

Estimating Pareto's Alpha with the Top-Wealth Database

preliminary results

Fabian Nemeczek, Tobias Albers
Marcelo Avila, Sebastian Kahl

10th December 2018



Outline

① Motivation

② Data Collection

Data Base – Orbis

Stratification Method

SOEP vs. Pretest

③ Pareto Distribution

④ Preliminary Results and Outlook

Application in Stata

Outlook

Motivation

- Data about high-worth individuals is necessary for science and politics
- **Problem:** little valid information available in Germany
 - No obligatory wealth-registration, no wealth tax
 - High-worth individuals are under-represented in surveys, i.e. SOEP
- **Solution:** new sampling strategy in order to identify high-worth individuals and survey them with SOEP instruments
- **Main assumption:** private wealth and private shares of business assets are highly positive correlated

Motivation (Cont.)

- **Aim:** Identification of high-worth individuals in available databases
- **Basis source:** *Orbis*
 - International business level database from *Bureau van Dijk (BvD)*
 - contains information of more than 270 million companies worldwide
 - information about financials and ownership structures/shareholdings of each firm



BUREAU VAN DIJK

A Moody's Analytics Company

Outline

- ① Motivation
- ② Data Collection
- ③ Pareto Distribution
- ④ Preliminary Results and Outlook

Database Orbis

The screenshot shows the Orbis database interface. At the top, there's a navigation bar with links for Companies, Contacts, News, M&A deals, Industry research, Global Reports, Royalty agreements, and Patents. Below the navigation bar is a search bar with placeholder text "Enter a company name or BvD ID number". To the right of the search bar are buttons for Alerts, Settings, Help, Contact us, and Log out.

The main content area displays information for "STATOIL ASA". It includes the address "4033 STAVANGER, Norway", status "Publicly quoted company", and note "The GO of this controlled subsidiary is GOVERNMENT OF NORWAY". It also shows the BvD ID number "NO923609016" and the latest account date "31 Dec 2011".

A sidebar on the right is titled "Back to list of results" and contains a tree view of search categories:

- Search
 - New search
 - Modify current search
 - Batch search
 - Define the format
 - Report format
 - Display a specific section
 - Customised section
 - PowerPoint templates
 - Contact information
 - Identification numbers
 - Legal & account information
 - Size & group information
 - Industry & overview
 - Industry research
 - Financial data
 - User defined variables
 - Segment data
 - Customers & competitors
 - Future outlook
 - Stock data
 - Earnings estimates & brokers recommendations
 - Directors, managers, contacts & advisors
 - Ownership data

Below the main content, there are two tables: "Shareholder information" and "Current shareholders". The "Shareholder information" table shows ownership details for three entities: GOVERNMENT OF NORWAY, OLJE- OG ENERGIDEPARTEMETET, and STATOIL ASA. The "Current shareholders" table shows a list of shareholders with their respective ownership percentages.

Figure: BvD's Database Orbis

Strategy

- ① Estimate consistent business values
- ② Transform business level data to personal level data
- ③ Rank by share values

Mickey Mouse dataset

company	estimated value	shareholder	share
company A	1,000,000 Euro	individual 1	0.5
		individual 2	0.5
company B	25,000,000 Euro	individual 2	0.8
		individual 3	0.2
company C	500,000 Euro	individual 1	1.0



individual	estimated share value	rank
individual 1	$(1,000,000 \text{ Euro} \times 0.5) + (500,000 \text{ Euro} + 1.0) =$ 1.000.000 Euro	3
individual 2	$(1.000.000 \text{ Euro} \times 0.5) + (25.000.000 \text{ Euro} \times 0.8) =$ 20.500.000 Euro	1
individual 3	$25.000.000 \text{ Euro} \times 0.2 =$ 5.000.000 Euro	2

Stratification Method

- **Aim:** Oversampling very high worth individuals to ensure that individuals at the very top of the wealth distribution are covered
- **Solution:** Order the population by net wealth, separate top 1% into 3 groups and draw a sample with disproportionate probabilities
Disproportionate stratification method

Stratification Method (Cont.)

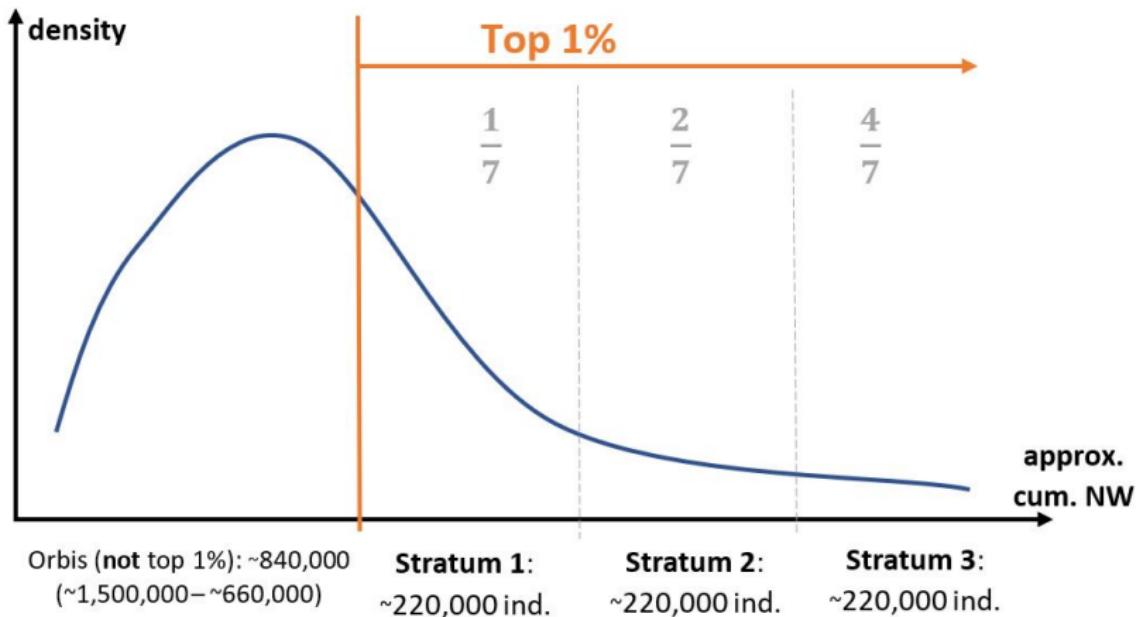


Figure: Stratification sketch

Stratification Method (Cont.)

- How to apply disproportionate stratified sampling:
 - Divide sample into k strata
 - Take **random and independent** samples from each of the k strata
 - Pool subsamples
 - Balance out the disproportionate probabilities by **reweighting** the drawn sample to the total population

From Gross Sample to Realized Interviews

Drawn Sample: 2000 ind.	
Stratum 1: 296 ind.; Stratum 2: 588 ind.; Stratum 3: 1119 ind.	
adjusted gross sample: 1652 ind.	address unavailable: 348 ind.
successfully contacted: 532 ind.	contact impossible: 1120 ind.
realised interviews: 124 ind.	non-participation: 408 ind.

Net Wealth of SOEP and Pretest in Comparison

Table: Descriptive Statistics of SOEP 2012 and Pretest 2017

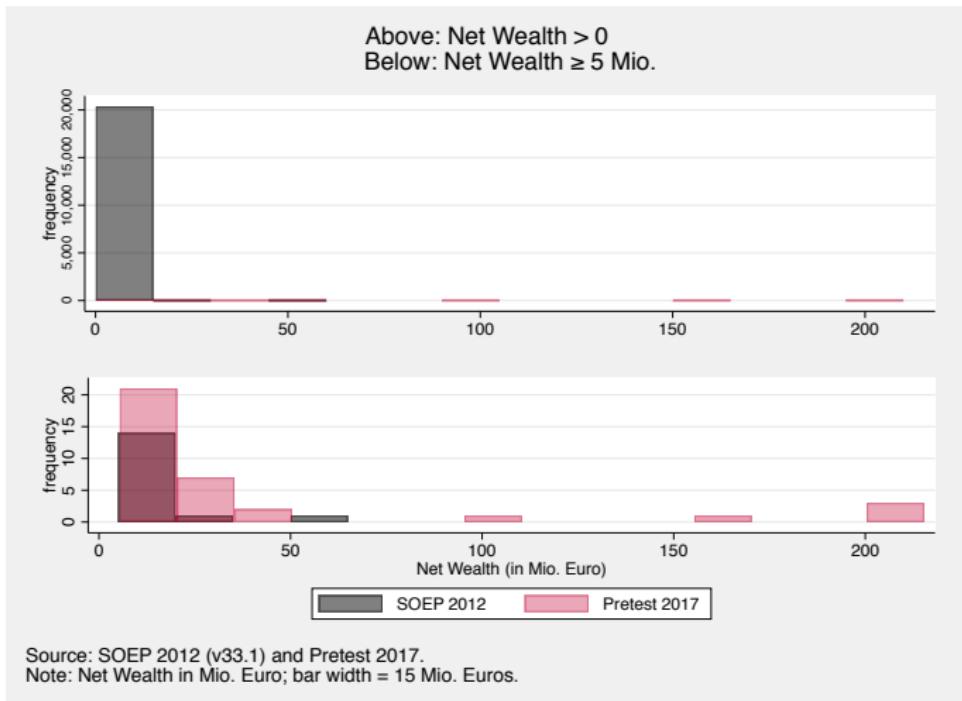
	count	mean	p50	p75	p90	p99	min	max
SOEP	27,948	0.09	0.02	0.10	0.22	0.88	-4.00	62.50
Pretest	124	10.22	2.21	6.02	27.89	156.56	-2.58	207.02
SOEP + Pretest	28,072	0.13	0.02	0.10	0.22	1.05	-4.00	207.02

Source: SOEP 2012 (v33.1) and Pretest 2017.

Note: Net Wealth in mio. Euro, for simplicity rounded.

SOEP and Pretest

Histograms of Net Wealth: SOEP vs. Pretest



From Samples to Population

Weighted Net Weath of SOEP and Pretest in Comparison

Table: Descriptive Statistics of SOEP 2012

	count	mean	p50	p75	p90	p99	min	max
SOEP	61,317,205	0.09	0.02	0.10	0.21	0.81	-4.00	62.50

Source: SOEP 2012 (v33.1).

Note: Net Wealth, weighted and displayed in mio. Euro, for simplicity rounded.

Table: Descriptive Statistics of Pretest 2017

	count	mean	p50	p75	p90	p99	min	max
Pretest	726,547	7.18	1.95	5.74	20.73	100.00	-2.58	207.02

Source: Pretest 2017.

Note: Net Wealth, weighted and displayed in mio. Euro, for simplicity rounded.

Outline

- ① Motivation
- ② Data Collection
- ③ Pareto Distribution
- ④ Preliminary Results and Outlook

Background

- Vilfredo Pareto (1897) presents a versatile functional relation that well describes wealth distribution across countries and centuries.
- Same concept is applied to several other fields and colloquially called *Pareto Principle*.
 - 80% of land owned by 20% of individuals (revenue ~ products; sales ~ clients; etc)
- Generally, it follows a *power law probability distribution*, where one measure varies constantly as an exponential of another, independently of initial values.
 - Example: if one increases the side length of a square by x , its area increases by x^2 , independently of initial area of square.

Functional Form

The **Pareto Distribution** is defined by

$$f(y, \underline{y}, \alpha) = \frac{\alpha \underline{y}^\alpha}{y^{\alpha+1}}, \quad 0 < \underline{y} \leq y$$

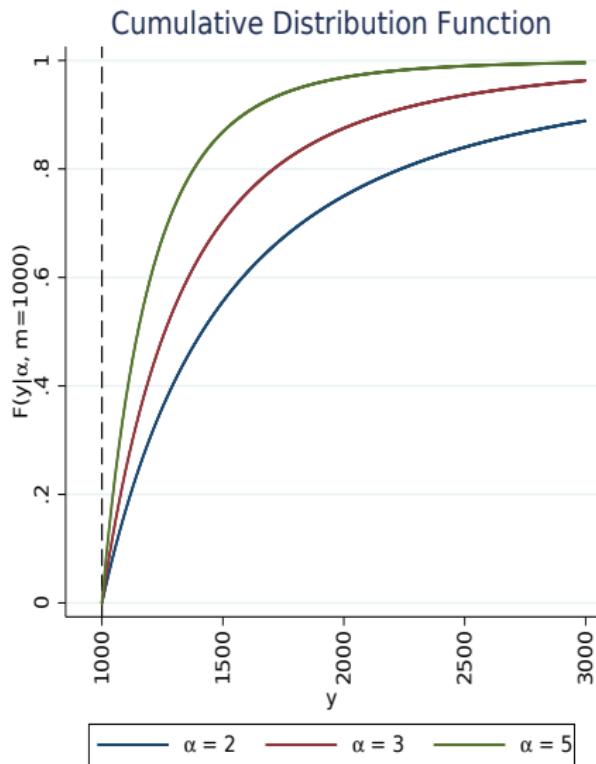
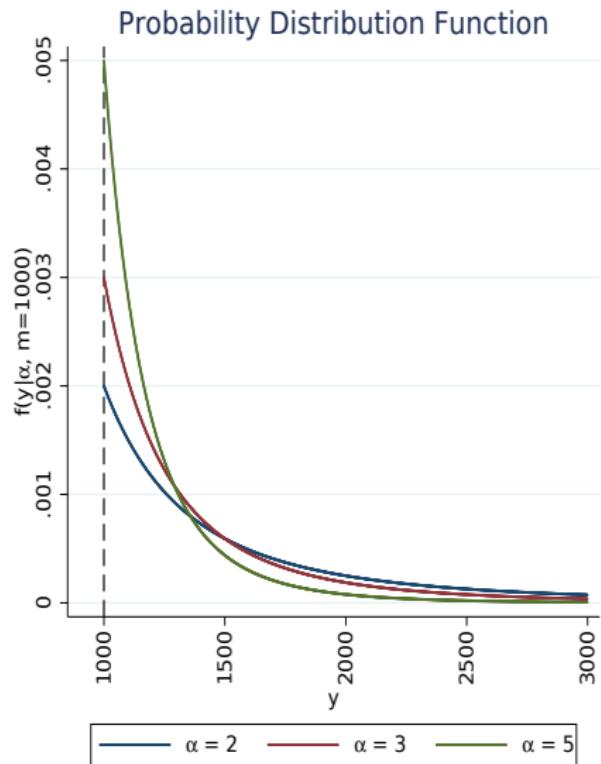
and

$$F(y, \underline{y}, \alpha) = 1 - \left(\frac{\underline{y}}{y} \right)^\alpha, \quad 0 < \underline{y} \leq y$$

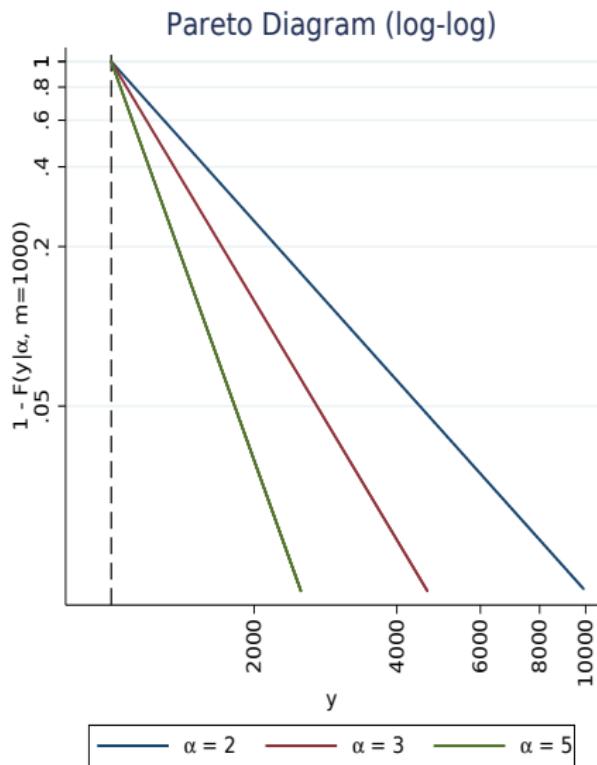
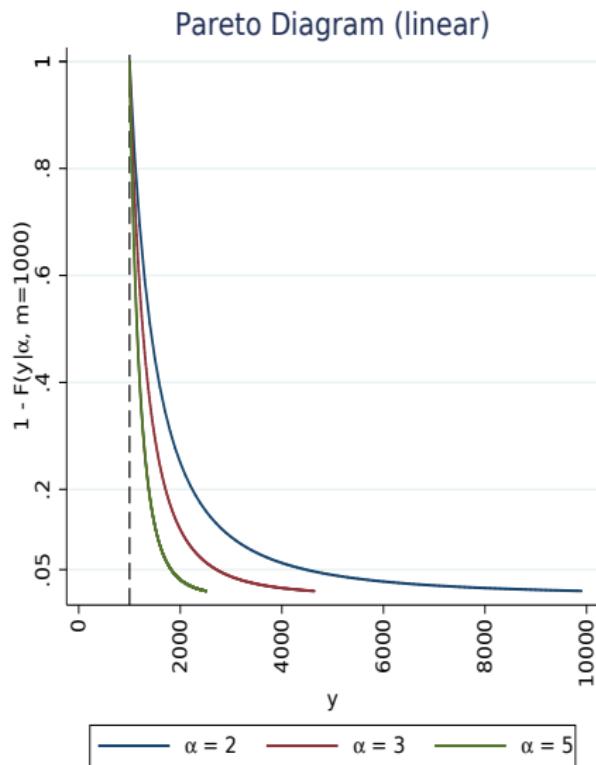
where:

- y : wealth measure
- α : Pareto's α (or *shape/slope parameter*)
- \underline{y} : lower bound (or *scale parameter* or *threshold value*)

Graphical Visualisation I



Graphical Visualisation II



Properties

Pareto's α :

- Sloppy interpretation: for a percentage increase in y , the proportion of *richer* individuals decrease by α percent.
- Higher α values \Rightarrow less inequality.¹
- Several inequality indices can be estimated based on α .
 - Example: Gini coefficient: $\frac{1}{2\alpha-1}$.

Possible problems:

- High flexibility on estimating the lower bound
- Sensibility of α due to choice of the lower bound \underline{y}

¹for inequality measures satisfying the *Weak Transfers Principle* (Cowell 2011, p. 93)

Outline

- ① Motivation
- ② Data Collection
- ③ Pareto Distribution
- ④ Preliminary Results and Outlook

Stata: Pareto Distribution I

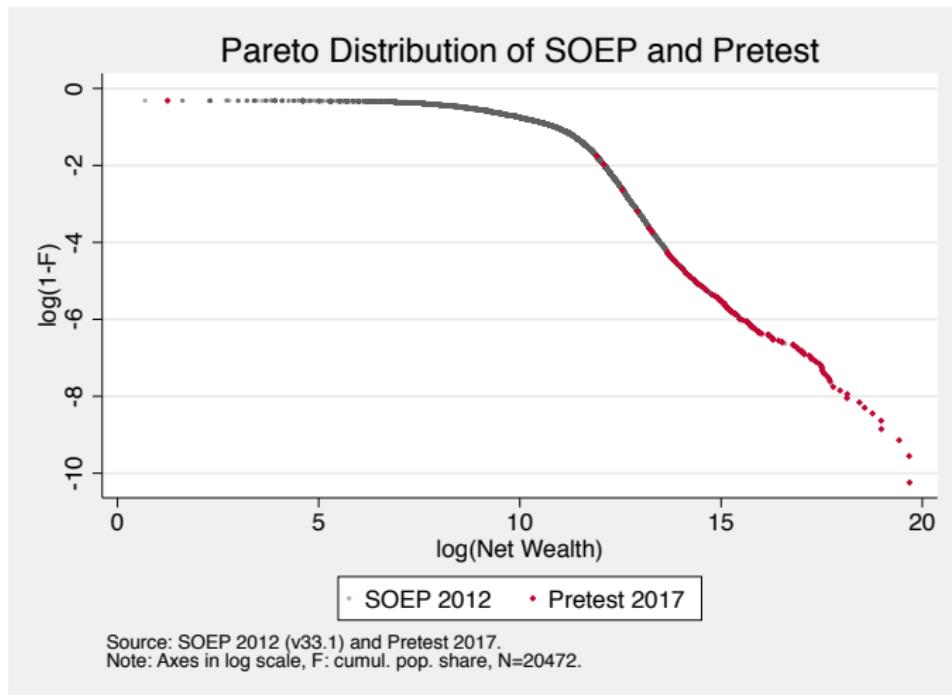
```
1 * Re-weighting
2 gen W_pretest = 0
3 sca sc_strata = 3
4
5 * Stratum 1
6 replace W_pretest = (7/1)/sc_strata if D_pretest==1 & schicht==1
7 * Stratum 2
8 replace W_pretest = (7/2)/sc_strata if D_pretest==1 & schicht==2
9 * Stratum 3
10 replace W_pretest = (7/4)/sc_strata if D_pretest==1 & schicht==3
11
12 * from sample to top 1% population
13 qui sum nw if D_pretest==1
14 replace W_pretest = W_pretest * (660000/r(N))
15
16 * check frequency weights
17 sum nw if D_pretest==0 [fw = round(W_soep)]
18 sum nw if D_pretest==1 [fw = round(W_pretest)]
19
20 * Scale weight between 0 and 1
21 gen nw_W = (nw - r(min)) / (r(max) - r(min)) if D_pretest==1
```

Stata: Pareto Distribution II

```
22
23 * preparation for pareto distribution
24 sort D_pretest nw
25 qui sum nw_W if D_pretest==1
26 sca sc_N = r(sum)
27 gen cum_pop_share = sum(nw_W)/sc_N if D_pretest==1
28 gen P = 1 - cum_pop_share
29 gen lnP = ln(P)
30 gen ln_nw = ln(nw)
31
32 * threshold1 = 100.000 Euros; threshold2 = 1.000.000 Euros
33 sca sc_thres1 = log(100000)
34 sca sc_thres2 = log(1000000)
35
36 * plot pareto distribution (SOEP and Pretest):
37 graph twoway (scatter lnP ln_nw if D_pretest==0, mcolor(gray) ///
38     (scatter lnP ln_nw if D_pretest==1, mcolor(red) ///
39     xline(`=sc_thres1', lcolor(green))      ///
40     xline(`=sc_thres2', lcolor(blue*1.25)))  ///
41     (lfit lnP ln_nw if ln_nw >= sc_thres1, lcolor(lime)) ///
42     (lfit lnP ln_nw if ln_nw >= sc_thres2, lcolor(blue))
```

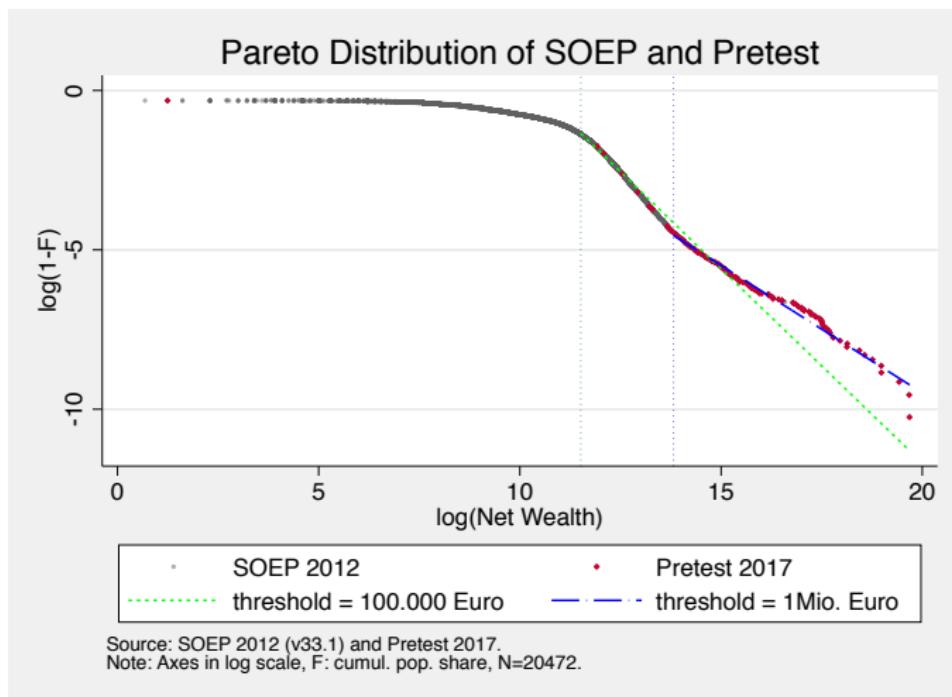
Preliminary Results and Outlook

Graph of Pareto Distribution



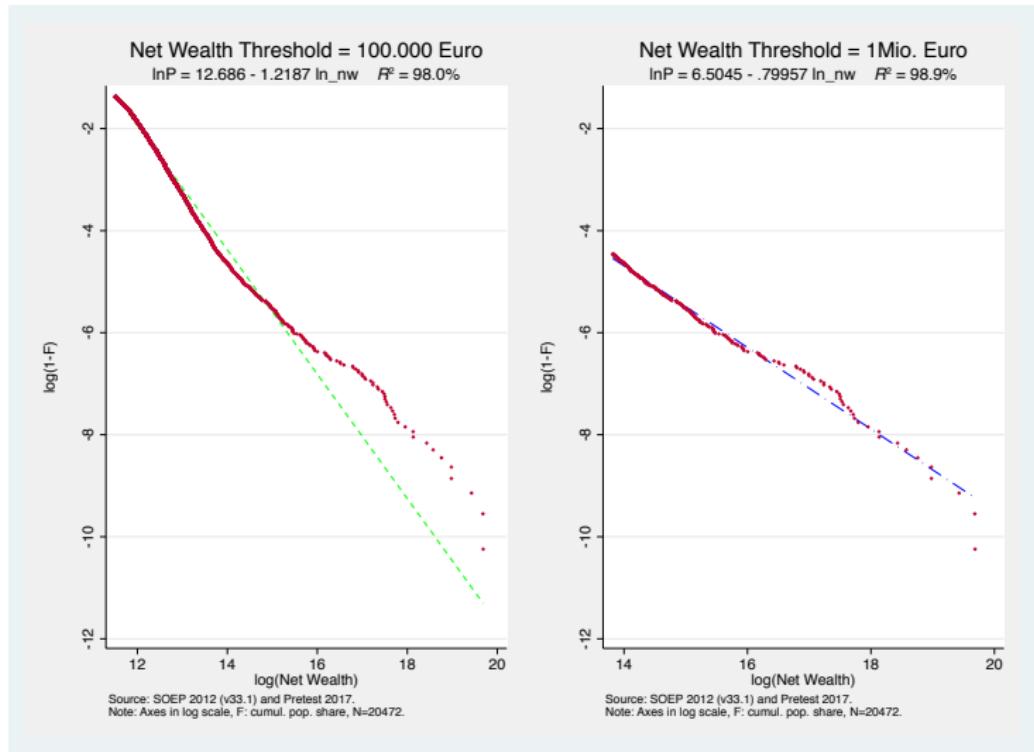
Preliminary Results and Outlook

Graph of Pareto Distribution



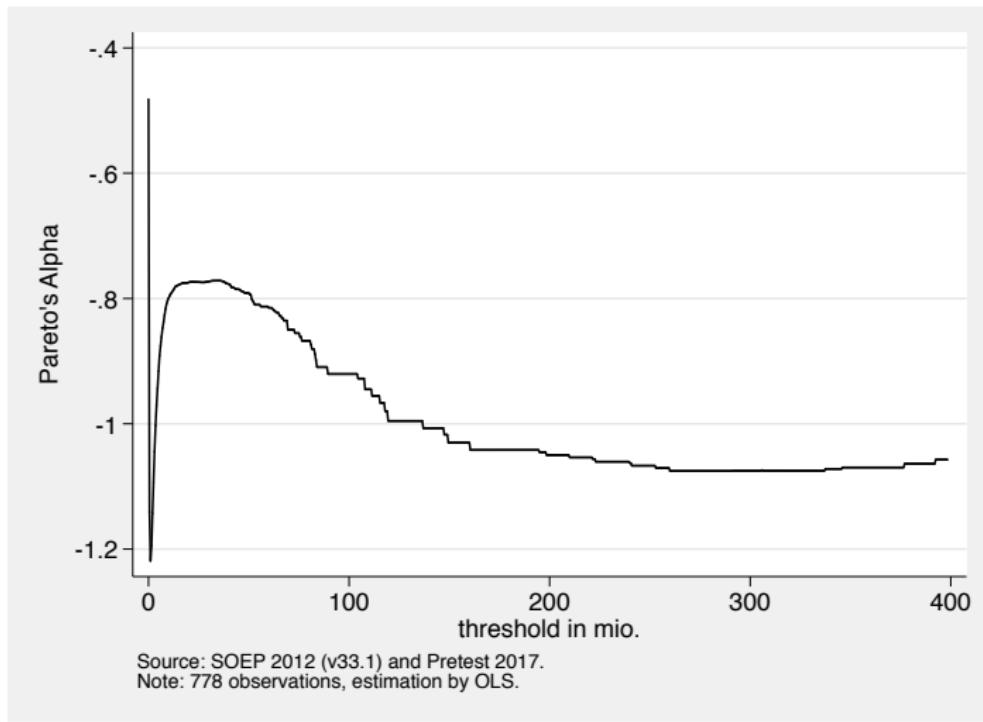
Preliminary Results and Outlook

Estimating Pareto's α



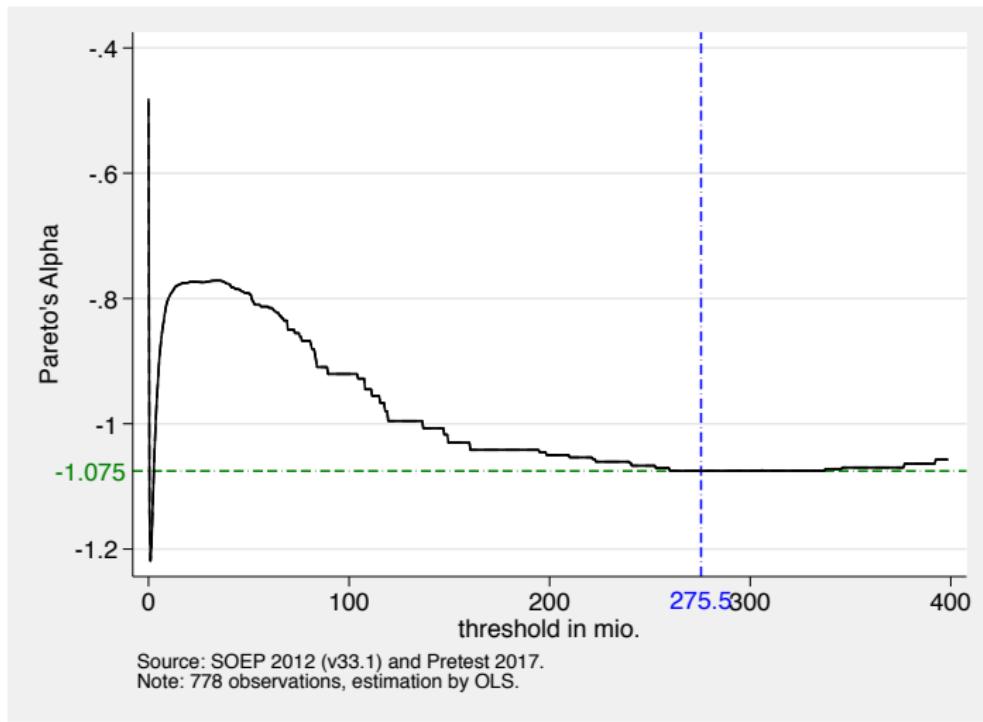
Preliminary Results and Outlook

Estimating Pareto's α



Preliminary Results and Outlook

Estimating Pareto's α



Outlook

- Robustness Check: Selection alpha and threshold
- Handling imputed net wealth variables
- Imputation of top-tails in SOEP
- Comparison SOEP with and without Pretest by applying inequality measures
- Weighting Pretest and SOEP together

Thank You for your Attention!

References

- Cowell, Frank A. 2011. *Measuring inequality*. 3rd ed. LSE perspectives in economic analysis. Oxford: Oxford University Press.
- Schröder, Carsten, et al. 2018. "A Novel Sampling Strategy for Surveying High-Worth Individuals – An Application Using the Socio-Economic Panel". *SOEPpapers on Multidisciplinary Panel Data Research* 2018 (978): 40p.
- Westermeier, Christian, and Markus Grabka. 2015. "Große statistische Unsicherheit beim Anteil der Top-Vermögenden in Deutschland". *DIW Wochenbericht* 82 (7): 123–133.