

HOW FAT IS THE TOP TAIL OF THE WEALTH DISTRIBUTION?

Philip Vermeulen
European Central Bank
DG-R MPRD

SEEK, Mannheim

16/May/2014

This slideshow should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the author and do not necessarily reflect those of the ECB.

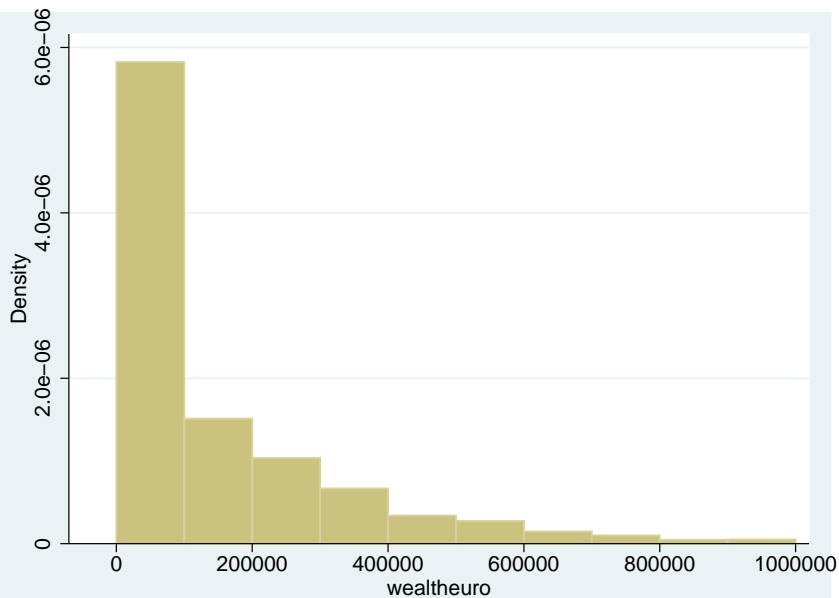
Wealth distribution and its tail

- 30 percent of wealth is held by 1 percent of the population (Kennickell, 2009)
- Piketty “*Capital in the Twenty-First Century*”
- Stiglitz “*The Price of Inequality: How Today’s Divided Society Endangers Our Future*”
- monetary and fiscal policy have first order distributional implications
- shift from representative agent model to heterogeneous agent models

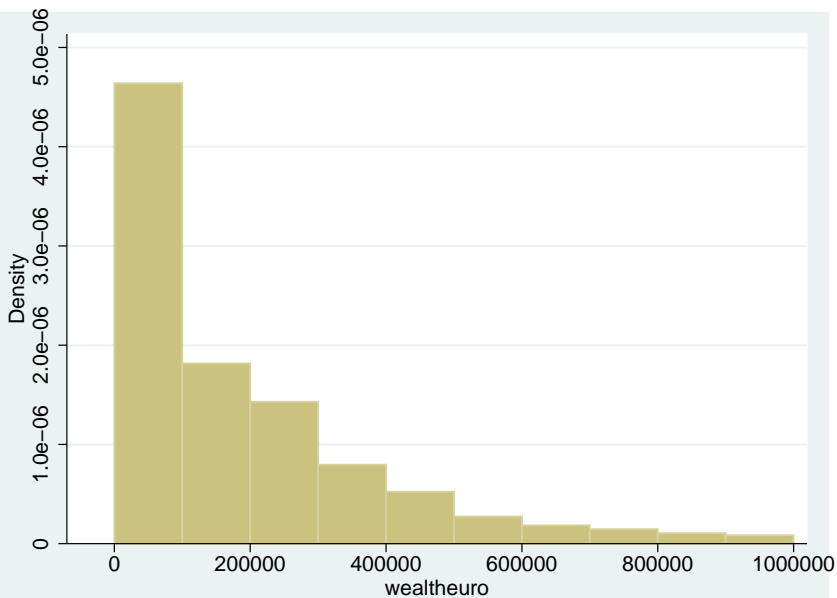
Wealth distribution data

- Household wealth surveys
- US Survey of consumer finances
- European Household Finance and Consumption Survey
- Complex survey designs: stratification, clustering, weighting
- representative of population

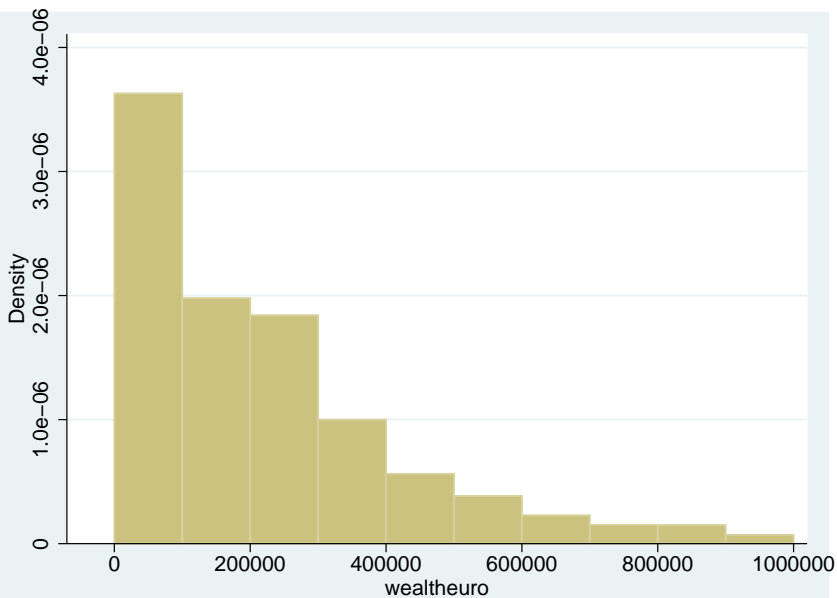
Part of the wealth distribution Germany

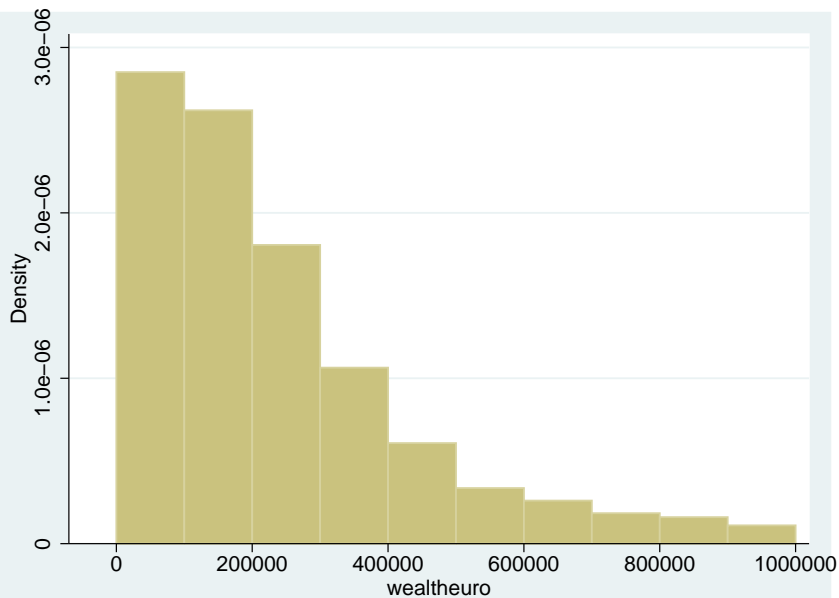


Part of the wealth distribution France



Part of the wealth distribution Italy



Part of the wealth distribution Spain

The wealth distribution is special

- Practically all familiar distributions: inflation, bond returns, stock returns, growth rates,
- “mean” and “variance” that makes sense (central tendency and fluctuations around it)
- wealth distribution in contrast:
- an astronomically long tail: roughly twenty thousand times larger than the first 95 percent of the distribution
- Has a “mean” and “variance” that don't make sense.
- NEED to sample especially from the tail to estimate its importance

Household wealth survey data: oversampling the rich

- a simple random sample would yield an inefficient estimate of the distribution of wealth
- very different strategies to oversample wealthy households
- US: income tax files
- Spain, France: wealth tax files
- Finland: income register data
- Germany, Belgium: income of regions
- Portugal, Austria: metropolitan areas
- Italy, Netherlands: No oversampling
- problem when survey weights cannot be corrected for *differential* non-response

Different oversampling strategies

Number of wealthy households in the samples > 2 million euro
 Absolute Number vs Pct in sample

	Sample size	Absolute number	Pct of sample
USA	6482	965	15
Germany	3565	85	2
France	15006	638	4
Italy	7951	78	1
Spain	6197	544	9
Netherlands	1301	2	0.15

Purpose of the research

Purpose of the research

- How much wealth is in the tail?

Purpose of the research

- How much wealth is in the tail?
- Does the survey data give a good account of tail wealth?

Purpose of the research

- How much wealth is in the tail?
- Does the survey data give a good account of tail wealth?
- Show the importance of *differential* non-response for tail wealth estimation

Purpose of the research

- How much wealth is in the tail?
- Does the survey data give a good account of tail wealth?
- Show the importance of *differential* non-response for tail wealth estimation
- Provide new method to estimate tail wealth in the presence of *differential* non-response.

Purpose of the research

- How much wealth is in the tail?
- Does the survey data give a good account of tail wealth?
- Show the importance of *differential* non-response for tail wealth estimation
- Provide new method to estimate tail wealth in the presence of *differential* non-response.
- New estimates of tail wealth: US, Germany, France, Spain, ...

*The extreme tail***The forbes list (billion euro)**

	Individuals	Total wealth	Pct country wealth
USA	396	978.6	2.3
Germany	52	183.3	2.4
France	11	60.1	0.9
Italy	14	46.6	0.7
Spain	12	28.3	0.6
Netherlands	3	4.8	0.4
Belgium	1	1.9	0.1
Austria	5	13.0	1.2
Finland	1	1.0	0.2
Portugal	2	4.1	0.7

*The Household data: the missing end of the tail***MIND THE GAP**

Max wealth Survey vs Min wealth at Forbes
Million euros

	Max wealth SCF/HFCS	Min wealth Forbes
USA	806	737
Germany	76	818
France	153	810
Italy	26	893
Spain	409	780
Netherlands	5	958
Belgium	8	1920
Portugal	27	1110
Austria	22	1560
Finland	15	958

Some observations

- large gap between the highest wealth household in HFCS sample and the poorest household in Forbes
- suggest combination of non-response and lack of effective oversampling
- suggest to estimate the tail of the distribution survey + Forbes data

How to estimate tail wealth

Survey Sample only

- Directly from the observations in the survey samples
- Assume tail is Pareto:(pseudo) maximum likelihood
- Assume tail is Pareto: Regression method on survey sample

Survey sample + Forbes

- Assuming tail is Pareto: Regression method on survey sample+Forbes data

A candidate distribution for the tail

- Pareto distribution (Power law)
- complementary CDF or tail function

$$P(W > w) = \left(\frac{w_{min}}{w}\right)^\alpha \quad (1)$$

- defined on the interval $[w_{min}, \infty[$ and $\alpha > 0$
- α tail index determines the fatness of the tail. The lower α , the fatter the tail.
- Klass et al. (2006), Ogwang (2011)

Method 1: Pseudo maximum likelihood

- Maximum likelihood is impossible, cannot write the likelihood function
- Pseudo maximum likelihood
- act "as if the data were a census"
- $L(\alpha, w_{min} | W) = \prod f(w_i)^{sw_i}$
- Does not deal with the *differential* non-response problem

An interesting property of the Pareto

- complementary CDF or tail function

$$P(W > w) = \left(\frac{w_{min}}{w}\right)^\alpha \quad (2)$$

- $n(w_i)$ the number of households that have wealth above w_i
(rank of the household)



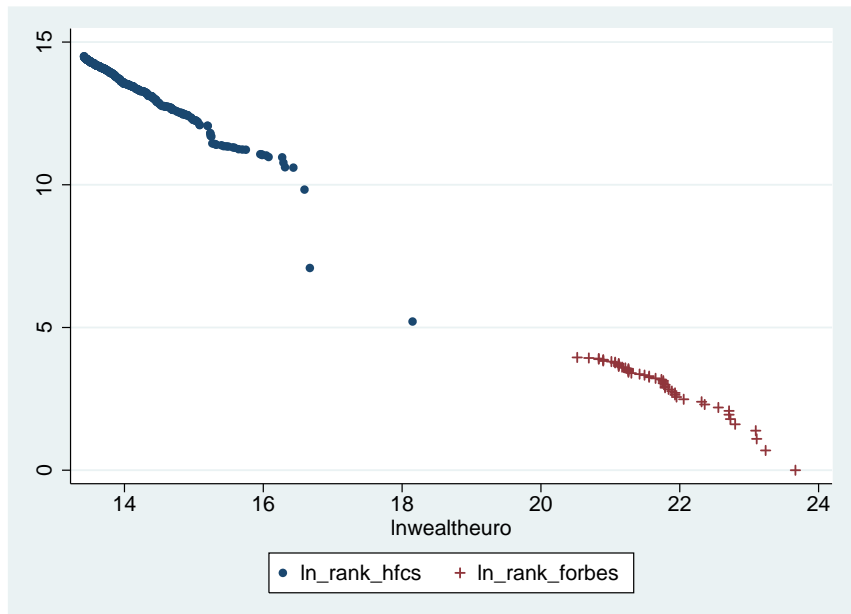
$$\frac{n(w_i)}{N} = \left(\frac{w_{min}}{w_i}\right)^\alpha \quad (3)$$

- $\log(n(w_i)) = \log(N) + \alpha \log(w_{min}) - \alpha \log(w_i)$
- log of the rank of the household and log of its wealth are on a straight line.

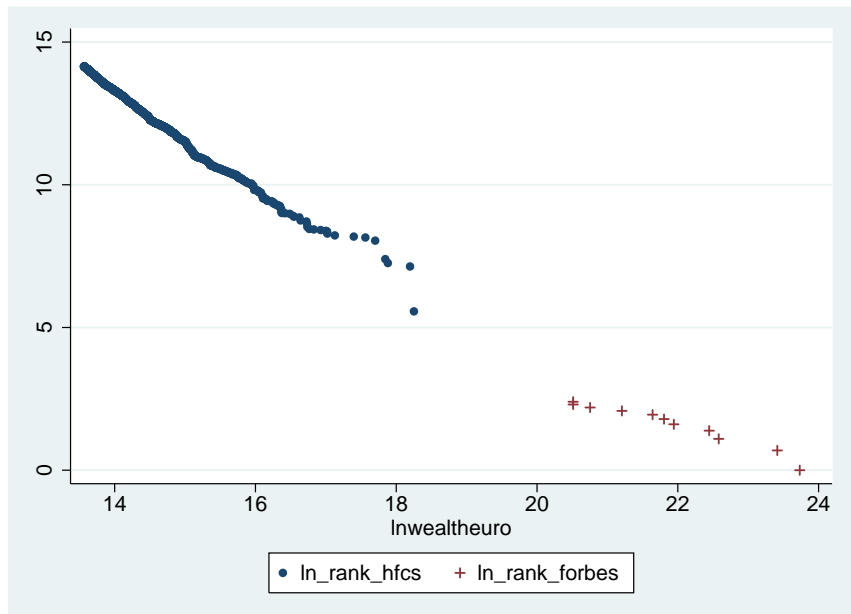
Method 2: Regression method

- $\log(n(w_i)) = \text{constant} - \alpha \log(w_i)$
- log of the rank of the household and log of its wealth are on a straight line.
- use survey data to construct $n(w_i)$
- also use Forbes data to construct $n(w_i)$
- α_{RH} : survey data only
- α_{RHF} : survey data + Forbes data

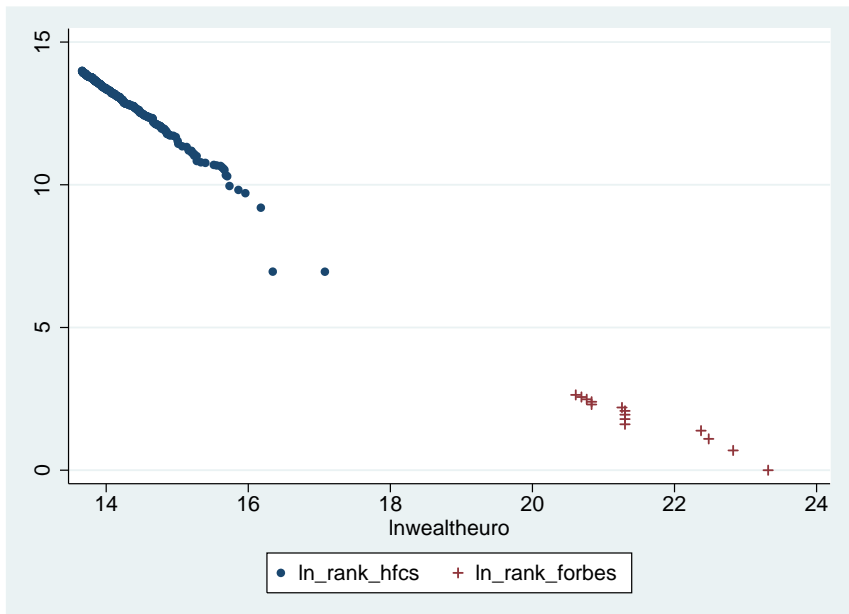
Log rank of the household and log wealth (Germany)



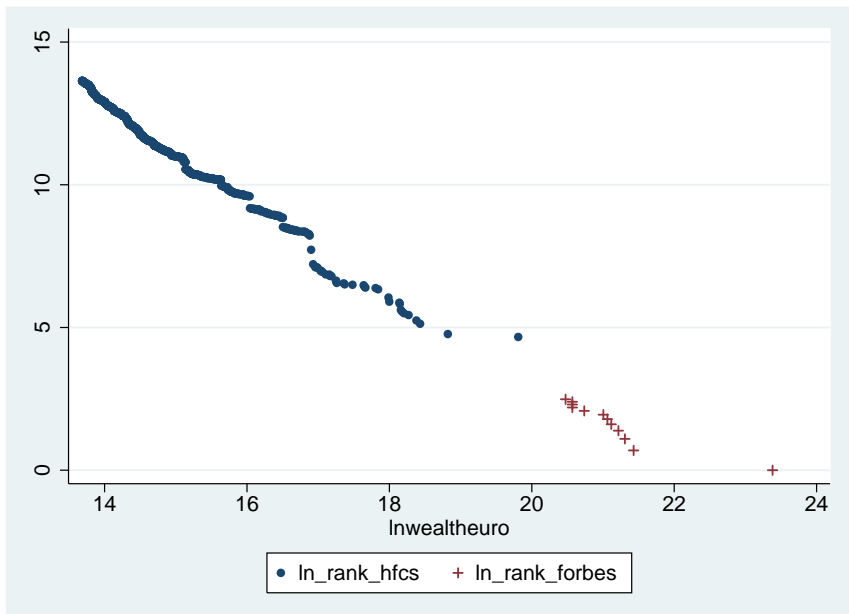
Log rank of the household and log wealth (France)



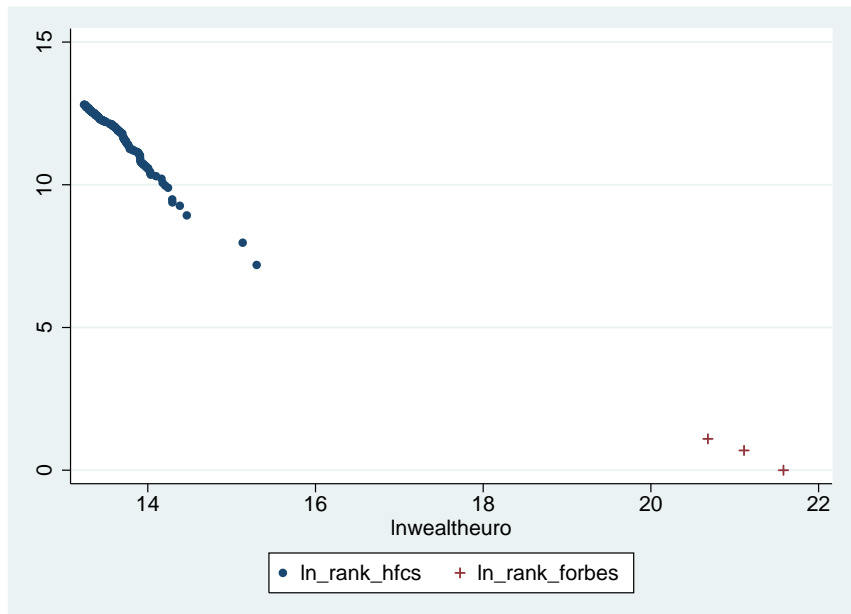
Log rank of the household and log wealth (Italy)



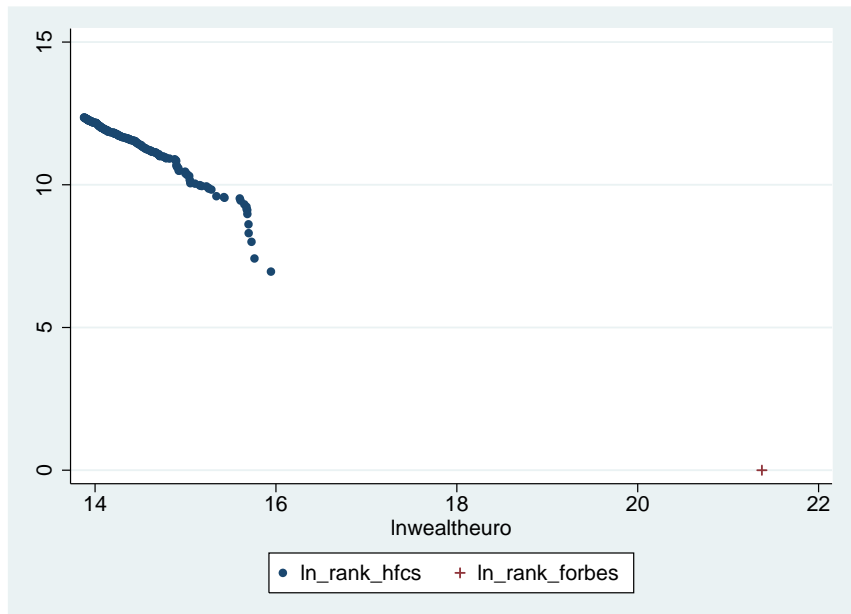
Log rank of the household and log wealth (Spain)



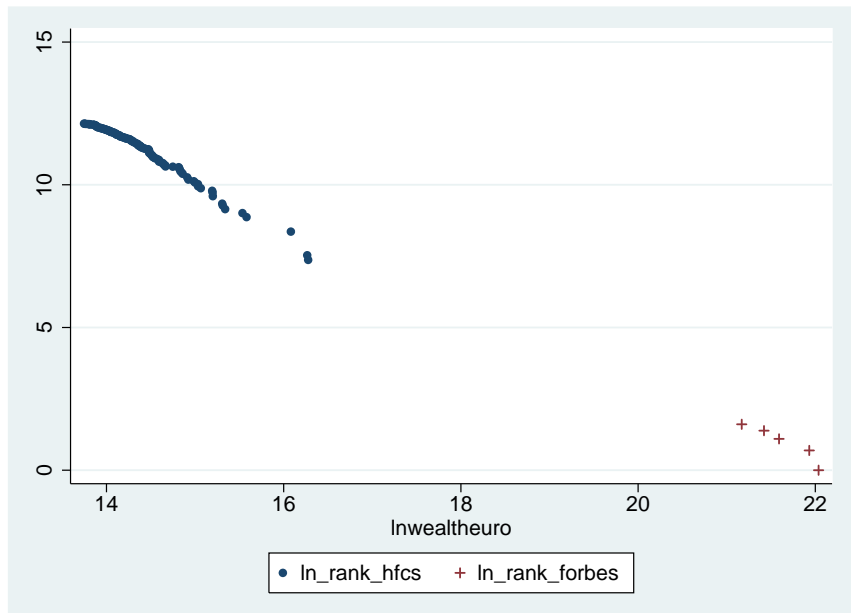
Log rank of the household and log wealth (Netherlands)



Log rank of the household and log wealth (Belgium)



Log rank of the household and log wealth (Austria)



Monte Carlo

- 1 million households draw wealth from Pareto distribution with tail index α
- rich list (households more than 740 million euro)
- A survey sample is drawn for 750 households
- non-response probability follows $0.097+0.0365*\ln(\text{Wealth})$
- Estimate tail index using pseudo maximum likelihood, regression method survey, regression method survey + rich list
- ten thousand Monte Carlo iterations

Pareto tail index estimates under differential non-response

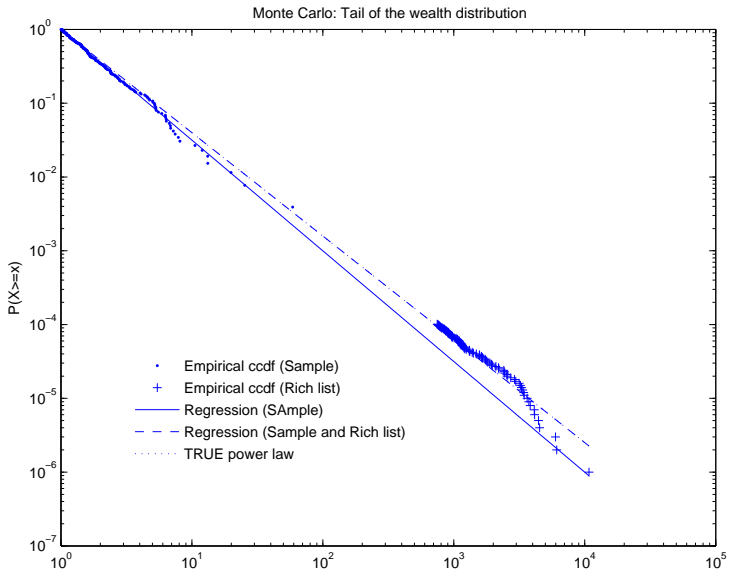
Monte Carlo estimates of Pareto tail index					
α	α_{ml}	α_{reg}	α_{regfor}	Resp obs	Rich list obs
(1)	(2)	(3)	(4)	(5)	(6)
1.30	1.41	1.38	1.30	277	186
	0.08	0.09	0.01	13	13
1.40	1.51	1.48	1.40	279	96
	0.09	0.10	0.02	13	10
1.50	1.61	1.57	1.50	280	50
	0.10	0.10	0.03	13	7
1.60	1.71	1.67	1.60	281	26
	0.10	0.11	0.04	13	5
1.70	1.81	1.77	1.69	282	13
	0.11	0.12	0.05	13	4

Pareto tail index estimates under differential non-response

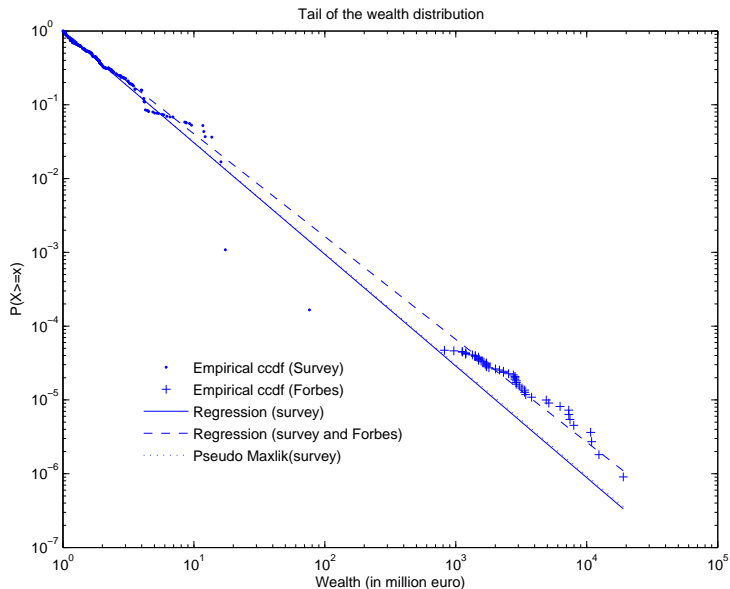
Monte Carlo estimates of tail wealth as a proportion of true tail wealth

α	survey est.	α_{ml}	α_{reg}	α_{regfor}
(1)	(2)	(3)	(4)	(5)
1.30	0.77	0.82	0.89	1.02
	0.41	0.14	0.20	0.06
1.40	0.83	0.87	0.92	1.01
	0.29	0.11	0.15	0.04
1.50	0.87	0.90	0.94	1.01
	0.20	0.09	0.12	0.03
1.60	0.90	0.91	0.95	1.01
	0.16	0.08	0.10	0.03
1.70	0.92	0.93	0.96	1.01
	0.18	0.07	0.09	0.04

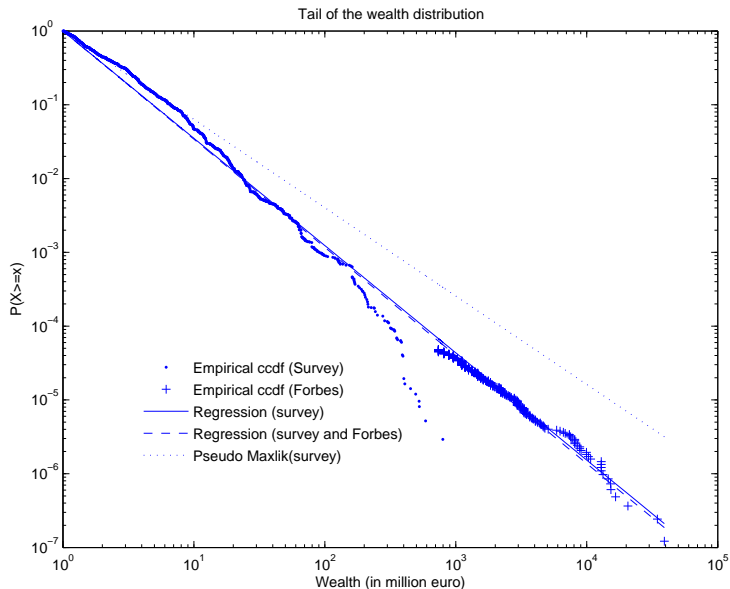
TAIL WEALTH DISTRIBUTION (MONTE CARLO EXAMPLE)



TAIL WEALTH DISTRIBUTION (Germany)



TAIL WEALTH DISTRIBUTION (USA)



Percentage wealth share of top 1 pct of households

		Regres	Δ
	SURVEY	incl Forbes	
Austria	23	33	+10
Germany	24	32	+8
Italy	14	21	+7
Belgium	12	17	+5
Netherlands	9	12	+3
Finland	12	15	+3
countries using tax records to oversample !			
France	18	20	+2
USA	34	35	+1
Spain	15	16	+1

Conclusion

- unobserved *differential* non-response creates substantially downward biased estimates of tail wealth
- oversampling (using tax records) allows to observe and address differential non-response
- combining Forbes tail wealth individual with survey data improves tail wealth estimates
- Germany, Austria, Italy, etc have substantially higher tail wealth than can be derived from the HFCS data only