

Estimating Pareto's Alpha with the Top-Wealth Database

final results

Fabian Nemeczek, Tobias Albers
Marcelo Avila, Sebastian Kahl

28th January 2019

Outline

① Recap

- Data

- Pareto distribution

- Correction of last presentation

② Results

- Estimated Pareto's Alpha

- Confidence Intervals

- Confidence Intervals

③ Robustness Checks

- Hausman Test

- Test Results

Recap

- **Problem:** High-worth individuals are typically under-represented in population surveys, i.e. SOEP for Germany
- **Solution:** DIW Berlin creates a new sampling strategy in order to identify high-worth individuals and survey them with SOEP instruments - Result: **Pretest** (N=124)
- **Research issue:** Does the top wealth data follow a Pareto distribution?

Data - SOEP vs. Pretest

- **SOEP:** Wealth module is collected every five years. We use the data from 2012
 - Personal net overall wealth (incl. property, business assets, etc. - minus debts, loans)
 - Missing values are already imputed
- **Pretest:** Same survey instrument is used as in the wealth module from the SOEP. Data are collected in 2017/2018.
 - Same variable characteristics as in SOEP 2012
- **Note:** Because of different weighting schemes in SOEP and Pretest and different years of observation, it is not possible to add the Pretest data to the SOEP

Descriptive statistics - SOEP vs. Pretest

	count	mean	p50	p90	p99	min	max
SOEP	27,948	0.09	0.02	0.22	0.88	-4.00	62.50
Pretest	124	10.22	2.21	27.89	156.56	-2.58	207.02

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

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

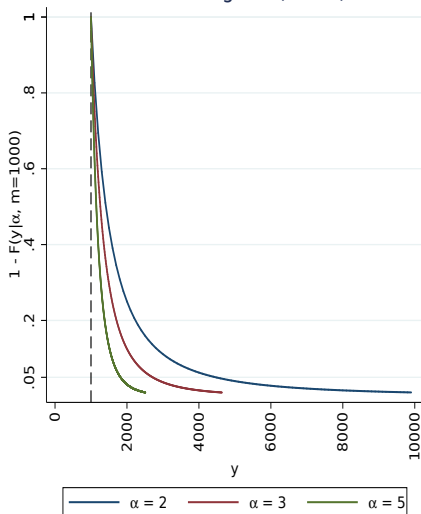
Table 1: Descriptive Statistics of SOEP 2012 and Pretest 2017

Pareto distribution I

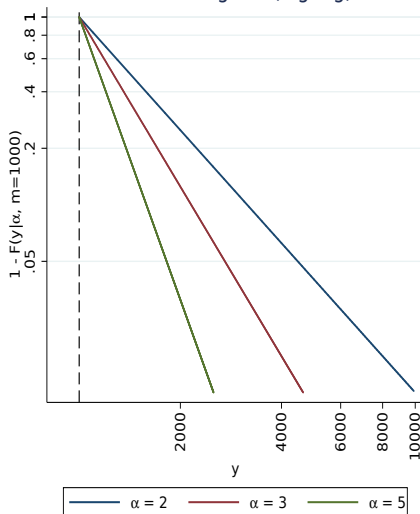
- Power-law probability distribution defined by Pareto (1897)
 - Describes the distribution of top-wealth and top-income in societies
 - From a defined lower bound threshold \underline{y} , the log-scaled cum. density function follows a straight line
- **Pareto's α :** Slope parameter of the Pareto distribution
 - graphical interpretation: Slope of the cum. density function
 - economical interpretation: With increasing α , decreases inequality
- **Empirical aim:** Estimate Pareto's α with the Pretest data
- **Methodology:** Use the p95 and p99 value of the SOEP as lower bound for the Pareto estimation with the Pretest

Pareto distribution II

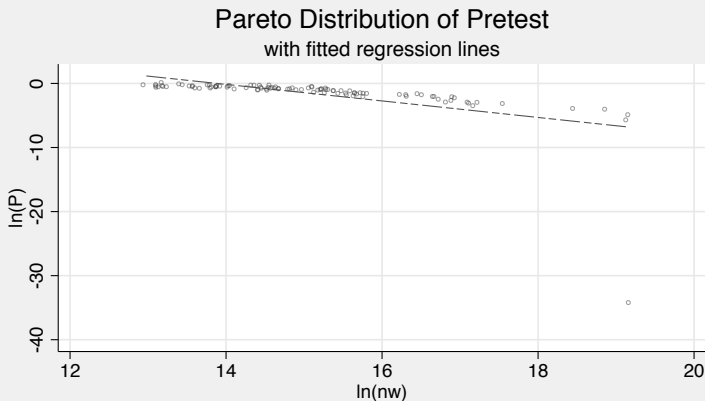
Pareto Diagram (linear)



Pareto Diagram (log-log)



Pareto distribution IV



Source: Pretest 2017.

Note: We display only 1 net wealth variable with own calculated re-weighting scheme.

Threshold at 95% percentile of the SOEP. Abbrev.: *nw*=net wealth, *pt*=Pretest.

Stata code - correction of last presentation

```
1 *** Re-weighting according to sampling probabilities - false method!  
2 * Stratum 1  
3 replace W_pt = (7/1)/sc_strata if D_pretest==1 & schicht==1  
4 * Stratum 2  
5 replace W_pt = (7/2)/sc_strata if D_pretest==1 & schicht==2  
6 * Stratum 3  
7 replace W_pt = (7/4)/sc_strata if D_pretest==1 & schicht==3  
8  
9 *** Re-weighting according to response rates - correct method.  
10 * Stratum 1  
11 replace W_pt = (124/23)/sc_strata if D_pretest==1 & schicht==1  
12 * Stratum 2  
13 replace W_pt = (124/41)/sc_strata if D_pretest==1 & schicht==2  
14 * Stratum 3  
15 replace W_pt = (124/60)/sc_strata if D_pretest==1 & schicht==3  
16 * ...  
17 * from sample to 1% population  
18 qui sum _1_nw if D_pretest==1  
19 replace W_pt = W_pt * (660000/124)
```

Outline

- 1 Recap
- 2 Results
- 3 Robustness Checks

Estimated Pareto's Alpha (1/2)

	$\hat{\alpha}_{SOEP}$	N_{SOEP}	$\hat{\alpha}_{Pretest}$	$N_{Pretest}$	<i>threshold</i>
net wealth 1	-2.047 (.04)	1019	-1.253 (.19)	104	341500
net wealth 2	-2.09 (.03)	1017	-1.265 (.19)	106	339500
net wealth 3	-1.965 (.04)	1021	-1.436 (.21)	104	345000
net wealth 4	-2.007 (.04)	1019	-1.319 (.2)	105	343000
net wealth 5	-2.048 (.04)	1019	-1.256 (.19)	104	343000

Table 2: Estimated Pareto Alphas (threshold at 95th percentile of SOEP)

Estimated Pareto's Alpha (2/2)

	$\hat{\alpha}_{SOEP}$	N_{SOEP}	$\hat{\alpha}_{Pretest}$	$N_{Pretest}$	<i>threshold</i>
net wealth 1	-2.818 (.17)	204	-1.61 (.28)	80	870000
net wealth 2	-2.894 (.16)	205	-1.714 (.29)	79	879250
net wealth 3	-2.645 (.2)	204	-2.027 (.32)	76	898000
net wealth 4	-2.92 (.19)	204	-1.777 (.3)	77	901250
net wealth 5	-2.911 (.18)	203	-1.647 (.29)	77	883999

Table 3: Estimated Pareto Alphas (threshold at 99th percentile of SOEP)

Dealing with imputed values


Analyzing the multiply-imputed data set according to *Rubin(1987)*
 (The data set consists of $i=\{1,\dots,n=124\}$ rows and $j=\{1,\dots,m=5\}$ columns)

- 1 Sample Variance of each row of imputations:

$$\sigma^2(\ln(NW)_j) = \frac{\sum_{i=1}^n (\ln(NW)_{i,j} - \overline{\ln(NW)}_j)^2}{n-1} \quad (1)$$

- 2 Variance of Alpha¹:

$$\sigma^2(\widehat{\alpha}_j) = \frac{\sigma^2(\ln(NW)_j)}{\sum_{i=1}^n (\ln(NW)_{i,j} - \overline{\ln(NW)}_j)^2} \quad (2)$$

¹According to the [Simple Linear Regression](#)'s formula of the estimator β 

Dealing with imputed values

Alternate calculation which is considering the SOEP's weighting scheme
(*not applicable for the pretest data, since there are no applied weights*)

- 1 Sample Variance of each row of imputations:

$$\sigma^2(\ln(NW)_j) = \frac{\sum_{i=1}^n ((w_i * \ln(NW)_{i,j}) - \frac{\sum_{i=1}^n (w_i * \ln(NW)_{i,j})}{\sum_{i=1}^n w_i})^2}{n-1} \quad (3)$$

- 2 Variance of Alpha²:

$$\sigma^2(\hat{\alpha}_j) = \frac{\sigma^2(\ln(NW)_j)}{\sum_{i=1}^n ((w_i * \ln(NW)_{i,j}) - \frac{\sum_{i=1}^n w_i * (\ln(NW)_{i,j})}{\sum_{i=1}^n w_i})^2} \quad (4)$$

²According to the [Simple Linear Regression](#)'s formula of the estimator β 

③ **Within Variance** of each row of imputations:

$$\sigma_W^2(\widehat{\alpha}_j) = \frac{\sigma^2(\widehat{\alpha}_j)}{n-1} \quad (5)$$

④ **Average Within Variance:**

$$\sigma_W^2(\widehat{\alpha}) = \frac{1}{m} \sum_{j=1}^{m=5} \sigma_W^2(\widehat{\alpha}_j) \quad (6)$$

5 Estimate of Alpha:

$$\bar{\hat{\alpha}} = \frac{1}{m} \sum_{j=1}^m \hat{\alpha}_j \quad (7)$$

6 **Between Variance:**

$$\sigma_B^2(\bar{\hat{\alpha}}) = \sum_{j=1}^{m=5} (\hat{\alpha}_j - \bar{\hat{\alpha}})^2 \quad (8)$$

7 From (4) and (6) we derive the **Total Variance:**

$$\sigma_T^2 = \underbrace{\frac{1}{m} \sum_{j=1}^{m=5} \frac{\sigma^2(\ln(NW)_j)}{n-1} + (1 - m^{-1}) * \sum_{j=1}^{m=5} (\hat{\alpha}_j - \bar{\hat{\alpha}})^2}_{\sigma_W^2(\bar{\hat{\alpha}}) + (1 - m^{-1}) * \sigma_B^2(\bar{\hat{\alpha}})} \quad (9)$$

6 Confidence Interval (95%):

Lower bound:

$$\bar{\hat{\alpha}} - [\mathcal{N}_{0.975}]_{df(1)} * \sigma_T \quad (10)$$

Upper bound:

$$\bar{\hat{\alpha}} + [\mathcal{N}_{0.975}]_{df(1)} * \sigma_T \quad (11)$$

CI of Pareto Alphas

	$\hat{\alpha}_{SOEP}$	N_{SOEP}	$\hat{\alpha}_{Pretest}$	$N_{Pretest}$	<i>threshold</i>
net wealth 1	-2.047 (.04)	1019	-1.253 (.19)	104	341500
net wealth 2	-2.09 (.03)	1017	-1.265 (.19)	106	339500
net wealth 3	-1.965 (.04)	1021	-1.436 (.21)	104	345000
net wealth 4	-2.007 (.04)	1019	-1.319 (.2)	105	343000
net wealth 5	-2.048 (.04)	1019	-1.256 (.19)	104	343000
CI lower (.025)	-2.235		-1.639		
CI upper (.975)	-1.828		-.972		

Table 4: Estimated Pareto Alphas with CI (threshold at 95th percentile)

CI of Pareto Alphas

	$\hat{\alpha}_{SOEP}$	N_{SOEP}	$\hat{\alpha}_{Pretest}$	$N_{Pretest}$	<i>threshold</i>
net wealth 1	-2.818 (.17)	204	-1.61 (.28)	80	870000
net wealth 2	-2.894 (.16)	205	-1.714 (.29)	79	879250
net wealth 3	-2.645 (.2)	204	-2.027 (.32)	76	898000
net wealth 4	-2.92 (.19)	204	-1.777 (.3)	77	901250
net wealth 5	-2.911 (.18)	203	-1.647 (.29)	77	883999
CI lower (.025)	-3.333		-2.465		
CI upper (.975)	-2.343		-1.045		

Table 5: Estimated Pareto Alphas with CI (threshold at 99th percentile)

Outline

- 1 Recap
- 2 Results
- 3 Robustness Checks

Robustness Check of Pareto's alpha

Question: Is the Pareto distribution a *good enough* approximation of the wealth distribution among Germany's top wealth individuals?

- We perform a generalized Hausman Tests (Hausman (1978) and Hausman and McFadden (1984)).
- Test is implemented in Stata's Seemingly Unrelated Estimation (suest) post-estimation method.
- Test statistic³:

$$\frac{(\hat{\beta}_i - \hat{\beta}_j)^2}{\text{var}(\hat{\beta}_i) - 2\text{cov}(\hat{\beta}_i, \hat{\beta}_j) + \text{var}(\hat{\beta}_j)} \sim \chi^2_1.$$

³as described in Stata 15 Manual: [suest - Seemingly unrelated estimation](#), pg. 07

Results of Hausman test (1/4)

	$\hat{\alpha}_{SOEP}$	N_{SOEP}	$\hat{\alpha}_{Pretest}$	$N_{Pretest}$	threshold	Hausman
net wealth 1	-2.047 (.04)	1019	-1.253 (.19)	104	341500	.164
net wealth 2	-2.09 (.03)	1017	-1.265 (.19)	106	339500	.151
net wealth 3	-1.965 (.04)	1021	-1.436 (.21)	104	345000	.423
net wealth 4	-2.007 (.04)	1019	-1.319 (.2)	105	343000	.248
net wealth 5	-2.048 (.04)	1019	-1.256 (.19)	104	343000	.162
CI lower (.025)	-2.235		-1.639			
CI upper (.975)	-1.828		-.972			

Table 6: Pareto Alphas and results of Hausman-Test (threshold at 95th percentile)

$$H_0 : \hat{\alpha}_{imp}^{SOEP} = \hat{\alpha}_{imp}^{Pretest}$$

Results of Hausman test (2/4)

	$\hat{\alpha}_{SOEP}$	N_{SOEP}	$\hat{\alpha}_{Pretest}$	$N_{Pretest}$	threshold	Hausman
net wealth 1	-2.818 (.17)	204	-1.61 (.28)	80	870000	.287
net wealth 2	-2.894 (.16)	205	-1.714 (.29)	79	879250	.311
net wealth 3	-2.645 (.2)	204	-2.027 (.32)	76	898000	.618
net wealth 4	-2.92 (.19)	204	-1.777 (.3)	77	901250	.351
net wealth 5	-2.911 (.18)	203	-1.647 (.29)	77	883999	.276
CI lower (.025)	-3.333		-2.465			
CI upper (.975)	-2.343		-1.045			

Table 7: Pareto Alphas and results of Hausman-Test (threshold at 99th percentile)

$$H_0 : \hat{\alpha}_{imp}^{SOEP} = \hat{\alpha}_{imp}^{Pretest}$$

Results of Hausman test (3/4)

	$\hat{\alpha}_{p95}$	N_{p95}	$\hat{\alpha}_{p99}$	N_{p99}	<i>Hausman</i>
net wealth 1	-2.047 (.035)	1019	-2.82 (.167)	204	.213
net wealth 2	-2.088 (.034)	1017	-2.884 (.157)	205	.19
net wealth 3	-1.966 (.039)	1021	-2.646 (.204)	204	.225
net wealth 4	-2.006 (.037)	1019	-2.916 (.189)	204	.177
net wealth 5	-2.048 (.036)	1019	-2.912 (.179)	203	.18
CI lower (.025)	-2.231		-3.319		
CI upper (.975)	-1.831		-2.352		

Table 8: Pareto's Alphas based on the SOEP

$$H_0 : \hat{\alpha}_{imp,p95}^{SOEP} = \hat{\alpha}_{imp,p99}^{SOEP}$$

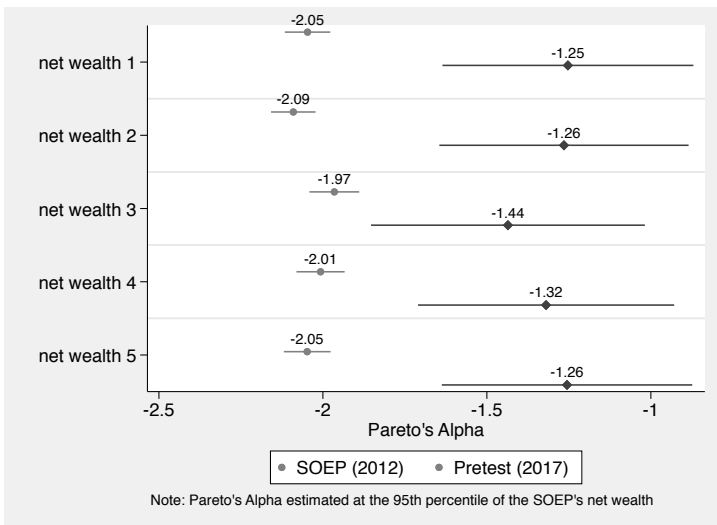
Results of Hausman test (4/4)

	$\hat{\alpha}_{p95}$	N_{p95}	$\hat{\alpha}_{p99}$	N_{p99}	<i>Hausman</i>
net wealth 1	-1.253 (.193)	104	-1.61 (.278)	80	.121
net wealth 2	-1.265 (.192)	106	-1.714 (.292)	79	.102
net wealth 3	-1.436 (.21)	104	-2.027 (.325)	76	.104
net wealth 4	-1.319 (.197)	105	-1.777 (.301)	77	.111
net wealth 5	-1.256 (.193)	104	-1.647 (.29)	77	.123
CI lower (.025)	-1.639		-2.465		
CI upper (.975)	-.972		-1.045		

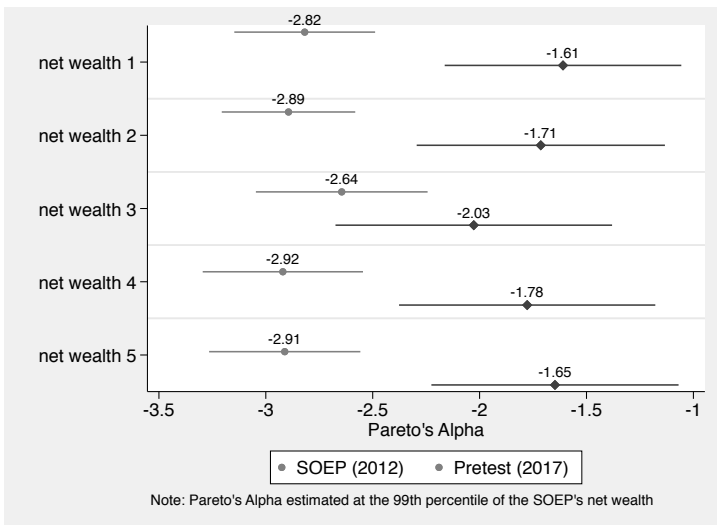
Table 9: Pareto's Alphas based on the Pretest

$$H_0 : \hat{\alpha}_{imp,p95}^{Pretest} = \hat{\alpha}_{imp,p99}^{Pretest}$$

Coefplot (1/2)



Coefplot (2/2)



Estimating the Top Wealths of the SOEP (1/2)

	<i>Top 5%</i>	<i>Top 2.5%</i>	<i>Top 1%</i>	<i>Top 0.1%</i>
net wealth 1	1244.552	2164.127	4496.786	28252.787
net wealth 2	1596.866	2762.137	5699.456	35186.422
net wealth 3	40.436	65.528	124.046	616.674
net wealth 4	294.669	498.281	997.835	5713.763
net wealth 5	1237.366	2149.067	4458.457	27901.178
mean	882.778	1527.828	3155.316	19534.165
N	3300000	1650000	660000	66000

Note: We estimated the Top percentages of the SOEP with $\hat{\alpha}_{pretest}$ and a threshold at the 95th percentile of the net wealth of the SOEP. Net wealths in mio Euros.

Table 10: Estimated net wealths for selected top percentages

Estimating the Top Wealths of the SOEP (2/2)

	<i>Top 5%</i>	<i>Top 2.5%</i>	<i>Top 1%</i>	<i>Top 0.1%</i>
net wealth 1	13721.229	21104.955	37288.393	155863.99
net wealth 2	6081.691	9113.403	15555.554	59621.496
net wealth 3	33.662	47.384	74.461	231.849
net wealth 4	3696.818	5460.02	9142.903	33396.908
net wealth 5	18778.842	28602.286	49883.767	201828.1
mean	8462.448	12865.61	22389.016	90188.467
N	3300000	1650000	660000	66000

Note: We estimated the Top percentages of the SOEP with $\hat{\alpha}_{Pretest}$ and a threshold at the 99th percentile of the net wealth of the SOEP. Net wealths in mio Euros.

Table 11: Estimated net wealths for selected top percentages

Final Results

- On the one hand, the α parameters seem to be robust.
- On the other hand, CIs of the SOEP and the Pretest are not overlapping. Thus, test results are not trustworthy.
- We expect that with a larger sample size of the high-worth individuals, we could achieve conclusive results.

Thank You for your Attention!