



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

# 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

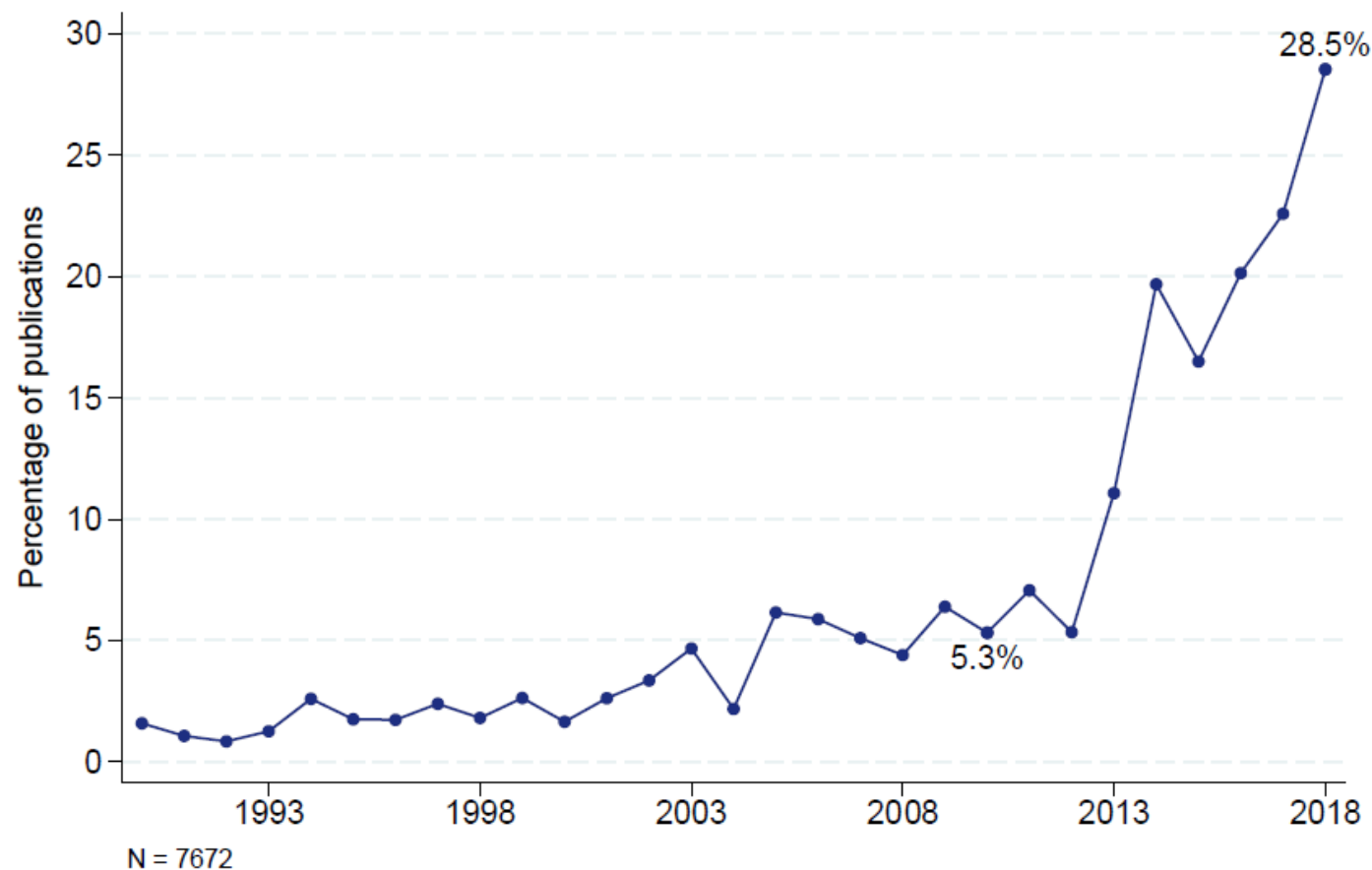
Political Economy, Public Finance





# 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

Source: M. Halla (2020), How did admin data change economic research output. A descriptive analysis.



# USE OF ADMIN DATA WILL EXPAND EVEN MORE



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

## (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

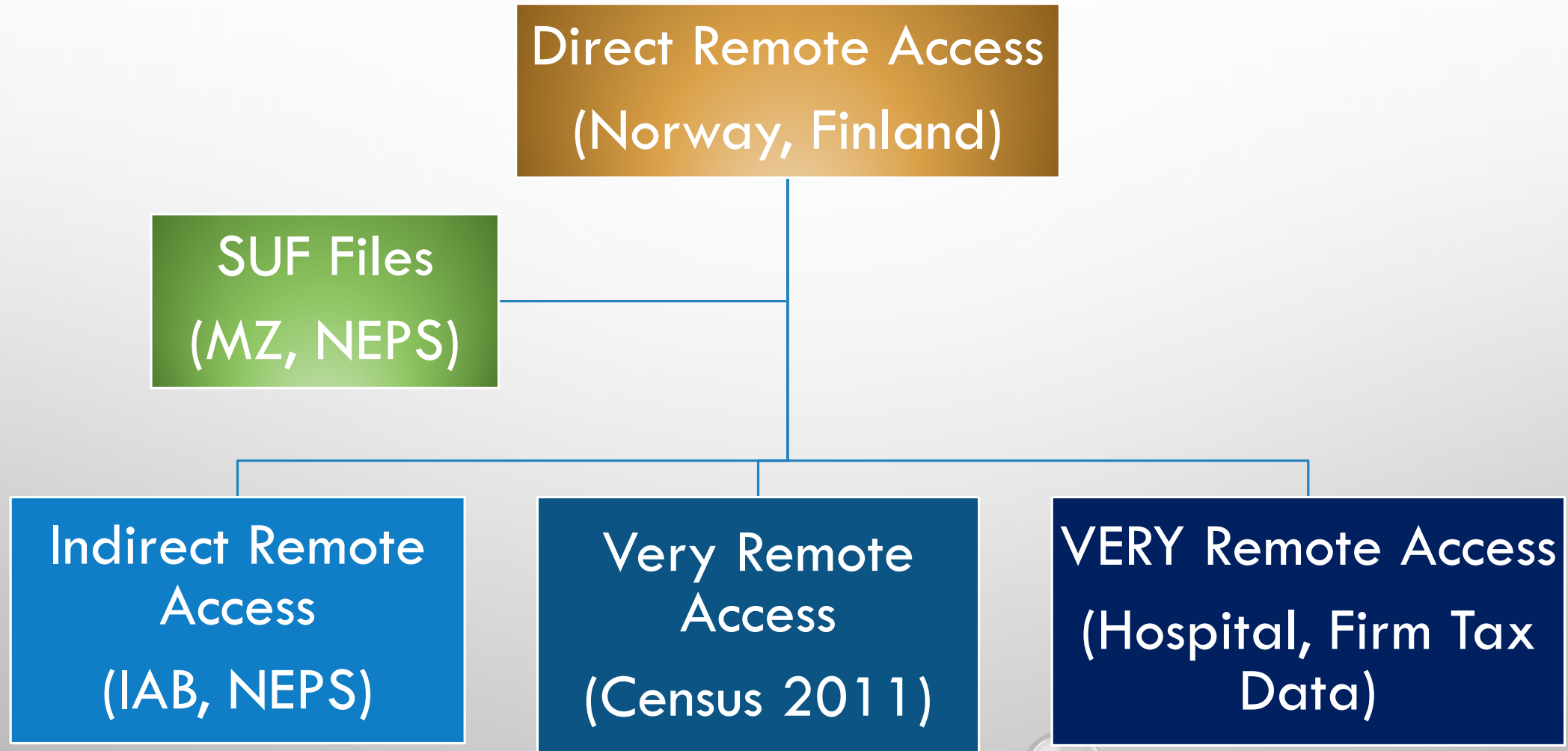
- a. LINK TO **SURVEY DATA** (ADD VARIABLES LIKE PERSONALITY TRAITS, FIRM USE OF ROBOTS ETC.)
- b. LINK TO **LAB OR FIELD EXPERIMENTS** (CASE WORKERS, INCENTIVES FOR TRANSFER RECIPIENTS)
- c. LINK ADMIN DATA **ACROSS COUNTRIES** (CROSS-BORDER MOBILITY)
- d. **PRIVATELY OWNED DATA** (SCANNER DATA, ONLINE DATA LIKE FACEBOOK)
- e. LINK TO **ONLINE OR BIG DATA**



# CURRENT USE OF DATA



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386





# SCIENTIFIC-USE FILES



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

- 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



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

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

- MATCHED EMPLOYER-EMPLOYEE DATA – **STATISTICS NORWAY**
- MATCHED EMPLOYER-EMPLOYEE + HEALTH DATA – **STATISTICS FINLAND**

Stata/IC 14.2 - /home/anna/data\_raw/analysis\_private1.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands here

#	Command	IC
1	use */home/anna/da...	
2	sum	

Results

Variable	Obs	Mean	Std. Dev.	Min	Max
lopnr_per-n	0				
pid	1,141,254	4963004	2874966	15	9940893
medlemsnr_r	0				
ygrk	0				
kjonn	0				
gender	488,145	1.273706	.4458601	1	2
ygr	0				
occ_stkodnho	488,145	2904.043	1521.493	1000	6909
mwage_nho	488,145	15723.01	5705.402	0	75000
tantieme_nho	488,145	80.76393	712.3167	0	33333
twage_nho	488,145	15803.77	5709.435	4500	77583
utd	0				
edu_nho	162,262	2.489042	1.620697	1	7
whours_nho	488,145	37.90898	1.823372	0	40
ygrf	0				
stkodf	488,145	2298.885	1750.797	0	6909
mwagef_nho	488,145	12108.66	7698.74	0	72720
tantiemef_o	488,145	62.9787	604.776	0	33333
twagef_nho	488,145	12171.64	7728.11	0	72720
utdf	0				
whoursf_nho	488,145	30.59087	15.05185	0	40
alder	1,141,254	39.09209	9.497809	18	61
firmsizegr_o	1,141,254	3.275663	.9801895	1	4
kodk	488,145	2.066052	1.010194	1	4
lavvik	485,811	25.49188	375.2742	0	29152
fnrkod	488,145	1.439644	.7949031	0	3

Variables

Filter variables here

#	Name	Label
	ybirth_child1	
	ybirth_child2	
	ybirth_child3	
	newbirth	
	newbirth_1	
	fsize_new	
	sector	
	sector_str	sector
	edu_years	years of education
	edu1	
	edu2	
	edu3	
	parttime	
	_merge	

Properties

Name: lopnr\_person

Label: Lopenummer person

Type: str8

Format: %8s

Value label: No notes

Notes: No notes

Data

EASY TO HANDLE

GOOD FOR EARLY CAREER

FLEXIBLE WORK SCHEDULE

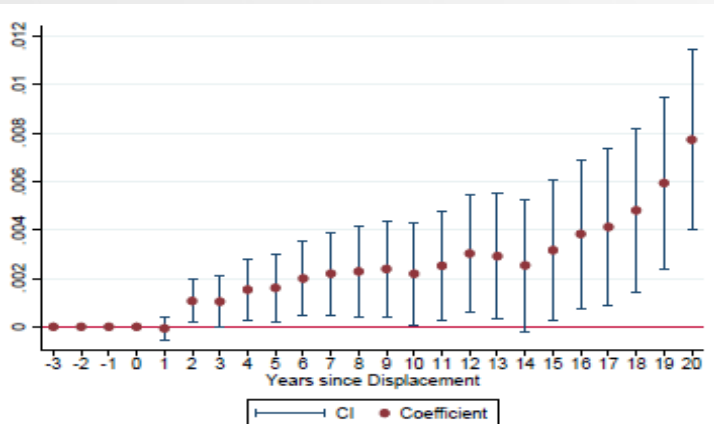


# HEALTH SPILLOVERS AFTER JOB DISPLACEMENT

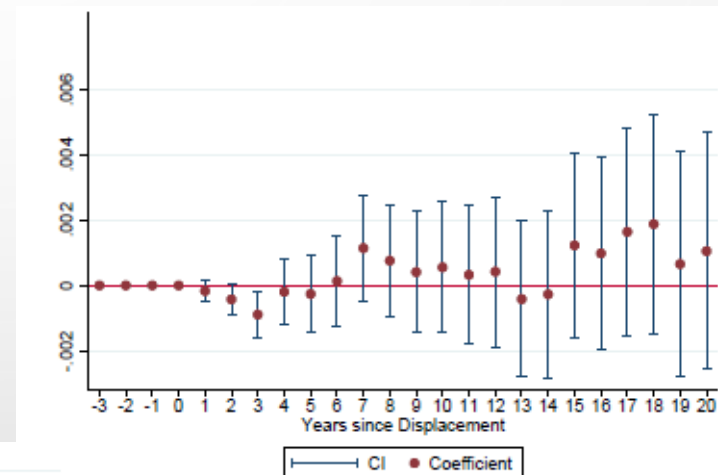
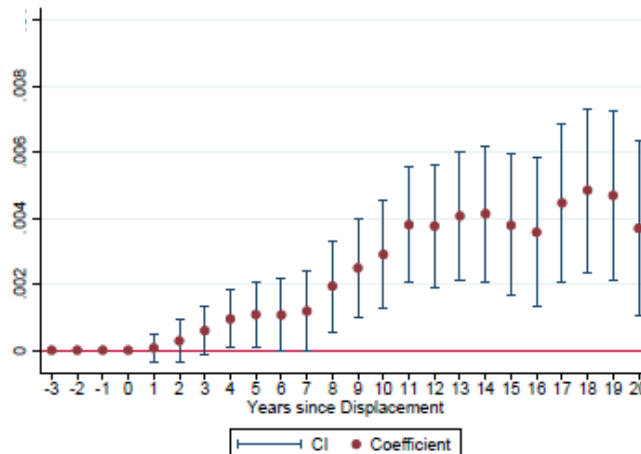


UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

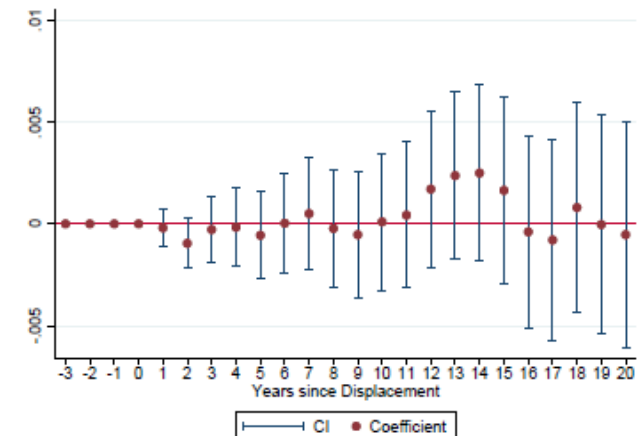
- MATCH RECORDS OF COUPLES (LIVING TOGETHER)
- MATCH LABOR MARKET INFORMATION + MORTALITY + HOSPITAL REGISTER DATA
- PANEL DATA OVER THREE DECADES



Men's Job Loss and  
Mortality



Women's Job Loss  
And Mortality





# INDIRECT REMOTE ACCESS



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

- 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



My Jobs Projects Job Submission You are: **doej9999**

Job List (Last 50)

Sent in	Project	Status	Censoring	Options
2016-06-10 12:42:05	FDZ9999	success	released	
2015-12-15 16:42:21	FDZ9999	failed	released	
2015-12-09 12:51:42 (presentation)	FDZ9999	success	released	

Showing 1 to 3 of 3 entries (filtered from 50 total entries)

GREAT DATA

MORE FLEXIBLE THAN RDC

MATCH EXTERNAL DATA



# INDIRECT REMOTE ACCESS



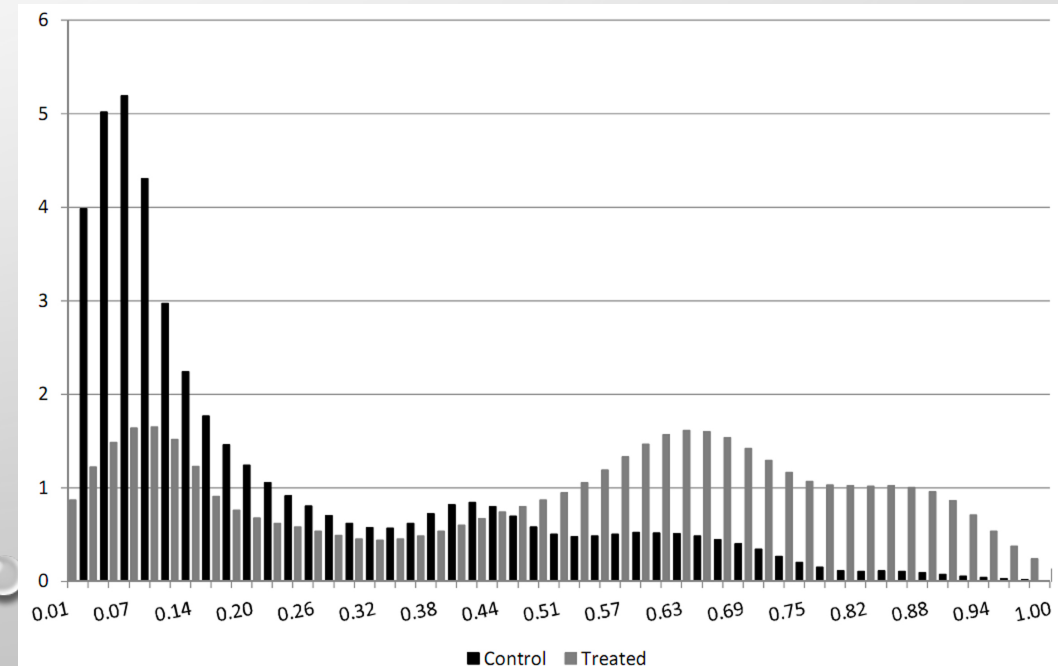
UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

- 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



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

- **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**



# VERY REMOTE ACCESS



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

- **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





UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

... PLEASE KEEP UP ...



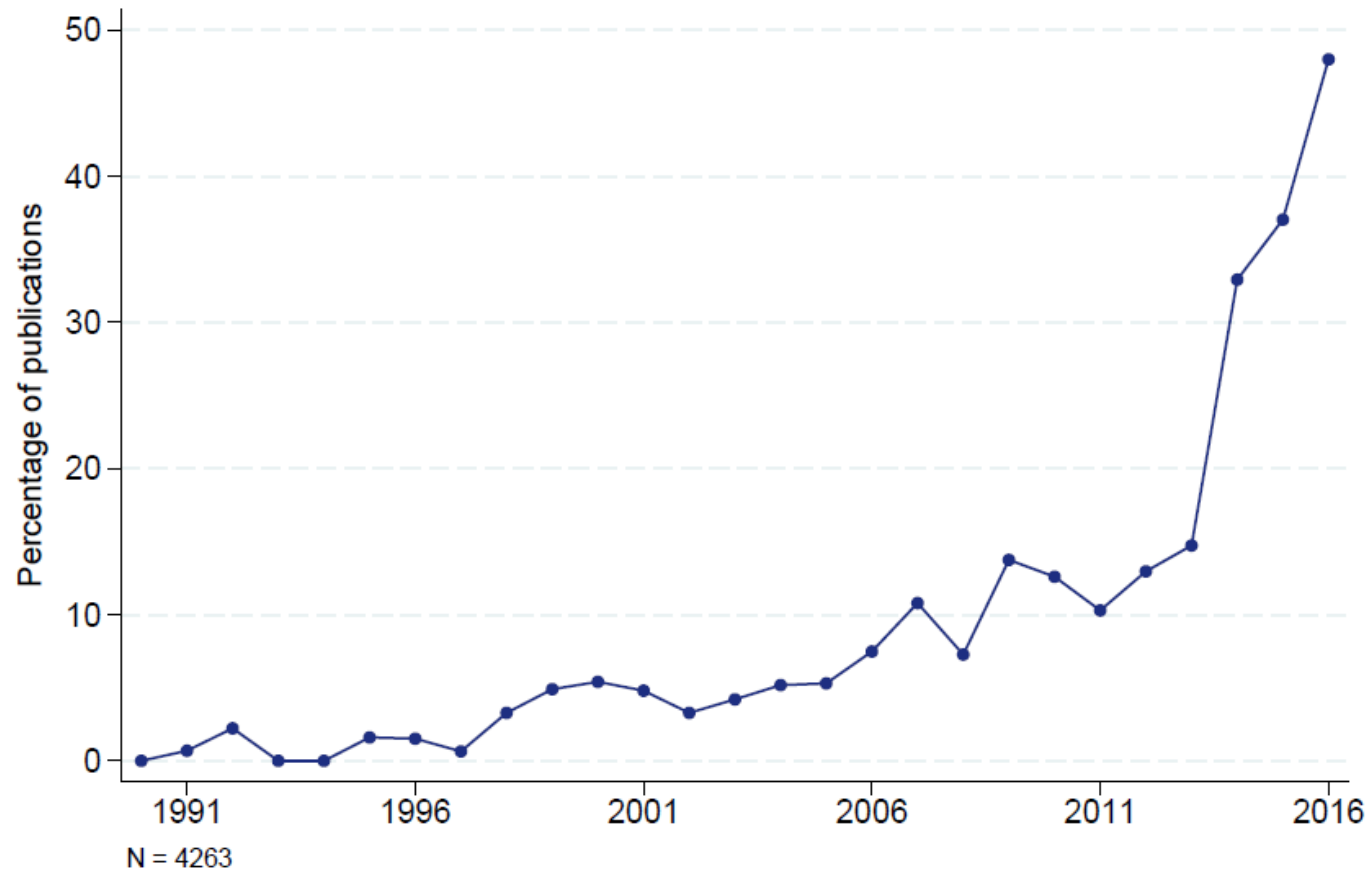


# GROWING IMPORTANCE OF ADMIN DATA



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

## Use of admin data in other journals



Source: M. Halla (2020), How did admin data change economic research output. A descriptive analysis.



# ECONOMICS BECOMES MORE EMPIRICAL



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

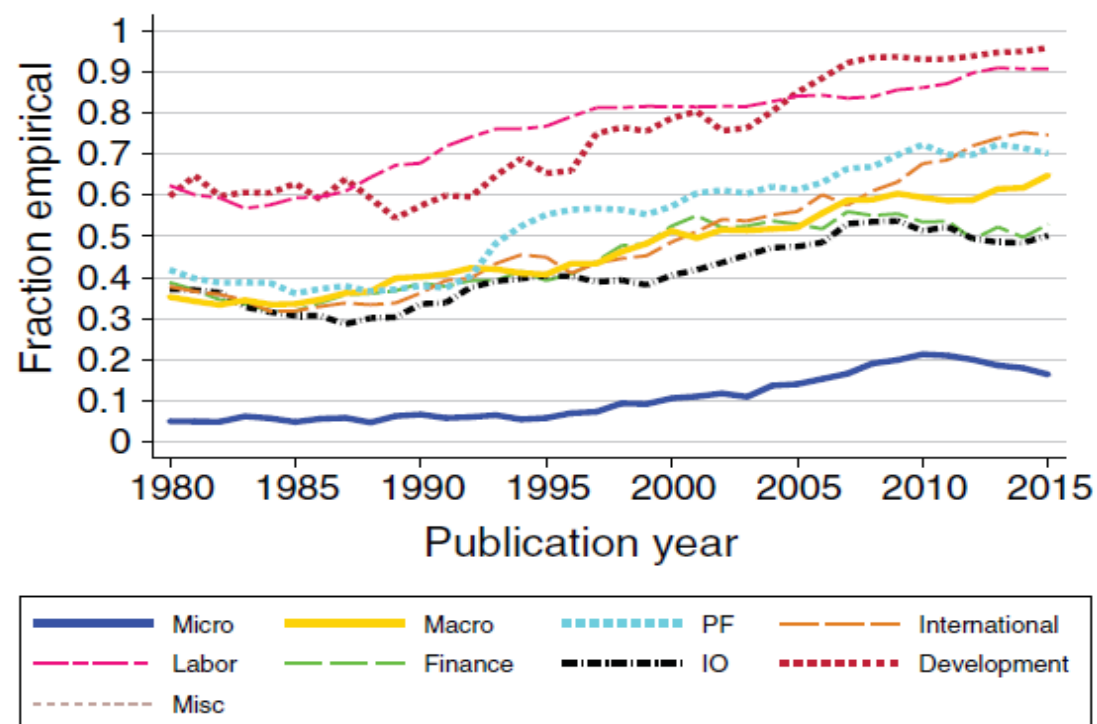


FIGURE 4. WEIGHTED FRACTION EMPIRICAL BY FIELD

*Note:* Five-year moving averages of the weighted fraction of publications in each field that are empirical.

Source: Angrist et al. (2017), AER P&P

N = 134,892 articles published in 80 journals 1980-2015