Freie Universität Berlin

A Joint Top Income and Wealth Distribution

Viktor Steiner
Junyi Zhu

School of Business & Economics
Discussion Paper

Economics

2021/3

A Joint Top Income and Wealth Distribution[1]

Viktor Steiner
Freie Universität Berlin

Junyi Zhu
Deutsche Bundesbank

January, 2021

Abstract

Top distributions of income and wealth are still incompletely measured in many national statistics, particularly when using survey data. This paper develops the technique of incorporating the joint distributional relationship to enhance the estimation of these two top distributions by using the best data available for Germany. We leverage the bivariate copula to extrapolate both income and wealth distributions from German PHF (Panel on Household Finance) data under the incidental truncation model. The copula modelling grants the separability in choosing the estimation domain as well as the parametric specification between the marginal distribution and dependence structure. One distinct feature of our paper is to complement the model fit with external validation. The copula estimate can help us to perform out-of-sample prediction on the very top of the tail distribution from one margin conditional on the characteristics of the other. The validation exercises show that our copula-based approach can approximate much closer to the top tax data and wealth "rich list" than those unconditional marginal extrapolations. The data and effectiveness of our copula-based approach also verify our presumption of incidental truncation and differential detectability in the top lists.

Keywords: income and wealth joint distribution, copula, heavy-tailed distributions, external consistency, incidental truncation

JEL codes: C52, C46, C81, D31

1 We are grateful to Irakli Sauer for his research assistance. Results and opinions expressed in this paper are those of the authors and do not necessarily reflect views of the Deutsche Bundesbank. Email: viktor.steiner@fu-berlin.de and junyi.zhu@bundesbank.de.

1. Introduction

The analysis of top income and wealth distributions on the basis of survey data is problematic due to severe under-coverage and measurement error. Respondents either under-report the figures or there is top-coding. People at the very top of the income and wealth distributions usually do not participate in such surveys and the survey administrators often do not have reliable information on the true statistical characteristics of non-respondents. Hence, most income and wealth surveys only sparsely cover the top given the nature of long tails. Researchers tackle this issue by extrapolation via conjecturing the wide observance of the univariate power-law distributions, in particular from the Pareto family (see, eg Vermeulen, 2014, Eckerstorfer, Halak, Kapeller, Schütz, Springholz and Wildauer, 2015 and Jenkins, 2017).

While most of the empirical applications refer to the estimation of either the top income or wealth distribution, the analysis of the joint top income and wealth distributions has not been the focus of much research so far.[2] In this paper, we contribute to this evolving literature by introducing the information from the top joint distribution between income and wealth. By postulating that the under-covered top bivariate distribution follows the same structure (copula) as those observed in the data as well as an incidental truncation model for the sample selection of the top lists, we extrapolate not only the marginal distributions but also the joint one. Rather accurate distributional information recovered by the income data from Panel on Household Finance (PHF) survey when benchmarking with tax administrative data also motivates the incorporation of joint distribution. Besides the novelty in leveraging the observed joint distribution, we propose a validation step by comparing the out-of-sample prediction accuracy of our copula-based (conditional) extrapolation to those from marginal (unconditional) Pareto ones. For this purpose, our benchmarks are the administrative income tax data (eg top 1,000 percentiles within top 1% - between p0 and p999) and the list of the 500 wealthiest people in Germany ("rich list") from the *Manager Magazin* (manager magazin, 2014).[3] Note the data accessed by this paper for both estimation and validation purposes have the highest quality for the distributional analysis in Germany. Our approach surpasses the existent marginal approach according to the distribution fit measure of predictive accuracy.

Out-of-sample validation and copula modelling are the highlights of this paper. The problem confronted by this literature has a pragmatic nature: fitting a distribution on the observed world which is supposed to generalize towards the missing tail; and this unobserved part is an empirical construct: ie the distribution of the richest population in terms of either wealth or income in a nation. Even after we obtain a rich list for the very top, we have to still impute the gap between the top of our survey wealth and the rich list. Treating the very top benchmark distributions as the hold-out validation data becomes a natural choice to judge the predictive accuracy.[4]

Copula modelling offers more flexibility by separating the estimation of dependence structure and the marginal distribution. Particularly, our incidental truncation benefits from the separability by ignoring the marginal distribution in analysing the

[2] Aaberge, Atkinson and Königs (2018) is moderately related as they inspect the top joint distribution between labor and capital incomes.

[3] https://www.manager-magazin.de/unternehmen/personalien/reichste-deutsche-quandt-und-klatten-fuehren-rangliste-an-a-1114555.html

[4] The problem confronted by us shares the nature of predictive modelling (Shmueli, 2010). Making casual inference and/or capturing the association is of the second order importance for this literature. Performing the out-of-sample prediction is the primary concern. It is then more proper than usual to demand a validation test. Accessing the predictive power from the data which the model was fitted often leads to the overfitting (Mosteller and Tukey, 1977). Accordingly, it is more standard to evaluate the prediction using a holdout dataset or cross-validation (Stone, 1974; Geisser, 1975; Breiman, 2001).

truncation effect. On the other hand, we can resort less to numerical integration or simulation common in this literature which alternatively adopts parametric assumptions of joint distribution on the truncation structure. Many of these assumptions can be rather unrealistic in dealing with our non-normal tail distribution. Thanks to the separability, our modelling can also be more data driven in the nature of extrapolation: more degree of freedom in choosing the different fitting sample and parametric / non-parametric structure independently adapted for the marginal distribution and dependence modelling.[5]

In general, we consider that the top wealth or income list to be extrapolated and benchmarked with is just an observed subsample of the complete underlying top distribution. Such a selection is described by an incidental truncation: only the wealth holders with income higher than some specific threshold and the income earners with wealth above a lower bound become conspicuous enough to be captured into either top list.[6] The top of the survey data is not high enough to be subject to such truncation which allows us to infer the (underlying marginal) tail distribution. Moreover, we assume that the copula $C(u, v)$ for the top distribution can be approximated by those estimated from some observed sample in the survey data (ie a parallel rank association assumption – copula structure is stable in the data with or without top tails). Our data provides the evidence to justify the assumptions of incidental truncation and parallel rank association.

In section 2 we present our data as well as external information on top tails. We outline the methodology in section 3 concerning Pareto distributions, copula and difference between marginal and copula-based extrapolations. Section 4 outlines the estimation of both top marginal and joint distributions from the survey data. Section 5 presents the external validation. A discussion of further work is given in section 6 which also concludes the paper.

2. Data

Our estimation comes from the Panel on Household Finance (PHF), the German component of the Household and Consumption Survey (HFCS) conducted by Deutsche Bundesbank. We pool income and wealth data for the two panel waves of 2009 and 2013 in order to increase sample size and the precision of estimates for the top of the respective distributions. Since PHF is multiply imputed, we average over five implicates to form the data used for the results presented. The top wealth tail distribution is drawn from the rich list collected by the *Manager Magazin* (manager magazin, 2014). We pooled each list of 500 richest Germans in 2010 and 2013. Dalitz (2016), Vermeulen (2018) and Bach, Thiemann and Zucco (2015) have described both data sets in detail. The top income distribution is retrieved from the administrative tax data for 2010 available from the Research Data Center of the Federal Statistical Office of Germany.[7] This data allows building the distribution of gross income for all the tax units (spouse or single) who file the tax return and someone who do not file the tax return. We construct the same concept of gross income for the tax units from the PHF.[8]

---

[5] See Trivedi and Zimmer (2007) for the detailed discussion on applying copula to the incidental truncation model.

[6] Given the construction of rich list and income tax data, it appears to be reasonable to treat our top lists as a select sample from the whole population (eg due to obstacle and incomplete information of reaching all the richest population by a voluntary (journalism) effort, tax evasion and institutional barrier to cover all the top income earners in the tax data).

[7] https://www.forschungsdatenzentrum.de/de/steuern/lest.

[8] The gross income concept used in PHF is the total amount of income (*Gesamtbetrag der Einkünfte*) according to the German income tax law (*EstG*). It consists of seven income categories: agriculture and forestry, business, self-employment, employment, capital income, renting and leasing, as well as other, including also tax-relevant capital gains less some expenses, allowances and losses. Our administrative tax data uses the '*adjusted gross income*' ('*Summe der Einkünfte*') which is simply the total amount of income deducted of the old-age lump-sum allowance and exemptions for single parents. The difference between these two concepts is rather little for the top distribution. Power-law distribution estimated using the

Table 1 demonstrates the comparison of the top 30% distributions of gross income for tax units from PHF (wave 1 - 2009) and income tax return data integrated with GSEOP (ITR-SOEP) 2008. Besides showing the percentile means, the last column provides the relative difference for each percentile. The PHF distribution can match with the ITR-SOEP one quite well, particularly in the top 10% distribution with the relative differences between two percentile means all below 5% except the top 1%.

3. Methodology

We discuss the choice of distribution specification for the power-law distribution and the search for its cutoff. The copula concept and relevant property is briefly introduced. We then rationalize the incidental truncation assumption and elaborate the role of copula modelling in connecting data to this assumption.

3.1 Pareto distribution

Pareto I distribution has been adopted widely as the power-law extrapolation. However, the inverted Pareto coefficient observed in the data has posed the challenge.[9] For type I Pareto model, inverted Pareto coefficient has a one-one mapping to the Pareto coefficient as the distribution parameter. Thus, its inverted Pareto curve is flat. Bach, Corneo and Steiner (2012) plots using German tax data (2005) the empirical inverted Pareto coefficient (see Figure 1). The flattening in the curves according to different income concepts is not obviously visible.[10]

This problem may be less severe in studies that use the more flexible (Generalized) Pareto type II distributions that allows non-constant inverted Pareto coefficient (see Blanchet et al., 2017). Jenkins (2017) discusses the comparison between type I and II Pareto models and how to determine the optimal threshold in estimating Pareto parameters.[11] As claimed by him using UK income data, (Generalized) Pareto type II distribution outperforms the Pareto I distribution in terms of goodness of fit.[12] He also shows the evidence that the choice of optimal threshold for estimating the Pareto I model is not clear: estimates are sensitive to the choice of threshold and optimal threshold in the type I model has more variability over the years. The optimal threshold estimated for type II model can be much lower than that for type I model. This feature is

'*adjusted gross income*' from PHF is almost same as the one appearing in our paper. We adopt the second concept for the tax data for the comparison with the validation results due to legacy reasons (ie. Zhu, 2016 which shared the same concept since the benchmarking statistics from the administration used such).
Since 2009, due to the introduction of flat rate capital income tax, capital income is not completely observable from the tax data. The coverage of capital income is also rather poor in PHF. This issue should be most serious in the top. Hence, our gross income concepts from either source are assumed to be close in the degree of covering the capital income.
The derivation of the total amount of income in PHF is facilitated via a tax microsimulation model (Zhu, 2016; also see Bach, Corneo and Steiner, 2013, for a similar construction and income concepts in the German tax system). The distribution is measured on the population of tax units for either data. As for the collection of tax non-filers and the coverage of this administrative data and PHF, see Flory and Stöwhase (2012) and Zhu (2016).

[9] Inverted Pareto coefficient is the ratio of the mean in the distribution above certain threshold and the threshold itself (see, eg Blanchet, Fournier and Piketty, 2017).

[10] There may be a flattening for the wage and professional income above around 100,000 euros. The flattening in the other two income concepts is actually not convincing and the starting points are much higher, if at all. Survey data in Germany suffer from very little sample size above these potential cutoffs to facilitate meaningful estimation (Bach et al., 2012)

[11] Also, he suggests the alternative estimation approaches regarding the choice of survey and external validation data.

[12] Jenkins (2017) calls the Generalized Pareto distribution $f(x) = \frac{1}{\sigma}\left[1 + \frac{\xi(x-\mu)}{\sigma}\right]^{-(\frac{1}{\xi}+1)}$ Pareto II (Lomax distribution) which results from the former under the restriction $\mu = 0$. We follow his notation throughout the paper.

particularly attractive for our application. It would be preferred to have larger sample when fitting our two dimensional model than the univariate one.

Atkinson (2017) shows inverted Pareto coefficient in type I model might not be constant over the top distribution even for a cross section using UK historical data. Blanchet et al. (2017) follow Atkinson (2017) and show that the inverted Pareto coefficient converges from below when the percentile rank of income distribution is near one using US and French data. Using tax returns data, Bach et al. (2012) also show for Germany that the inverted Pareto coefficient looks more like U shaped than a flat curve and converges from below when income grows to the very high end. Based on these findings, we estimate the Pareto II distribution for the use of both marginal and copula-based extrapolations.[13]

Next to the choice of the functional form in the power-law estimation, the pick of threshold is also demanding. Many of the current practices in this literature often identify such cutoff in a subjective way which can be prone to the empirical uncertainty. For instance, since the relatively small sample with rather high income or wealth can be observed in survey data, to necessitate enough size of the fitting sample, this threshold parameter is regularly set to be too low to be true. As we show next such low cutoff is contradicted with the evidence from tax data that supposedly covers the tail well.

Our choice of the cutoff in Pareto distribution results from searching the subsample with the best goodness-of-fit for estimation through a downward expansion process (Clauset, Young, and Gleditsch, 2007 and Clauset et al., 2009). If we lower the cutoffs to expand the sample (from the maximum values in the survey), we are potentially reducing the impact of statistical fluctuation in the estimation and bias due to small sample size as long as the distribution structure is stable. However, when we expand too much and include some lower part of the sample generated from a different distribution, our estimation of Pareto distribution would be biased.

3.2 Copula

Copulas provide a functional link between multivariate distribution functions and their univariate margins (Sklar, 1959). This allows modelling the underlying dependence structure separately from the margins, a result which explains the growing popularity of these functions (Nelsen, 2006). A property of the conditional distribution (with some specific dependence structure) relevant in this paper is stochastic increasing (SI): Y is stochastically increasing in X if $P(Y > y|X = x)$ is increasing in x for all y. As we will see, this SI is equivalent to the SI on their marginal cumulative distributions, which simply becomes a property of copula itself.

3.3 Incidental truncation

Observations are passively collected by the investigative journalism into the rich list and/or the tax auditing into top income tax list. In our joint distribution context, we assume the conditioning variable – income / wealth – broadly summarizes the detectability for the dependent variable - top wealth / income. It is rather conceivable from the perspective of journalists or tax auditors who has to infer the essentially obscure wealth / income from some highly correlated features. Hence, we formulate this detectability by a simple incidental truncation. Only the wealth holders with income higher than some specific

[13] Our estimation is executed by the EXTREME package in Stata (Roodman, 2016).

threshold and the income earners with wealth above a lower bound become conspicuous enough to be captured into either top lists.[14]

As we will see, the gap between the traditional marginal extrapolation (distribution) and our copula-based one lie in a weighting factor which transforms the former into the latter. Using copula modelling, the difference of this factor between extrapolated points is dependent only on the dependence structure of income and wealth distributions but not on the marginal distributions. Take the extrapolation of wealth for example. Formally, let the copula be $\mathrm{C(u,v)}$, where u and v are the cumulative distributions of income Y and wealth W respectively, ie $\mathrm{u = P(Y \leq y_b)}$, $\mathrm{v = P(W \leq w)}$. Here y_b is the income lower bound for the population in some top wealth distribution (eg the rich list) and w is the specific wealth value for someone among them. Then the copula-based probability density function of wealth W, conditional on the income Y all above a lower bound y_b is an incidental truncation: $\mathrm{P(W = w|Y > y_b)} = \frac{\mathrm{P(W=w,Y>y_b)}}{\mathrm{P(Y>y_b)}} = \frac{\mathrm{P(W=w)P(Y>y_b|W=w)}}{\mathrm{P(Y>y_b)}} = \frac{\mathrm{P(W=w)[1-P(Y\leq y_b|W=w)]}}{\mathrm{P(Y>y_b)}} = \frac{\mathrm{P(W=w)[1-\frac{P(Y\leq y_b,W=w)}{P(W=w)}]}}{\mathrm{P(Y>y_b)}} = \mathrm{P(W = w)}\frac{1-\frac{\frac{\partial}{\partial v}\mathrm{C(u,v)}\frac{dv}{dw}}{\mathrm{P(W=w)}}}{1-\mathrm{u}} = \mathrm{P(w)}\frac{1-\frac{\partial}{\partial \mathrm{v}}\mathrm{C(u,v)}}{1-\mathrm{u}}$. The last step is achieved due to the construction of copula linking exactly marginal cumulative probabilities u and v to the joint distribution.[15] $\frac{1-\frac{\partial}{\partial \mathrm{v}}\mathrm{C(u,v)}}{1-\mathrm{u}}$ is our weighting factor mentioned above which is effectively detached from the modelling of marginal distributions. The analysis of the discrepancy between the marginal extrapolation and two top list in the following can benefit from such a separability by concentrating on the property of copula only.

Our weighting factor $\frac{1-\frac{\partial}{\partial \mathrm{v}}\mathrm{C(u,v)}}{1-\mathrm{u}}$ is also equal to $\frac{\mathrm{P(Y>y_b|W=w)}}{\mathrm{P(Y>y_b)}}$. In general, we expect this conditional distribution $\mathrm{P(Y > y_b|W = w)}$ should be increasing with wealth, ie income is stochastically increasing in wealth given the top joint distribution between them.[16] Fixing the truncation $\mathrm{y_b}$, equivalently, this weighting factor should be larger when wealth grows. We will construct the empirical counterpart of the weighting factor as the density proportion of the marginal extrapolation and observed distribution for each point in the top list. If our incidental truncation assumption represents the true data generation process of the top list, this empirical weighting factor should consequently be increasing with wealth. Furthermore, owning to the separability as described above, it may be feasible to improve the degree of approximation accuracy between the copula-based extrapolation and the top list by only choosing the right copula. The interaction from the marginal distribution can be ignored by design.

---

[14] Vermeulen (2018) highlights the existence of differential unit non-response in the wealth and income surveys when the richer the less they are participating the survey. Such evidence for wealth is available from Kennickell and Woodburn (1997) by linking an income tax based wealth estimate to the response rate in SCF (Survey of Consumer Finance). Kennickell and McManus (1993) observe that non-response is correlated with financial income in the SCF. Likewise, there can be a missingness mechanism widespread among the target populations of top rich list and income tax list. In contrast to the spontaneous survey, the observability in our top lists has a passive characteristic.

[15] This is allowed by Sklar's Theorem. Other forms of joint distribution representation (eg parametric joint distribution) would not produce such a result.

[16] Alternatively speaking, the richer in wealth for someone the more likely her/his income is above some threshold $\mathrm{y_b}$.

4. Estimating top marginal and joint distributions

We describe the fitting of Pareto II distributions on both top income and wealth distributions as well as the copula estimation.

4.1 Top marginal distributions

Besides searching the threshold as the minimizer of the goodness-of-fit as proposed, there are other procedures for validation (Jenkins, 2017). Mainly, an intuitive approach is to plot the estimated parameters against thresholds and pick the one above which the estimates are flat. In addition to the threshold where the model starts to apply, the Pareto II model has the other two parameters, a shape parameter ξ and the scale parameter σ, whose estimates are simultaneously determined with each candidate threshold. Jenkins (2017) suggests that for UK data the potential values of the threshold may range between the 95 to the 99 percentile of the income distribution. Alternatively, the empirical inverted Pareto coefficient observed in some external data can be the benchmark of our estimate for the income Pareto distribution.

Figure 2 plots the KS measure for the income observed for PHF between 35,000 and 290,000 euros, which is far below p90 (as a robustness check) and somewhat larger than p99.5. P95 to p99.5 is the common searching range and the optimal thresholds for UK income distribution lie within this range (Jenkins, 2017). The author also suggests this finding might not be universally applicable in other countries and variables.[17] The minimum is reached at somewhere below or at p90. Figure 3 provides the curves of estimates against threshold. It justifies our choice by showing that the above area is most stable. The exact minimizer of KS is 70,271 euros. However, the inverted Pareto coefficient *b* based on the estimates using this threshold converges from above as rank *p* approaches to 100%.[18] This also goes against the Figure 1. Our favoured threshold is 86,957 euros which is 7th smallest thresholds and has the KS measure being 0.0135. The KS measure for 70,271 is 0.0131.[19]

Figure 4 and Figure 5 are the inverted Pareto coefficients by income and probability for the top income when the Pareto II follows the most favoured estimate: threshold=86,957, shape=.42 and scale=34,241. They both converge from below which implies the tail is getting fatter as income becomes higher. It aligns with the behaviour of *b(p)* in US, France and UK for the recent decades (see Fig 1 in Blanchet et al., 2017, and the discussion in Atkinson, 2017). Most interestingly, comparing with Figure 1, we can observe in Figure 4 our inverted Pareto coefficient curve is very close to the grey counterpart (ie the coefficient for the wage and professional income in the income tax data) which lies between around 1.68 and 1.72 and converges at around 1.72. PHF does cover the distribution and aggregate of labor income very well when benchmarking with the income tax data and national account aggregate (coverage rate of 98%). But the coverage of capital income is just about 60%.[20] It is reasonable to postulate that our gross income is closest to the wage and professional income concept in the tax data. Given these facts, this estimate of Pareto II distribution looks convincingly to be close to the true distribution.

---

[17] If our threshold searched becomes low and let our fitting sample to include some non-power-law distribution, KS statistic will become large than the optimal (minimum) one (See p.671 of Clauset et al. (2009); generally, we should expect to see the U curves in Figure 2 and Figure 6 in our paper if we search the whole sample).
[18] Blanchet et al. (2017) provides the formula for *b(p)*.
[19] The 6th smallest threshold is 70,400 euros. And all the first six thresholds together with their parameter estimates can only deliver the downward converging inverted Pareto coefficient.
[20] See Zhu (2014).

Figure 6 plots the KS measure for the wealth observed for PHF between 100,000 and 5,000,000 euros, which is again far below p90 and somewhat larger than p99.5. The minimum 245,160 euros is located between somewhere around the half of p90 and somewhere a bit higher than p90. Figure 7 provides the estimated curves against the thresholds. For the modified sigma, this area does look most stable. For ξ, the most stable area lies between p90 and p99. But these estimates, at around 0.6, are close to that obtained when threshold is at about the half of p90. Given these results, we decide to take this exact minimizer 245,160 euros which can offer us more sample size in considering fitting a top joint distribution.

Figure 8 and Figure 9 are the inverted Pareto coefficients by income and probability for the top wealth when the Pareto II follows the most favoured estimate: threshold=245,160, shape=.6 and scale=197,206. They both converge from above which implies the tail is getting thinner as wealth becomes higher. We do not have direct evidence from other countries or Germany to verify this profile. However, the Pareto indices (alpha in Pareto I distribution) are 1.52 for the Manager Magzin rich list and 1.47 for PHF when the most favoured MLE estimation is adopted for these two data separately (see the very first row in Table 3 of Dalitz, 2016). It implies the inequality is not so high in the top wealth distribution as that below the very top in the rich list for Germany. This is compatible with the profile of *b(p)* for the most favoured estimate.

4.2 Top joint distribution

We conjecture there exists an invariant copula structure for the top income and wealth joint distribution. The approach to search the cutoffs for fitting sample is simply a bi-dimensional extension from that used for power-law estimation discussed above. The product of such estimation is labelled as *best fitting* copula.

There are the other two top copulas we proposed for the copula-based extrapolation as sensitivity analysis. Besides the plausible extrapolability as argued below, they will assist us in exploring the usefulness of the *best fitting* copula in terms of the explanatory power of the incidental truncation. The a priori candidate is the copula for the *full sample*. The underlying copula we want to exactly extrapolate under our incidental truncation should have the marginal distributions following these two favoured Pareto II specifications with the cutoffs identified above. In what follows, the other natural candidate is the one estimated from the upper right rectangular sample of the point defined by these two cutoffs. This fitting sample is simply the observed part of the data behind our target underlying copula. The other argument to go for this second alternative copula (fitting sample) is to gauge the net contribution of introducing copula purely from the modelling comparision perspective. Since we are comparing the extrapolation accuracy of marginal and copula-based approach, we should estimate the copula from (almost) the same sample which fits the marginal distributions. Furthermore, the copula should be estimated based on the parametrically estimated marginal cumulative distribution instead of the empirical one. Otherwise, the worse prediction accuracy from the marginal approach may be partially attributed to more data noise during the estimation given that copula estimation could be based on larger sample size or empirical marginal distribution is more proper estimate for a skewed tail distribution. This contending subsample is labelled as *high income - high wealth*.

Estimations for these three cases are presented in Appendix A. Table 2 summaries these three cases as well as the accompanying marginal specification. The copulas estimated in the samples defined in Table 2 belong to either (Student) t or BB8 families. Both are symmetric copula. For the BB8 copula, either dimension is stochastically increasing in the other. However, the degree of SI can be quantitatively very small for the very top of the distribution. Student t copula does not always have SI. For instance, $\mathrm{P}(\mathrm{U} > \mathrm{u}|\mathrm{V} = \mathrm{v})$ can be decreasing in v for small u under a t copula $C(U, V)$. However, the t copula demonstrates SI under the parameterization in this paper.[21]

5. External validity

We carry out the validation exercise by comparing the marginal extrapolation and our copula-based extrapolation with the incidental truncation on our holdout datasets – top lists on wealth and income. The empirical effectiveness of our *best fitting* case is exhibited followed by the sensitivity analysis. We then connect these results with the differential detectability which is built on the incidental truncation assumption.

5.1 Comparison framework

The competing specification is benchmarked with the real distribution of extrapolation targets (top rich list or top income tax list) through the p-p plot. The predictive accuracy is assessed by the distance between the extrapolated distributions and the observed ones. The extrapolated probability for each data points is calculated by using the observed income or wealth values according to the specification in either marginal or copula-based approaches. One advantage of this distribution comparison tool is to minimize the disruption from skewness. The plots draw the empirical cumulative distributions of the top lists against the extrapolated one at each of the observations in the top lists. The former lies on the x-axis and the latter on the y-axis. In our setting, we plot two extrapolated distributions (distinguished by the case numbers 1 and 2) and each one contains both the marginal and copula-based approaches as specified in Table 2. They are denoted as *marginal 1*, *copula 1*, *marginal 2* and *copula 2* respectively. This allows us to compare not only between marginal and copula-based approaches within cases (specifications) but also between cases.

In our incidental truncation framework, we still have an unobservable parameter – the income or wealth lower bound for the top wealth-richest population or income tax payers. In this paper, we roughly estimate it so that copula-based extrapolation can achieve the visually best fit with the top lists by a trial - and - error grid search. To analyse the function of the copula structure in incidental truncation, we perform a sensitivity analysis by adopting another potential lower bounds when fixing the copula. They are our naïve lower bounds which are the maximum values in PHF, ie 76.3 million euros for wealth and 2.9707 million euros for income.[22] We present these specifications under each p-p plot and particularly highlight the contrasting components in italics. Except the contrasting pairs, all the remaining components are kept the same between case *1* and *2*.

[21] The analytic proof for the SI of BB8 as well as some numeric examples under t using the parameterization in this paper can be provided upon request.

[22] This assumes the minimum wealth of all the top 1% income distribution (as in the tax data) is at least 76.3 million euros and the minimum income of all the observations in the wealth rich list is at least 2.9707 million euros. The real bound for the former should be lower than 76.3 million since PHF does cover some distribution above top 1% percentile. And the real bound for the latter should be higher than 2.9707 million euros since the maximum wealth in PHF is still much below the minimum in the rich list.

5.2 Results

By trial and error, Figure 10 shows such an "optimal" conditioning income lower bound as roughly 20 million. Figure 11 presents the "optimal" conditioning wealth lower bound at about 0.7 million.[23] Claiming the minimum income for the observations in the rich list as 20 million is not inconceivable given the literature in investigating the German economic elites' gross income using tax administrative data (Bach et al, 2013)[24] Nevertheless, it is rather plausible that the top 1% income earners have at least 0.7 million wealth. Moreover, the tax data is more reliable than the rich list and prediction with top 1% distribution is fairly accurate.[25] In general, the copula-based approach can be reasonably verified by the real holdout data.

We provide the results of sensitivity analysis as varying the truncation lower bounds and copula in Appendix B. The lower bound does matter for the predictive accuracy in various cases which is conjectured by our incidental truncation assumption. For the extrapolation of the rich list, we observe that BB8 copula always underperform against the t copula in terms of predictive accuracy given different lower bounds of income. The predictive accuracy from both is similar for the extrapolation of the top income tax list under appropriate lower bounds of wealth. Next, we will illustrate these validation outcomes by linking our incidental truncation mechanism to the data and copula structures.

5.3 Differential detectability

Another advantage of the p-p plot is to allow us to visualize the weighting factor in the incidentally truncated distribution for the top lists. The slope of the marginally extrapolated distribution in the p-p plot is simply the inverse of the empirical weighting factor. If the incidental truncation is indeed underlying the top lists, the x-axis increment of any two neighbouring observations along the top lists is the density of the richer observation according to the incidentally truncated distribution and the y-axis counterpart is the density of the richer observation with the marginally extrapolated distribution. The former is a product of the latter and the weighting factor as presented above.

The marginal extrapolation in all the p-p plots are concave. This implies that empirical weighting factor is increasing with our dependent variable in the incidental truncation. Equivalently, either dimension is stochastically increasing in the other one for the joint distribution of income and wealth. As claimed this serves as the evidence endorsing our incidental truncation assumption. Under the student t copula, as discussed above, SI is maintained for either dimension in the range of the marginal cumulative probabilities of top lists and the neighbourhood around the roughly optimal truncation variable. However, the level of SI, eg measured by the second order partial derivative of the copula, under BB8 copula, can be quantitatively quite small in the parameterization for the extrapolation of rich list as this list lies in very top of our estimated

[23] These seem to be the unique "optimal" conditioning value such that the extrapolated distributions almost overlap the true ones only once when we vary the conditional value.
[24] In Table A2 from 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.
[25] See Bach et al. (2015) and Chakraborty and Waltl (2018) for the inaccuracy of rich list reporting.

power-law wealth distribution. Nevertheless, such level for BB8 copula is significant and similar to that of t copula in the parameterization for the extrapolation of top income tax list.[26]

Given the inferior validation outcome on the rich list by our *best fitting* (BB8) copula in relative to the *full sample* (t) copula, can we reject the capability of our principled top copula selection procedure? Probably not. First, as we have argued the credibility of the rich list as a representative sample of the top wealth distribution is not convincing. Second, probably without some small statistical fluctuation, we should have achieved the top copula structure same as those from this *full sample* t copula in our estimation. One manifestation is that these two copula structures are rather close in the top 30% as we align their contour curves together in Figure 20. This covers the area we are interested: the rich list has the wealth values above percentile 99.997488% in the wealth marginal cdf and top income tax list has the income values above percentile 83%; optimal wealth lower bound .7 mill is at about .77 in the wealth margin cdf and optimal income lower bound (for t copula) 20 mill is at about .999998 in income marginal cdf. Third, a more convincing validation data for wealth should cover a long range top tail instead of the current rich list over a very tiny tip. Imagining obtaining a top representative distribution with a probability spread like our top 1% income list, our *best fitting* (BB8) should have performed well given the validation evidence for the top income tax data and the fact BB8 is symmetric in the top. In fact, this vision can come true soon with the materialization of the new top tail sampling scheme designed and implemented recently by the German Social Economics Panel – GSOEP from German Institute for Economic Research (Schröder, Bartels, Grabka, König, Kroh, & Siegers, 2019).

6. Conclusions

This paper proposes a copula-based joint extrapolation for the top income and wealth distributions which are commonly under-covered in the survey data. When benchmarking with the empirical top rich distribution from administrative tax data and wealth rich list, the copula-based extrapolation does always perform better than the marginal one when the copula is estimated under a range of plausible scenarios, especially the *full sample* / *high income* - *high wealth* cases, as long as the conditioning lower bound is sensibly set. For example, if credibly assuming the minimum wealth for the population in top 1% income distribution to be about 0.7 million, our copula-based extrapolation using a copula estimated from the whole PHF sample can almost exactly replicate the distribution of top 1% tax data. We also propose a principled search of the underlying top copula which in validation exercise can beat the marginal extrapolation for the more credible holdout data – top income tax data. The presumption of an incidental truncation and differential detectability rationalizes the observed gap between the existing marginal extrapolation and top lists as well as the effectiveness of our copula-based approach.

The adoption of joint association has a positive net contribution beyond the marginal approach. The extrapolation can be improved by simply adding only a copula estimation step using the existing parametrically estimated marginal power law and almost exact sample (but jointly).

Our validation on the rich list has revealed the uncertainty of our in-sample copula fitting under our principled approach. Therefore, in the future, we need a hypothesis test to tell the validity of our (copula) model selected. On the other hand, it

[26] BB8 copula still has some slight stochastic monotonicity in the very top if we zoom in these lines of marginal extrapolation in the p-p plot for the rich list. The cumulative probability specified by the estimated power-law distribution for the minimum point in the rich list is already as high as 0.99997. The counterpart in the top income tax list is 0.83.

seems to be imperative to formally estimate the lower bound in our incidental truncation setting, which simply treats it as an additional parameter and include our holdout data as fitting sample too. Thus, incorporating such a complete incidental truncation estimation into a bootstrap hypothesis test can be a solution (see section 4 in Clauset et al. (2009)). This approach would simulate many samples from the estimated model and build the sampling distribution of their fit towards the estimate. Then we compare the empirical fit from the one-shot estimation with that distribution to obtain the p-value. Since we will lose the holdout data in this complete estimation, for the purpose of model selection between marginal extrapolation and copula-based one (and between copulas), we can rely on other techniques such as cross-validation or likelihood ratio test as proposed in section 5 of Clauset et al. (2009).[27]

[27] Shalizi (2016; p.70) emphasizes that cross-validation can help to produce unbiased generalization error beyond avoid overfitting.

References

Aaberge, R., Atkinson, A. B., & Königs, S. (2018). From Classes to Copulas: Wages, capital, and top incomes. *The Journal of Economic Inequality*, *16*(2), 295-320.

Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica*, *84*(334), 129-156.

Bach, S., Corneo, G., & Steiner, V. (2012). Optimal top marginal tax rates under income splitting for couples. *European Economic Review*, *56*(6), 1055-1069.

Bach, S., Corneo, G., & Steiner, V. (2013). Effective taxation of top incomes in Germany. *German economic review*, *14*(2), 115-137.

Bach, S., Thiemann, A., & Zucco, A. (2015). The top tail of the wealth distribution in Germany, France, Spain, and Greece. Discussion paper 1502, Deutsches Institut für Wirtschaftsforschung, Berlin, 2015.

Blanchet, T., Fournier, J., & Piketty, T. (2017). Generalized Pareto Curves: Theory and Applications.

Brechmann, E., & Schepsmeier, U. (2013). Cdvine: Modeling dependence with c-and d-vine copulas in r. *Journal of Statistical Software, 52*(3), 1-27.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–215. MR1874152

Chakraborty, R., & Waltl, S. (2018). *Missing the wealthy in the HFCS: micro problems with macro implications* (No. 2163). European Central Bank.

Clauset, C. R. Shalizi, and M. E. Newman, “Power-law distributions in empirical data,” SIAM Review, vol. 51, no. 4, pp. 661–703, 2009.

Clauset, A., Young, M., & Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, *51*(1), 58-87.

Dalitz, C. (2016). "Estimating Wealth Distribution: Top Tail and Inequality." Technischer Bericht Nr. 2016-01, Hochschule Niederrhein, Fachbereich Elektrotechnik und Informatik

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.

Flory, J., & Stöwhase, S. (2012). MIKMOD-ESt: A static microsimulation model of personal income taxation in Germany. *International Journal of Microsimulation*, *5*(2), 66-73.

Geisser, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70** 320–328.

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2017). copula: Multivariate dependence with copulas. *R package version 0.999-17, https://cran.r-project.org/web/packages/copula/copula.pdf*

Jenkins, S. P. (2017). Pareto models, top incomes and recent trends in UK income inequality. *Economica*, *84*(334), 261-289.

Joe, H. (1997). Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Joe, H. and Xu, J. (1996). “The Estimation Method of Inference Functions for Margins for Multivariate Models.” Technical Report 166, Department of Statistics, University of British Columbia.

Kennickell, A.and D. McManus, “Sampling for Household Financial Characteristics Using Frame Information on Past Income,” Paper presented at the 1993 Joint Statistical Meetings, Atlanta, GA, 1993.

Kennickell, A. B., & Woodburn, R. L. (1999). Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth. *Review of Income and Wealth*, *45*(2), 193-215.

Kojadinovic, I. (2017). Some copula inference procedures adapted to the presence of ties. *Computational Statistics & Data Analysis*, *112*, 24-41.

Kojadinovic, I., & Yan, J. (2010). Modeling multivariate distributions with continuous marginals using the copula R package. *Journal of Statistical Software*, *34*(9), 1-20.

manager magazin (2014). Die reichsten Deutschen. Top 500. manager magazin SPEZIAL, October:16–30

Mosteller, F. and Tukey, J. W. (1977). Data Analysis and Regression. Addison-Wesley, Reading, MA.

Nelsen, R. B. (2006), *An Introduction to Copulas*. 2nd edition. New York: Springer.

Neri, A., & Ranalli, M. G. (2012). To misreport or not to report? The measurement of household financial wealth. *The Measurement of Household Financial Wealth (July 26, 2012). Bank of Italy Temi di Discussione (Working Paper) No*, *870*.

Roodman, D. (2016). EXTREME: Stata module to fit models used in univariate extreme value theory. Statistical Software Components archive. Boston, MA: Boston College. https://ideas.repec.org/c/boc/bocode/s457953.html

Schröder, C., Bartels, C., Grabka, M. M., König, J., Kroh, M., & Siegers, R. (2019). A Novel Sampling Strategy for Surveying High Net - Worth Individuals—A Pretest Application Using the Socio - Economic Panel. Review of Income and Wealth.

Shalizi, C. R. (2016). Advanced data analysis from an elementary point of view. *http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV*.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229-231.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T., ... & Killiches, M. (2017). Package 'VineCopula', https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf

Shmueli, G. (2010). To explain or to predict?. *Statistical science*, *25*(3), 289-310.

Stone, M. (1974). Cross-validatory choice and assesment of statistical predictions (with discussion). *J. Roy. Statist. Soc. Ser. B*, **39** 111–147. MR0356377

Trivedi, Pravin K., and David M. Zimmer. "Copula modeling: an introduction for practitioners." Foundations and Trends in Econometrics 1, no. 1 (2007): 1-111.

Vermeulen, P. (2018). How fat is the top tail of the wealth distribution?. *Review of Income and Wealth*, *64*(2), 357-387.

Yan, J. (2007). Enjoy the joy of copulas: with a package copula. Journal of Statistical Software, 21(4), 1-21.

Zhu, J. (2014). Bracket Creep Revisited-with and without r> g: Evidence from Germany. *Journal of Income Distribution*, *23*(3), 106-158.

Zhu, J. (2016). Asking Flexible Disaggregate Income - Gross or Net Income - vs Single Total Question. Memo.

Table 1: German gross income for the top 30% by percentile mean: PHF (2009) and ITR-SOEP (2008)

Percentile	PHF 09	ITR-SOEP 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: ITR-SOEP 08 - income tax return data *(Geschäftsstatistik*, 2008) integrated with GSEOP data to account for the missing tax non-filers which is considered to distributionally represenative for Germany (see Bach et al., 2013; we thank Stefan Bach for providing us this integrated data for 2008). *Geschäftsstatistik* is also used for the Taxpayer-Panels micro data offered by the Research Data Centre of the Federal Statistical Office: https://www.forschungsdatenzentrum.de/de/steuern/tpp;
PHF 09 - wave one of the Panel on Household Survey (Bundesbank). Both are measured as the gross income for the tax unit population.

Table 2: Candidate copula estimates and the marginal distribution specification/estimates by different fitting samples

	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

Note: Each case is defined by the fitting sample above the income and wealth cutoffs and the copula estimated. *Best fitting* case has the fitting sample as the upper right rectangle of the grid with best bivariate KS statistic in our grid search of optimal copula estimation (or sample cutoffs). *Full sample* case simply takes the whole data for fitting. The marginal distribution is used together with the copula estimate in each case to construct conditional probability. Only the couple for the case of high income-high wealth is also estimated under the same fitting sample as copula estimate. Note that, due to the joint restriction of cutoffs, the exact income and wealth lower bounds (ie threshold parameters specified in the Pareto II distribution) are 87,000 and 245,680 which are slightly higher than the cutoffs 86,957 and 245,160.

Figure 1: Empirical inverted Pareto curve for the couples' income in the German tax data (2005) - Fig 2 in Bach et al. (2012)

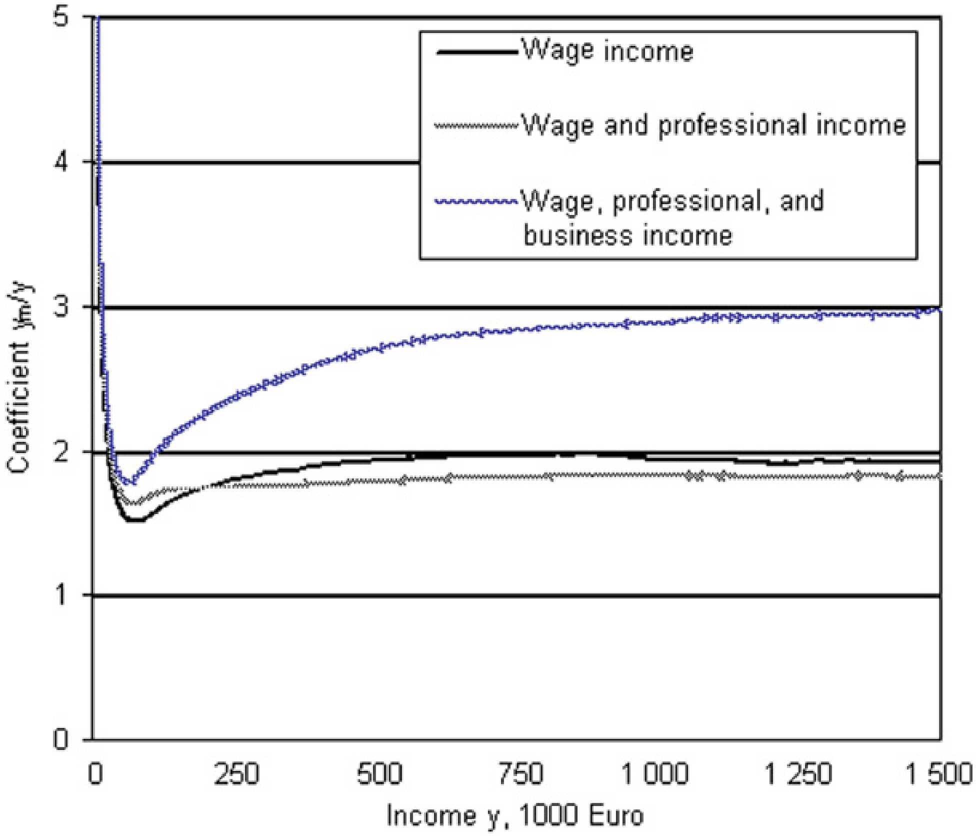


Figure 2: Goodness-of-fit criteria after Kolmogorov-Smirnov (KS) as a function of threshold for the averaged PHF income data (dashed lines: p90, p95, p99, p99.5)

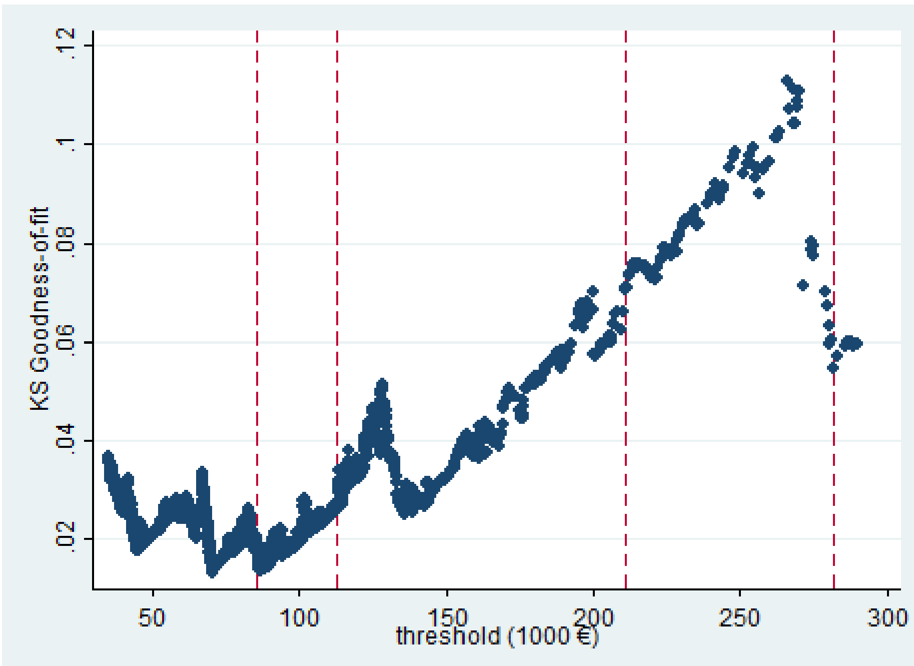


Figure 3: Pareto II parameter estimates by threshold for the averaged PHF income data (dashed lines: p90, p95, p99, p99.5)

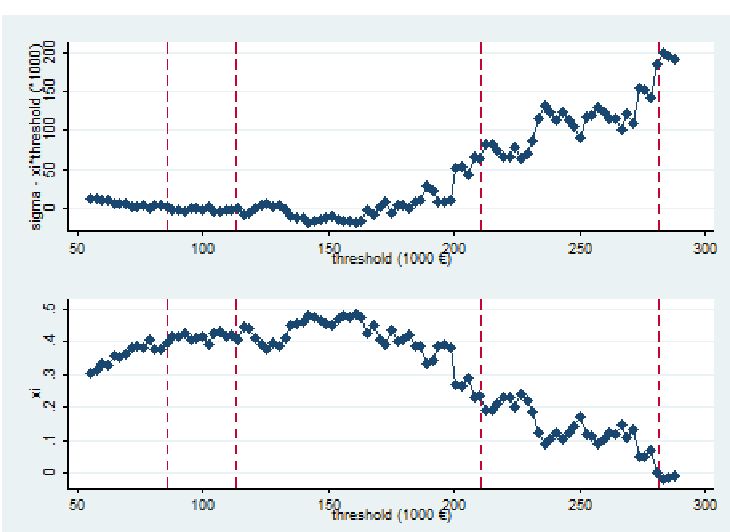


Figure 4: Inverted Pareto coefficient w.r.t. income for the Pareto II distribution (threshold=86,957, shape=.42 and scale=34,241)

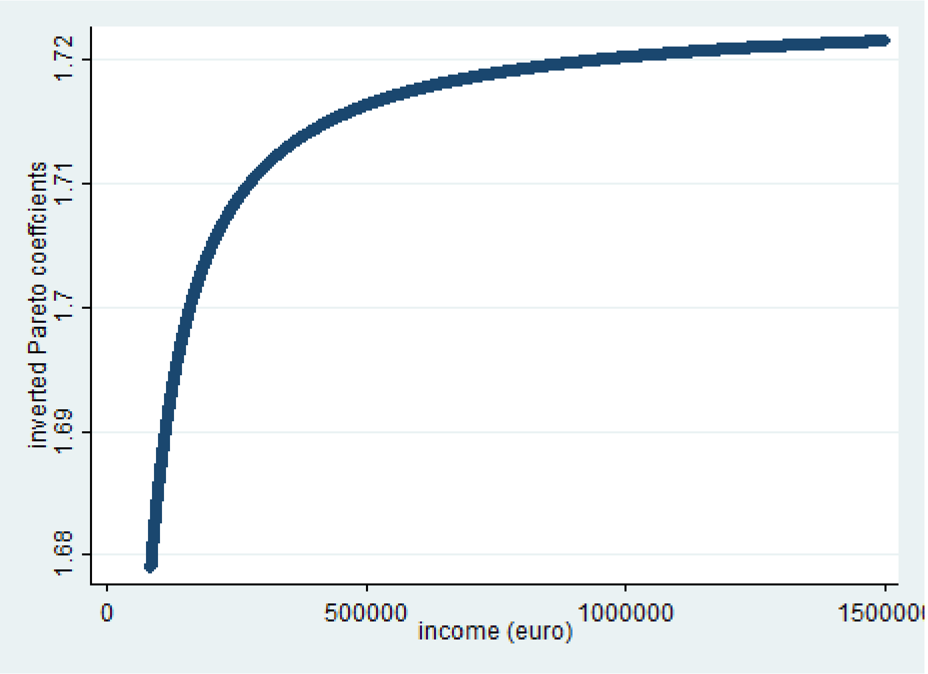


Figure 5: Inverted Pareto coefficient w.r.t. probability for the Pareto II distribution (threshold=86,957, shape=.42 and scale=34,241)

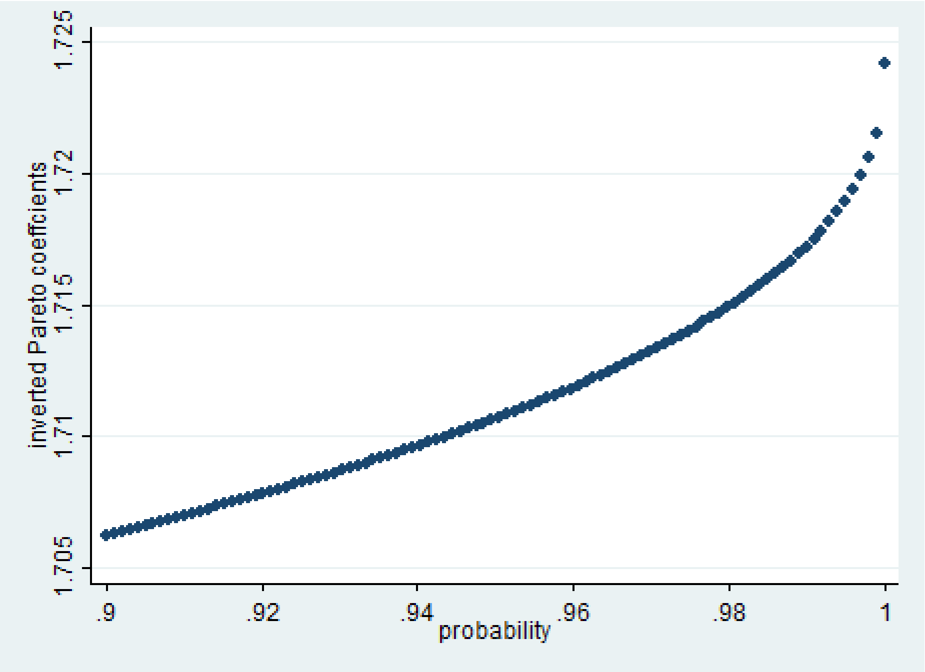


Figure 6: Goodness-of-fit criteria after Kolmogorov-Smirnov (KS) as a function of threshold for the averaged PHF wealth data (dashed lines: p90, p95, p99, p99.5)

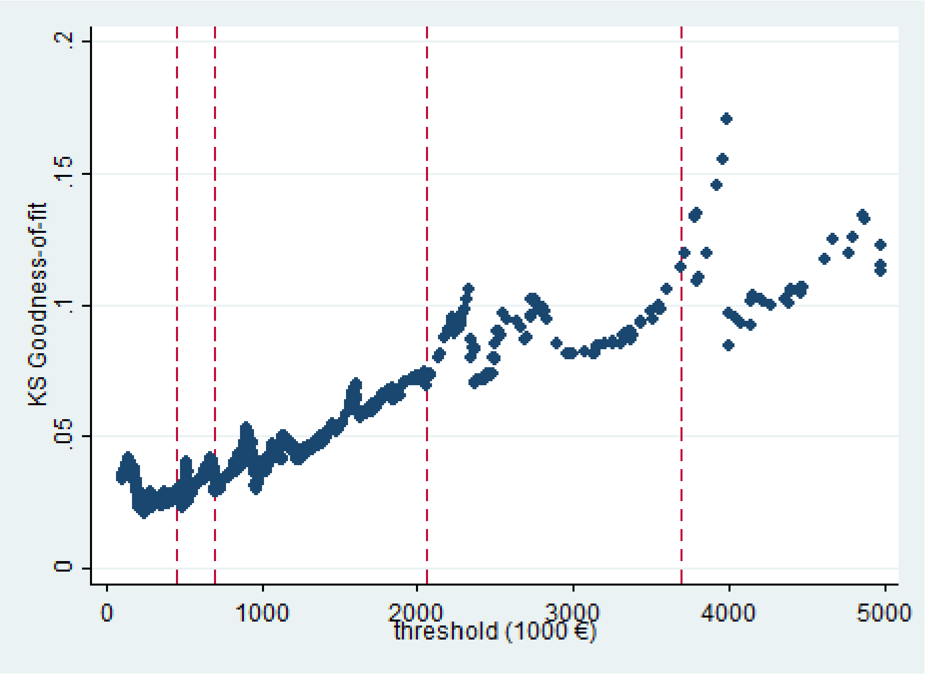


Figure 7: Pareto II parameter estimates by threshold for the averaged PHF wealth data (dashed lines: p90, p95, p99, p99.5)

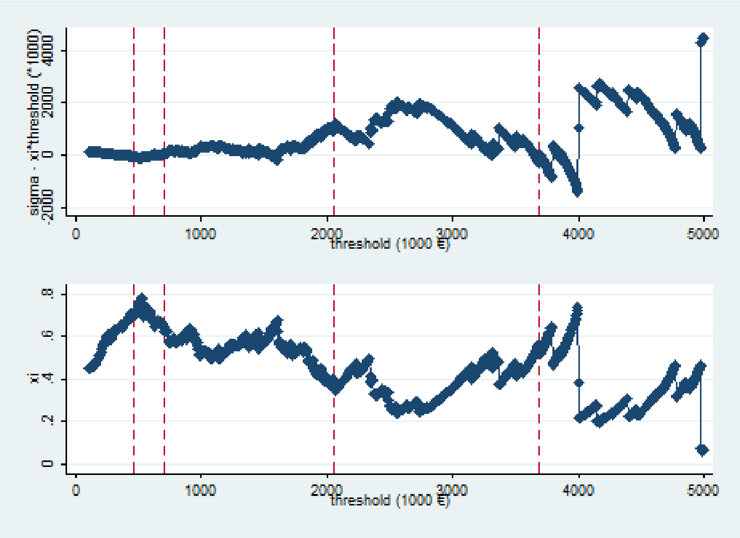


Figure 8: Inverted Pareto coefficient w.r.t. wealth for the Pareto II distribution (threshold=245,160, shape=.6 and scale=197,206)

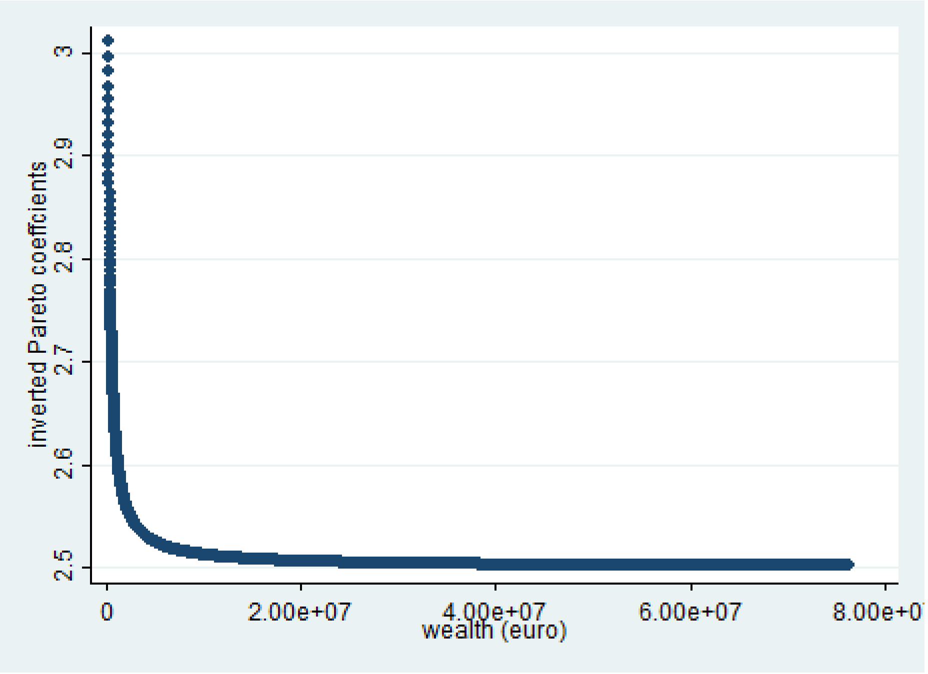


Figure 9: Inverted Pareto coefficient w.r.t. probability for the Pareto II distribution (threshold=245,160, shape=.6 and scale=197,206)

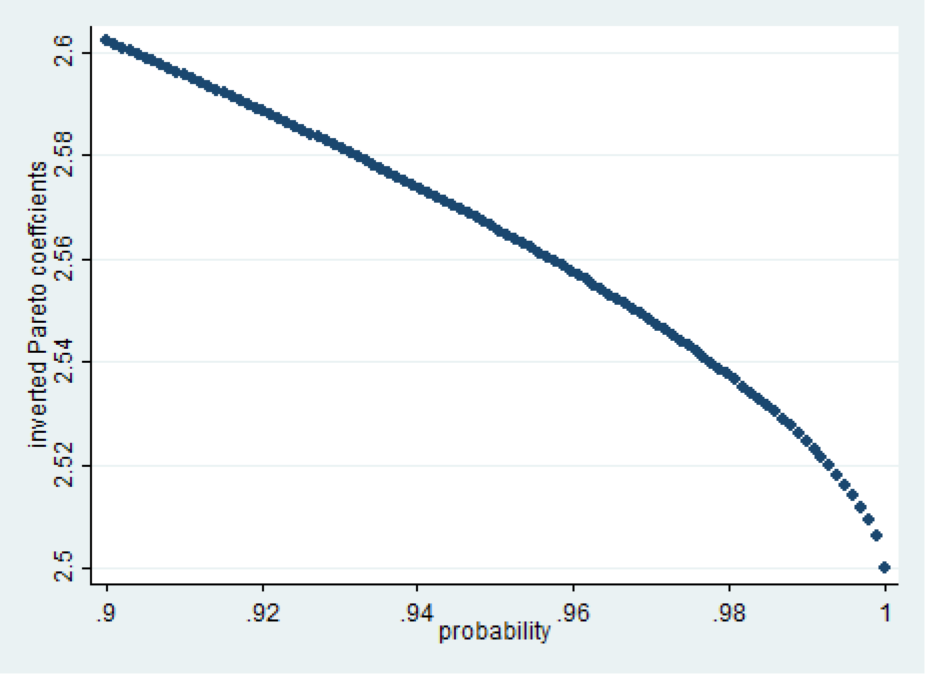
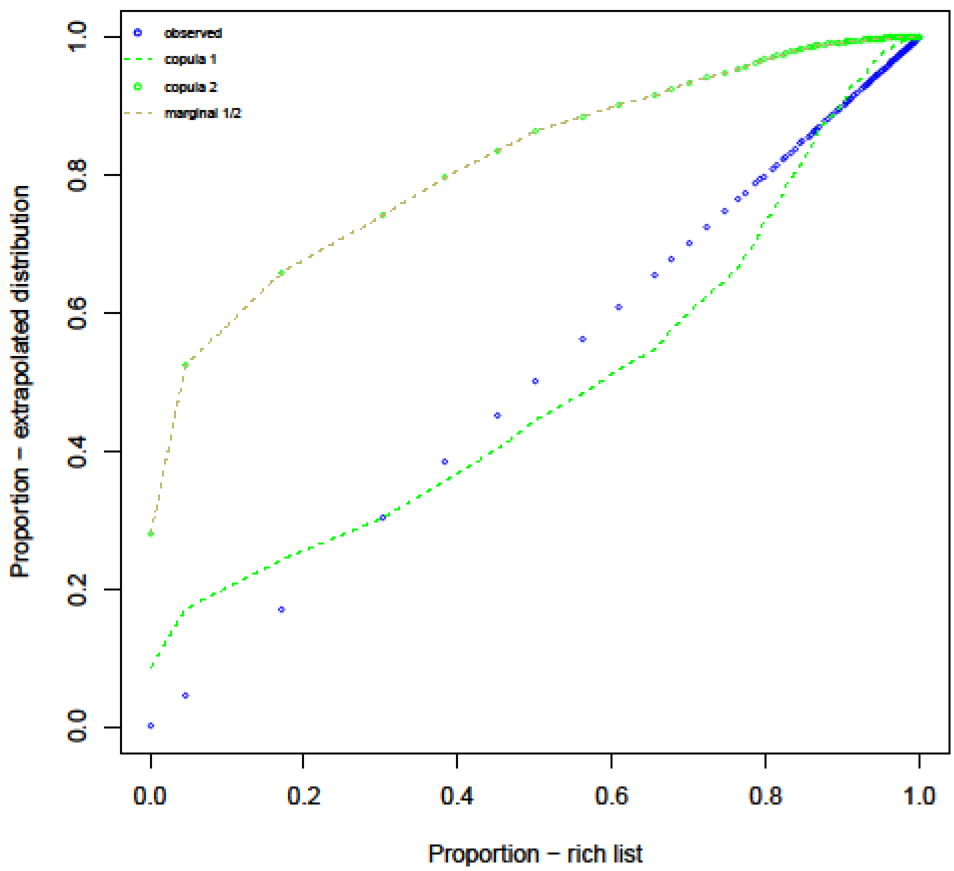
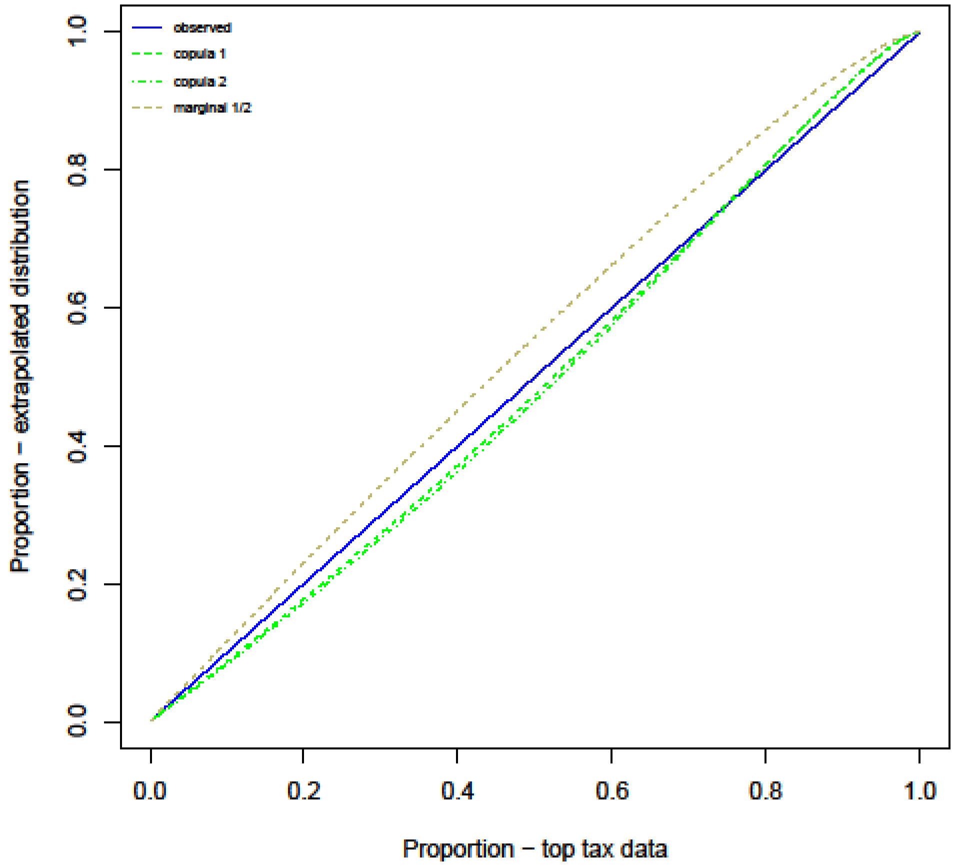


Figure 10: P-p plot benchmarking marginal and copula-based extrapolated distributions with top rich list – comparing *full sample* (t) and *best fitting* (BB8) copulas under roughly optimal conditioning income



1 – *full sample* case and 2 – *best fitting* case
marginal 1/2 – wealth distribution - Pareto II (threshold=245160, shape=.6, scale=197203)
copula 1: - *Student t (parameter 1=0.589, parameter 2=6.2)*, conditioning income lower bound=20,000,000 and income distribution - Pareto II (threshold=86957, shape=.41, scale=34612)
copula 2: - *BB8 (parameter 1=1.62, parameter 2=0.97)*, conditioning income lower bound=20,000,000 and income distribution - Pareto II (threshold=86957, shape=.41, scale=34612)
Dots in unconnected lines (eg observed) correspond to the exact observations which reflect the dispersion of the sample.

Figure 11: P-p plot benchmarking marginal and copula-based extrapolated distributions with top income tax data – comparing *full sample* (t) and *best fitting* (BB8) copulas under roughly optimal conditioning wealth



1 – *full sample* case and 2 – *best fitting* case
marginal 1/2 – income distribution - Pareto II (threshold=86957, shape=.41, scale=34612)
copula 1: - *Student t (parameter 1=0.589, parameter 2=6.2)*, conditioning wealth lower bound=700,000 and wealth distribution - Pareto II (threshold=245160, shape=.6, scale=197203)
copula 2: - *BB8 (parameter 1=1.62, parameter 2=0.97)*, conditioning wealth lower bound=700,000 and wealth distribution - Pareto II (threshold=245160, shape=.6, scale=197203)
It is not necessary to show the dispersion of this sample because they are 1,000 percentile points within top 1%.

Appendix A. Copula estimation

A.1 Best fitting copula

A grid of 10,000 points is imposed in the full sample, which are the intersections of 100 empirical weighted percentiles for either dimension. We hope these subsamples from the upper right rectangle of each grid point can sufficiently represent the universe of copula structures of all the rectangular subsets in the coordinate system formed by the whole sample.[28] The copula structure is then estimated parametrically for each of these subsamples.[29] The inputting marginal probabilities are the empirical weighted cumulative probability. It fits a comprehensive set of copulas via maximum likelihood estimation. Then the Akaike Information Criteria are computed for all of them. The one with the minimum value is selected.[30] Similar to the last step in identifying the cutoff for power-law estimation, we finally calculate the two dimensional Kolmogorov-Smirnov (KS) statistic for the fittings of each subsample and pick the copula estimate with the smallest KS statistic.

Figure 12 illustrates the bivariate KS statistics at these 10,000 grid points. Our strategy does work: there is a plateau in the bottom left rectangle bordered by approximately 30% percentile line in wealth and about 45% percentile line in income and a trench along these border lines.[31] The grid with minimum KS statistic is at the intersection of 7% percentile in income (5,300) and 32% percentile in wealth (29,500). The copula estimated at this point is a BB8 with two parameters (1.62, 0.97).[32] We observe the copulas estimated in this trench area (with estimated copulas fitted closest to the empirical ones) are almost all BB8 with similar parameter values. This evidence is in favour of our stability presumption for the top copula − otherwise we may see a variety of copula estimates all with the almost best goodness-of-fit but distinct with family and / or parameters.

A.2 Copula for sensitivity analysis

The copula estimation for the *full sample* case takes the same approach as that implemented for the *best fitting* case.[33] For both the weighted empirical cdfs are the marginal inputs. However, we adopt the two-stage estimation method called inference functions for margins (IFM) in *high income-high wealth* case. The IFM method firstly estimates the parameters of marginal distributions via a MLE and then estimates the copula parameters via a separate MLE using the parametric cdfs calculated based on the estimates in the first step. The second step is same as the semi-empirical estimation method used for the first three cases except the formation of marginal cdfs (See Yan, 2007, and Joe and Xu, 1996).

---

[28] The subsample in upper right rectangle of each grid point contains all the observations with both income and wealth above the values of that grid point. Alternatively, the copulas for the subsamples among the data points in the neighbour of these grid points are expected to be not much different from the counterparts from these grid points.

[29] We use BiCopSelect() from the VineCopula package in R to choose the best copula using weights. Copula package is also used for eg plotting. For the description of each R packages (ie VinceCopula and Copula), we suggest to read Brechmann and Schepsmeier (2013), Schepsmeier, Stoeber, Brechmann, Graeler, Nagler, Erhardt and Killiches (2017) and Hofert, Kojadinovic, Maechler and Yan (2017). A comprehensive theoretical and computational coverage for the Copula package as well as copula modelling itself is available in Yan (2007) and Kojadinovic and Yan (2010).

[30] The package also allows using Bayesian Information Criteria when the models with higher number of parameters are to be penalized which we can experiment in the future.

[31] The plateau is more visible in the perspective plot with the assistance from the color scale and the trench is more distinguishable from the level curve from the contour plot.

[32] It is also named as Joe-Frank (Joe, 1997) with one parameter larger or equal to one and the other between zero and one.

[33] Namely, we also use BiCopSelect() with MLE together with AIC.

To check the goodness-of-fit, we compare the contour curves from both the empirical copula with those from the estimated one in Figure 13, Figure 14 and Figure 15.[34] When the margins adopt the empirical distribution, the contour curves are compared in the unit square. When the margins are parametrically specified, they are ploted in the bivariate supports of fitting sample. All the contour lines from both copulas seem to lie close to each other.

Appendix B. Results of sensitivity analysis

Under the naïve lower bounds, in Figure 16, we compare the extrapolation on wealth rich list between the *full sample* and *best fitting* cases where only the copulas (between t and BB8) are contrasted. The same is done on the top tax data in Figure 17. In terms of the predictive accuracy, copula-based extrapolation using *full sample* (t) copula beats *best fitting* (BB8) copula for the top wealth distribution and it is the opposite for the top income extrapolation.

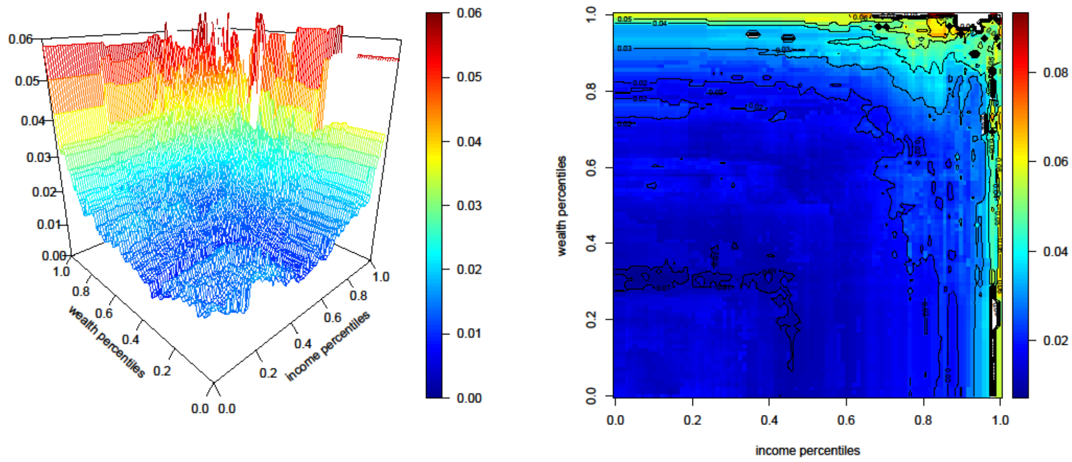
For the top wealth list in Figure 16, copula-based extrapolation with *full sample* (t) copula outperforms the marginal approach. The conditioning income lower bound 2.9707 million has positive probability to be true given the farther-out minimum wealth in the rich list. This superiority over the marginal approach cannot be observed for the best fitting (BB8) copula in Figure 17 for the top income distribution. The conditioning wealth lower bound being as high as 76.3 million seems to be simply unrealistic.

The net contribution from incorporating the joint structure is presented via our validation using the *high income - high wealth* case. By trial-and-error to locate the roughly optimal lower bound, Figure 18 presents its benchmark on the top rich list when income cutoff is 60 million and Figure 19 displays the benchmark on the top tax data with a wealth threshold at 0.7 million. Again, similar to the *full sample* (t) copula, the *high income - high wealth* case have the predictive power superior to the marginal approaches under plausible lower bound assumptions.[35]

[34] As noted before, the first two cases use the unit square as margins since the inputs for copula estimation are the marginal CDFs. The last case is plotted on the support spaces of income and wealth distributions since they are directly the inputs for (IFM) estimation.

[35] At about 40 – 50 million of the income lower bound, the copula-based extrapolation in the *high income - high wealth* case can already approximate the top rich list well. Note the fit from this case looks better than the one under roughly optimal income lower bound for the *full sample* case in Figure 19. Again in Table A2 from Bach et al. (2013), in 2005, the threshold to access the top 0.0001% (45 persons) is 57 million.

Figure 12: Bivariate Kolmogorov-Smirnov (KS) statistics of copula estimates at grids on the unit square with empirical weighted marginal cumulative probabilities of income and wealth in the *full sample*: (a) 3D perspective, (b) contour



(a) (b)

Note: 1. Unit square is composed of the empirical ranks of income and wealth in the whole PHF data considering weight. There are 10,000 grids defined at the intersections of 100 (weighted) percentiles of each dimension. 2. The color level appears in both plots with the scale shown in the right bar. 3. We cut off some outliers only in the 3D perspective plot in (a) to avoid the visual distortion. Therefore, the maximum color scale in (a) is smaller than that in (b).

Figure 13: Contour curves of fitted and empirical copula over marginal CDFs of income and wealth – *best fitting* case (income cutoff = 5,300, wealth cutoff = 29,500, copula family, 1st and 2nd parameters – BB8, 1.62, 0.97); joint CDFs are marked above each contour curves

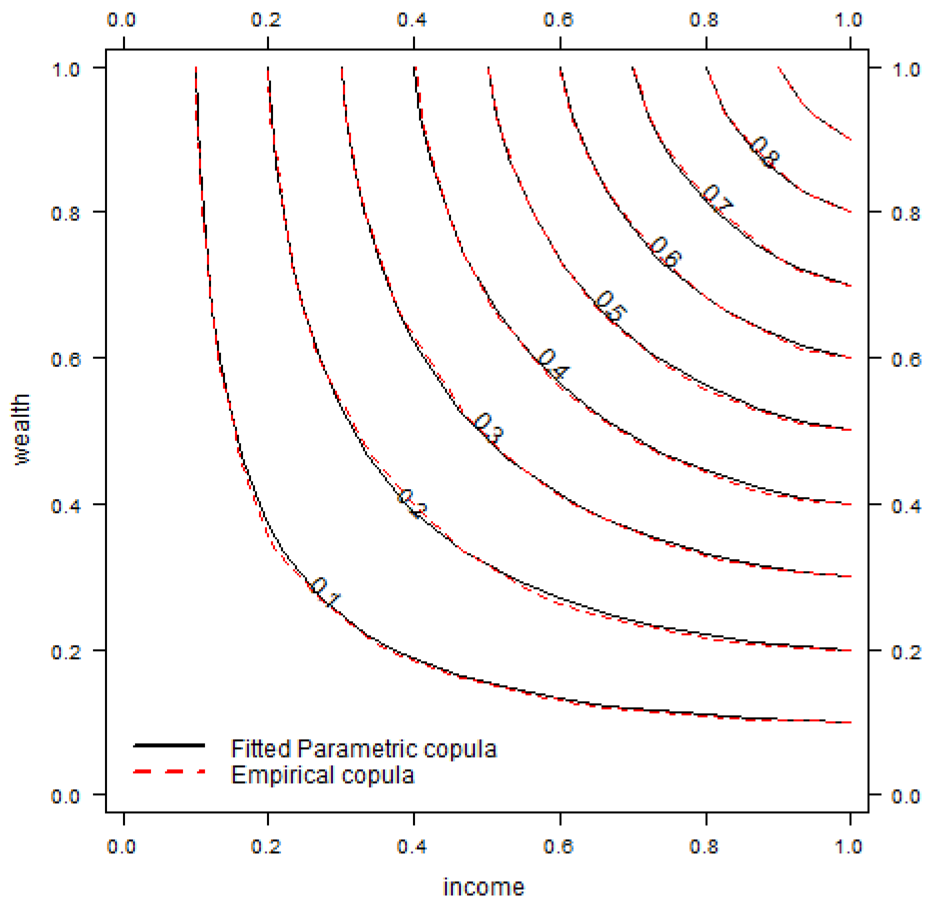


Figure 14: Contour curves of fitted and empirical copula over marginal CDFs of income and wealth – *full sample* case (copula family, 1st and 2nd parameters – t, 0.589, 6.2); joint CDFs are marked above each contour curves

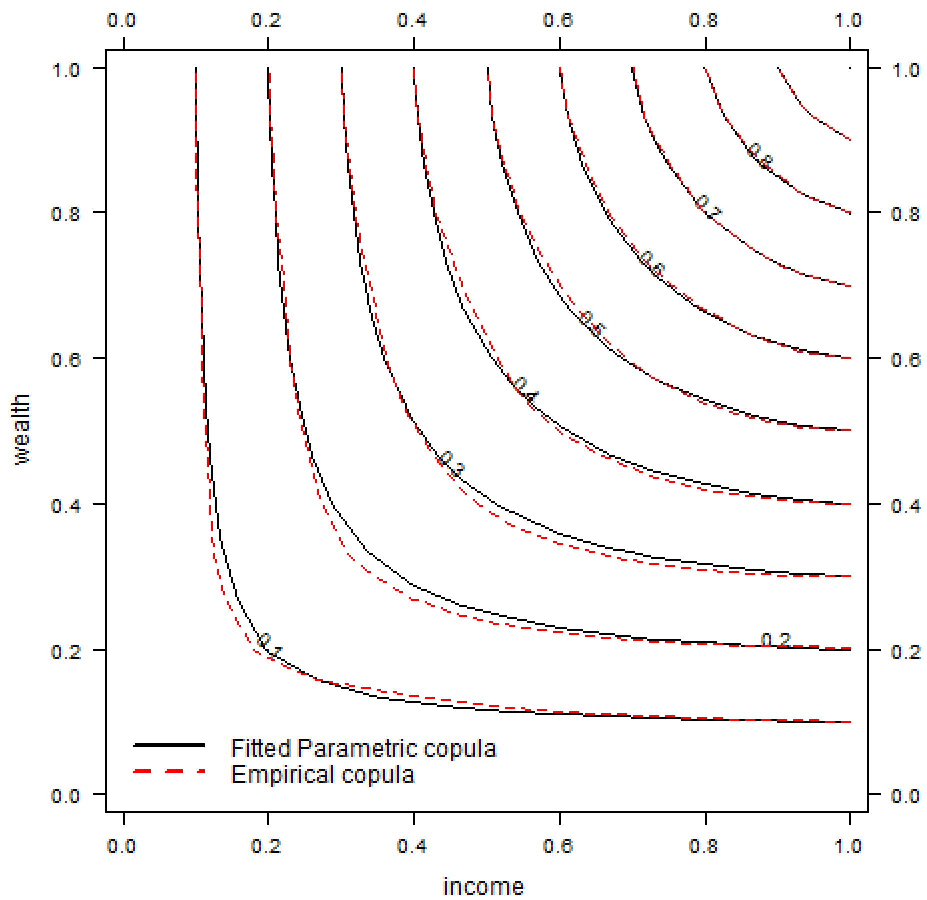


Figure 15: Contour curves of fitted and empirical copula over margins of income and wealth (unit: 1,000 euros; log-scaled axes) – *High income - high wealth* case (income cutoff = 86,957, wealth cutoff = 245,160, copula family, 1st and 2nd parameters – t, 0.275, 30) ; joint CDFs are marked above each contour curves

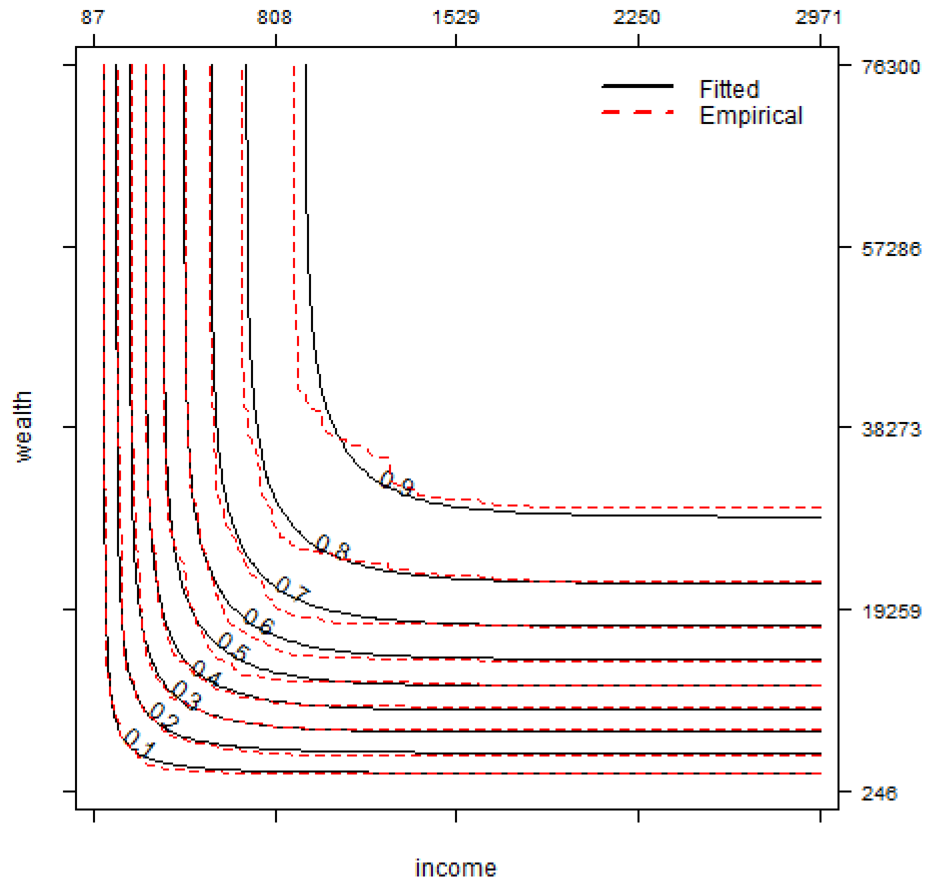
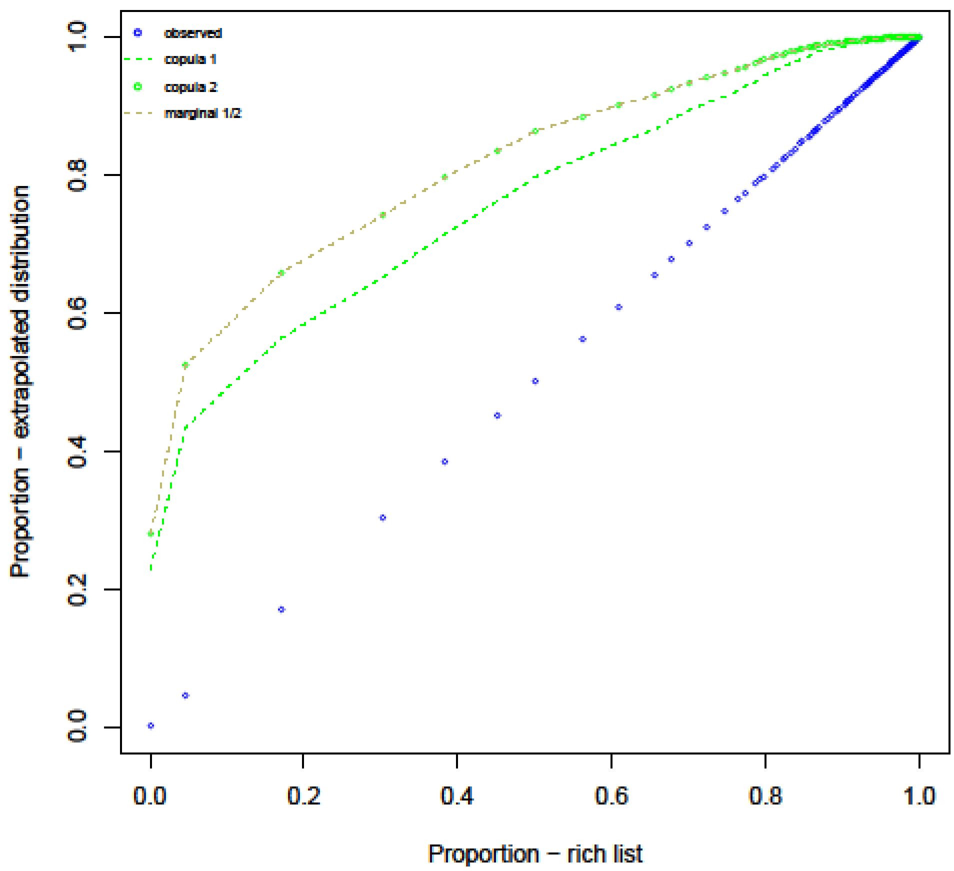
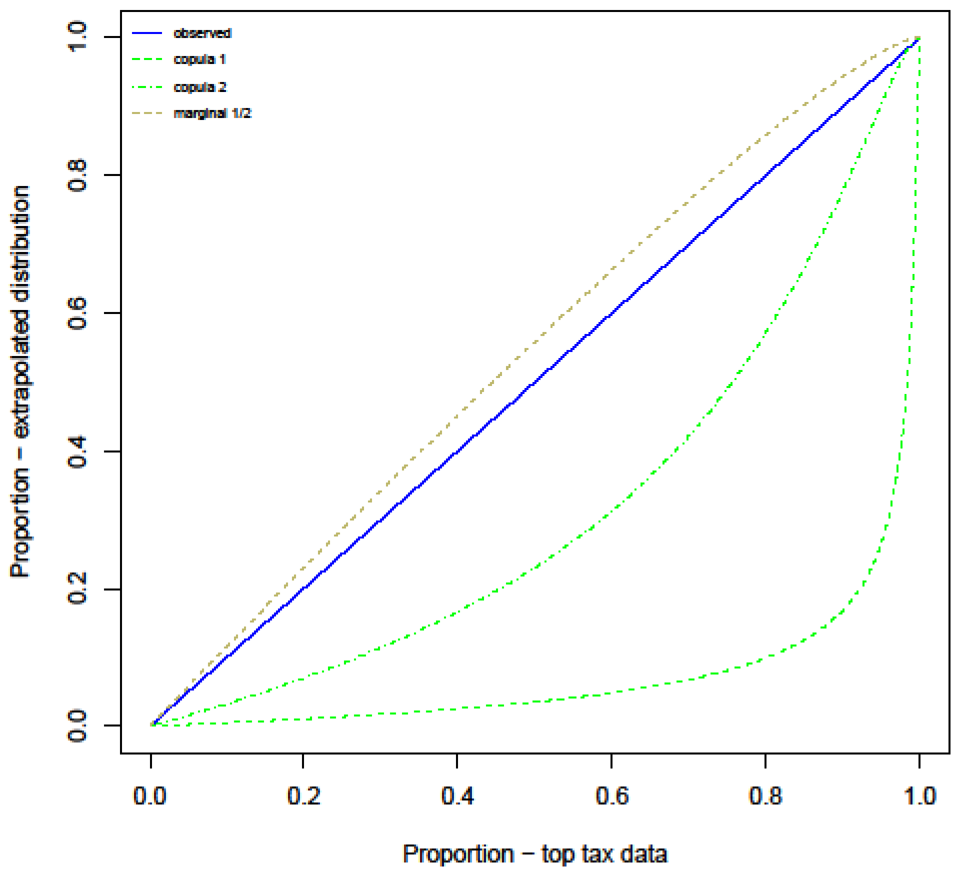


Figure 16: P-p plot benchmarking marginal and copula-based extrapolated distributions with top rich list – comparing *full sample* (t) and *best fitting* (BB8) copulas under naïve lower bound



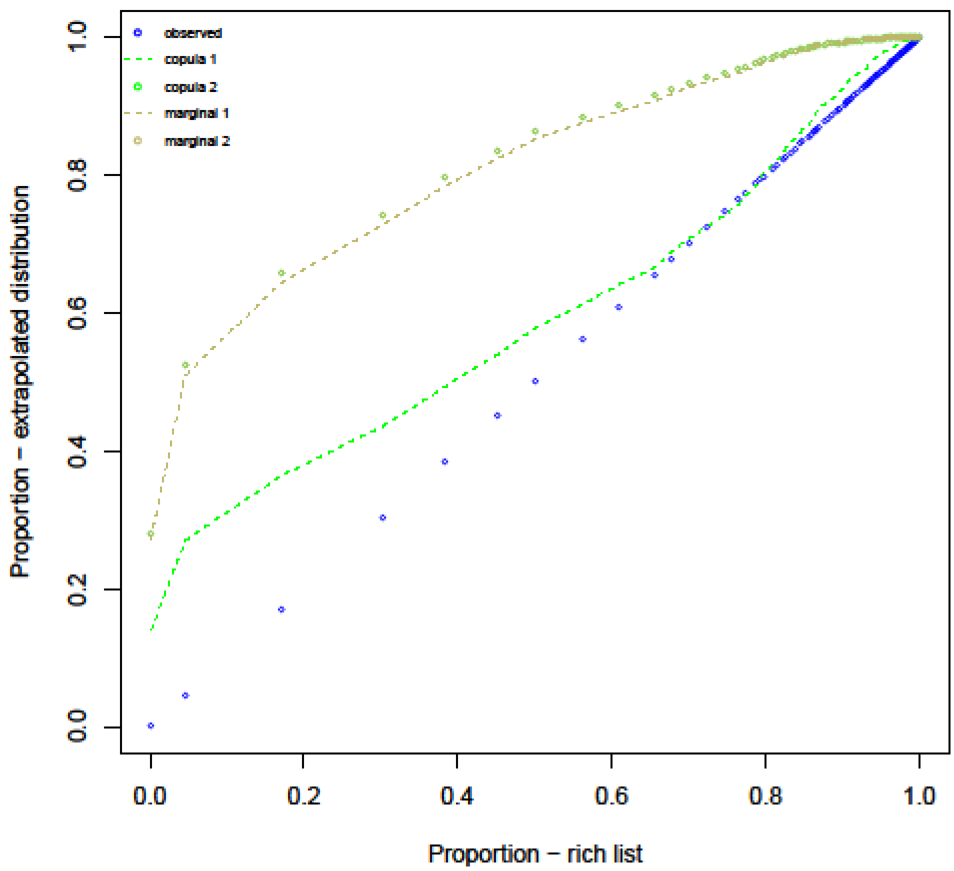
1 – *full sample* case and 2 – *best fitting* case
marginal 1/2 – wealth distribution - Pareto II (threshold=245160, shape=.6, scale=197203)
copula 1: - *Student t (parameter 1=0.589, parameter 2=6.2)*, conditioning income lower bound=2,970,700 and income distribution - Pareto II (threshold=86957, shape=.41, scale=34612)
copula 2: - *BB8 (parameter 1=1.62, parameter 2=0.97)*, conditioning income lower bound=2,970,700 and income distribution - Pareto II (threshold=86957, shape=.41, scale=34612)
Dots in unconnected lines (eg observed) correspond to the exact observations which reflect the dispersion of the sample.

Figure 17: P-p plot benchmarking marginal and copula-based extrapolated distributions with top income tax data - comparing *full sample* (t) and *best fitting* (BB8) copulas under naïve lower bound



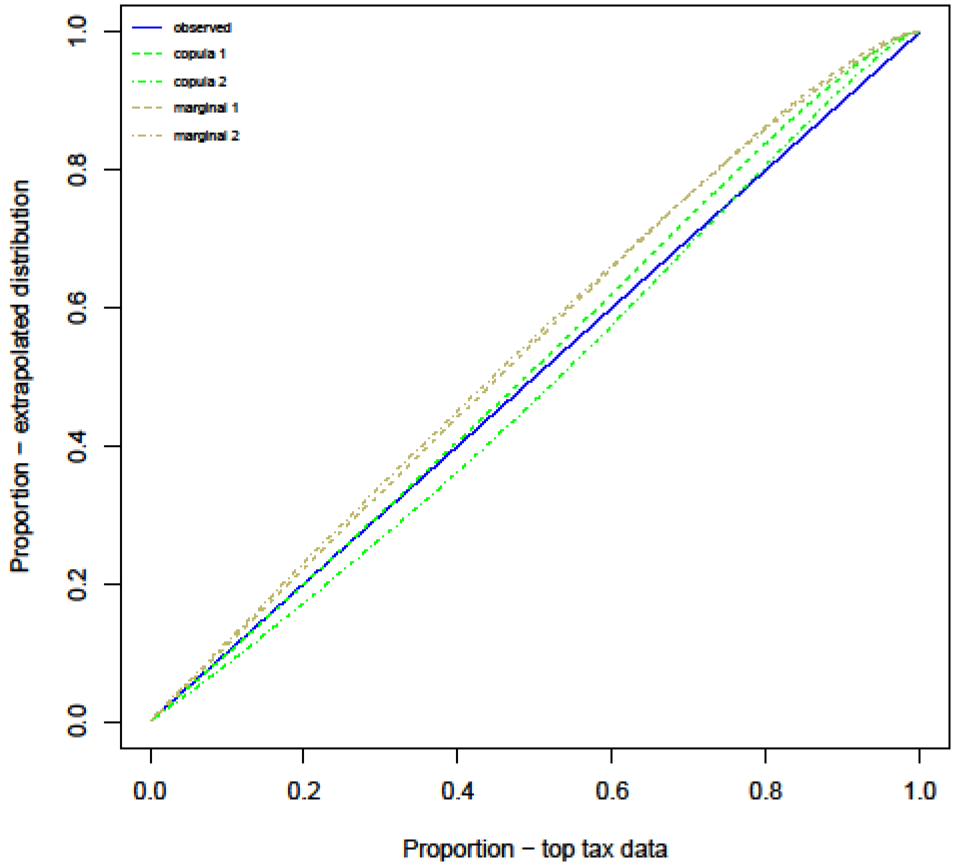
1 – *full sample* case and 2 – *best fitting* case
marginal 1/2 – income distribution - Pareto II (threshold=86957, shape=.41, scale=34612)
copula 1: - *Student t (parameter 1=0.589, parameter 2=6.2)*, conditioning wealth lower bound=76,300,000 and wealth distribution - Pareto II (threshold=245160, shape=.6, scale=197203)
copula 2: - *BB8 (parameter 1=1.62, parameter 2=0.97)*, conditioning wealth lower bound=76,300,000 and wealth distribution - Pareto II (threshold=245160, shape=.6, scale=197203).
It is not necessary to show the dispersion of this sample because they are 1,000 percentile points within top 1%.

Figure 18: P-p plot benchmarking marginal and copula-based extrapolated distributions with top rich list – comparing *high income-high wealth (t)* and *best fitting (BB8)* copulas under roughly optimal conditioning income



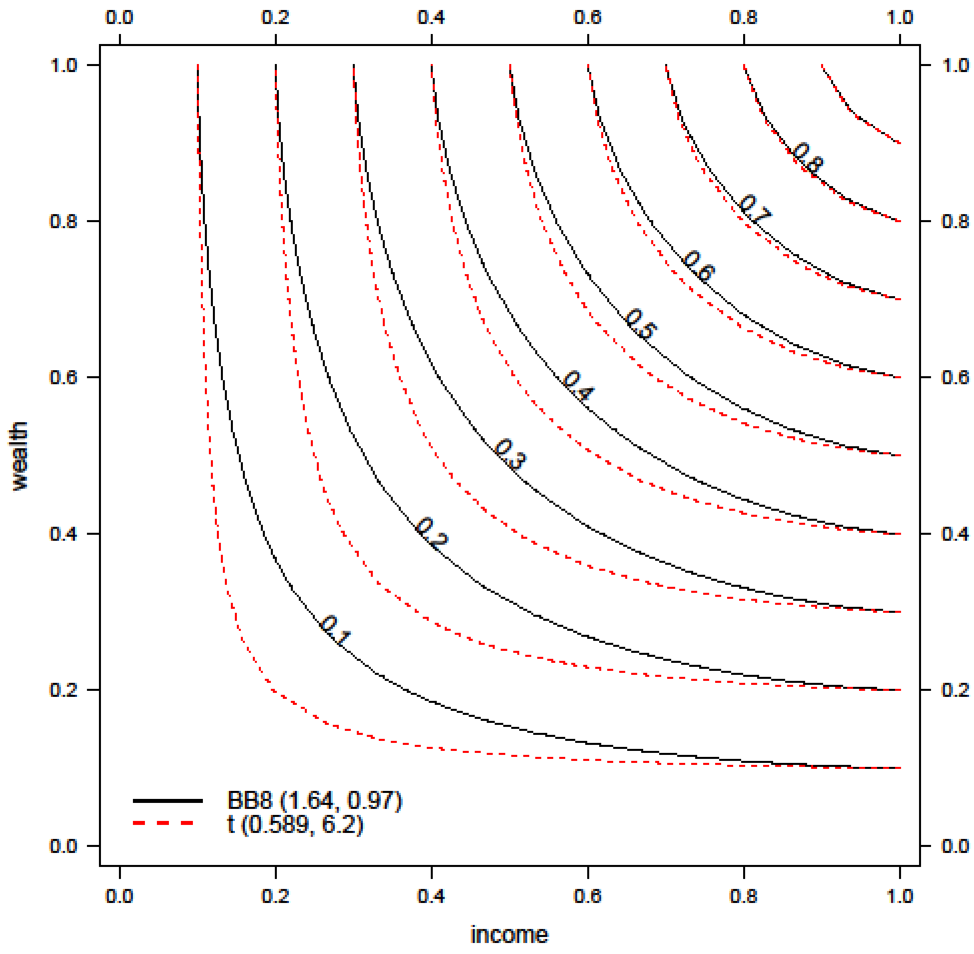
1 – *high income-high wealth case* and 2 – *best fitting* case
marginal 1 – wealth distribution - Pareto II (*threshold=245680, shape=.63, scale=370165*)
marginal 2 – wealth distribution - Pareto II (*threshold=245160, shape=.6, scale=197203*)
copula 1: - *Student t* (*parameter 1=0.275, parameter 2=30*), conditioning income lower bound=60,000,000 and income distribution - Pareto II (*threshold=87000, shape=.34, scale=45208*)
copula 2: - *BB8* (*parameter 1=1.62, parameter 2=0.97*), conditioning income lower bound=60,000,000 and income distribution - Pareto II (*threshold=86957, shape=.41, scale=34612*)
Dots in unconnected lines (eg observed) correspond to the exact observations which reflect the dispersion of the sample.

Figure 19: P-p plot benchmarking marginal and copula-based extrapolated distributions with top income tax data – comparing *high income-high wealth (t)* and *best fitting (BB8)* copulas under roughly optimal conditioning wealth



1 – *high income-high wealth case* and 2 – *best fitting case*
marginal 1 – income distribution - Pareto II (*threshold=87000, shape=.34, scale=45208*)
marginal 2 – income distribution - Pareto II (*threshold=86957, shape=.41, scale=34612*)
copula 1: - *Student t (parameter 1=0.275, parameter 2=30)*, conditioning wealth lower bound=700,000 and wealth distribution - Pareto II (*threshold=245680, shape=.63, scale=370165*)
copula 2: - *BB8 (parameter 1=1.62, parameter 2=0.97)*, conditioning wealth lower bound=700,000 and wealth distribution - Pareto II (*threshold=245160, shape=.6, scale=197203*)
It is not necessary to show the dispersion of this sample because they are 1,000 percentile points within top 1%.

Figure 20: Contours for BB8 (1.62, 0.97) and t (0.589, 6.2)



Diskussionsbeiträge - Fachbereich Wirtschaftswissenschaft - Freie Universität Berlin
Discussion Paper - School of Business & Economics - Freie Universität Berlin

2021 erschienen:

2021/1	HÜGLE, Dominik: Higher education funding in Germany: a distributional lifetime perspective *Economics*
2021/2	ISHAK, Phoebe W.: Murder Nature: Weather and Violent Crime in Brazil *Economics*