On Top of the Top:
Top-Corrected Wealth Distributions for Eurozone Countries

This Version: Jan 2021

Franziska Disslbacher[a,b]
Miachel Ertl[b]
Emanuel List[a,d]
Patrick Mokre[b,c]
Matthias Schnetzer[b]

[a] Vienna University of Economics and Business (WU)
[b] Vienna Chamber of Labour
[c] The New School
[d] Macroeconomic Policy Institute (IMK)

The missing wealth(y) in survey data

The wealthiest are missing in survey data (Davies and Shorrocks, 2000)

1. **Differential non-response:** Survey non-response is correlated with wealth (D'Alessio et al., 2002; Kenickell, 2019)
2. **Differential under-reporting:** Richer households are more likely to underreport wealth (Vermeulen, 2016)

Sampling errors: Even in absence of non-sampling errors, wealth at the top will be underestimated in survey data (Davies, Sandström, et al., 2010; Kenickell, 2019)

The distribution of household wealth is top-heavy. Thus, the top tail is crucial for understanding wealth inequality

Semi-Parametric remedy:

- Re-estimate the frequency **above a cut-off point** w_{min} according to a functional form (Pareto type I distribution)
- Combination of re-estimated tail and the observational frequency **below the cutoff** w_{min}

Our contribution

- We account for the wealthy missing in survey data for **fourteen euro-zone countries**
- We combine **survey data** on household net wealth (HFCS) with information on top wealth from of national **richlists** from the new European Rich List Database (ERLDB)
- We propose a **unified regression approach** to estimate all aspects of the standard
    1. two-parameter **Pareto type I distribution**, and
    2. the more flexible three-parameter **Generalized Pareto (GP)** distribution
- We close the gap between survey data and rich lists, account for sampling errors and non-sampling errors, and show how severely surveys underestimate the wealth of the super-rich

Data

A. Household Finance and Consumption Survey (HFCS 2017)

- Surveys household's balance sheets and provides information on **net wealth at the household level**
- **Item non-response** is addressed by a multiple imputation strategy
- **Unit non-response and sampling error** are addressed by oversampling of wealthy households
  - Large variation in type of external information used for oversampling and hence in the effectiveness of oversampling

B. European Rich List Database (ERLDB)

- Compiled by journalists
- We use country-specific rich lists that contain more observations than Forbes list of world billionaires (GE: Manager Magazin, FR: Challenges, NL: Quotes,...)
- In most countries: Highest wealth according to HFCS is substantially smaller than smallest fortune according to rich lists
- ⇒ App with many, many lists is in preparation

Data

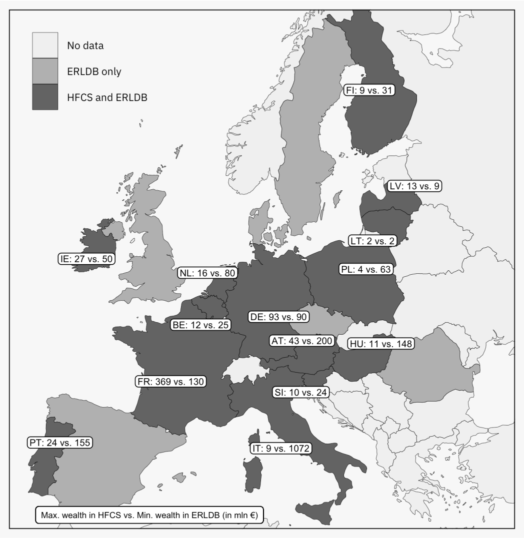

Figure 1: Countries Covered in HFCS and ERLDB

A parametric model for the top tail: The Pareto type I distribution

Pareto type I distribution is a model for heavy-tailed distributions and describes top-tail of the wealth distribution (Davies, Sandström, et al., 2010; Gabaix, 2016) with CDF and PDF given by

$$F(w) = 1 - \left(\frac{w_{min}}{w}\right)^{\alpha} \quad \text{and} \quad f(w) = \frac{\alpha w_{min}^{\alpha}}{w^{\alpha+1}} \quad \text{for } w \geq w_{min} \tag{1}$$

Two parameters: α and w_{min}

- α is a scale parameter inversely related to inequality
  - Many estimators for α have been proposed, we stick to those relying on linearization of the CCDF as suggested by Vermeulen (2018)
- w_{min} locates the distribution (lower bound of observations following a Pareto type I distribution)
  - Previous work has set w_{min} by best guessing or visual inspections of plausible values

The Pareto type I distribution has clear links to economic theory (Jones, 2015; Benhabib et al., 2018)

Estimation of α and w_{min}: A unified regression approach i

Linear relationship between CCDF, $log(1-(F(w))$, and $log(w)$

$$log(1-F(w)) = \alpha log(w_{min}) - \alpha log(w) \tag{2}$$

Empirical CCDF for $w_i > w_{min}$

$$1 - F(\tilde{w}) = \frac{n(w_i)}{n} \tag{3}$$

- $n(w_i)$ is the rank of observations, with $i = 1$ being the richest household, and n is the number of observations in the sample

  Note: Rank i is proportional to the empirical CCDF: $1 - F(\tilde{w})) = \frac{i}{N} \propto i$

$$log(i) = C - \alpha log(w_i) + \epsilon_i \quad \text{where } C = log(n) + \alpha log(w_{min}) \tag{4}$$

- Gabaix and Ibragimov (2011) suggest using $log(i - 0.5)$ to avoid bias towards leading rank
- Vermeulen (2018) shows how to incorporate survey weights
- Chakraborty et al. (2018) observed that median quantile regression (Koenker et al., 1978) is more robust to outliers than OLS estimator of equation 4

Estimation of α and w_{min}: A unified regression approach ii

α: Median Quantile Regression

$$log((i-0.5)\frac{\bar{N_{fi}}}{N} = \underbrace{log\frac{\bar{N_i}}{N} + \alpha log(w_{min})}_{Constant} - \alpha log(w_i) \tag{5}$$

i is a decreasing ranking of wealth, N is the sum of total weights, $\bar{N}$ is the average weight of an observation, and $\bar{N_{fi}}$ denotes the average weight of the first highest i observations

Step 1: Choosing a w_{min}

- For setting the location parameter w_{min} we make use of the interpretation of the root mean squared error (RMSE) as a measure of linearity Langousis et al. (2016)
- We estimate α for many potential w_{min} and choose the w_{min} that minimizes the RMSE

Step 2: Estimating α

- Given w_{min}, we choose the corresponding α

Step 3: Detecting w_0

Under-reporting in the top percentiles implies that the (weighted) mass of observations in the estimated Pareto tail is lower than in the population

w_0 indicates a threshold in the tail above which survey data cannot be trusted (Note: $w_0 > w_{min}$; w_0 corresponds to the transition from continuous to discrete survey observations (Eckerstorfer et al., 2016))

We require continuity of the approximate kernel function of the observations $\hat{f}_{kern}(w)$ instead of continuity of survey data (Dalitz, 2016)

$\Rightarrow$ w_0 is the point where the (kernel) density function of the data falls below the theoretical probability density function

Note: The distance between $\hat{w}_{min}$ and $\hat{w}_0$ is an indicator of how well surveyors are able to battle differential biases among the richest households

Simulating tail observations

Step 4: Simulating tail observations We keep survey observations with $w_i < w_0$, but simulate a new tail above w_0

- Tail length: Calculated number of households with wealth above w_0 according to a $Pareto(\hat{\alpha}, \hat{w}_{min})$ is based on extrapolating the number of households between $\hat{w}_{min}$ and $\hat{w}_0$ with the cumulative density function above $\hat{w}_0$ (Note: $1 - F(\hat{w}_{min})$ gives the theoretical share of tail observations above $\hat{w}_0$)
- To each synthetic observation we assign a wealth value w_i and a uniform household weight of 1

We ensure that the sum of weights of the new distribution corresponds to the sum of weights according to the survey

Generalized Pareto (GP) Framework

Atkinson (2017) has called for a richer functional form for the top tail in the spirit of Pareto (1965 [1896])

We introduce the three parameter Generalized Pareto distribution which is more flexible and has a CCDF given by:

$$\left(1+\xi\frac{W-\mu}{\sigma}\right)^{\frac{-1}{\xi}} \tag{6}$$

GP parameters

$\mu = w_{min}$ is a location parameter that has the same interpretation as w_{min}, i.e. it indicates the threshold above which household net wealth approximately follows a Generalized Pareto (GP) distribution

$\xi = \alpha_{GP}$ is a shape parameter and relates to Pareto type I α such that $\xi = \frac{1}{\alpha}$ (Jenkins, 2017)

σ is a scale parameter that determines the drift towards the end of the tail and defines a higher or lower wealth concentration

The two-parameter Pareto distribution is a nested special case of the GP distribution with $w_{min} = \frac{\sigma}{\xi}$ and therefore no drift

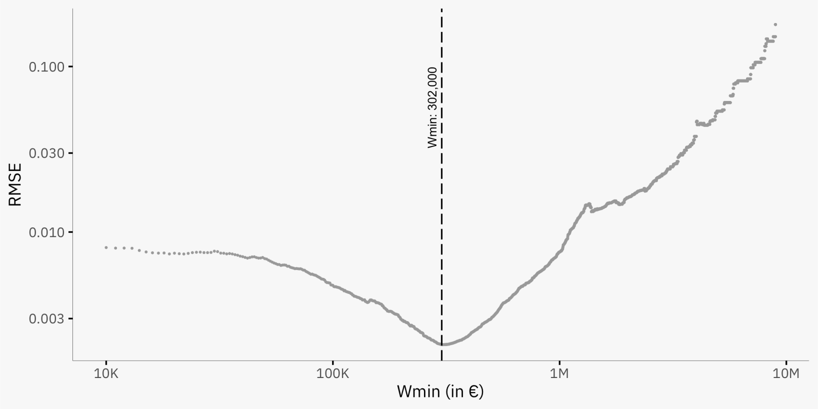

Figure 2: Germany: Choosing w_{min} and corresponding α

Case: Germany

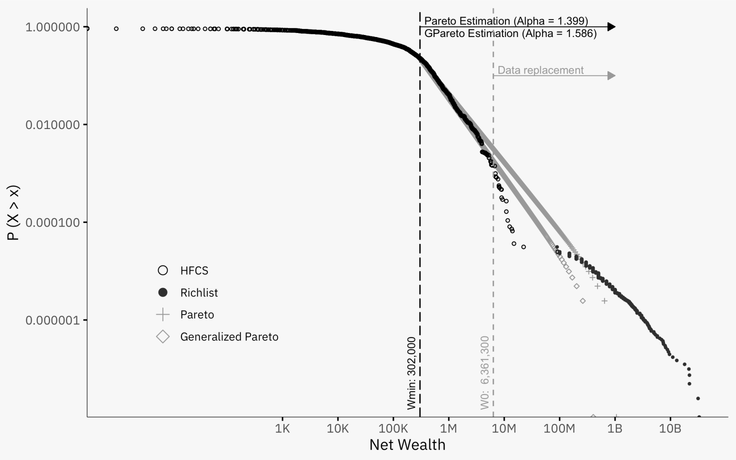

Figure 3: Germany: Tail adjustment based on Pareto and GP distribution

Figure 4: w_{min} and w_0. Mean estimate across five implicates

Results: Shape and scale parameters

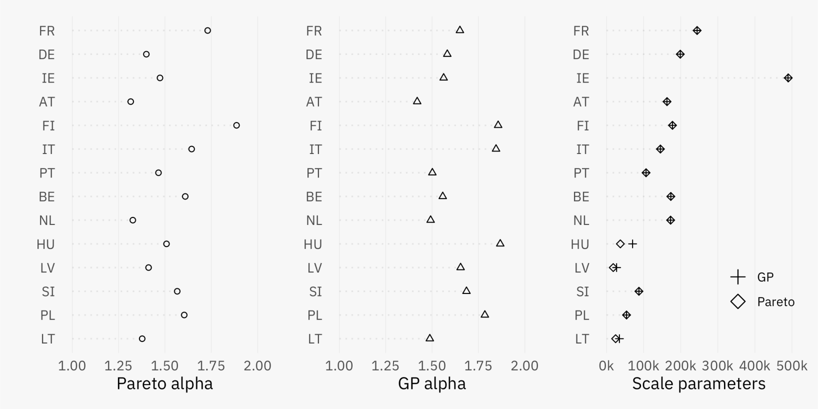

Figure 5: Shape and scale parameters. Mean estimate across five implicates

Example: Change in wealth share of Top 1%

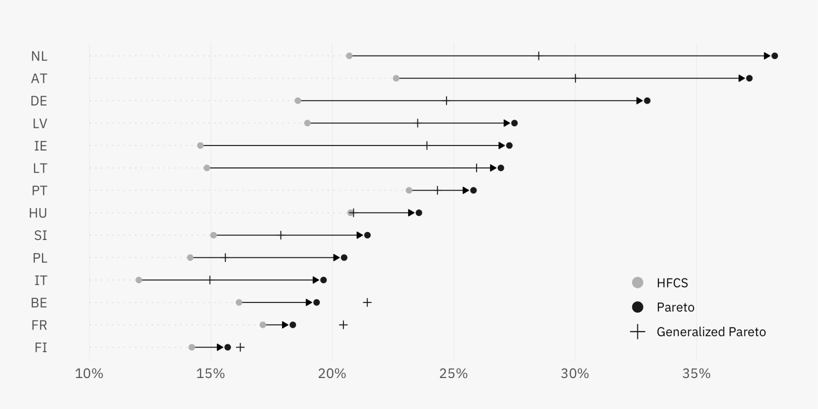

Figure 6: Mean estimate across five implicates

Conclusions

The variety of best-fit location parameters highlights the advantage of a unified and rule-based approach over arbitrary choices of w_{min} and w_0

The better the survey data is able to cover the top, the higher is the threshold for data replacement, w_0. We argue that a rule-based approach is of particular importance in a cross-country setting

Estimates of the shape parameter are consistent with the assertion in Gabaix (2016) that α is around 1.5

The quality of the data used used for oversampling is crucial

For countries where wealth-correlated information is limited, ex post adjustments by means of national rich lists significantly increase aggregate wealth, top shares, and other measures of inequality

The strength of the Generalized Pareto is to intervene when the survey deviates from the Pareto distribution

Our non-discretionary algorithmic approach proves to be suited to correct for differences in the methodological differences in the surveys

Appendix: w_0

$$\hat{w}_0 = w_0 : \hat{f}_{kern}(w_0) = \frac{1}{Nh} \sum_i n(w_i) K(\frac{w_0 - w_i}{h}), \tag{7}$$

$$\hat{f}_{kern}(w_0) - \underbrace{\alpha w_{min}^{\alpha} \frac{1}{N} \sum_{w_i > w_{min}} n(w_i)}_{\text{normalizing constant C}} \times w_0^{-(\alpha+1)} = 0, \tag{8}$$

Let $n(w_i)$ be the weight of some household i, and h be the bandwidth for the kernel estimation, which we choose using Sheather et al. (1991)'s procedure.

Note that the procedure includes a normalizing constant C, which adjusts the number of tail observations such that the sum of weights in the population before and after re-estimation remains the same (Eckerstorfer et al., 2016).

C shifts the theoretical PDF up or down, and is crucial for finding the intersection of theoretical and empirical densities.

Appendix: Tail length

Definition of tail length:

$$\sum_{w_i > w_0} n(w_i) = [\sum_{w_i \in (w_{min}, w_0)} n(w_i)] * \frac{1 - F(w_0)}{F(w_0)} \tag{9}$$

Assignment of wealth levels:

$$w_i = w_{min} \left(\frac{\sum_{w_i > w_{min}} n(w_i)}{\sum_{w_j > w_i} n(w_j)} \right)^{1/\alpha}. \tag{10}$$

w_0 and effectiveness of oversampling

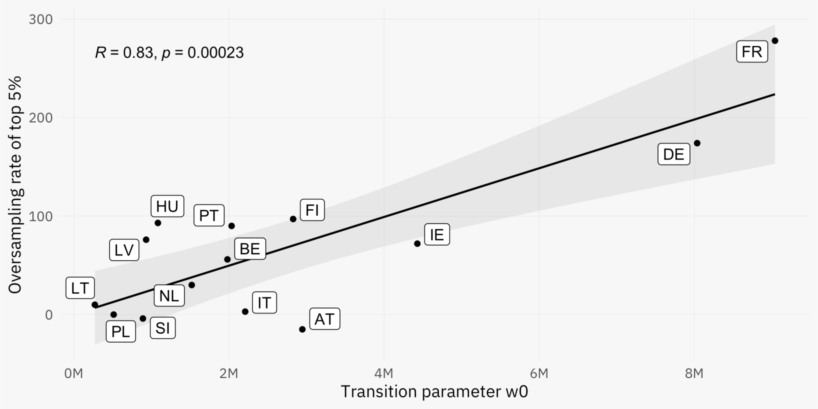

Figure 7: Correlation between w_0 and the effective oversampling rate

If the scaled excess of a random variable of some location parameter w_{min} follows a Generalized Pareto distribution, the scaled excesses for any threshold $u \geq w_{min}$ are also Generalized Pareto distributed with the same shape parameter $\frac{1}{\alpha_{GP}}$ (Langousis et al., 2016)

The scale parameters σ_u depends linearly on the scale parameter of the threshold w_{min}, the shape parameter, and the excess over u.

The scaled excess of a random variable over any threshold u is defined as $e(u) = E[W - u \mid W > u]$.

Equation 11 gives the linear relationship for σ_u, equation 12 the expected value of the excess over u.

$$\sigma_u = \sigma_\mu + \frac{1}{\alpha_{GP}}(u - w_{min}) \tag{11}$$

$$e(u) = E[W - u \mid W > u] = \frac{\sigma_u}{1 - \frac{1}{\alpha_{GP}}} = \frac{\sigma_\mu + \frac{1}{\alpha_{GP}}(u - w_{min})}{1 - \frac{1}{\alpha_{GP}}} = \beta_0 + \beta_1 u \tag{12}$$

Linearization

The linear relationship in equation for the expected value of the excess over u allows for a linear regression based estimation of both the scale and shape parameters, since

$\beta_1 = \frac{1}{\alpha_{GP}}/(1 - \frac{1}{\alpha_{GP}})$

and

$\beta_0 = (\sigma_u - \frac{1}{\alpha_{GP}} w_{min})/(1 - \frac{1}{\alpha_{GP}})$

Thus:

$\frac{1}{\alpha_{GP}} = \beta_1/(1 + \beta_1)$

$\sigma_{w_{min}} = \beta_0(1 - \frac{1}{\alpha_{GP}}) + \frac{1}{\alpha_{GP}} w_{min}.$

Weighting

We estimate the weighted mean excesses $e(w) = E[W - u \mid W > u]$ above different thresholds $u_i = W_{i,n}$ with $i = 1, 2, ..., n - 20$, to ensure that mean excess are calculated based on at least 20 observations

Effectively, we pair every observation w_i with a mean excess value $e(w_i) = E[W - w_i \mid W > w_i]$

For each observation w_i, $i = 1, 2, ..., n - 20$, we calculate the conditional weighted excess variance $Var[W - w_i \mid W > w_i]$ to account for the increasing estimation variance of $e(w_i)$ in w_i

The weights v_i are calculated as $v_i = (N - i)/(Var[W - w_i \mid W > w_i]$

Finally, we perform a weighted least squares and weighed median quantile regression estimation corresponding to equation 12 with weights v_i.

Aggregate Wealth

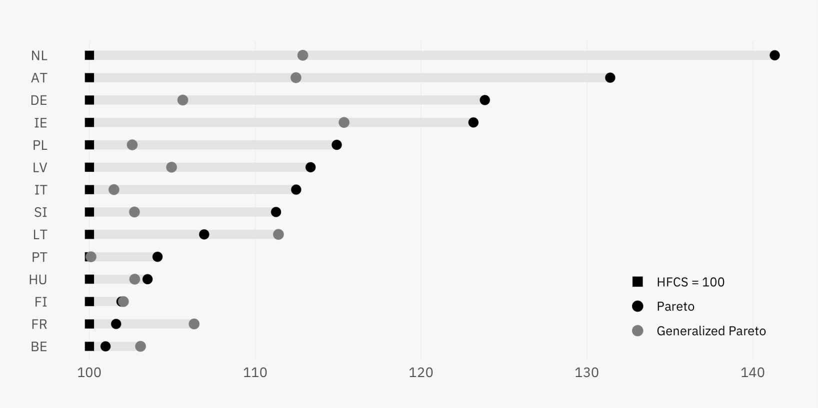

Figure 8: Aggregate Household Net Wealth

Comparison to Wealth Totals from National Accouns

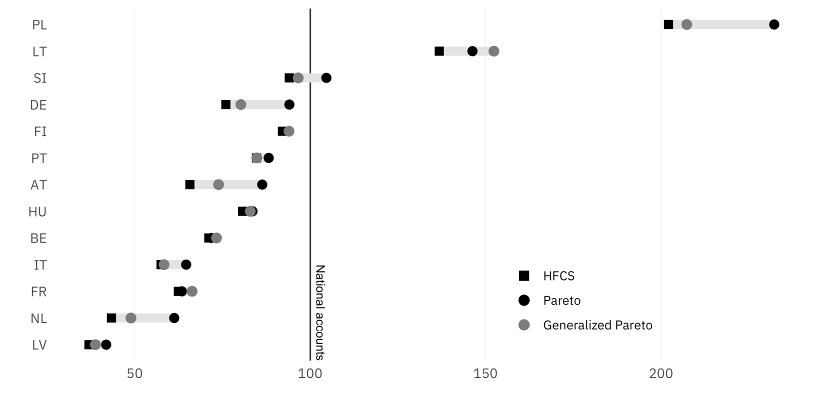

Figure 9: In % of NA Aggregates

References

Atkinson, Anthony B. (2017). “Pareto and the Upper Tail of the Income Distribution in the UK: 1799 to the Present”. In: *Economica* 84.334, pp. 129–156.

Benhabib, Jess and Alberto Bisin (2018). “Skewed Wealth Distributions: Theory and Empirics”. In: *Journal of Economic Literature* 56 (4), pp. 1261–1291.

Chakraborty, Robin and Sofie R. Waltl (2018). *Missing the wealthy in the HFCS: micro problems with macro implications*. Working Paper Series. European Central Bank.

D’Alessio, Giovanni and Ivan Faiella (2002). *Non-response behaviour in the Bank of Italy’s Survey of Household Income and Wealth*. Working Paper Series 462. Banca D’Italia.

- Dalitz, Christoph (2016). *Estimating Wealth Distribution: Top Tail and Inequality*. Tech. rep. 2016-01. Hochschule Niederrhein, Fachbereich Elektrotechnik & Informatik.
- Davies, J.B., Susanna Sandström, et al. (2010). “The level and distribution of global household wealth”. In: *The Economic Journal* 121 (3), pp. 223–254.
- Davies, J.B. and Anthony E. Shorrocks (2000). “The Distribution of Wealth”. In: *Handbook of Income Distribution Volume 1*. Ed. by Anthony B. Atkinson and François Bourguignon. North Holland, pp. 605–675.
- Eckerstorfer, Paul et al. (2016). “Correcting for the Missing Rich: An Application to Wealth Survey Data”. In: *Review of Income and Wealth* 62 (4), pp. 605–627.
- Gabaix, Xavier (2016). “Power Laws in Economics: An Introduction”. In: *Journal of Economic Perspectives* 30 (1), pp. 185–206.
- Gabaix, Xavier and Rustam Ibragimov (2011). “Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents”. In: *Journal of Business Economics and Statistics* 29 (1), pp. 24–39.

Jenkins, Stephen P. (2017). “Pareto Models, Top Incomes and Recent Trends in UK Income Inequality”. In: *Economica* 84 (334), pp. 261–289.

Jones, C.I. (2015). “Pareto and Piketty: The Macroeconomis of Top Income and Wealth Inequality”. In: *Journal of Economic Perspectives* 29 (1), pp. 29–46.

Kenickell, A.B. (2019). “The tail that wags: differences in effective right tail coverage and estimates of wealth inequality”. In: *The Journal of Economic Inequality* 17 (4), pp. 443–459.

Koenker, Roger and Gilbert Bassett (1978). “Regression Quantiles”. In: *Econometrica* 46 (1), pp. 33–50.

Langousis, Andreas et al. (2016). “Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database”. In: *Water Resources Research* 52.4, pp. 2659–2681.

Pareto, Vilfredo (1965 [1896]). “La Courbe de la Repartition de la Rihhesse”. In: *Écrits sur la courbe de la répartition de la richesse*. Ed. by Giovani Busino. Vol. 3. Librairie Droz, pp. 1–15.

Sheather, Simon J. and Chris M. Jones (1991). "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 53 (3), pp. 683–690.

Vermeulen, Philip (2016). "Estimating the Top Tail of the Wealth Distribution". In: *American Economic Review* 106 (5), pp. 646–650.

– (2018). "How Fat Is the Top Tail of the Wealth Distribution?" In: *Review of Income and Wealth* 64 (2), pp. 357–387.