March 2023

Danish Register Data

A quick guide for researchers in BigFi and at Department of Finance, CBS

Julie Marx

Agenda

1. What’s in the data?
2. How to work with the data?
3. How to apply for access?

BIGFI

What is register data?

- Micro data on all individuals and firms in Denmark
- Collected by authorities, e.g
  - The Civil Register (CPR)
  - Tax authorities
  - The Central Bank
  - Ministry of Education
  - Ministry of Employment
  - Health authorities
  - The Housing Registry (BBR)
- Stored, anonymized, and made accessible to researchers by Statistics Denmark (DST)

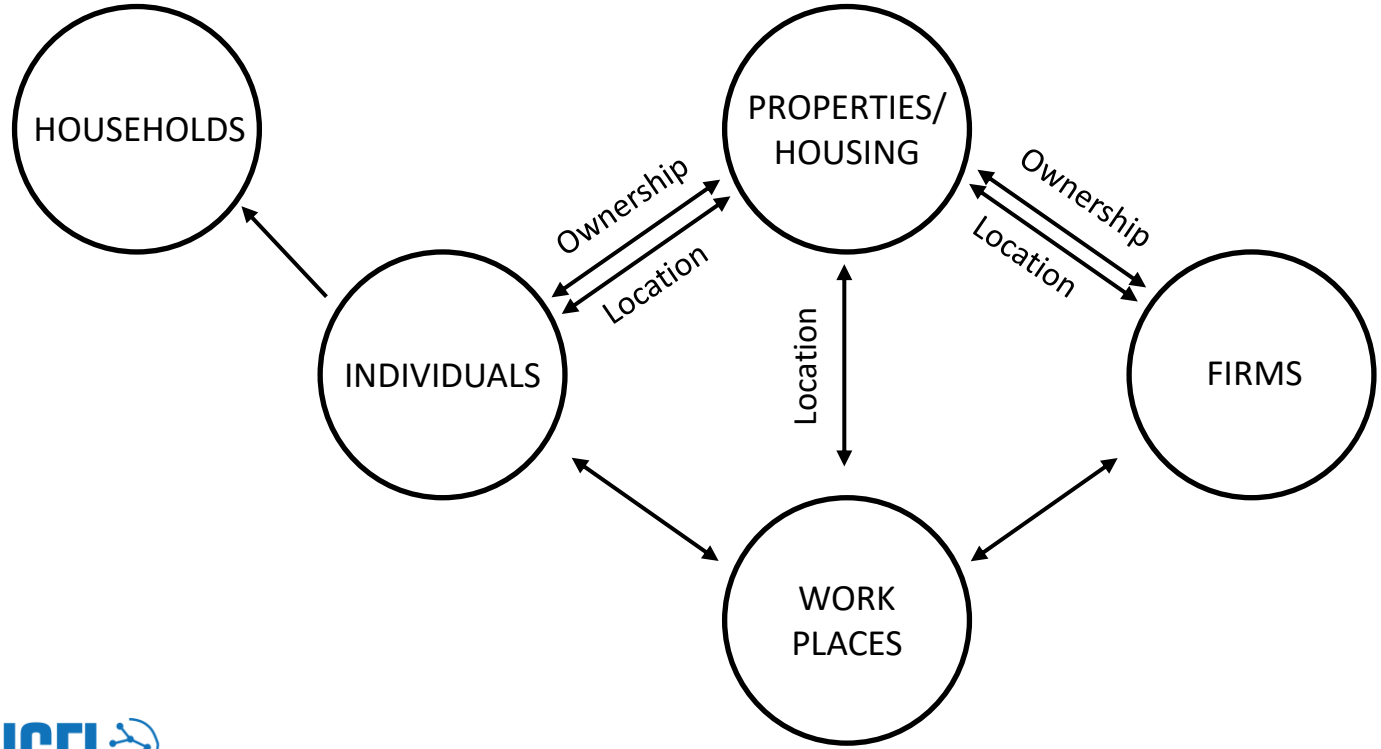
BIGFI

Units of observation

- Individuals / households
  - Annonymised CPR: Pnr
  - Household ID: Familie_id
- Firms and work places
  - Firm ID: CVRnr / SEnr
  - Work place ID: Lbnr
- Properties / housing
  - Property ID: Ejdnr
  - Address: Opgikom / bopikom

BIGFI

Linking the units



Individuals/households

- Demographics (age, gender, household type, children, location, citizenship, etc.)
- Income (total, labor, capital, pension, subsidies, disposable, tax, fringe benefits, and much more)
- Education (degrees and ongoing, type, length, institution, grades…)
- Employment (status, position, workplace(s), job-specific salary, experience…)
- Health (diagnoses, prescriptions, doctor visits, hospitalizations, costs, death causes…)
- Crime (reports, charges, sentences, imprisonments, types of crime)
- Wealth (bank deposits, stock holdings (total), bonds, bank debt, mortgage debt…)
- Mortgages (interest rate, maturity, LTV…)
- Interests (account-specific payments, bank affiliation, total payments and income)
- Pensions (payments, income, holdings)
- Portfolios (forthcoming) (trades of individual stocks and mutual funds)
- Real estate ownership (owner-property link, duration, ownership share)
- Cars (type, age, owner, user, estimated market value, CO2-emissions and many other details)

BIGFI

Firms / Work places

- Accounting information (assets, liabilities, revenue, earnings, investments… everything in the books)
- Sales and purchases
- Other firm characteristics (number of employees, industry, sector, growth, import/export…)
- Account-specific interest rates + bank affiliation
- Green goods and services (and reductions of non-green inputs)
- Entrepreneurs (identification of new firms and their founders)
- Energy use
- Compensations for self-employed
- Details on foreign trade
- Details on goods sold and purchased by manufacturing firms
- Covid wage compensation
- Firm IT use (survey)

BIGFI

Housing

- Location (adress code, zipcode, shire, municipality, region, rural/urban)
- Property characteristics (building size, lot size, rooms, floors, number of bathrooms, age, type of roof, type of sewer and much more like…)
- Property ownership (link to owners, owner type)
- Property sales (price, date, type of property)
- Property valuations (tax assessed value)

BIGFI

The full list of data

Grunddataoversigten

24,614 variables in total

Registre i Forskningsservices grunddatabank

Bestillingsliste

Opdateringsfrekvens Luk betyder registeret ikke opdateres.

Register	Registertitel	Særlige forhold	Referencetype	Opdateringsfrekvens	Første år	Start måned	Seneste år	Slut måned
ADOP	Adoptioner fra CPR		Hændelse	Luk	1988		2009	
AEFB	Ældredokumentation forebyggende hjemmebesøg		Status	År	2008		2021	
AEFV	Ældredokumentation visiteret hjemmehjælp frit valg		Status	År	2008		2021	
AEHJSP	Ældredokumentation - hjemmesygepleje i eget hjem		Hændelse	År	2016		2021	
AELH	Ældredokumentation leveret hjemmehjælp		Status	År	2011		2021	
AEPB	Ældredokumentation visiteret hjemmehjælp plejebolig		Status	År	2008		2021	
AEPI	Ældredokumentation borgere på plejehjem bolig (imputeret)		Status	Luk	2008		2016	
AEREHAB	Ældredokumentation - Rehabilitering		Status	Luk	2017		2018	
AERH	Ældredokumentation Rehabilitering månedsangivet		Status	År	2019		2021	
AETR	Ældredokumentation genoptræning vedligeholdelsestræning		Status	År	2008		2021	
AKAS	A-kasse medlemmer		Status	År	2007		2020	
AKM	Arbejdsklassifikationsmodulet		Status	År	1976		2020	
AMFO	Arbejdsmarkedsforanstaltninger (AMFORA)		Status	Luk	1994		2006	
AMRTN	Arbejdsmarkedsregnskab med timenormering		Status	År	2008		2020	
AMRUN	Arbejdsmarkedsregnskab uden timenormering		Status	År	2008		2020	
AMRUN_PARTIEL	Arbejdsmarkedsregnskab uden timenormering		Status	År	2019		2020	
ATF	Små og mellemstore virksomheders adgang til finansiering		Status	År	2010		2018	
BBRB	Bygge og boligregister (BED) - bygninger	JA	Status	Luk	2005	01	2008	12
BBRBYGNING	BBR bygninger		Status	År	2009	01	2020	12
BBRE	Bygge og boligregister (BED) - enheder		Status	Luk	2005	01	2008	12
BBREJENDOM	BBR Ejendom		Status	År	2009	01	2020	12
BBRENHED	BBR enhed		Status	År	2009	01	2020	12
BBRJORD	BBR Jord		Status	År	2010	12	2020	12
BBR_ADG	Adgangsadressereference		Forløb	År	1968	12	2020	12
BBR_EJD	Ejendomsreference		Hændelse	År	1968	12	2020	12
BBR_ENH	Boligenhvervenheds adressereference		Hændelse	År	1968	12	2020	12
BC	Beskæftigede CRAM		Status	Luk	1985		2007	
BEF	Befolkningen		Status	Kvt	1985	12	2022	03

BIGFI

Note:

- Data is high quality, but not perfect
- Some datasets requires a lot of cleaning
- Some variables are very poorly documented, if documented at all
- Still large fixed costs of getting to know the data

Other data

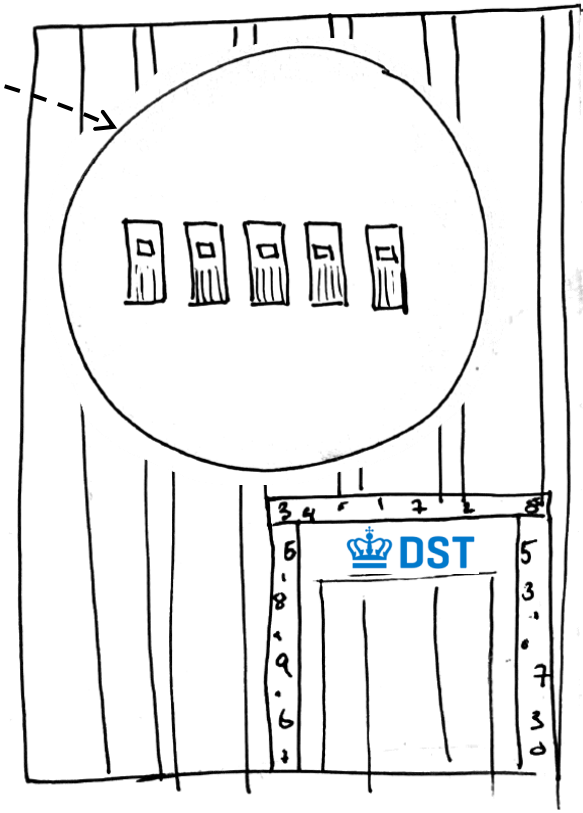
- External data can be uploaded and linked to register data
  - As long as it has a relevant ID (CPR, Name + birthday, CVR, SE-number, etc…)
  - Sensitive data can be send directly from the data provider to DST

- Data collection: Surveys and experiments
  1) Extract respondents from the CPR-registry - Run survey/experiment externally - Upload data and link to registries.
  2) DST runs surveys/experiments. See more here

BIGFI

Accessing the data

- We have 5 servers located at Statistics Denmark
- Remote access
- Connecting from abroad: use CBS VPN
- Servers contains all FI data and all FI projects
- Available software now
    - STATA, SAS, Excel, Python, R, Matlab.
- Guides for remote connection here



Working with the data in projects

- Work is organized in projects within DST (like a locked folder)
- Individual projects to be registered and approved by DST
- Set up project in one of the servers
- Access to relevant data only

- Right now: FI has 58 researchers in 13 projects on 5 servers.

Data security is important

- Respect the confidentiality - we are all in there
- No micro data can ever leave the server!
    - no statistic covering less than 5 observations, no percentiles, no minimums or maximums, no scatter plot, no ID-numbers - even if anonymized, careful of histogram tails.
- Exports of results allowed by FI employees only
- Work only from office or home and lock computer when leaving it
- Connect through CBS VPN
- If rules are violated, sanctions are immediate and tough
- DST's Rules for export of results

BIGFI

Publishing

- Cannot provide data.
- Be open about it from the beginning
- Provide codes and careful descriptions
- Archive the project in the server for a given amount of time.

- In the future: Maybe journals will be able to access the servers to check the data?

How to start a project

1) Register project with CBS
2) Get a DST researcher account (Julie orders it)
3) Send project application and variable list to Statistics Denmark (via Danmarks Datavindue)
    - Title
    - Purpose
    - Description
    - Importance to society
    - Population
    - Data sets and variables
4) When project approved, Julie prepares the data and DST transfers it to your project.

Julie will guide you through the process.

ETA: Depends, but more than a month…

BIGFI

Price estimates (2023)

- Starting a new project: app. 3,200 DKK
- Adding standard data: 3,000-6,000 DKK
- Other: 1,568 DKK/hour

- Survey / experiments:
    - No standard price
    - Depends on respondent rate, survey method, linkage to registry data and other…

- See also DST priser

More info

- For more information, go to
  - SPA4.12
  - Jma.fi@cbs.dk
  - https://www.dst.dk/da/TilSalg/Forskningsservice

Example: household data

Participation Constraints in the Stock Market: Evidence from Unexpected Inheritance Due to Sudden Death

Steffen Andersen
Copenhagen Business School

Kasper Meisner Nielsen
Hong Kong University of Science and Technology

We use a natural experiment to investigate the impact of participation constraints on individuals' decisions to invest in the stock market. Unexpected inheritance due to sudden deaths results in exogenous variation in financial wealth, and allows us to examine whether fixed entry and ongoing participation costs cause non-participation. We have three key findings. First, windfall wealth has a positive effect on participation. Second, the majority of households do not react to sizeable windfalls by entering the stock market, but hold on to substantial safe assets—even over longer horizons. Third, the majority of households inheriting stock holdings actively sell the entire portfolio. Overall, these findings suggest that participation by many individuals is unlikely to be constrained by financial participation costs. (*JEL* D14, G11)

BIGFI

Example: linked to external data

American Economic Review 2022, 112(10): 3398–3440
https://doi.org/10.1257/aer.20191766

Reference Dependence in the Housing Market†

By Steffen Andersen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai*

We quantify reference dependence and loss aversion in the housing market using rich Danish administrative data. Our structural model includes loss aversion, reference dependence, financial constraints, and a sale decision, and matches key nonparametric moments, including a "hockey stick" in listing prices with nominal gains, and bunching at zero realized nominal gains. Households derive substantial utility from gains over the original house purchase price; losses affect households roughly 2.5 times more than gains. The model helps explain the positive correlation between aggregate house prices and turnover, but cannot explain visible attenuation in reference dependence when households are more financially constrained. (*JEL* D12, D91, G51, R21, R31)

BIGFI

Example: linked to external data

How do Interest-only Mortgages Affect Consumption and Saving over the Life Cycle?

Linda Sandris Larsen Claus Munk Rikke Sejer Nielsen Jesper Rangvid

July 12, 2022

Abstract

Using a unique data set with detailed information on Danish households and their mortgages, we show that young and old households are more likely to use IO mortgages compared to middle-aged households. Young households use IO mortgages because they expect higher future income, old households because IO mortgages allow them to circumvent an otherwise binding liquidity constraint. Through different channels, IO mortgages thus facilitate consumption smoothing for young and old households. Our detailed data also allow us to examine how households with IO mortgages differ from households with repayment mortgages in terms of leverage, debt and asset composition, and pension contributions.

Double Jeopardy: Households' consumption responses to shocks in stock *and* mortgage markets

Linda Sandris Larsen Rikke Sejer Nielsen Ulf Nielsson Jesper Rangvid

November 4, 2022

Abstract

Households reduce consumption following negative shocks to their stock holdings. Households also lower consumption following exogenous increases in mortgage debt payments. But what is the impact of simultaneous adverse shocks in both markets, such as in the 2008 financial crisis? Using detailed Danish household data we find that the reduction in consumption doubles if households are highly exposed to both the stock and the mortgage market. We also find that the negative effects persist over time. It has a severe effect on consumption as households with a high-risk profile in the asset market also tend to have high exposure in the debt market. We discuss underlying reasons behind our results and their implications for macroprudential policies.

BIGFI

Example: field experiment

Econometrica, Vol. 79, No. 3 (May, 2011), 651–692

UNWILLING OR UNABLE TO CHEAT? EVIDENCE FROM A TAX AUDIT EXPERIMENT IN DENMARK

BY HENRIK JACOBSEN KLEVEN, MARTIN B. KNUDSEN, CLAUS THUSTRUP KREINER, SØREN PEDERSEN, AND EMMANUEL SAEZ[1]

This paper analyzes a tax enforcement field experiment in Denmark. In the base year, a stratified and representative sample of over 40,000 individual income tax filers was selected for the experiment. Half of the tax filers were randomly selected to be thoroughly audited, while the rest were deliberately not audited. The following year, threat-of-audit letters were randomly assigned and sent to tax filers in both groups. We present three main empirical findings. First, using baseline audit data, we find that the tax evasion rate is close to zero for income subject to third-party reporting, but substantial for self-reported income. Since most income is subject to third-party reporting, the overall evasion rate is modest. Second, using quasi-experimental variation created by large kinks in the income tax schedule, we find that marginal tax rates have a positive impact on tax evasion for self-reported income, but that this effect is small in comparison to legal avoidance and behavioral responses. Third, using the randomization of enforcement, we find that prior audits and threat-of-audit letters have significant effects on self-reported income, but no effect on third-party reported income. All these empirical results can be explained by extending the standard model of (rational) tax evasion to allow for the key distinction between self-reported and third-party reported income.

KEYWORDS: Tax evasion, field experiment, tax enforcement.

BIGFI

Example: external survey (epinion)

Subjective Unemployment Expectations

Ida Maria Hartmann[1] Søren Leth-Petersen[1]

December 2022

Abstract

We study how individual unemployment expectations are shaped and updated using a unique longitudinal survey data set with subjective unemployment expectations. The survey data is linked with third-party reported administrative data on unemployment realizations, such that we are able to examine how prediction errors lead individuals to update their unemployment expectations. We find that people are constantly uncertain about their unemployment prospects. The uncertainty causes them to adjust their unemployment expectations when their predictions turned out to be incorrect. As a result, people's expectations concerning future unemployment are not constant and they are heterogeneous across the population at any given point in time. We document that unemployment expectations and prediction errors are important determinants of economic decisions, such as how much to save or whether to insure against earnings losses. Subjective unemployment expectations can thus help explain why people, who are observationally similar, make differently economic decisions.

BIGFI

Example: lab experiment (external)

ASSET INTEGRATION AND ATTITUDES TOWARD RISK: THEORY AND EVIDENCE

Steffen Andersen, James C. Cox, Glenn W. Harrison, Morten I. Lau,
E. Elisabet Rutström, and Vjollca Sadiraj*

Abstract—We provide evidence that choices over small-stakes bets are consistent with assumptions of some payoff calibration paradoxes. We then exploit the existence of detailed information on individual wealth of our experimental subjects in Denmark and directly estimate risk attitudes and the degree of asset integration. We discover that behavior is consistent with partial, rather than full, asset integration. The implied risk attitudes from estimating these specifications indicate risk premiums and certainty equivalents that are a priori plausible. This theory and evidence suggest one constructive solution to payoff calibration paradoxes.

Example: lab experiment (run by DST)

Beliefs and the Disposition Effect[•]

Steffen Andersen
Copenhagen Business School and CEPR
san.fi@cbs.dk

Jimmy Martínez-Correa
Copenhagen Business School
jima.eco@cbs.dk

Tobin Hanspal
WU Vienna University of Economics and Busine
tobin.hanspal@wu.ac.at

Kasper Meisner Nielsen
Copenhagen Business School
kmn.fi@cbs.dk

Abstract

This study examines whether heterogeneous beliefs contribute to the incidence of the disposition effect. We measure individuals' optimism using elicited beliefs from incentivized experiments and surveys and link these measures to investment decisions using administrative register data. We find that optimistic beliefs lead to both a higher allocation of wealth to risky assets and a higher incidence of the disposition effect. We further find that subjective beliefs explain heterogeneity in the disposition effect which are not captured by risk preferences. Overall, our study documents that individuals' beliefs play an important role in the prevalence of the disposition effect.

Exampel: Firm data + survey

HOW WORKER PRODUCTIVITY AND WAGES GROW WITH TENURE AND EXPERIENCE:
THE FIRM PERSPECTIVE

Andrew Caplin
Minjoon Lee
Søren Leth-Petersen
Johan Sæverud
Matthew D. Shapiro

Working Paper 30342
http://www.nber.org/papers/w30342

How worker productivity evolves with tenure and experience is central to economics, shaping, for example, life-cycle earnings and the losses from involuntary job separation. Yet, worker-level productivity is hard to identify from observational data. This paper introduces direct measurement of worker productivity in a firm survey designed to separate the role of on-the-job tenure from total experience in determining productivity growth. Several findings emerge concerning the initial period on the job. (1) On-the-job productivity growth exceeds wage growth, consistent with wages not being allocative period-by-period.(2) Previous experience is a substitute, but a far less than perfect one, for on-the-job tenure. (3) There is substantial heterogeneity across jobs in the extent to which previous experience substitutes for tenure. The survey makes use of administrative data to construct a representative sample of firms, check for selective non-response, validate survey measures with administrative measures, and calibrate parameters not measured in the survey.

BIGFI

Example: Firm data + survey

Preserving job matches during the COVID-19 pandemic:
firm-level evidence on the role of government aid*

Morten Bennedsen† Birthe Larsen‡

Ian Schmutte§ Daniela Scur¶

June 20, 2020

[Latest version here]

Abstract

We analyze the impact of the COVID-19 pandemic and government policies on firms' aid take-up, layoff and furlough decisions. We collect new survey data for 10,642 small, medium and large Danish firms, and match to government records of all aid-supported furloughed workers during the pandemic as well as administrative accounting data. This is the first representative sample of firms reporting the pandemic's impact on their revenue and labor choices, showing a steep decline in revenue and a strong reported effect of labor aid take-up on lower job separations. Relative to a normal year, 30 percent more firms have experienced revenue declines. Comparing firms' actual layoff and furlough decisions to their reported counterfactual decisions in the absence of aid, we estimate 81,000 fewer workers were laid off and 285,000 workers were furloughed. Our results suggest the aid policy was effective in preserving job matches at the start of the pandemic.

BIGFI

Thank you!