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

<table>
<thead>
<tr>
<th>Country</th>
<th>Sample size</th>
<th>Absolute number</th>
<th>Pct of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>6482</td>
<td>965</td>
<td>15</td>
</tr>
<tr>
<td>Germany</td>
<td>3565</td>
<td>85</td>
<td>2</td>
</tr>
<tr>
<td>France</td>
<td>15006</td>
<td>638</td>
<td>4</td>
</tr>
<tr>
<td>Italy</td>
<td>7951</td>
<td>78</td>
<td>1</td>
</tr>
<tr>
<td>Spain</td>
<td>6197</td>
<td>544</td>
<td>9</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1301</td>
<td>2</td>
<td>0.15</td>
</tr>
</tbody>
</table>
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 \textit{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)

<table>
<thead>
<tr>
<th>Country</th>
<th>Individuals</th>
<th>Total wealth</th>
<th>Pct country wealth</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>396</td>
<td>978.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Germany</td>
<td>52</td>
<td>183.3</td>
<td>2.4</td>
</tr>
<tr>
<td>France</td>
<td>11</td>
<td>60.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Italy</td>
<td>14</td>
<td>46.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Spain</td>
<td>12</td>
<td>28.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Netherlands</td>
<td>3</td>
<td>4.8</td>
<td>0.4</td>
</tr>
<tr>
<td>Belgium</td>
<td>1</td>
<td>1.9</td>
<td>0.1</td>
</tr>
<tr>
<td>Austria</td>
<td>5</td>
<td>13.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Portugal</td>
<td>2</td>
<td>4.1</td>
<td>0.7</td>
</tr>
</tbody>
</table>
The Household data: the missing end of the tail

MIND THE GAP
Max wealth Survey vs Min wealth at Forbes
Million euros

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

\[ P(W > w) = \left( \frac{w_{\text{min}}}{w} \right)^{\alpha} \] (1)

- Defined on the interval \([w_{\text{min}}, \infty[\) and \(\alpha > 0\)
- \(\alpha\) tail index determines the fatness of the tail. The lower \(\alpha\), the fatter the tail.
- Klass et al. (2006), Ongwang (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”
- \( L(\alpha, w_{min} \mid W) = \prod f(w_i)^{sw_i} \)
- Does not deal with the differential non-response problem
An interesting property of the Pareto

- complementary CDF or tail function

\[ P(W > w) = \left( \frac{w_{\text{min}}}{w} \right)^\alpha \]  \hspace{1cm} (2)

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

\[ \frac{n(w_i)}{N} = \left( \frac{w_{\text{min}}}{w_i} \right)^\alpha \]  \hspace{1cm} (3)

- \( \log(n(w_i)) = \log(N) + \alpha \log(w_{\text{min}}) - \alpha \log(w_i) \)

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 \( n(w_i) \)
- \( \hat{\alpha}_{RH} \): survey data only
- \( \hat{\alpha}_{RHF} \): survey data + Forbes data
Log rank of the household and log wealth (France)
Tail distributions

Log rank of the household and log wealth (Italy)

![Graph showing log rank of the household and log wealth (Italy)]
Log rank of the household and log wealth (Spain)
Tail distributions

Log rank of the household and log wealth (Netherlands)
Log rank of the household and log wealth (Belgium)
Tail distributions

Log rank of the household and log wealth (Austria)

Philip Vermeulen
Monte Carlo

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

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\alpha_{ml}$</th>
<th>$\alpha_{reg}$</th>
<th>$\alpha_{reg for}$</th>
<th>Resp obs</th>
<th>Rich list obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.30</td>
<td>1.41</td>
<td>1.38</td>
<td>1.30</td>
<td>277</td>
<td>186</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.09</td>
<td>0.01</td>
<td></td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>1.40</td>
<td>1.51</td>
<td>1.48</td>
<td>1.40</td>
<td>279</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.10</td>
<td>0.02</td>
<td></td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>1.50</td>
<td>1.61</td>
<td>1.57</td>
<td>1.50</td>
<td>280</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>0.03</td>
<td></td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>1.60</td>
<td>1.71</td>
<td>1.67</td>
<td>1.60</td>
<td>281</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.11</td>
<td>0.04</td>
<td></td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>1.70</td>
<td>1.81</td>
<td>1.77</td>
<td>1.69</td>
<td>282</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.12</td>
<td>0.05</td>
<td></td>
<td>13</td>
<td>4</td>
</tr>
</tbody>
</table>
Pareto tail index estimates under differential non-response

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

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>survey est.</th>
<th>$\alpha_{ml}$</th>
<th>$\alpha_{reg}$</th>
<th>$\alpha_{regfor}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.30</td>
<td>0.77</td>
<td>0.82</td>
<td>0.89</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>0.41</td>
<td>0.14</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td>1.40</td>
<td>0.83</td>
<td>0.87</td>
<td>0.92</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
<td>0.11</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>1.50</td>
<td>0.87</td>
<td>0.90</td>
<td>0.94</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.09</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>1.60</td>
<td>0.90</td>
<td>0.91</td>
<td>0.95</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.08</td>
<td>0.10</td>
<td>0.03</td>
</tr>
<tr>
<td>1.70</td>
<td>0.92</td>
<td>0.93</td>
<td>0.96</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
<td>0.07</td>
<td>0.09</td>
<td>0.04</td>
</tr>
</tbody>
</table>
TAIL WEALTH DISTRIBUTION (MONTE CARLO EXAMPLE)

Monte Carlo: Tail of the wealth distribution
TAIL WEALTH DISTRIBUTION (Germany)

Tail of the wealth distribution

- Empirical ccdf (Survey)
- Empirical ccdf (Forbes)
- Regression (survey)
- Regression (survey and Forbes)
- Pseudo Maxlik(survey)

Figure:
Philip Vermeulen
How fat is the top tail
16/May/2014 32 / 34
TAIL WEALTH DISTRIBUTION (USA)

Tail of the wealth distribution

Wealth (in million euro)
P(X ≥ x)

Empirical cccdf (Survey)

Empirical cccdf (Forbes)

Regression (survey)

Regression (survey and Forbes)

Pseudo Maxlik(survey)

Figure:

Philip Vermeulen

16/May/2014
### Percentage wealth share of top 1 pct of households

<table>
<thead>
<tr>
<th>Country</th>
<th>Surveys</th>
<th>Surveys incl Forbes</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>23</td>
<td>33</td>
<td>+10</td>
</tr>
<tr>
<td>Germany</td>
<td>24</td>
<td>32</td>
<td>+8</td>
</tr>
<tr>
<td>Italy</td>
<td>14</td>
<td>21</td>
<td>+7</td>
</tr>
<tr>
<td>Belgium</td>
<td>12</td>
<td>17</td>
<td>+5</td>
</tr>
<tr>
<td>Netherlands</td>
<td>9</td>
<td>12</td>
<td>+3</td>
</tr>
<tr>
<td>Finland</td>
<td>12</td>
<td>15</td>
<td>+3</td>
</tr>
<tr>
<td>France</td>
<td>18</td>
<td>20</td>
<td>+2</td>
</tr>
<tr>
<td>USA</td>
<td>34</td>
<td>35</td>
<td>+1</td>
</tr>
<tr>
<td>Spain</td>
<td>15</td>
<td>16</td>
<td>+1</td>
</tr>
</tbody>
</table>

*countries using tax records to oversample!*
**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