A Joint Top Income and Wealth Distribution

Junyi Zhu

(Deutsche Bundesbank)

Viktor Steiner

(Free University, Berlin)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 4D: Distributional Diversity in the National Accounts

WEDNESDAY, AUGUST 22, AFTERNOON

A Joint Top Income and Wealth Distribution[1]

Viktor Steiner
Free University of Berlin

Junyi Zhu
Deutsche Bundesbank

Preliminary draft – 2018.08.09

Abstract

We use the marginal and bivariate parametric distributions to extrapolate both income and wealth distributions from German PHF data. The methodology developed recently in estimating the Pareto distribution incorporating top tail information is adopted in extrapolating the marginal distribution. We then fit a copula model for the top joint distribution observed in PHF. This fully parametric model can perform out-of-sample prediction on the very top of the conditional tail distribution. Results from both marginal and joint extrapolation approaches are compared against the top rich list from the Manager Magazin for the wealth and fine top distribution from the administrative income tax data. There are evidences that our copula-based extrapolation outperforms the one using marginal distribution only.

1 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

There are under-coverage problems among many survey data for both income and wealth distribution. Respondents either under-report the figures or the survey designer has a top-coding of these responses. Alternatively, to well sample the top distributions is a difficult job: the survey administrators often do not have the reliable information on how to access these respondents and the true statistical characteristics of top distributions, responses can only sparsely cover the top given the nature of long tail and covering the very end of the distribution may not be tasked by design (Jenkins, 2016). Researchers tackle this issue by extrapolation via parametric distributions (eg. Pareto family). The literature in fitting top distribution has evolved to incorporate refined restrictions and top tail information (e.g rich list; see Eckerstorfer, Halak, Kapeller, Schütz, Springholz and Wildauer (2015) and Dalitz (2016)). They propose to estimate the Pareto parameters using goodness-of-fit criteria as well as requiring the continuity of Pareto distribution with observed density function. Jenkins (2016) discusses the comparison between type I and II Pareto models and how to search the optimal threshold in estimating Pareto parameters.

In this paper we further introduce the information from the top joint distribution between income and wealth available through survey data. By assuming that the under-covered top bi-dimensional distribution follows the same structure (copula) as those observed in the data, we extrapolate not only the marginal distributions but also the joint one. There are two folds of contributions. First, this approach provides an alternative method to extrapolate the marginal distributions using enlarged information set. For instance, if we are certain about some characteristics of the income distribution for the top tail of wealth distribution, we can construct this wealth distribution by conditional on the income information given the joint distribution. Second, to examine the joint distribution can allow us to answer a broader set of economic questions. Due to the growing top shares of income and wealth, for instance, it becomes more imperative that the public can comprehensively evaluate the individual tax burden (distribution) over the whole distribution when imposing both income and wealth tax.

On the other hand, there is top tail information available: rich list for the wealth and top distribution accessible via tax data (eg. top 1000 percentiles within top 1%). In the current stage, we focus on using them for external validation but not directly in the estimation procedure. This provides a cleaner ground to investigate the benefit of our joint approach. We will perform both marginal and joint fitting. Then two extrapolations from each will be benchmarked with this external information.

It is reasonable to assume the very top tail of the wealth holders (ie. those in the rich list) have their income well above the maximum observed in the survey data since the gap of the minimum wealth in the rich list and the maximum wealth in the survey is also huge (see, eg. Dalitz (2016)). After fitting the top bi-dimensional distribution both marginally and jointly, we parametrically calculate the top wealth distribution conditional on their income being larger than the maximum in survey data. We then construct the density using the wealth values observed in the rich list and this estimated top (conditional distribution). The same density is built by only using marginal extrapolation. There are evidences that our density estimated from the conditional distribution outperform the one from estimated marginal distribution in comparing with the true density observed in the rich list.

In the next steps, we will perform the same validation exercise on the top income distribution. Following Jenkins (2016), Eckerstorfer et al. (2015) and Dalitz (2016), we can extend the one dimensional goodness-of-fit criteria to a two dimensional one in order to pin down the optimal thresholds in fitting joint distribution. Besides the thresholds, we have the flexibility in searching within a range of parametric setting beyond the parameter values themselves: marginal distributions (Pareto type I vs II) and copula class. As discussed, by combining the estimated top distribution and observed one from micro data, we can further explore the application on inequality evaluation as well as public finance in the future.

In section 2 we present our data as well as external information on top tails. Section 3 outlines the implementation of estimating both top marginal and joint distributions from the survey data. Section 4 presents the external validity check. Section 5 concludes the paper.

2. Data

Our estimation comes from the Panel on Household Finance (PHF), the German component of Household and Consumption Survey (HFCS). We pool two waves of income and wealth data together in order to expand the sample to achieve better goodness-of-fit for the top. They refer to the years of 2009 and 2013. 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 Manger Magazin. Dalitz (2016), Vermeulen (2014) and Bach, Thiemann and Zucco (2015) have described both data in detail. The top income distribution is retrieved from the administrative tax data available from the Research Data Center of the Federal Statistical Office of Germany (Bach, Corneo and Steiner, 2012). 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.

3. Estimating top marginal and joint distributions

We fit both Pareto I and II distributions on both top income and wealth distributions observed in the PHF pooled data.

3.1 Top marginal distributions

As claimed by Jenkins (2016) using UK income data, Pareto II (Generalized Pareto) outperforms type I distribution in terms of goodness of fit. 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 II model has more variability across the years. The optimal threshold estimated for type II model can be much lower than that for type I model. This feature is particularly attractive for our application. It would be preferred to have larger training sample when fitting two dimensional model than single dimension one.

Furthermore, Atkinson (2017) shows Pareto coefficient in type I model might not be constant over the top distribution even for a cross section using UK historical data (see the 'baronial' and 'regal' regimes). Blanchet, Fournier and Piketty (2017) follow Atkinson (2017) and show that the inverted Pareto coefficient converges upward when the rank of income distribution is close to one using US and France data. The German inverted Pareto curve is simply Figure 1 (ie. Fig 2 in Bach et al. (2012)). They provide the evidence that typically inverted Pareto coefficient is U shaped and converges upward when income grows to the very right end. For type I model, inverted Pareto coefficient has a one-one mapping to the Pareto coefficient. Thus, its inverted Pareto curve is flat.

Based on these arguments, we mainly present the estimation using type II model following the implementation in Jenkins (2016). The software used is the EXTREME STATA module by Roodman (2015). For the purpose of completeness in the external validation, we also perform the estimation using type I model. The estimation is carried out using maximum likelihood after choosing the optimal threshold based on the goodness of fit between model estimate and empirical distribution as done in Eckerstorfer et al. (2015) and Dalitz (2016).[2]

We adopt the "more objective" approach to determine the optimal threshold: measuring distance between estimated parametric and observed distributions using the Kolmogorov-Smirnov (KS) statistic, i.e. the maximum distance between their cumulative distribution functions. The optimal one is supposed to minimize this distance. Alternatively, an intuitive approach is to plot the estimated parameters against thresholds and pick the one above which the estimate is flat. Besides the threshold where the model starts to apply, Pareto II model has two parameters whose estimates are theoretically constant after a minimum of thresholds: shape as *xi* and (modified) scale as *sigma*xi – threshold*.

Figure 2 plots the KS measure for the income observed for PHF between 35,000 and 290,000 euros, which is far below p90 and somehow larger than p99.5 (common searching range across many countries). The minimum is reached at somewhere below or at p90. Figure 3 provides the estimate curves against threshold. It justifies our choice by showing that the above area is most stable. The exact minimizer of KS is 70,271 euro. However, the inverted Pareto coefficient *b* based on the estimates using this threshold converges downward as rank *p* approaches to 100%.[3] 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.[4]

Figure 4 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 somehow larger than p99.5. The minimum is somewhere between the half of p90 and wealth a bit higher than p90. Figure 5 provides the estimate curves against threshold. For the modified sigma, this area does look most stable. For xi, the most stable area lies between p90 and p99. But the estimate is close to that obtained when threshold is at about the half of p90, ie. about .6. Given these evidences, we decide to take the exact minimizer 245,160 euro which can offer us more sample size in considering fitting a top joint distribution.

3.2 Top joint distributions

One advantage of copula modelling is that we can separately estimate the marginal and joint (copula) distributions. The previous section has actually already achieved the

[2] Only estimates will be presented in the section on external validation. We do not elaborate the exact procedure here since it follows the other two papers.
[3] Blanchet et al. (2017) provides the formula for *b(p)*.
[4] 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.

marginal part. Then, to fit a copula, we only need to know the empirical cumulative distributions (ecdf) from either dimension as well as a joint one about these ecdf's. It boils down to the same problem marginal fitting has to confront with in the first place: how to select the optimal threshold in both dimensions to cut off a top sample? Generally, we can postulate the similar approaches: minimizing a bi-dimensional distance measure and checking the stability of copula (both structure and parameters) over a range of thresholds. Currently, we pick up the thresholds in an ad-hoc way as we want to focus on the comparison between the marginal and joint (conditional probability) extrapolation.

After determining the thresholds, we use BiCopSelect() from the VineCopula package of R to choose the best copula using weight. It fits all available copulas via maximum likelihood estimation. Then the Akaike Information Criteria are computed for all of them. The one with the minimum value is selected.[5] To examine the property and goodness of fit, commands from the Copula package of R are implemented. Using the fitted copula, a sample with two marginal ranks (ie. the two variates of copula) is randomly generated.[6] We can then compare the Spearman's rho from this sample and the empirical one. And the contour curves over two marginal ranks are drawn for both samples and overlaid for a contrast. For the description of each R packages, we suggest to read Schepsmeier et al. (2017) and Hofert et al. (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).

This section presents three cases of copula fitting which are used for the validation exercises in section 4. The marginal distributions of both dimensions are fitted either by type I or II Pareto distributions.[7] We simply construct the fitting sample for copula estimation by selecting all the observations with income above the same income threshold as the optimal one in the marginal fitting discussed in the section 3.1. Among them, we drop those with negative wealth (ie. in debt) which are ruled out by the specification of Pareto model. The size of those fully indebted is negligible in all cases.[8] We decide to not further restrict the sample by the optimal wealth threshold for two reasons: first, the size of fitting sample can be very small; second, imposing such a

[5] The package also allows using Bayesian Information Criteria when the models with higher number of parameters are to be penalized.

[6] Sample size is the sum of (observation) weight in the top cut-off sample.

[7] We present two cases of Pareto I fitting because the most favoured one estimated for income distribution using purely PHF data cannot match with the characteristics of very top tail (eg. shape parameter of Pareto I model) known in the literature according to the administrative tax data. The other selected one by combining both PHF and ten decile points within top one percent retrieved from tax data is also chosen since its estimates are close to those of the empirical top tail.

[8] About 0.1 or 0.01% of sample total population.

wealth threshold in estimating a top joint distribution and performing extrapolation would be equivalent to assume the correlation between top income and wealth distributions is high. This is actually invalid in reality. One piece of evidence presented below is that there is a considerable size of subsample within the fitting sample with wealth significantly smaller than the optimal marginal fitting threshold for wealth.

Table 1 collects estimation in both copula and marginal fitting, characteristics of copula fitting sample and goodness of fit (Spearman's rho comparison) for three cases. Case 1 uses type II margins and case 2 and 3 use type I ones. Case 2 is our alternative best type I estimate while using some external information in fitting: the shape parameter in fitting the income type I distribution is 1.69 which is close to those estimated using top tax data (see Bach et al. (2012)). [9] Case 3 has the optimally chose type I model for income but the shape parameter being 2.49 unmatchable with those true top estimate.

We only go over the characteristics of copula fitting sample and goodness of fit. As discussed, the proportion with wealth well below the wealth threshold estimate for all three cases is considerable: ranging from 0.35 to 0.86.[10] Both cases 1 and 3 have large enough fitting sample size (ie. 10% of the total population) which facilitates a better fitting – eg. the Spearman's rho from both simulated and fitting sample is almost same. This is not the situation for the case 2. There is a significant gap between two Spearman's rho's (0.15 vs 0.02) because the size of fitting sample is far too small: it is only 0.3% of the population (actually it contains only 80 observations).

Figure 6, Figure 7and Figure 8 present the surface of the joint density over the two margins of income and wealth based on the estimated parametric copula from case 1 to 3. They illustrate the distributional nature of the corresponding copulas. We will discuss the implications together with the results in validation exercise. Figure 9, Figure 10 and Figure 11 provide the contour plot of joint cumulative distribution from both estimated parametric copula and empirical copula over these two margins. Both sets of contours almost overlay each other except in case 2 where the size of fitting sample is two small and thus the empirical contours cannot be smooth.

[9] See footnote 7 for the details of the external information.

[10] Among them, quite a few have wealth considerably below the optimal wealth threshold in marginal fitting. For instance, the median of the fitting sample can be only one quarter of the optimal threshold for case 3.

4. External validity

Can incorporating the joint structure between income and wealth enhance the extrapolation accuracy in either dimension? We illustrate this validation exercise discussed previously for three cases of estimates. Namely, densities are drawn according to empirical distribution in the rich list, estimated marginal (Pareto) wealth distribution and probability of wealth conditional on the income being larger than the maximum observed in PHF derived using estimated copula (with the same marginal wealth distribution as the marginal extrapolation) respectively.[11] To avoid the influence from the choice of sampling and kernel density construction, they are calculated on all the same wealth values observed in the rich list and bandwidth selection method is "SJ" for all in the R function density (Sheather and Jones, 1991).

Figure 12, Figure 13 and Figure 14 display these density comparisons for three cases. Interestingly, all the copula-based (conditional probability) densities are closer to the empirical density although the extent can vary. Note that the whole estimation relies on almost none of the external information (ie. rich list or top tail from income tax data).[12]

As observed from Figure 6, Figure 7and Figure 8, besides a mass of both income and wealth rich, there is also a mass of income and wealth (relatively) poor in the top. Copula extrapolation incorporates this piece of information so that the estimated conditional density has thinner tail and more mass on the relatively poor among the top rich which is a closer portrait of the observed rich list.

5. Conclusion

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 – rich list, the conditional probability derived using fitted copula performs better than marginal extrapolation

[11] Let the copula be $C(u,v)$, where u and v are the CDFs of income and wealth respectively. Then the conditional probability based on copula $P(w|y>y_{\max}) = \frac{P(w,y>y_{\max})}{P(y>y_{\max})} = \frac{P(w)P(y>y_{\max}|w)}{P(y>y_{\max})} = \frac{P(w)P(U>u'|V=v')}{P(U>u')} = \frac{P(w)}{1-u'}\int_{u'}^{1} c(x,v')dx = \frac{P(w)}{1-u'}\left[\frac{\partial}{\partial v}C(1,v') - \frac{\partial}{\partial v}C(u',v')\right] = P(w)\frac{1-\frac{\partial}{\partial v}C(u',v')}{1-u'}$, where y is income, w is wealth, $y_{\max}$ is the maximum income observed in PHF, $u' = P(y \leq y_{\max})$, $v' = P(W \leq w)$ and $c(u,v)$ is the copula density. u' and v' are calculated using the estimated marginal (Pareto) distributions. Then $P(w|y>y_{\max})$ can be fully derived analytically for each w in the rich list.

[12] The only exception is the fitting of Pareto distribution for income in case 2 which only uses 10 data points from top income tax data.

widely adopted in the literature among the estimation choices selected.[13] The same external validation exercise will be done on using the top income distribution accessible via the administrative tax data. In the next step, we will develop the more rigorous criteria in fitting the copula analogous to those applied to the marginal distribution fitting (ie. how to choose the optimal threshold in forming fitting sample).

[13] We do observe the situations when the copula approach does not prevail over the marginal approach (eg., when choosing some particular fitting samples).

References

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., 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.

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.

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.

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.

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

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

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 683-690.

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.

Figure 1 Inverted Pareto curve for the couples' income in the German tax data (2005), Fig 2 in Bach et al. (2012)

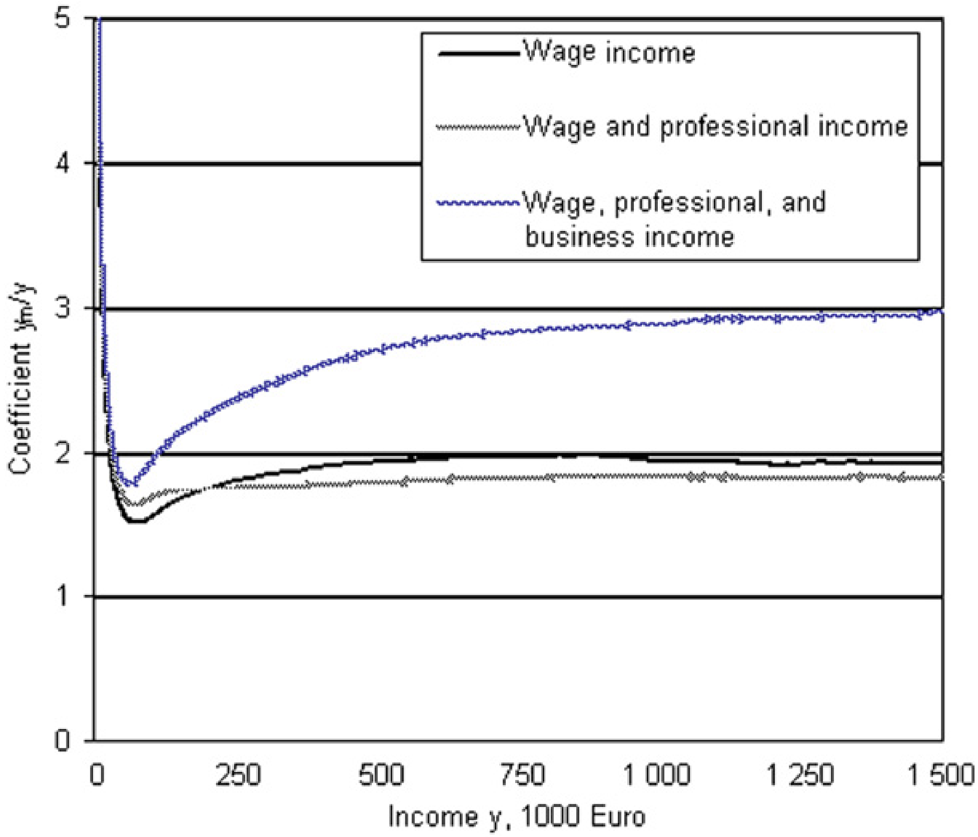


Figure 2 Goodness-of-fit criteria after Kolmogorov-

Smirnov (KS) as a function of threshold for the averaged PHF income data (dashed lines: p90, p95, p99, p99.5)

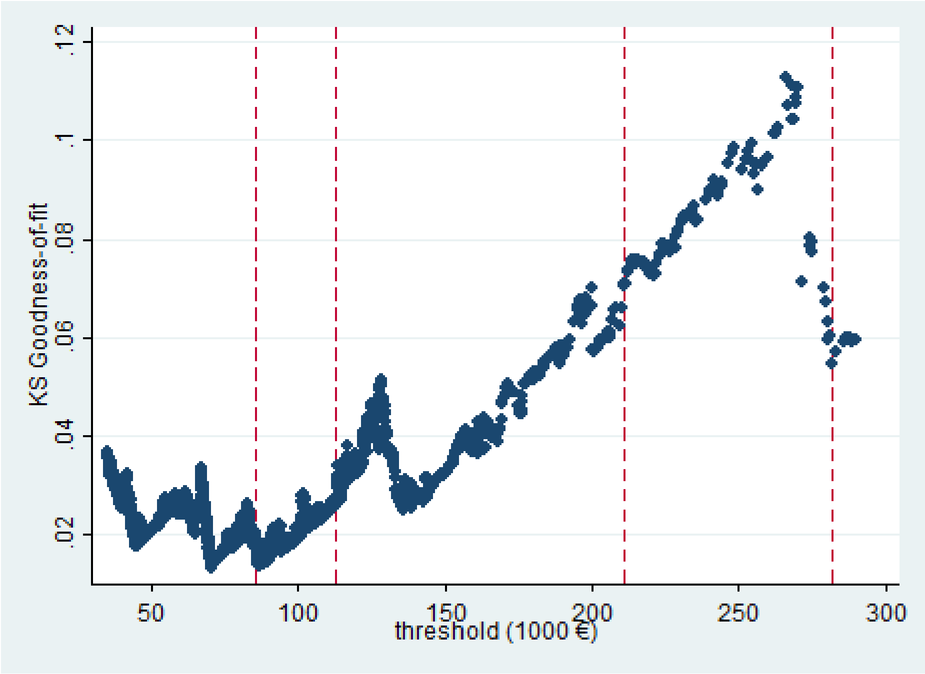


Figure 3 Pareto II parameter estimates by threshold for the averaged PHF income data (dashed lines: p90, p95, p99, p99.5)

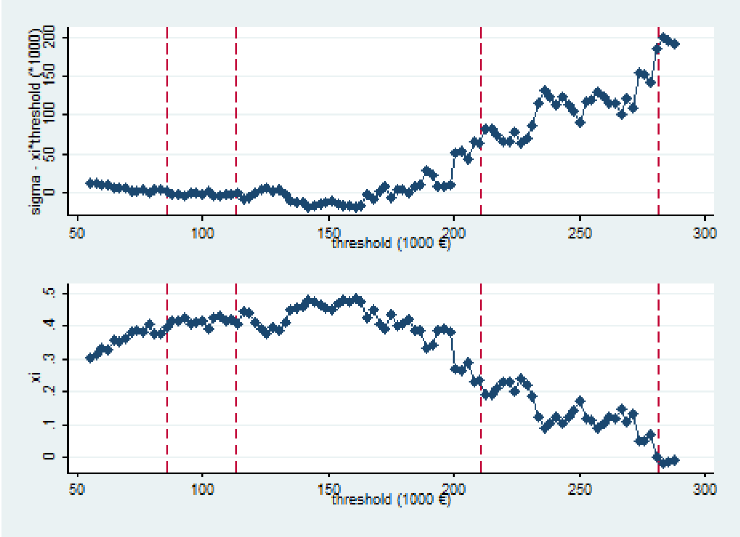


Figure 4 Goodness-of-fit criteria after Kolmogorov-Smirnov (KS) as a function of threshold for the averaged PHF wealth data (dashed lines: p90, p95, p99, p99.5)

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

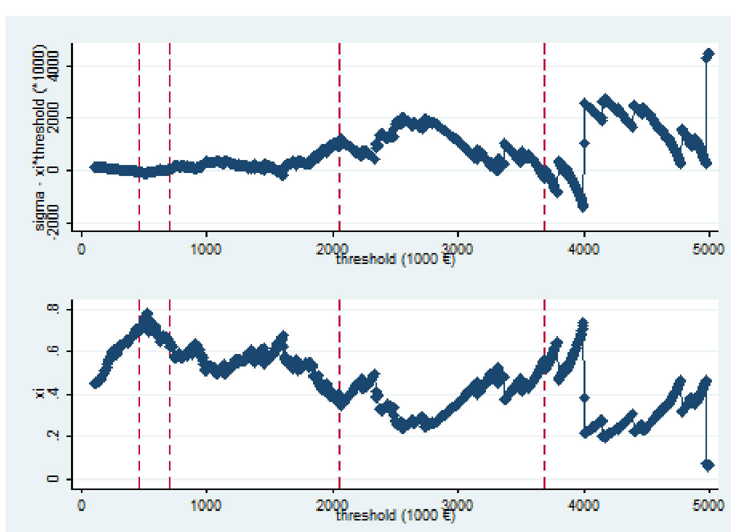


Table 1 Estimates in copula and marginal fitting, characteristics of copula fitting sample and goodness of fit (Spearman's rho comparision)

	Case	1	2	3
Copula fitting	income threshold in forming fitting sample (euro)	86,957	362,860	86,470
	wealth threshold in forming fitting sample	0	0	0
	copula family	Student - t	Student - t	Student - t
	1st parameter - copula	0.35	0.16	0.34
	2nd parameter - copula	8	30	8.18
	proportion with wealth smaller than the threshold of pareto model in the fitting sample	0.35	0.6	0.86
	proportion of fitting sample in the population	0.1	0.003	0.1
	Spearman's rho - copula simulation	0.33	0.15	0.32
	Spearman's rho - fitting sample	0.33	0.02	0.32
Marginal fitting	Pareto type - income	II	I	I
	threshold -income (euro)	86,957	362,860	86,470
	shape parameter - income	0.42	1.69	2.49
	scale parameter - income	34,241		
	Pareto type - wealth	II	I	I
	threshold - wealth (euro)	245,160	1,226,520	1,226,520
	shape parameter - wealth	0.6	1.53	1.53
	scale parameter - wealth	197,206		

Figure 6 Surface of the copula density over marginal CDFs of income and wealth – Case 1 (Student –t copula, 1^{st} parameter=0.35, 2^{nd} parameter=8)

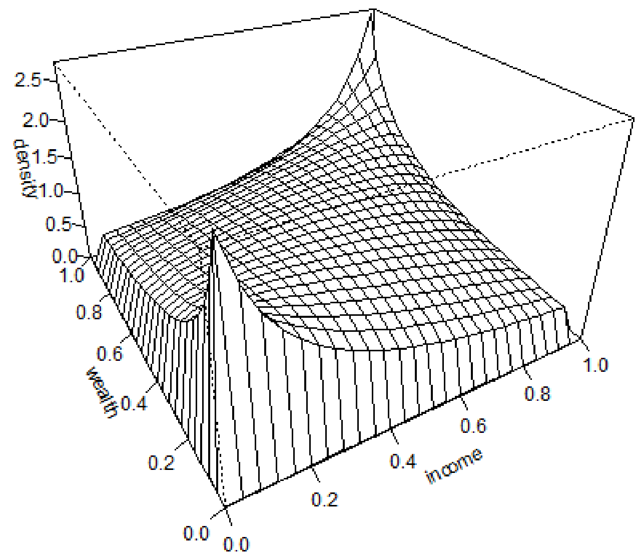


Figure 7 Surface of the copula density over marginal CDFs of income and wealth – Case 2 (Student –t copula, 1^{st} parameter=0.16, 2^{nd} parameter=30)

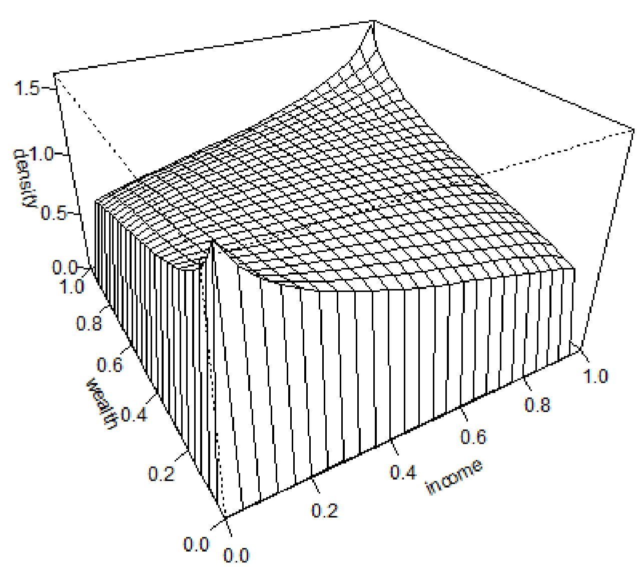


Figure 8 Surface of the copula density over marginal CDFs of income and wealth – Case 3 (Student –t copula, 1st parameter=0.34, 2nd parameter=8.18)

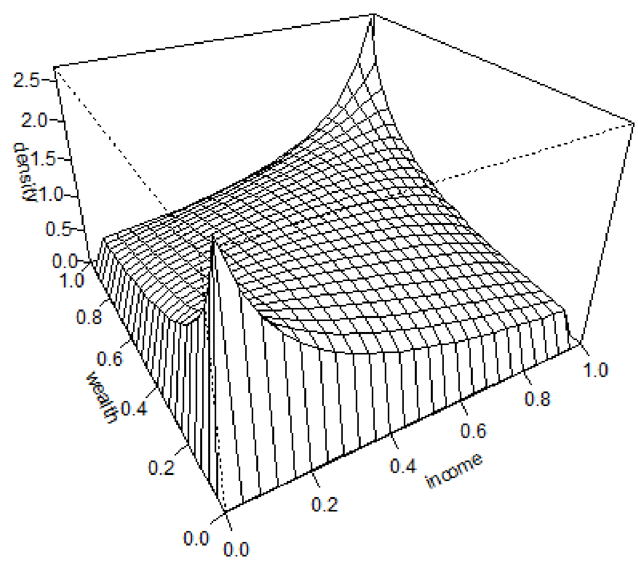


Figure 9 Contour curves of fitted and empirical copula over marginal CDFs of income and wealth – Case 1

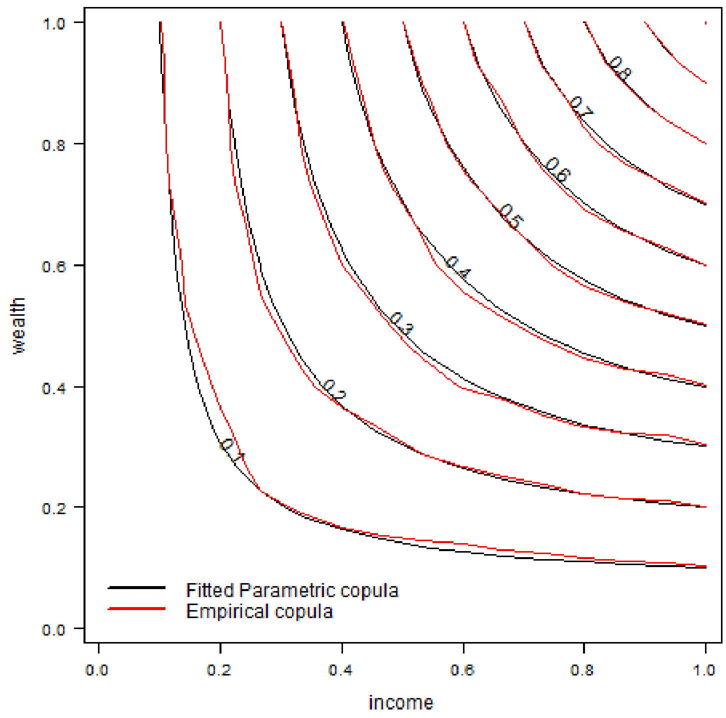


Figure 10 Contour curves of fitted and empirical copula over marginal CDFs of income and wealth – Case 2

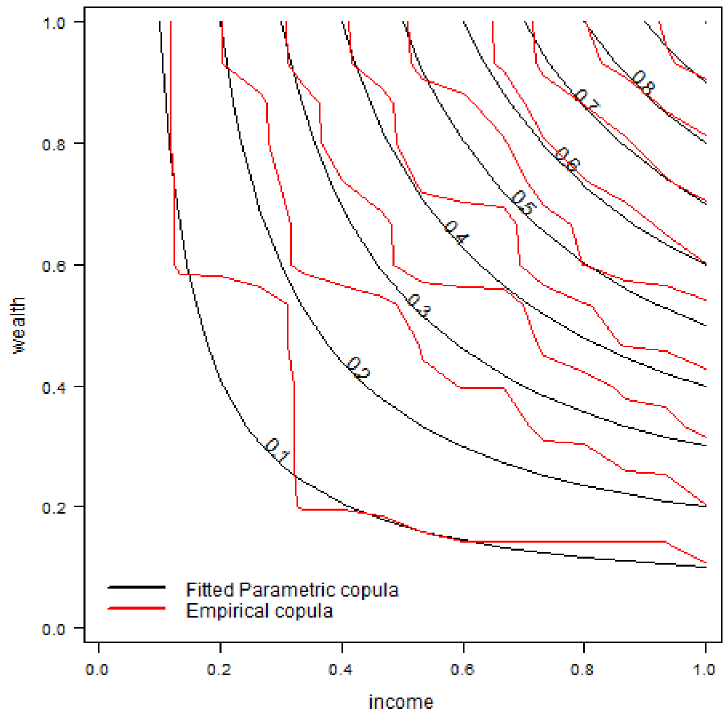


Figure 11 Contour curves of fitted and empirical copula over marginal CDFs of income and wealth – Case 3

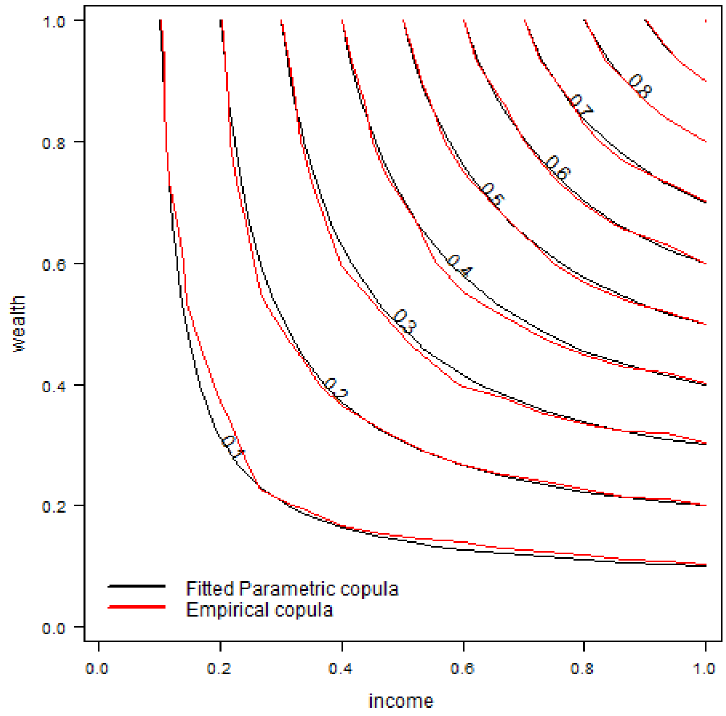
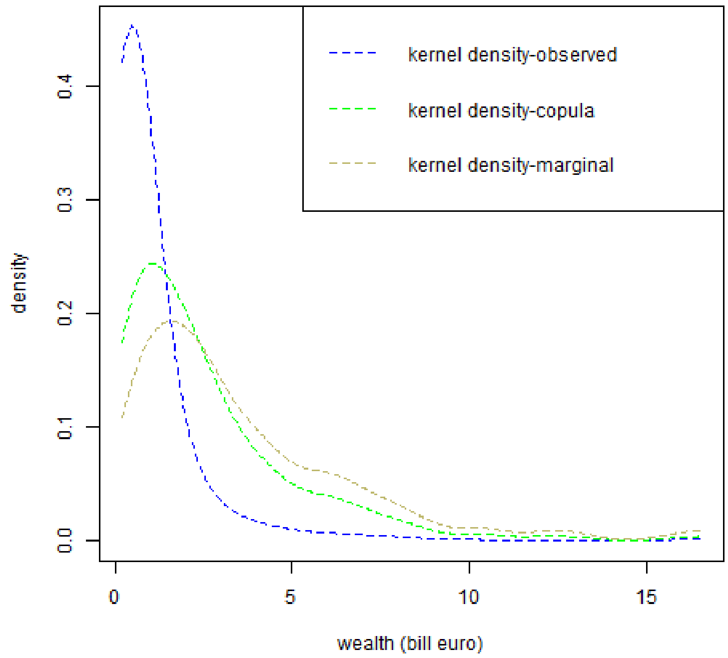
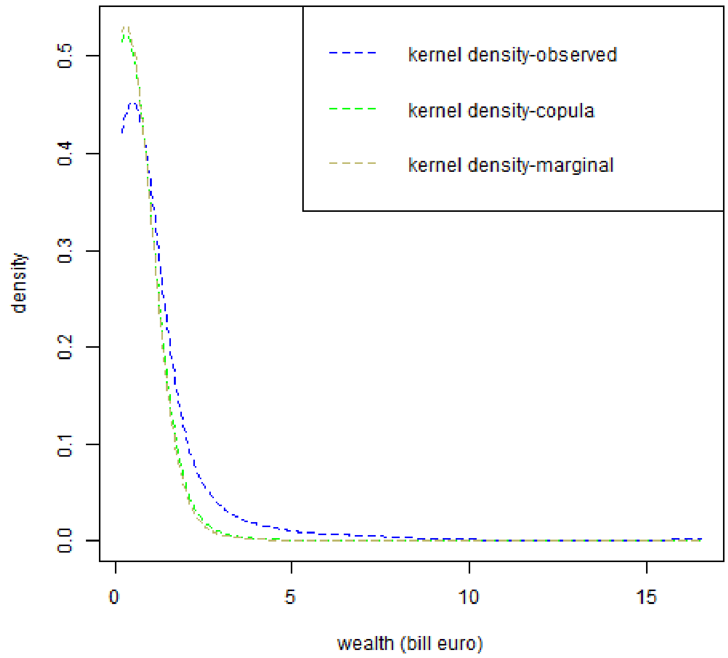


Figure 12 Kernel density from observed rich list, estimated marginal distribution and conditional distribution of estimated copula – Case 1



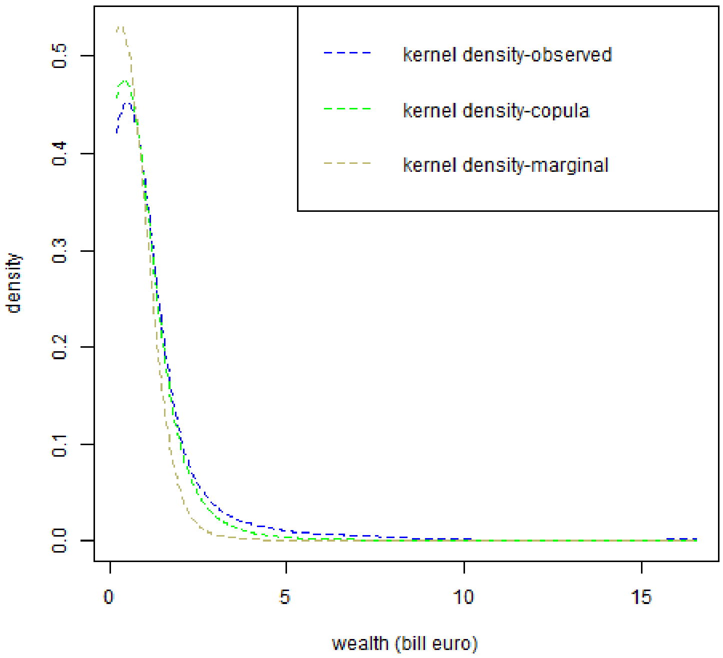
Note: conditional distribution of estimated copula is the probability of wealth conditional on income all above the observed maximum in PHF given the estimated copula

Figure 13 Kernel density from observed rich list, estimated marginal distribution and conditional distribution of estimated copula – Case 2



Note: conditional distribution of estimated copula is the probability of wealth conditional on income all above the observed maximum in PHF given the estimated copula

Figure 14 Kernel density from observed rich list, estimated marginal distribution and conditional distribution of estimated copula – Case 3



Note: conditional distribution of estimated copula is the probability of wealth conditional on income all above the observed maximum in PHF given the estimated copula