Working with Admin Data in Germany:
- (Still) Work in Progress
CHRISTINA GATHMANN

University of Heidelberg
and CESifo, IZA, ZEW

Research Interests:
Labor Economics, Migration, Policy Evaluation
Health, Education, Family Policy
Political Economy, Public Finance
GROWING IMPORTANCE OF ADMIN DATA

Average use of admin data in Top-5

28.5% OF ALL ARTICLES PUBLISHED IN TOP5 JOURNALS IN 2018
(ALMOST 50% OF EMPIRICAL PAPERS)

48% OF ALL ARTICLES PUBLISHED IN SECOND TIER JOURNALS IN 2018

USE OF ADMIN DATA WILL EXPAND EVEN MORE

(1) **OTHER FIELDS AND DISCIPLINES**

MANY FIELDS JUST BEGIN TO USE ADMIN DATA

EXA: ENVIRONMENTAL ECONOMICS, ECONOMIC HISTORY, MACROECONOMICS

(2) **RECENT DATA AND METHODOLOGICAL DEVELOPMENTS**

a. **LINK TO SURVEY DATA** (ADD VARIABLES LIKE PERSONALITY TRAITS, FIRM USE OF ROBOTS ETC.)

b. **LINK TO LAB OR FIELD EXPERIMENTS** (CASE WORKERS, INCENTIVES FOR TRANSFER RECIPIENTS)

c. **LINK ADMIN DATA ACROSS COUNTRIES** (CROSS-BORDER MOBILITY)

d. **PRIVATELY OWNED DATA** (SCANNEER DATA, ONLINE DATA LIKE FACEBOOK)

e. **LINK TO ONLINE OR BIG DATA**
CURRENT USE OF DATA

Direct Remote Access
(Norway, Finland)

SUF Files
(MZ, NEPS)

Indirect Remote Access
(IAB, NEPS)

Very Remote Access
(Census 2011)

VERY Remote Access
(Hospital, Firm Tax Data)
SCIENTIFIC-USE FILES

• **INDIVIDUAL SOCIAL SECURITY RECORDS (SIAB + BHP)**
  - VERY DETAILED CLASSIFICATIONS (INDUSTRY OR OCCUPATIONAL) NOT AVAILABLE
    Exa: Evaluation of the crafts and trade reform: need 4-digit occupation code
  - SMALL SAMPLE SIZES FOR DETAILED REGIONAL ANALYSIS
    Exa: College Openings at District Level: Education x Age Groups x District x Year

• **NATIONAL EDUCATIONAL PANEL SURVEY (NEPS)**
  - STATE INFORMATION NOT AVAILABLE
  - SOURCE COUNTRY INFORMATION NOT AVAILABLE
    Exa: Effects of Citizenship Reforms on Skill Development

• **MIKROZENSUS**
  - PANEL INFORMATION
  - BIRTH MONTH

EASY TO WORK WITH
GOOD FOR EARLY CAREER
FLEXIBLE WORK SCHEDULE
EASY REPLICABILITY
DIRECT REMOTE ACCESS

CLIENT (SSH+X-WIN32) ACCESS TO REGISTER DATA ABROAD

• MATCHED EMPLOYER-EMPLOYEE DATA – STATISTICS NORWAY

• MATCHED EMPLOYER-EMPLOYEE + HEALTH DATA – STATISTICS FINLAND

EASY TO HANDLE
GOOD FOR EARLY CAREER
FLEXIBLE WORK SCHEDULE
HEALTH SPILLOVERS AFTER JOB DISPLACEMENT

- MATCH RECORDS OF COUPLES (LIVING TOGETHER)
- MATCH LABOR MARKET INFORMATION + MORTALITY + HOSPITAL REGISTER DATA
- PANEL DATA OVER THREE DECADES

Men's Job Loss and Mortality

Women's Job Loss And Mortality
MANY SOCIAL SECURITY RECORDS

• SENSITIVE VARIABLES AVAILABLE
• MATCHED EMPLOYER-EMPLOYEE DATA (LIAB)
• OTHER MATCHED DATA (SURVEY + SOCIAL SECURITY RECORDS)

NATIONAL EDUCATIONAL PANEL STUDY (NEPS)

• SENSITIVE VARIABLES AVAILABLE

GREAT DATA
MORE FLEXIBLE THAN RDC
MATCH EXTERNAL DATA
INDIRECT REMOTE ACCESS

**RESTRICTIONS ON USE**

- ONE JOB PER DAY / ONE EXPORT PER WEEK
- TIME CONSUMING ESP. EARLY IN THE PROJECT
- EARLY CAREER RESEARCHERS NOT AS VERSATILE IN HANDLING LARGE DATA
- PROBLEMATIC ESP. FOR DOCTORAL STUDENTS WITH FIXED-TERM CONTRACTS

- EMPIRICAL RESEARCH MORE DATA-INTENSIVE
  - RESEARCH WORKS IN CYCLES
    (WEEKS OF INTENSE WORK, REVIEW PROCESS)
  - GRAPHS OFTEN TEDIOUS TO PRODUCE
**VERY REMOTE ACCESS**

**HOSPITAL STATISTIC**

- **DISSERTATION PROJECT**: ABOLISHING MILITARY/CIVIL SERVICE -> EFFECT ON HOSPITALS/PATIENTS?

- POPULATION OF HOSPITALS (TYPES OF HOSPITALS, INPUTS, COSTS)
- MATCHED TO INDIVIDUAL PATIENT DATA (DIAGNOSES, LENGTH OF STAY)
- RICH DATA

- GUEST RESEARCH STAY IN ONE OF THE RESEARCH DATA CENTERS
- „CONTROLLED REMOTE ACCESS“:
  - PREPARE ANALYSIS FILES (STATA) USING TEST DATASET
  - BUT VARIABLE NAMES AND DATA STRUCTURE DIFFERS FROM ACTUAL DATA
  - SEND DO-FILE TO EMPLOYEE -> RUNS DO-FILE ON PC -> CHECKS OUTPUT -> SENDS LOG-FILE
• **VALUE ADDED AND BUSINESS TAX DATA**

  • **DISSERTATION PROJECT**: HOW DO FIRMS REACT TO (ACTUAL) TAX AUDITS?
  • MATCH BOTH DATA SOURCES, DATA SPANS 2001-2014

  • **EACH STATE OFFICE HAS TO AGREE** (LÄNDERABSTIMMUNG) -> VETO ON OUTPUT CHECKS
  • „CONTROLLED REMOTE ACCESS“:
    • PREPARE ANALYSIS FILES (STATA) USING „DATA STRUCTURE FILES“ (FAKE DATA, N=100)
    • BUT: **FAKE DATA DOES NOT REFLECT ACTUAL COMPLEXITY OF 4 MIO. DATA POINTS**
    • SEND DO-FILE TO EMPLOYEE -> RUNS DO-FILE ON LOCAL PC -> CHECKS OUTPUT -> SENDS LOG-FILE IF OK
WHY DIRECT REMOTE ACCESS IS IMPORTANT

- **INCREASE IN FLEXIBILITY AND SPEED**
  - Flexible Work Schedule and Timing of Work

- **IMPORTANT FOR JUNIOR RESEARCHERS**
  - Less versatile, but time-constrained

- **COMPETITION HIGH-QUALITY DATA – RESEARCHERS**
  - Top-notch researchers go where the good data are
  - German social security data has generated enormous interest and fantastic research on German labor markets
    - ... influential social science research
    - ... informative evidence-based policy-advice
GROWING IMPORTANCE OF ADMIN DATA

Use of admin data in other journals

ECONOMICS BECOMES MORE EMPIRICAL

Source: Angrist et al. (2017), AER P&P
N = 134,892 articles published in 80 journals 1980-2015