A Joint Top Income and Wealth Distribution

ASSA 2022, Jan 7th-9th

Viktor Steiner, Freie Universität Berlin
Junyi Zhu, Deutsche Bundesbank

Top tail and extrapolation

Under-coverage and measurement error

Marginal extrapolation: assuming univariate power-law, esp. Pareto distributions (Vermeulen 2014, Eckerstorfer et al., 2015 and Jenkins, 2017)

Little analysis of the joint top income and wealth distribution – a remote starter: Aaberge, Atkinson and Königs (2018)

What we do

Incorporating the top joint distribution:

Postulations:

- Parallel rank association assumption – copula structure is invariant in the data with or without top tails
- Incidental truncation model for the top – the other dimension large enough to be captured

Contributions:

- Potentially less reliance on specific marginal distributions or the estimation domain
- External consistency – benchmark the extrapolation with administrative income tax data (eg top 1000 percentiles within top 1%) and the list of the 500 wealthiest people in Germany (”rich list”)
- Developing a principled approach to estimate top copula

Influences

Essentially a generalization problem – predictive modelling:

- Current practice to fit in a combination of survey and top lists -> overfitting (Mosteller and Tukey, 1977)
- Holdout data or cross-validation (Stone, 1974; Geisser, 1975; Breiman, 2001)

Many „wishful thinking“ but less objective, little testing on reality and predictive validation (Clauset, Shalizi and Newman, 2009)

Feedback from the predictive modelling to explanatory modelling (Shmueli, 2010)

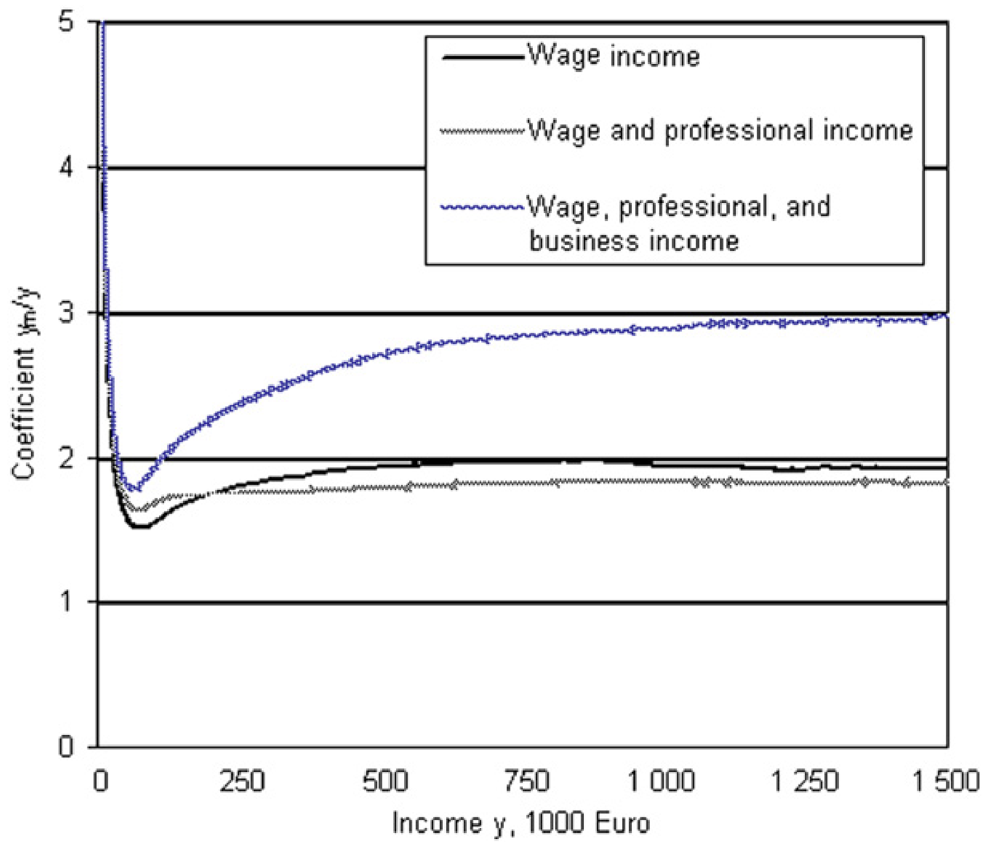
Evolution of the marginal approach - Pareto I

Pareto type I distribution (two parameters):

Strong dependence of the estimated Pareto parameter on the threshold value chosen

- too low and unrealistic
- contradicting the observation from top administrative (tax) data <- constant inverted Pareto coefficient $E(y|y > y_0)/y_0$ for Pareto I

Constant inverted Pareto coefficient?
- Bach, Corneo and Steiner (2012; Fig 2)



Evolution of the marginal approach - Pareto II

(Generalized) Pareto II distribution (three parameters; Lomax distribution):

- *More flexible*: non-constant inverted Pareto coefficient (Blanchet, Fournier and Piketty, 2017)
- *Better fitting than Pareto I* (Jenkins, 2017)
- *Lower threshold than Pareto I* (Jenkins, 2017)
- Denying Pareto I from the real data:

Estimates are *sensitive* to the threshold and optimal threshold has more *variability* over the years (UK tax data; Jenkins, 2017)

Pareto coefficient in type I model might *not be constant* over the top distribution (UK historical data, Atkinson, 2017)

Inverted Pareto coefficient *converges from below* when the percentile rank of income distribution is near one (US and French data; Blanchet et al., 2017; DE tax data, Bach et al., 2012)

Data

Fitting sample: Panel on Household Finance (PHF) – pooling w1 and w2 to expand the size (household)

External validation:

- Top wealth tail – rich list from Manager Magazin (2014)

- Top income tail – (1000 percentiles) top 1% gross income from administrative tax data (2010) available from the Research Data Center of the Federal Statistical Office of Germany (tax filing unit)

Unit: for the top household and tax filing unit should be almost equivalent

PHF coverage of labor income and capital income (wrt national account): 98% and 34% (Zhu, 2014)

German gross income for the top 30% by percentile mean: PHF (2009) and Income tax statistics (Est, 2008)

Percentile	PHF 09	Est 08	Relative difference (%: (PHF - Est)/Est*100)
71	40,769	46,517	-12
72	41,742	46,918	-11
73	42,591	48,414	-12
74	43,585	49,539	-12
75	44,728	49,563	-10
76	45,677	51,544	-11
77	46,481	51,697	-10
78	47,901	53,006	-10
79	49,634	54,557	-9
80	50,956	55,681	-8
81	52,409	56,771	-8
82	53,805	59,125	-9
83	55,064	59,680	-8
84	56,650	61,686	-8
85	58,259	63,085	-8
86	60,414	64,515	-6
87	62,663	67,368	-7
88	65,488	69,412	-6
89	68,027	71,715	-5
90	71,711	73,340	-2
91	74,938	76,061	-1
92	78,888	78,858	0
93	83,021	82,120	1
94	88,231	88,214	0
95	94,183	93,449	1
96	104,036	100,041	4
97	114,847	111,194	3
98	132,376	127,176	4
99	158,500	162,689	-3
100	277,084	531,812	-48

Note: Est 08 - income tax return data in 2008 from Income tax statistics (Einkommensteuerstatistik; source: Research Data Centre of the Federal Statistical Office of Germany); PHF 09 - wave one of the Panel on Household Survey (Bundesbank). Both are measured as the gross income for the tax unit population.

Copula

The copula is the function that binds together two marginal distributions and is defined by $C(u,v) = H\left(F^{-1}(u), G^{-1}(v)\right)$, $u \times v \in [0,1]^2$, where $F^{-1}(u)$ and $G^{-1}(v)$ are the inverse cumulative distribution function, and $H(\cdot)$ is the joint cumulative distribution on the support of these two marginal distribution

Sklar's Theorem states the existence and uniqueness of the above mapping function for continuous distributions -> advantage of copula:

- Transforming the joint distribution to a "disciplined" uniform support $[0,1]^2$
- Estimation of marginal and joint structure (copula) can be separate
- Marginal distributions can come from different family
- Insensitive to the tail properties of the marginal distributions (only rank-rank association matters)

Postulations

- Parallel rank association assumption – the copula structure is probably quite stable regarding the ranks with respect to the richest individual (the king) in the society (or subsociety)

  "Topology (copula density on the rank-rank support) between you and Lhasa (Garmisch - Partenkirchen) when you are standing at Shanghai (Frankfurt) is invariant with that between you and Mt Everest (Zugspitze) when you are standing at Lhasa (Garmisch - Partenkirchen)"

- Incidental truncation model for the observed top
  - *Top only responds / observable if income or wealth is high enough (missing eg prudent rich - Washington Post, 1982)*
  - *(Differential) detectability – investigative journalism and tax auditing*
  - *Not truncated for the top in survey*

The relationship between copula-based and marginal extrapolations

Copula-based probability density function of wealth W, conditional on the income Y all above a lower bound:

$$P(w|Y > y_b) = P(w)\frac{1 - \frac{\partial}{\partial v}C(u, v)}{1 - u}$$

$$u = P(Y \leq y_b),\ v = P(W \leq w) \text{ and } C(u, v) \text{ is the copula density}$$

Difference btw two approaches: $\frac{1-\frac{\partial}{\partial v}C(u,v)}{1-u}$, which is just $\frac{P(U>u|V=v)}{P(U>u)}$

Stochastic increasing in the conditional distribution (differential detectability) - the richer in wealth the more likely the income is above some threshold (gap btw two approaches is smaller)

Discovering the copula

Clauset, Shalizi and Newman (2009) – typical bias-variance tradeoff:

- expanding the sample from the top - reducing the impact of statistical fluctuation in the estimation
- Too much expansion and include some lower part of the sample with a different joint structure – estimation deviates
- An optimal cutoff

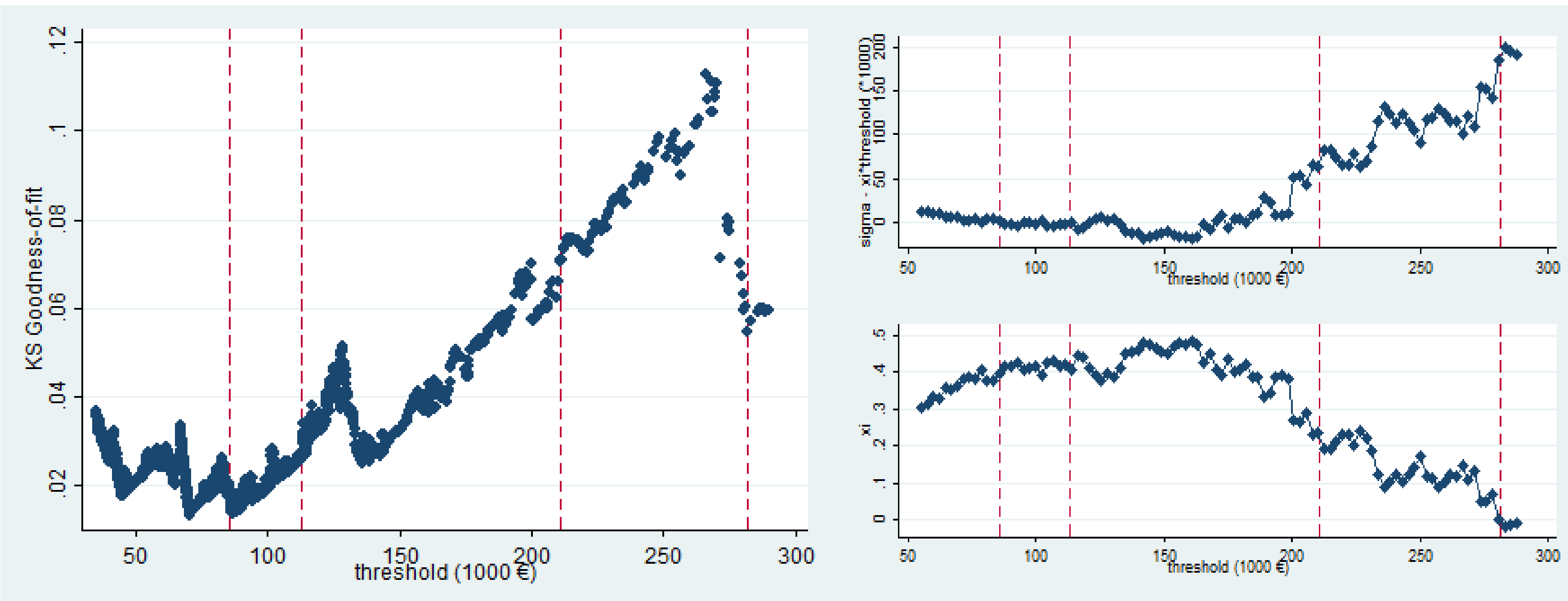
Estimation – Pareto II (Jenkins, 2017)

Clauset, Shalizi and Newman (2009) for the cutoff

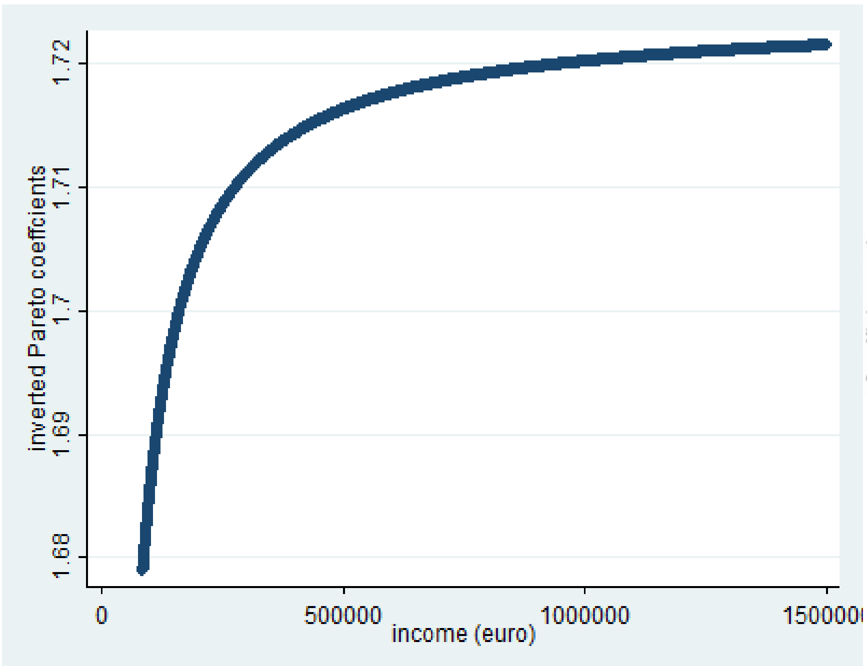
Plot the estimated parameters against thresholds and pick the one above which the estimate starts to be flat for some area

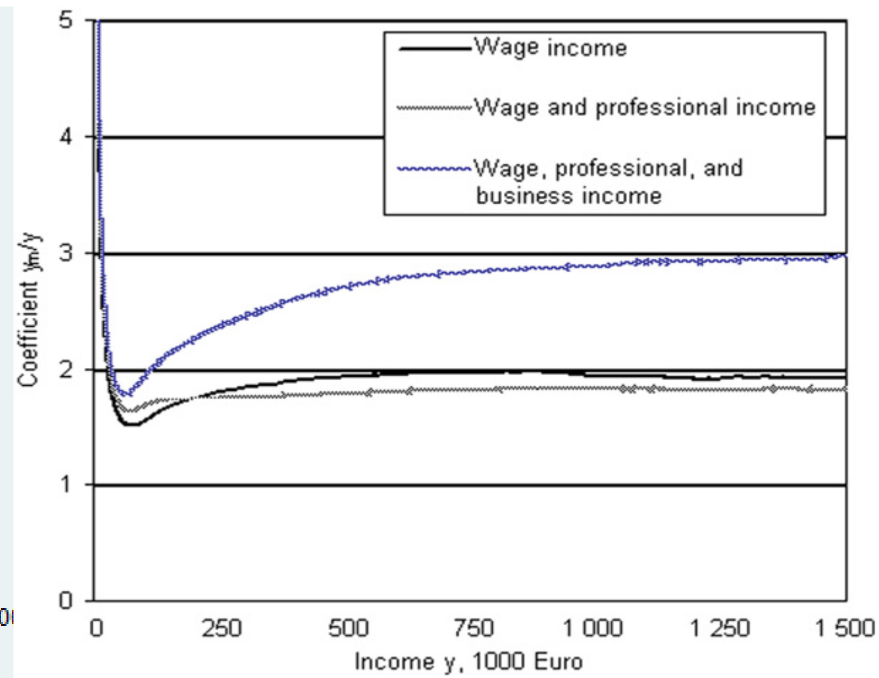
Kolmogorov-Smirnov (KS; left) and estimated parameters (right) for income threshold (dashed lines: p90, p95, p99, p99.5)

Our favoured threshold is 86,957



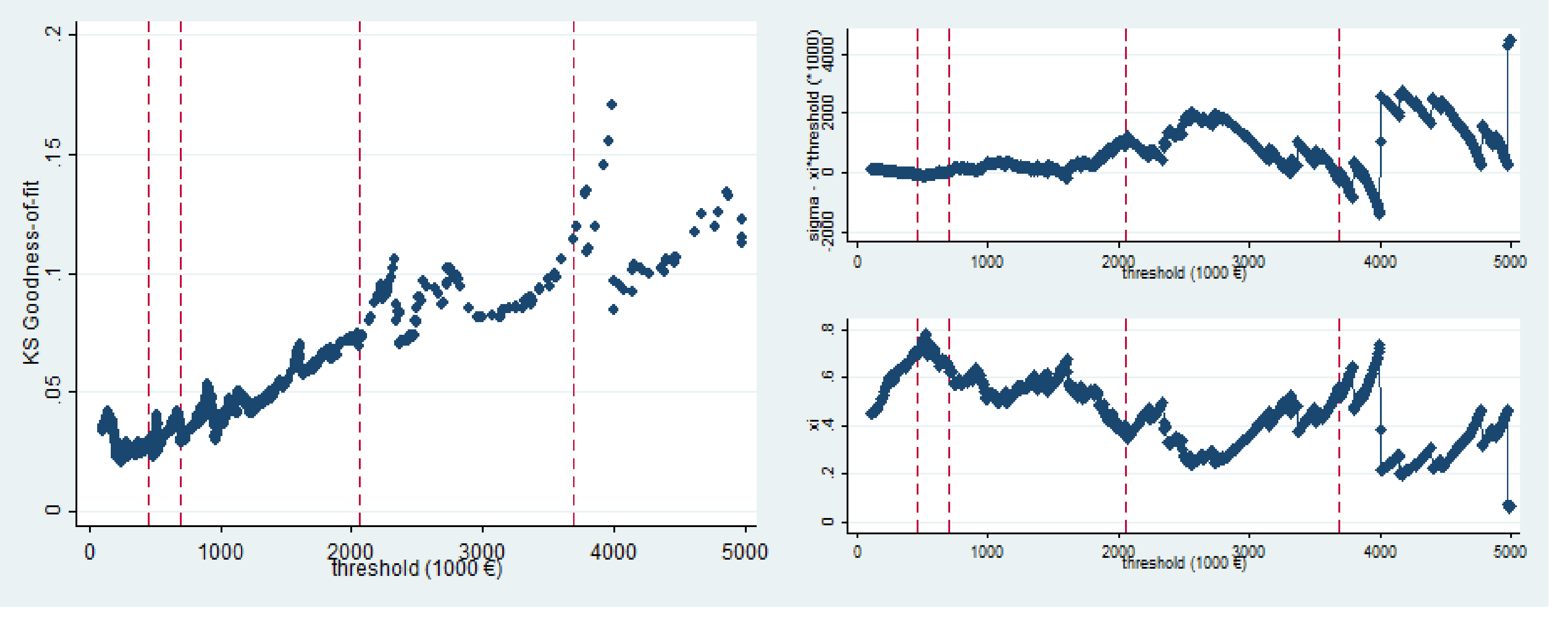
Inverted Pareto coefficient w.r.t. income (left: our favourite estimate; right: empirical one from tax data - Bach, Corneo and Steiner (2012; Fig 2))



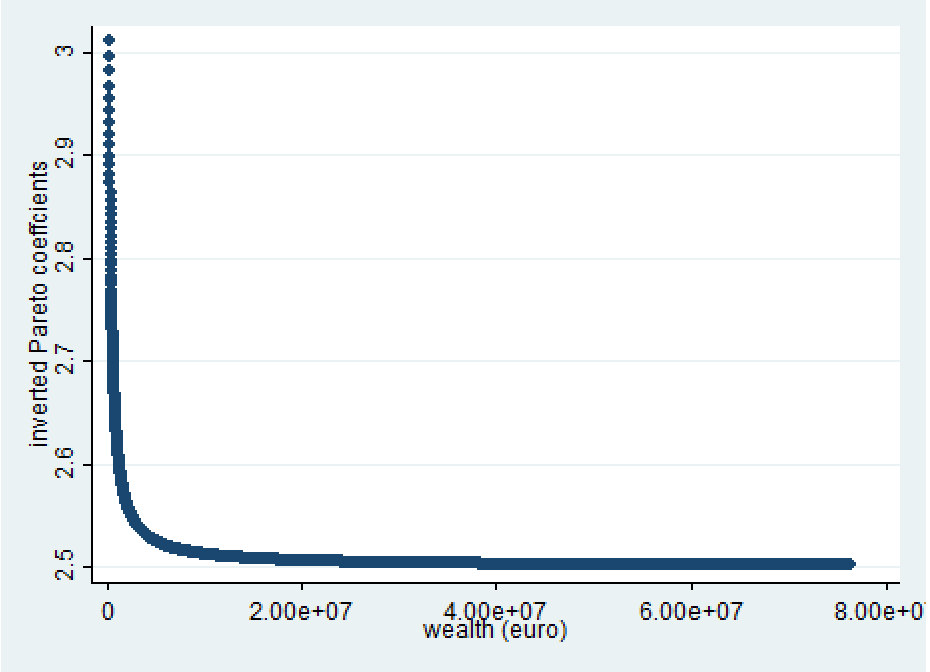


Kolmogorov-Smirnov (KS; left) and estimated parameters (right) for wealth threshold (dashed lines: p90, p95, p99, p99.5)

Our favoured threshold is 245,160



Inverted Pareto coefficient w.r.t. wealth from our favourite estimate



The tail is getting thinner as wealth becomes higher

One evidence: Pareto indices (alpha in Pareto I distribution) are 1.52 for the Manager Magzin rich list and 1.47 for PHF (Dalitz, 2016)

Estimation – copula

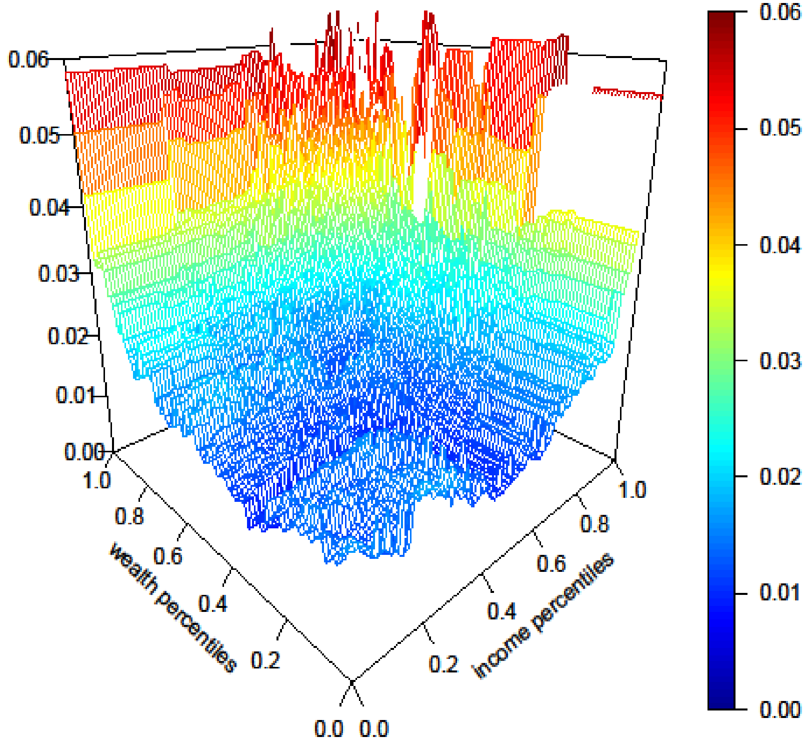
Grid points on the rectangle to search the optimal cutoff (min bivariate KS)

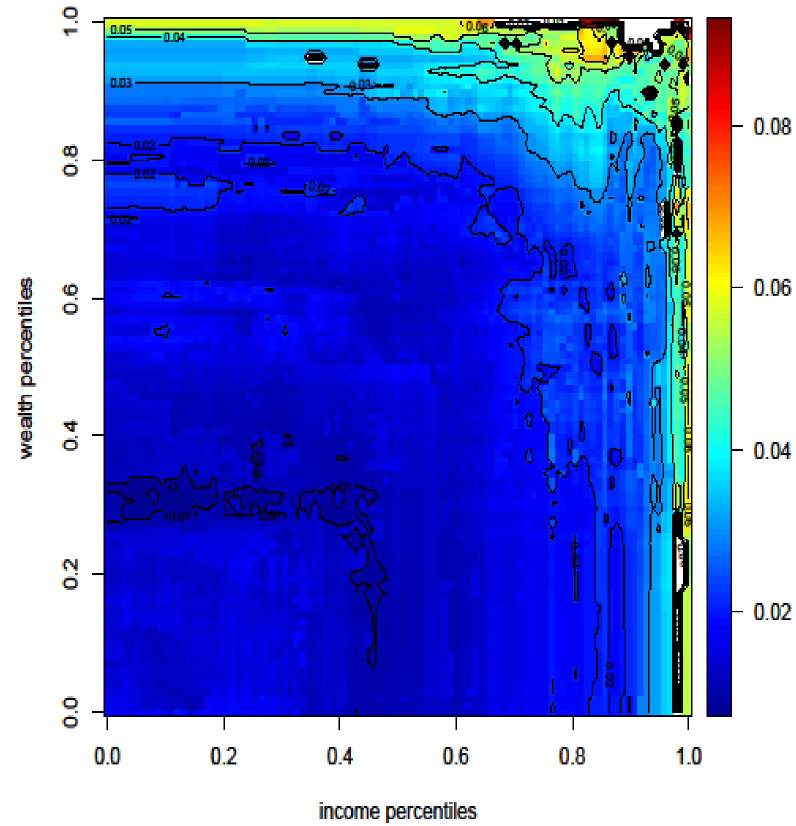
BiCopSelect() from the VineCopula in R (Brechmann and Schepsmeier, 2013 and Yan, 2007): MLE across a rich set of parametrical families and selected using Akaike Information Criteria

Goodness-of-fit: we compare the contour curves of both the empirical copula (from the fitting sample) and the estimated one

Best fitting copula

Bivariate Kolmogorov-Smirnov (KS) statistics of copula estimates at grids on the unit square: perspective and contour plots

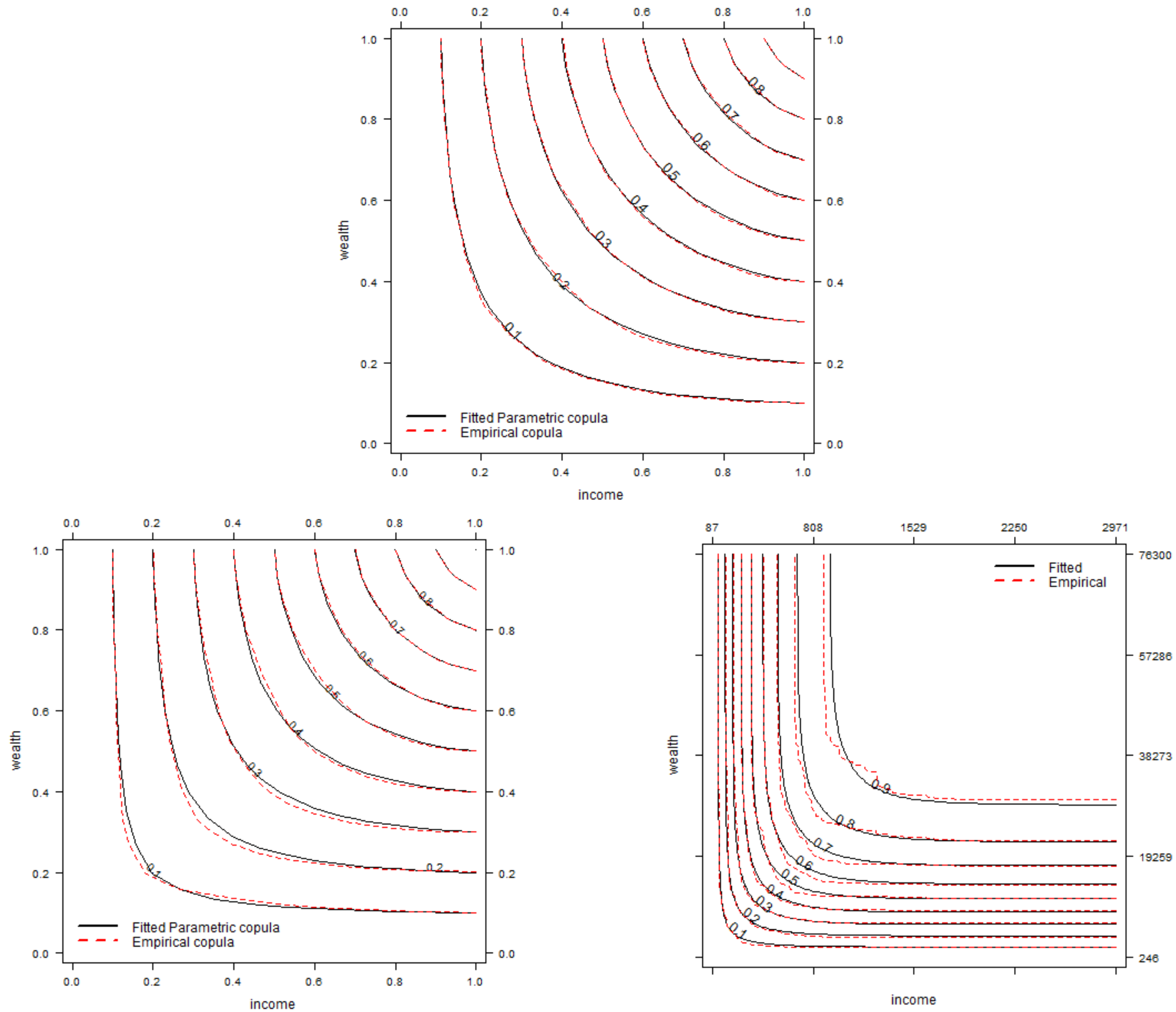




Competing copula and marginal distributions

	Case	Best fitting	Full sample	High income - high wealth
Fitting sample for copula	income cutoff in forming fitting sample	5,300		86,957
	wealth cutoff in forming fitting sample	29,500		245,160
Copula estimate	copula family	BB8	t	t
	1st parameter	1.62	0.589	0.275
	2nd parameter	0.97	6.2	30
Marginal distribution (Pareto II)	threshold -income	86,957	86,957	87,000
	shape parameter - income	0.41	0.41	0.34
	scale parameter - income	34612	34612	45208
	threshold - wealth	245,160	245,160	245,680
	shape parameter - wealth	0.60	0.60	0.63
	scale parameter - wealth	197203	197203	370165

Goodness-of-fit (top best fitting, bottom left: full sample and bottom right: high income – high wealth)



External validation and comparison

P-P plot

Two parametric settings and each containing both marginal and copula-based approaches in one plot

Roughly optimal lower bound

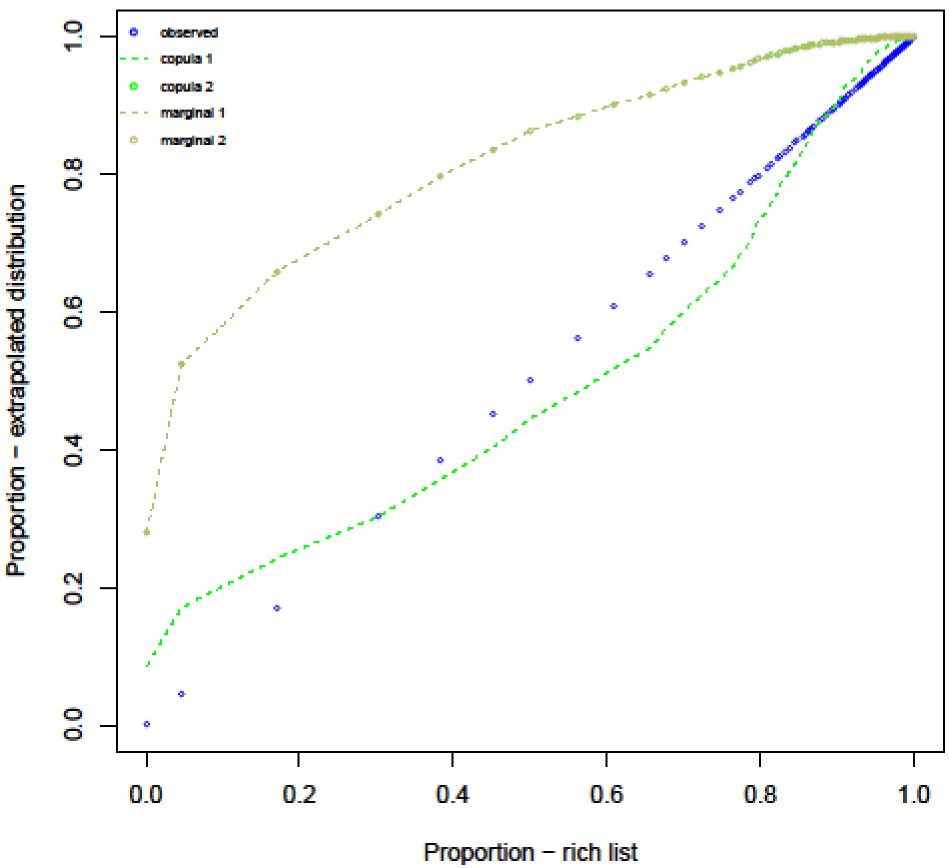


Fig 19. **top rich list**
1 – full sample case and 2 – best fitting case

Copula 1 – t with **conditioning income lower bound *20,000,000***
Copula 2 – BB8 with conditioning income lower bound *20,000,000*

20 million is almost twice the p999 within the top 1% tax data – 11.31 million. Bach et al. (2013), in 2005, the top 0.001% tax payers (about 450 persons) at least receive 11 mill and the average gross income in this top range is about 36 mill.

Most approximating settings

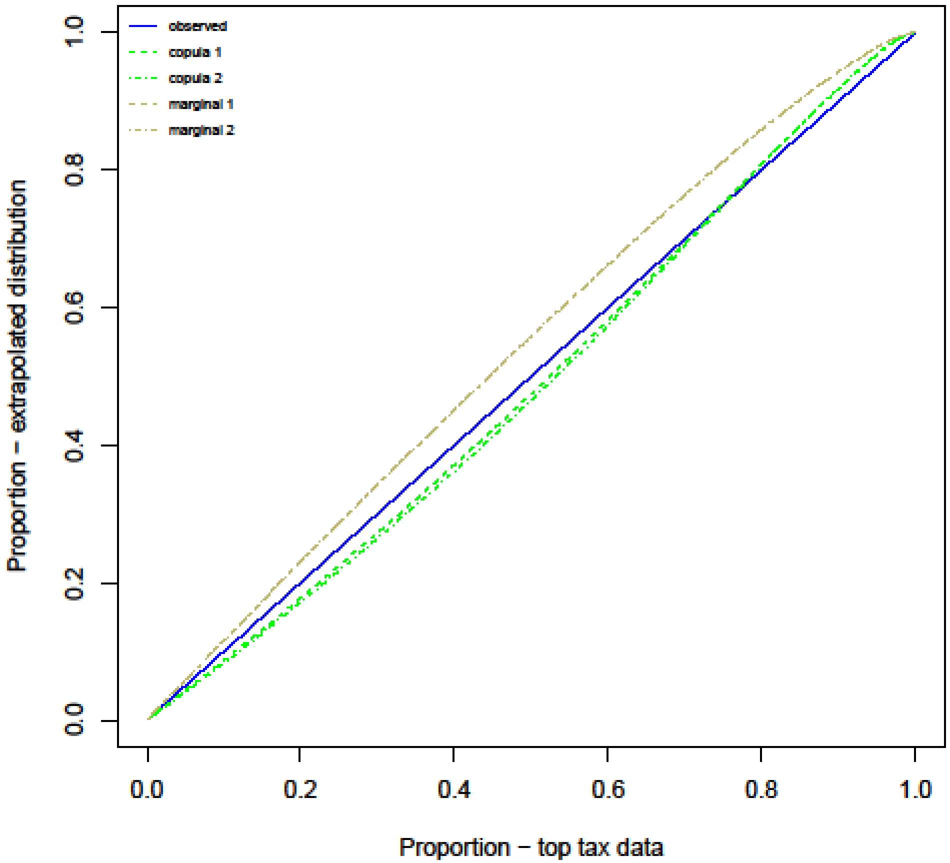


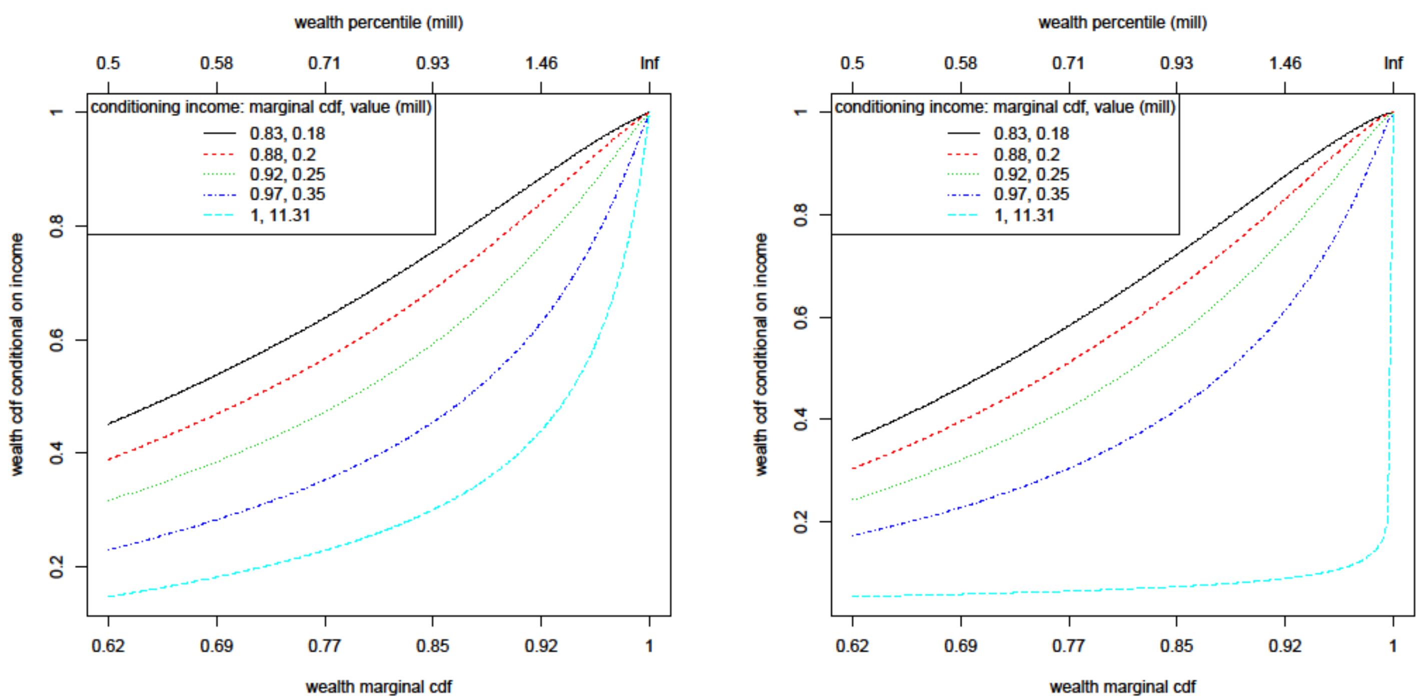
Fig 20. **top income tax data –** 1 – full sample case and 2 – best fitting case

Copula 1 – t with **conditioning wealth lower bound *700,000***
Copula 2 – BB8 with conditioning wealth lower bound ***700,000***

It is quite plausible that the top 1% income earners have at least 0.7 million wealth!

Differential detectability – top income tax list

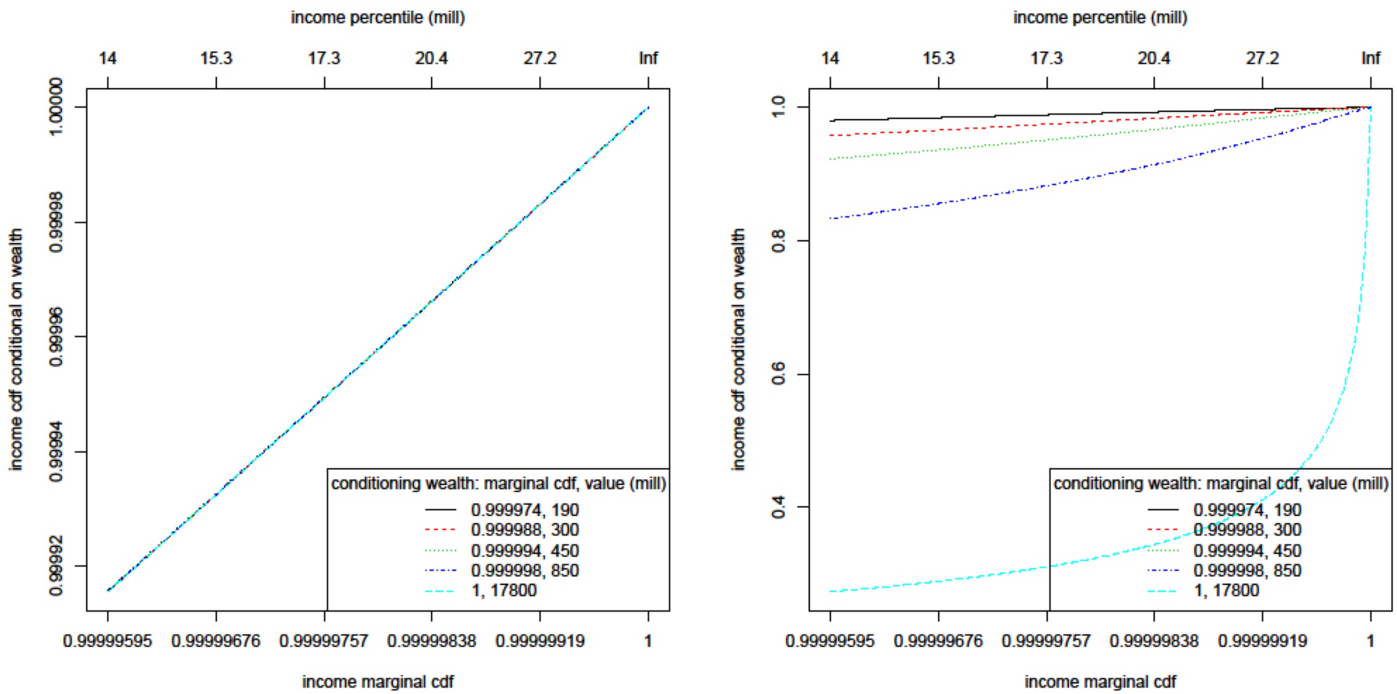
Link our data and estimates with incidental truncation – stochastic monotonicity from the conditional distribution



copula-based – best fitting case (left) and full sample case (right) - wealth distribution conditional on income in the range of top 1% income distribution and the neighbourhood around the roughly optimal wealth for the extrapolation (0.7 mill)

Differential detectability – rich list

Some evidence for Aaberge et al (2018)

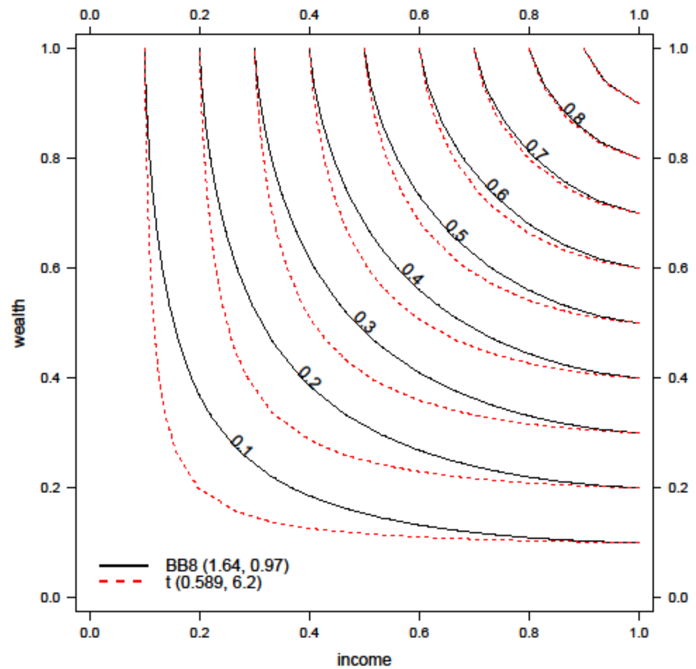


copula-based – best fitting case (left) and full sample case (right) - income distribution conditional on wealth in the range of top wealth rich list and the neighbourhood around the roughly optimal wealth for the extrapolation (20 mill)

Reject the best fitting case (copula)?

Probably no:

- credibility of the rich list as a representative sample of the top wealth distribution
- small statistical fluctuation – top 30% of full sample (t) and best fitting (BB8) copula close
- stochastic monotonicity for best fitting (BB8) copula in the income tax data can emerge in a more representative/not –very- top wealth list (SOEP-TS from DIW)



Contours for BB8 (1.62, 0.97) and t (0.589, 6.2)

Conclusion

- This paper proposes a copula-based joint extrapolation for the top income and wealth
- Benchmarking with the real external top tails, the copula-based extrapolation is never worse off than the marginal one
- The adoption of joint association has a positive net contribution beyond the marginal approach (high income – high wealth case)
- The approximation towards the true top distribution can be almost exact (eg. if credibly assuming the minimum wealth for the top 1% income is around 0.7 million)
- Our exercises justify to acquire the least information from the other dimension which can be cheap to implement
- The extensions:

Wave 3 PHF and rich list from the SOEP-TS

Estimate lower bound with bootstrapping hypothesis test and cross-validation/LR test - Multivariate extension of Clauset et. al. (2009)

Non-parametric / empirical copula

Test over historical and/or cross-country data (some with one margin completely observed)