</td>
<td>5.055</td>
<td>0.816</td>
</tr>
<tr>
<td>2010</td>
<td>4.813</td>
<td>4.627</td>
</tr>
<tr>
<td>2016</td>
<td>5.108</td>
<td>4.023</td>
</tr>
</tbody>
</table>

Source: IAB employment history (BeH)

Manufacturing employment, number in full-time equivalents and share in total workforce

Employment in 1 million fte
Why do we care about manufacturing?

Source: Statistics department of the BA and BeH

Unemployment rates and manufacturing in German counties

W. Dauth

Empirical International and Regional Economics
Regional disparities - Unemployment

Highest and Lowest Values:

Western Germany
- Eichstätt: 1.4%
- Gelsenkirchen: 13.1%

Eastern Germany
- Hildburghausen: 3.2%
- Uckermark: 11.2%

Source: BA monthly labor market reporting

Unemployment rates, July 2018
Regional disparities - Employment growth

Highest and Lowest Values:

Western Germany
Pirmasens  -31.7%
Eichstätt  75.4%

Eastern Germany
Cottbus  -49.8%
Teltow-Fläming  25.1%

Source: IAB employment history (BeH)

Emp. growth rates 1992-2016
Regional disparities - Wages

Highest and Lowest Values:

Western Germany
Wittmund 88.68 €
Erlangen 200.44 €*

Eastern Germany
Saale-Orla-Kreis 77.92 €
Berlin 115.78 €

Source: IAB employment history (BeH).
*Above social security contribution ceiling in 2010 of 125 €
Regional disparities - Manufacturing employment

Highest and Lowest Values:

Western Germany
Frankfurt 7.0%
Tuttlingen 59.99%

Eastern Germany
Potsdam 2.6%
Sonneberg 41.9%

Source: IAB employment history (BeH)
Regional disparities - Trade exposure

Figure: Imports

Figure: Exports
0.3 Causality
Causal Effects

The aim of empirical research is to find a *causal effect* of $x$ on $y$.

E.g.:

- the effect of exporter status on wages
- the effect of city size on wages
- the effect of import competition on (un)employment
- the effect of place based policies on (un)employment
Causality vs. correlation

Be aware of the difference between *correlation* and *causality*!

Would you conclude that more cops increase the crime rate?

Source:
Identification

How can we identify the causal effect?

- **Gold standard:** Randomized experiment
- Regression analysis
- Difference-in-differences
- Instrumental Variables

It is still useful to think about “What would be the ideal experiment?” before starting an econometric analysis.

This shapes the terminology in econometrics: Let’s call $y$ the “outcome” and $x$ the “treatment”!
0.4

Data Sources
Federal Statistical Office (Destatis)

www.destatis.de

- First address for general official statistics
- Especially useful for national accounts, population, specific sectors, …
- Also regional statistics (Regionaldatenbank)
- Often needs manual work to import to Stata. (Researchers apparently not main target group)
Process data from Federal Employment Agency from EACH person subject to social security from 1978-2010

Data often restricted to information relevant for administrative purposes: start/end dates, wages, qualification, place of work, industry, etc.

Reporting error, change of classifications, censoring at contribution ceiling

But still highly reliable, high coverage, high disaggregation possible (in theory)

Specialized datasets for each research question, additional surveys of establishments or people
Institute for Employment Research (IAB) cont’d

**BHP** Establishment History Panel: Aggregation of employment data at the establishment level. Perfect to characterize local labor markets

**SIAB** Sample of Integrated Labour Market Biographies: 2%-sample of all social security notifications (employment, marginal employment, UI benefit receipt, job-seeking behavior, participation in programs)

**Establishment Panel** Annual representative survey on various topics such as the determinants of labor demand on 16,000 plants

**LIAB** Matched data set of individual biographies of workers in plants from the Establishment Panel
UN Commodity Trade Statistics Database (Comtrade)

http://comtrade.un.org/

- 3.1 billion trade records starting from 1962
- Extremely detailed bilateral trade data at the level of 6-digit commodity codes, e.g.
  - 011.1 Meat of bovine animals, ...
  - 011.11 ... with bone in
  - 011.12 ... boneless
- Useful for a quick glimpse (e.g. who is Trinidad and Tobago’s biggest trading partner?), but otherwise servers often overloaded
- For bulk downloads, subscription (121 US-§) allows to run batch jobs
Others

- **KLEMS**: Industry level data for the analysis of growth and productivity patterns around the world
- **Penn World Table (PWT)**: data on capital, productivity, employment and population
- Tons of others, unfortunately often costly, not nicely prepared, or no good coverage
0.4

Types of Datasets
## Cross Sectional Data

<table>
<thead>
<tr>
<th>county</th>
<th>name</th>
<th>population</th>
<th>crime type 435 cases</th>
<th>crime type 435 solved</th>
<th>crime type 50 cases</th>
<th>crime type 50 solved</th>
</tr>
</thead>
<tbody>
<tr>
<td>01001</td>
<td>Flensburg</td>
<td>85,300</td>
<td>264</td>
<td>67</td>
<td>595</td>
<td>88</td>
</tr>
<tr>
<td>01002</td>
<td>Kiel</td>
<td>233,039</td>
<td>492</td>
<td>58</td>
<td>3784</td>
<td>188</td>
</tr>
<tr>
<td>01003</td>
<td>Lübeck</td>
<td>212,754</td>
<td>367</td>
<td>72</td>
<td>1637</td>
<td>109</td>
</tr>
<tr>
<td>01004</td>
<td>Neumünster</td>
<td>78,951</td>
<td>283</td>
<td>39</td>
<td>851</td>
<td>53</td>
</tr>
<tr>
<td>01051</td>
<td>Dithmarschen</td>
<td>137,428</td>
<td>218</td>
<td>42</td>
<td>708</td>
<td>42</td>
</tr>
<tr>
<td>16077</td>
<td>Altenburger Land</td>
<td>109,304</td>
<td>243</td>
<td>22</td>
<td>344</td>
<td>170</td>
</tr>
</tbody>
</table>

Source: Federal Criminal Office

- Sample of persons/firms/regions/countries/... observed at the same point in time
- In case of persons or firms: typically a random sample
## Time Series Data

<table>
<thead>
<tr>
<th>Year</th>
<th>GDP</th>
<th>Public Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>2,809,480</td>
<td>2,147,028</td>
</tr>
<tr>
<td>2012</td>
<td>2,749,900</td>
<td>2,160,951</td>
</tr>
<tr>
<td>2011</td>
<td>2,699,100</td>
<td>2,087,747</td>
</tr>
<tr>
<td>2010</td>
<td>2,576,220</td>
<td>2,059,174</td>
</tr>
</tbody>
</table>

Values in Million Euros. Source: German Federal Bank

- Represent the evolution of variables over time
- Time is a relevant factor: Past influences future, order of observations matters!
- Peculiarity: Seasonality, business cycle
Panel Data

<table>
<thead>
<tr>
<th>id</th>
<th>year</th>
<th>female</th>
<th>educ</th>
<th>wage</th>
<th>region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2008</td>
<td>0</td>
<td>4</td>
<td>18.8</td>
<td>674</td>
</tr>
<tr>
<td>1</td>
<td>2009</td>
<td>0</td>
<td>4</td>
<td>24.5</td>
<td>674</td>
</tr>
<tr>
<td>1</td>
<td>2010</td>
<td>0</td>
<td>5</td>
<td>50.2</td>
<td>674</td>
</tr>
<tr>
<td>2</td>
<td>2008</td>
<td>0</td>
<td>3</td>
<td>47.1</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>2009</td>
<td>0</td>
<td>3</td>
<td>46.3</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>2010</td>
<td>0</td>
<td>3</td>
<td>44.9</td>
<td>32</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>789</td>
<td>2008</td>
<td>1</td>
<td>5</td>
<td>50.3</td>
<td>56</td>
</tr>
<tr>
<td>789</td>
<td>2009</td>
<td>1</td>
<td>5</td>
<td>55.6</td>
<td>40</td>
</tr>
<tr>
<td>789</td>
<td>2010</td>
<td>1</td>
<td>5</td>
<td>56.7</td>
<td>40</td>
</tr>
</tbody>
</table>

Fake SIAB-Data

- Repeated (yearly) observations of each unit
### Spell Data

<table>
<thead>
<tr>
<th>id</th>
<th>begepi</th>
<th>endepi</th>
<th>plantID</th>
<th>female</th>
<th>educ</th>
<th>wage</th>
<th>region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01.01.2008</td>
<td>27.09.2008</td>
<td>34265564</td>
<td>0</td>
<td>2</td>
<td>50.26</td>
<td>74</td>
</tr>
<tr>
<td>1</td>
<td>28.09.2008</td>
<td>30.11.2008</td>
<td></td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>01.12.2008</td>
<td>31.12.2008</td>
<td>67643232</td>
<td>0</td>
<td>2</td>
<td>36.87</td>
<td>74</td>
</tr>
<tr>
<td>1</td>
<td>01.01.2009</td>
<td>31.12.2009</td>
<td>67643232</td>
<td>0</td>
<td>2</td>
<td>37.57</td>
<td>74</td>
</tr>
<tr>
<td>789</td>
<td>01.01.2008</td>
<td>31.03.2008</td>
<td>12560778</td>
<td>1</td>
<td>6</td>
<td>50.36</td>
<td>40</td>
</tr>
<tr>
<td>789</td>
<td>01.04.2008</td>
<td>30.06.2008</td>
<td>67868433</td>
<td>1</td>
<td>6</td>
<td>15.71</td>
<td>40</td>
</tr>
<tr>
<td>789</td>
<td>01.04.2008</td>
<td>30.06.2008</td>
<td>12560778</td>
<td>1</td>
<td>6</td>
<td>50.36</td>
<td>40</td>
</tr>
<tr>
<td>789</td>
<td>01.07.2008</td>
<td>31.12.2008</td>
<td>12560778</td>
<td>1</td>
<td>6</td>
<td>50.36</td>
<td>40</td>
</tr>
<tr>
<td>789</td>
<td>01.07.2008</td>
<td>31.12.2008</td>
<td>12560778</td>
<td>1</td>
<td>6</td>
<td>51.46</td>
<td>40</td>
</tr>
</tbody>
</table>

**Fake SIAB-Data**

- Repeated observations of each unit with daily precision