

# The invisible rich: A new approach to differential nonresponse in wealth survey data

EEA-ESEM Congress

25 August 2021

Dr. Rafael Wildauer



joint work with Prof. Jakob Kapeller

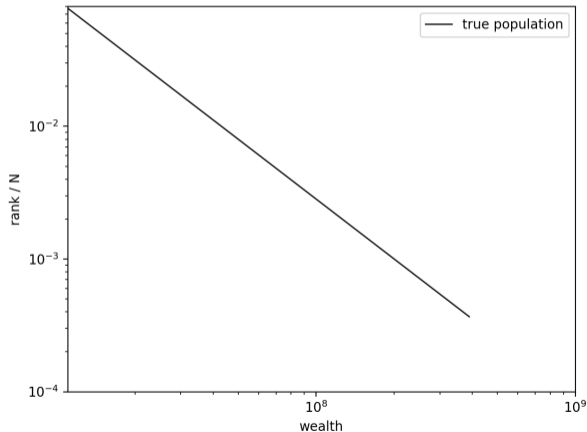
# Motivation

# Motivation

- survey data: contextual information and representative
- income and wealth surveys: poor tail coverage (HFCS, WAS) because
  - ▶ **nonobservation bias** [Eckerstorfer et al. \(2016\)](#)
  - ▶ **differential nonresponse bias** ([Bricker et al. 2016](#), [D'Alessio & Faiella 2002](#), [Osier 2016](#))
- three standard approaches:
  - ① do nothing, trust data (**raw data approach**)
  - ② fit Pareto tail (**Pareto correction, PC**)
  - ③ add rich list data and fit Pareto tail ([Vermeulen 2018](#)) (**rich list correction, RL**)
- this paper proposes a fourth approach: **rank correction, RC**

## A Pareto tail in income or wealth

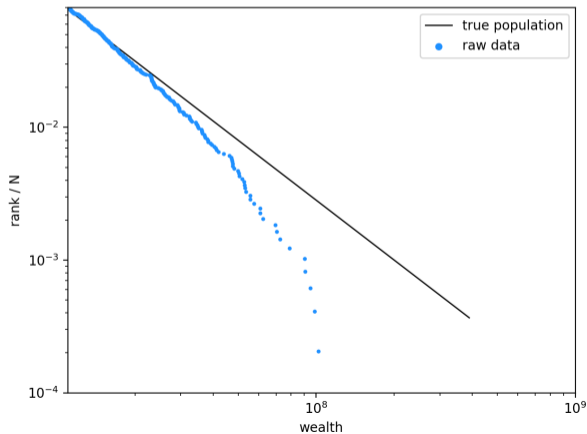
rank: households ranked in descending order (rank 1 = richest household)



## A Pareto tail in income or wealth

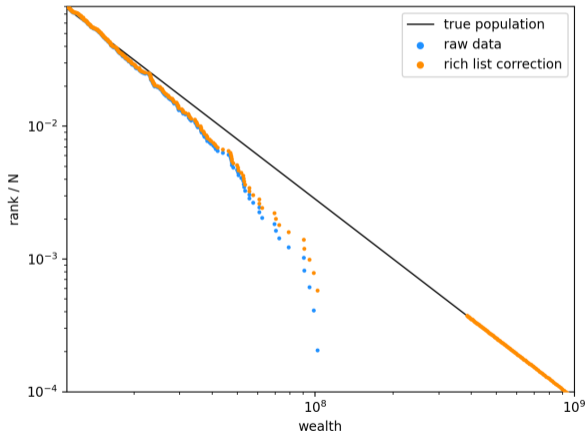
richest 1600 households excluded from sample design due to privacy concerns

**(binary differential nonresponse)**



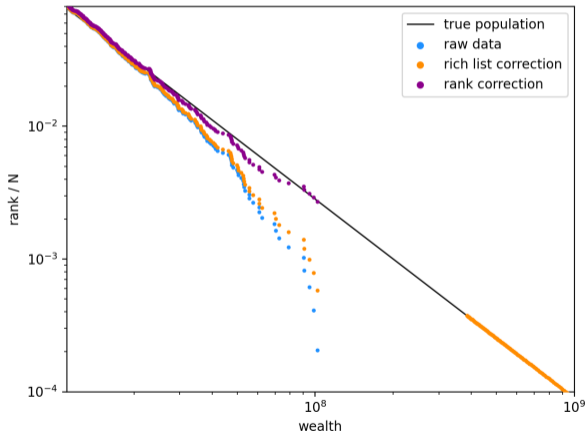
## A Pareto tail in income or wealth

amend sample by richest 100 observations from rich list



# A Pareto tail in income or wealth

correcting the ranks of the observed households





## The rank correction approach

## The rank correction approach

Standard approach of fitting Pareto tail to survey data: regress

$$\ln(\text{rank}_i) = c_1 - \alpha \ln(\text{wealth}_i) + \epsilon_i$$

Rank correction approach: regress

$$\ln(\text{rank}_i + u) = c_1 - \alpha \ln(\text{wealth}_i) + \epsilon_i$$

where  $u$  is the correction factor to correct household ranks

## How to find $u$ ?

- in some cases we know  $u$ : SCF ( $u=400$ )
- often we don't
- search for  $u$  which yields best fit, exploiting linear log-log relationship
- find  $u$  which minimizes RMSE:

$$\min_u \sqrt{\frac{1}{n} \sum_{i=1}^n (\ln(\text{rank}_i + u) - \hat{c}_1 + \hat{\alpha} \ln(\text{wealth}_i))^2} \quad (1)$$

- one-dimensional optimization problem solved numerically

# Monte Carlo Simulations

## Monte Carlo: Set up

- Pareto tail distribution of 1 million households with wealth of at least 1 million Euros for
  - ▶ for three different shape parameters  $\alpha = (1.25, 1.5, 1.75)$
  - ▶ and four different net sample sizes  $(s_1, \dots, s_4) = (0.3\%, 0.8\%, 2\%, 6\%)$
- further we look at two different **nonresponse mechanisms**:
  - ▶ binary differential nonresponse: richest 1600 households excluded
  - ▶ general differential nonresponse: response probability declines in wealth:  
$$ResProb(wealth_i) = 0.903 - 0.036594 \cdot \ln(wealth_i)$$
- draw 1000 samples for each of these combinations

## Monte Carlo: Binary differential non-response

- top 1600 households excluded from sample frame due to privacy
- net sample sizes HFCS: FR 0.4‰, IT 0.3‰, DE 0.1‰

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
true $\alpha$	sample size (‰)	data WD	PC $\hat{\alpha}$	PC WD	RC $\hat{\alpha}$	RC WD	RC $u$	RLS $\hat{\alpha}$	RLS WD	RLL $\hat{\alpha}$	RLL WD
1.5	0.20	-11.4	1.547	-5.5	1.493	1.2	1800	1.523	-2.8	1.503	-0.3

## Monte Carlo: General differential non-response

- $ResProb(wealth_i) = 0.903 - 0.036594 \cdot \ln(wealth_i)$
- net sample sizes HFCS: FR 0.4‰, IT 0.3‰, DE 0.1‰

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
true alpha	sample size	data WD	PC $\hat{\alpha}$	PC WD	RC $\hat{\alpha}$	RC WD	RC u	RLS $\hat{\alpha}$	RLS WD	RLL $\hat{\alpha}$	RLL WD
1.5	0.20	-18.2	1.643	-14.6	1.597	-10.6	1300	1.58	-9.1	1.508	-1.1

## Application to wealth survey data



## Application to wealth survey data

- applied methods to three household wealth surveys: HFCS (EU), WAS (UK), SCF (US)
- tail defined as top 3% and top 1% for most effective oversampling (US, FR and ES)
- four estimators applied to each country:
  - ① raw data
  - ② Pareto correction
  - ③ rich list correction
  - ④ rank correction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Country	$\alpha$ PC	$\alpha$ RC	$\alpha$ RL	raw	PC/raw	RC/raw	RL/raw	U	RL
US <sub>2013</sub> *	1.789	1.671	1.496	66.8	98.	102	110	360	442
UK	1.958	1.949	1.547	6.59	100	100	110	100	37
Austria	1.404	1.390	1.342	1.00	110	111	113	100	10
Belgium	2.348	1.403	1.687	1.58	99	122	108	11,780	3
Germany	1.597	1.397	1.340	8.50	100	113	119	10,760	85
Spain*	1.724	1.334	1.582	4.77	100	112	102	1,560	16
Finland	2.140	2.012		0.51	100	102		100	
France*	1.525	1.423	1.351	7.03	103	106	110	1,200	43
Greece	3.382	2.146	1.401	0.44	98	105	127	10,300	3
Italy	2.417	2.288	1.372	5.59	100	101	126	1,300	35
Luxembourg	1.578	1.306		0.16	104	122		100	
Malta	1.291	1.168		0.06	132	152		64	
Netherlands	3.082	1.867	1.419	1.15	99	108	124	22,800	6
Slovenia	1.267	1.126		0.11	141	171		300	
Estonia	1.727	1.607		0.06	101	105		100	
Hungary	1.719	1.552		0.21	102	107		900	
Poland	2.193	2.027		1.31	100	102		1,920	

## Top wealth shares - before and after corrections

Country	raw	PC	RC	RL
	Top 1%	Top 1%	Top 1%	Top 1%
US	35.4%	34.1%	36.2%	40.6%
UK	15.1%	14.8%	14.9%	21.6%
Austria	25.4%	29.9%	30.6%	31.8%
Germany	23.6%	25.3%	32.6%	35.6%
Spain	16.3%	16.1%	24.2%	17.9%
Finland	13.3%	13.4%	14.5%	NA
France	18.7%	20.7%	23.1%	25.4%
Greece	9.2%	9.1%	14.0%	26.9%
Italy	11.7%	11.6%	12.3%	27.8%
Luxembourg	18.8%	20.7%	30.1%	NA
Malta	19.9%	30.9%	36.8%	NA
Slovenia	22.9%	33.3%	42.5%	NA
Estonia	21.2%	22.3%	24.9%	NA
Hungary	17.2%	18.6%	22.2%	NA
Poland	11.7%	12.1%	13.3%	NA

# Reconciliation I

	country	data and method	(1) top 0.1%	(2) top 1%	(3) top 10%
(1)	France	World Inequality Database	8.2	23.4	55.3
(2)	France	rank correction estimator	11.6	23.1	53.4
(3)	France	uncorrected survey data (HFCS)	7.3	18.7	50.8
(4)	USA	World Inequality Database	20.3	37.0	73.2
(5)	USA	rank correction estimator	14.4	36.2	75.3
(6)	USA	uncorrected survey data (SCF)	13.1	35.4	75.0
(7)	UK	World Inequality Database		19.9	51.9
(8)	UK	rank correction estimator	5.6	15.1	45.7
(9)	UK	uncorrected survey data (WAS)	4.8	14.9	45.7
(10)	DE	<a href="#">Schröder et al. (2020)</a>		35.3	67.3
(11)	DE	rank correction estimator	16.8	32.6	64.0
(12)	DE	uncorrected survey data (HFCS)	6.3	23.6	59.8

## Reconciliation II

Country	(1) raw	(2) PC	(3) RC	(4) RL
US	43	220	391	904.
UK	0	1	1	19
Austria	0	11	12	12
Germany	0	20	80	117
Spain	0	4	44	8
Finland	0	0	0	0
France	0	19	36	56
Greece	0	0	0	3
Italy	0	0	0	55
Luxembourg	0	1	3	0
Malta	0	1	2	0
Slovenia	0	3	4	0
Estonia	0	0	0	0
Hungary	0	0	0	0
Poland	0	0	0	0

## Conclusion

# Conclusion

- ① Rank correction: new tool for Pareto tail fitting
- ② Key features:
  - ▶ no additional data (rich list) required
  - ▶ easy to implement
  - ▶ consistently outperforms raw data and standard Pareto correction
- ③ no deus ex machina: better oversampling required in data production

Thank you!

r.wildauer@gre.ac.uk

 @RafaelWildauer



## References I

- Bricker, J., Henriques, A., Krimmel, J. & Sabelhaus, J. (2016), 'Measuring income and wealth at the top using administrative and survey data', *Brookings Papers on Economic Activity* **2016**(1), 261–331.
- D'Alessio, G. & Faiella, I. (2002), 'Non-response behaviour in the bank of italy's survey of household income and wealth', *Temi di discussione (Bank of Italy Economic working papers)* **2002**(462).
- Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B., Springholz, F. & Wildauer, R. (2016), 'Correcting for the missing rich: An application to wealth survey data', *Review of Income and Wealth* **62**(4), 605–627.

## References II

- Osier, G. (2016), 'Unit non-response in household wealth surveys: Experience from the eurosystem's household finance and consumption survey', *European Central Bank Statistics Paper Series* **2016**(15).
- Schröder, C., Bartels, C., Göbler, K., Grabka, M. M. & König, J. (2020), 'Millionaires under the microscope: Data gap on top wealth holders closed; wealth concentration higher than presumed', *DIW Weekly Report* .
- Vermeulen, P. (2018), 'How fat is the top tail of the wealth distribution?', *Review of Income and Wealth* **64**(2), 357–387.