Session 6D: Income and Wealth: Theory and Practice II
Time: Thursday, August 9, 2012 PM

Paper Prepared for the 32nd General Conference of
The International Association for Research in Income and Wealth

Boston, USA, August 5-11, 2012

Modelling the Joint Distribution of Income and Wealth

Markus Jäntti, Eva Sierminska and Philippe Van Kerm

For additional information please contact:

Name: Philippe Van Kerm
Affiliation: Center CEPS/INSTEAD, Luxembourg

Email Address: philippe.vankerm@ceps.lu

Modelling the joint distribution of income and wealth*

Markus Jäntti
SOFI, University of Stockholm

Eva Sierminska
CEPS/INSTEAD, Luxembourg

Philippe Van Kerm
CEPS/INSTEAD, Luxembourg

July 2012

Paper prepared for comments and discussion
at the 32nd IARIW General Conference,
August 5–11 2012, Boston, USA.

Session 6D: Income and Wealth: Theory and Practice II
Organizers: Ruth Meier and Tom Priester (Federal Statistical Office, Switzerland)

*This research is part of the WealthPort project (*Household wealth portfolios in Luxembourg in a comparative perspective*) financed by the Luxembourg 'Fonds National de la Recherche' (contract C09/LM/04) and by core funding for CEPS/INSTEAD from the Ministry of Higher Education and Research of Luxembourg.

This paper considers a parametric model for the joint distribution of income and wealth. The model is used to analyze income and wealth inequality in five OECD countries using comparable household-level survey data. We focus on the dependence parameter between the two variables and study whether accounting for wealth and income jointly reveals a different pattern of social inequality than the traditional 'income only' approach.

Keywords: income; wealth; inequality; copula; multivariate Gini

1 Introduction

Inequality in living conditions within industrialized countries is almost always gauged on the basis of annual household income data, the distribution of which is typically summarized in coefficients such as the Gini index (Jenkins and Van Kerm, 2009). Far less is known about the distribution of other measures of economic well-being, such as consumption expenditure or wealth and asset holdings. The latter is of particular complementary interest as it captures long-term economic resources better than do income flows and represent resources that people are able to draw on to face adverse shocks. The sources of reliable household- or individual-level data on wealth, asset holdings and debt however remain limited.

While there are obvious links between income and net wealth accumulation via savings and borrowing constraints, the dependency between these two covariates cannot be summarized in a simple way. Relatively little is known empirically about this dependence, especially outside the United States (Kennickell, 2009), although some work has been conducted on the basis of the Luxembourg Wealth Study database (Sierminska et al., 2006, Jäntti et al., 2008). The relationship is mitigated by, e.g., wealth portfolio allocation choices, life-cycle effects, intergenerational transfers (inheritance), past income streams and their volatility, etc. It is not entirely clear –empirically and theoretically– if there is some trade-off between them or if they tend to be positively associated thereby reinforcing social inequality overall. Better knowledge about the joint distribution of income and wealth is also relevant for the design of taxation and redistribution policies as well as for better identification and targeting of vulnerable population groups.

Some measurement issues make wealth inequality and the joint distribution of income and wealth somewhat more challenging to examine than the standard analysis of income or consumption. These include the presence of a substantial fraction of negative values in most sample data (when debts exceed the value of assets), and the skewness and long tail of the distributions leading to extreme data. Negative net worth, in particular, makes some traditional measures of relative inequality inadequate (Jenkins and Jäntti, 2005). The presence of negative net worth also invalidates parametric statistical models typically used to describe income distributions, as virtually all such models are defined for positive data only (Dagum, 1990).

The objective of this paper is to elaborate methods to analyze inequality in the joint distribution of income and wealth, and illustrate this approach based on data for five OECD countries. We develop and fit a simple yet flexible parametric model for the bivariate distribution. The model handles specificities of the distribution of wealth, in particular the presence of zero and negative values. Specifically, our approach to studying the joint distribution of income and wealth is based on separate estimation of univariate marginal distribution models for income and for wealth, and estimation of parametric copula functions to capture the dependence between income and wealth (Genest and McKay, 1986, Trivedi and Zimmer, 2007). Combination of estimates for the marginal distributions and the copula provides flexible estimates of the joint distribution of income and wealth. This approach has the advantage of separating problems of estimation of marginal distributions and the dependence structure between the two variates.

Endowed with estimates of the joint distribution of income and wealth for five OECD countries, we compare the degree of dependence between the two variates in the different countries and construct various counterfactual distributions to estimate the implications of variations on the dependence parameter on a bi-dimensional version of the Gini coefficient proposed by Koshevoy and Mosler (1997). We find that cross-country variations in the association parameter effectively accounts only for a small fraction of cross-country differences in the bi-dimensional Gini. The index appears primarily driven by differences in inequality in the net worth distribution.

We describe the parametric model for the joint distribution of income and wealth in Section 2. Section 3 describes the datasets used and provides a preliminary inspection of the data. Estimation results and analysis of the bi-dimensional Gini coefficient are presented in Section 4. Section 5 concludes.

2 A parametric model for the joint distribution of income and wealth

As mentioned in the Introduction, our approach to modelling the bivariate distribution of income and wealth involves separate specification of models for the marginal distributions and for dependence parameter(s). This modelling strategy relies on Sklar's

(1959) theorem which shows that (continuous) multivariate distributions can be uniquely expressed as a function of a copula and univariate marginal distributions,

$$F(y, w) = C(F_Y(y), F_W(w)) \tag{1}$$

where F is the joint cumulative distribution function of income and wealth, F_Y and F_W denote the marginal cumulative distribution functions of income (Y) and wealth (W), and C is a copula. We model F parametrically by specifying separate models for each of F_Y, F_W and C.

We describe our specifications for each of the three components in turn.

The marginal distribution of income

Specifying a parametric model for the marginal distribution of income is unproblematic. A variety of specifications are available; see, for example, McDonald (1984). We follow common practice and rely on a Singh-Maddala specification (Singh and Maddala, 1976). The Singh-Maddala distribution is a three-parameter model for unimodal distributions allowing varying degrees of skewness and kurtosis and dealing with the heavy tails typical of income and earnings distributions. It is used, e.g., in Brachmann et al. (1996) and Biewen and Jenkins (2005) to model income distributions. The cumulative distribution function is

$$F_Y(y) = \mathrm{SM}(y; a, b, q) = 1 - \left[1 + \left(\frac{w}{b}\right)^a\right]^{-q} \tag{2}$$

where $b > 0$ is a scale parameter, $q > 0$ is a shape parameter for the upper tail, $a > 0$ is a shape parameter affecting both tails (Kleiber and Kotz, 2003).

The marginal distribution of wealth

Specifying the marginal distribution of wealth is more difficult. Although it is customary to analyze specific components of wealth such as assets or debts, the literature on wealth inequality most often focuses on the concept of *net worth*, defined as the value of total assets (financial and non-financial) minus total debts. Consequently, it is relatively common to observe data with zero or negative net worth. Our parametric model must therefore be able to accommodate negative data. This rules out virtually all specifications typically used for modeling income distributions, since these size distributions have positive density only over R_+.

To accommodate zero and negative data, we follow Dagum (1990) (also see Jenkins and Jäntti, 2005) and use a finite mixture model where negative, zero and positive data are modeled separately with an exponential distribution (negatives), a point-mass at zero and a Singh-Maddala distribution (positives):

$$F_W(w) = \begin{cases} \pi_1 \ \exp(\theta w) & \text{if } w < 0 \\ \pi_1 + \pi_2 & \text{if } w = 0 \\ \pi_1 + \pi_2 + (1 - \pi_1 - \pi_2)\,\mathrm{SM}(w; \alpha, \beta, \gamma) & \text{if } w > 0 \end{cases} \tag{3}$$

where π_1 and π_2 are the shares of negatives and zeros in the data, α, β and γ are interpreted as above and $\theta > 0$ is shape parameter for the negative distribution with lower values for $\theta > 0$ leading to thicker negative tail.[1]

Copula function specification

The third ingredient of the specification of the joint distribution is the shape of the copula C which captures the rank-order association between the two marginal distributions. In the absence of clear guidance from earlier research as to the most appropriate specification for the copula in this context, we experiment with four alternative specifications. See, e.g., Trivedi and Zimmer (2007) for a review of different copula functions.

Our first specification is a Plackett copula (Plackett, 1965):

$$C_P(u, v; \tau) = \frac{\left((1 + (\tau - 1)(u + v)) - \sqrt{(1 + (\tau - 1)(u + v))^2) - 4uv(\tau - 1)\tau}\right)}{2(\tau - 1)} \tag{4}$$

where $\tau \in [0, \infty \setminus \{1\}$ is a dependence parameter. $\tau > 1$ leads to positive dependence and the higher τ, the higher is the dependence. The Plackett copula exhibits symmetric upper-tail and lower-tail dependence. It is used by, e.g., Bonhomme and Robin (2009) in a model of earnings dynamics.

[1] Note that we depart from Dagum's (1990) original model by using a Singh-Maddala distribution for positive data whereas Dagum (1990) uses a Dagum type I distribution. The two distributions have similar shapes. We use the Singh-Maddala distribution for coherence with the income distribution model.

Our second specification is a Clayton copula

$$C_C(u, v; \tau) = \Big(u^{-\tau} + v^{-\tau} - 1\Big)^{-\frac{1}{\tau}} \tag{5}$$

where $\tau \in [-1, \infty \setminus \{0\}$ is the dependence parameter. $\tau > 0$ leads to positive dependence. The Clayton copula exhibits stronger lower tail dependence.

The third specification attempts to capture stronger upper tail dependence by rotating the Clayton copula:

$$C_R(u, v; \tau) = u + v - 1 + \Big((1-u)^{-\tau} + (1-v)^{-\tau} - 1\Big)^{-\frac{1}{\tau}} \tag{6}$$

where, again, $\tau \in [-1, \infty \setminus \{0\}$ is the dependence parameter.

As we show in Section 3, our sample data do not systematically exhibit clear lower-tail or upper-tail dependence. We therefore also consider a fourth and final specification whihch is a more flexible 3-parameter specification that flexibly captures upper- and lower-tail dependence in an asymmetric way by mixing a Clayton and a rotated Clayton copula (see Chau, 2010, for a similar specification):

$$C_M(u, v; \tau_1, \tau_2, p) = p\, C_C(u, v; \tau_1) + (1-p)\, C_R(u, v; \tau_2) \tag{7}$$

Estimation

Parameters of the three components F_Y, F_W and C can be estimated separately. In this paper, all parameters for F_Y and F_W were first estimated by conventional maximum likelihood using the built-in Newton-Raphson optimizer of Stata™(StataCorp, 2011). Maximum likelihood estimation of the copula parameter is done in a second stage based on the sample values of $(\hat{F}_Y(y_i), \hat{F}_W(w_i))$ with $\hat{F}_Y$ and $\hat{F}_W$ based on the first stage parameter estimates.

3 Data and preliminary inspection

We estimate the model on household survey data for five countries: the United States, Germany, Italy, Luxembourg and Spain. The sources for cross-country comparable income and wealth data are few.[2] In this paper we use the conceptual framework developed by the Luxembourg Wealth Study (LWS) for creating harmonized variables of wealth and income (Sierminska et al., 2006). The data contain information on multiple income sources and detailed information on financial, non-financial assets and debts. We construct a variable of total disposable household income by aggregating income sources and deducting taxes. This is done for all countries except Spain, where only gross income is available. The variable of net worth is constructed by adding up available assets and deducting debts. All values are expressed in 2009 US dollars.[3]

The goal of LWS is to provide users to the extent possible with comparable data through ex-post harmonization. This means that a thorough examination of wealth and income components is performed in order to identify the most comparable aggregated measure, inherently not all comparability issues can be addressed due to variations in data collection techniques and survey collection goals. For example, in some countries an oversampling of the very wealthy takes place. We leave this intact in order to highlight how distributional differences affect our results.[4] For these purposes also, we do not apply equivalence scales as is standard in the income literature.

The data for the United States come from the 2007 Survey of Consumer Finances (SCF), for Italy the 2008 Survey of Household Income and Wealth (SHIW), for Germany the 2007 wealth module of the Socio-Economic Panel (SOEP), for Luxembourg from the 2007 wealth module of the PSELL-3/EU-SILC and for Spain the 2008 Spanish Survey of Household Finances (EFF). Sample sizes can be found in Table 1.

Figure 1 shows kernel density estimates of the marginal distributions of income and wealth. To help visualize the density at small and negative values, the x-axis is scaled by an inverse hyperbolic sine transformation. While income distributions have broadly similar shapes in the five countries, the wealth distributions exhibits substantially more

[2]The existing data collected based on ex ante harmonization are not available for the whole population. For example, the Survey of Health, Ageing and Retirement in Europe (SHARE) captures individuals over 50. The forthcoming data European Household Finance and Consumption Survey (EHFCS) coordinated by the European Central Bank will be available in 2013 for euro-zone countries.

[3]We use the national price deflators for personal consumption to express currencies in 2009 prices and then convert them to international dollars using PPPs for personal consumption (OECD 2011).

[4]While we do not adjust for extreme data on wealth, we delete observations with income less than or equal to zero as these are not considered in our income distribution model. They only represent a tiny fraction of our samples.

variation across countries. The wealth distribution is typically bimodal with a first mode at zero, and is more stretched out over positive values, thereby exhibiting more inequality. Further sense of the difference in the shape of the wealth and income distributions in the five countries is derived from Figure 2, which shows a relative quantile-difference plot.[5] Figure 2 shows steep upward sloping curves at low percentiles, where wealth quantiles are lower than corresponding income quantiles. There is also a lot of variation across countries in the extent to which wealth is lower than income at the bottom of the distributions. The equality between income and wealth levels occurs around the 25th percentile, except for Spain (20th percentile) and Germany (50th percentile). The positive upward slope for most of percentile values above the 25th percentile and followed by a steep increase for the highest percentiles is another reflection of the higher dispersion of net worth and an indication of the presence of very large values for net worth at the top (beyond the 95th percentile). Ignoring differences at the very bottom of distributions (at negative or zero net worth), it is in Germany and the US that the distributions of income and net worth are most similar. Note however that the relative quantile curves compares the marginal distributions of income and net worth; it does not yet tell us much about the association between the two variates.

The fraction of negative and zero net worth in our samples vary dramatically across countries. Table 1 indicates that about 20% of the sample in Luxembourg has zero or negative net worth, 7% in the US, 27% in Germany, 10% in Italy and 4% in Spain. This is driven, at least partly, by discrepancies in data collection (the five surveys do not collect exactly the same sets of assets and debts). This illustrates the necessity of a strategy to deal with such observations, but also the difficulty of gathering internationally harmonized micro-data on net worth.

The descriptive statistics reported in Table 1 show that the average level of wealth is much higher than the average level of annual income. The ratio of means for wealth and income varies substantially across countries however, from about 4 in Germany to almost 10 in Luxembourg (ratios of medians range from 0.8 in Germany to 8.1 in Luxembourg

[5] A relative quantile-difference plot shows the difference between the values of a variable at each percentile of its distribution and the corresponding values of another distribution, as a percent of the values for one of the distributions (Kennickell, 2009).

Table 1: Sample descriptives

	US	Germany	Italy	Luxembourg	Spain
Observations	4,232	10,907	7,899	3,651	5,013
NW>0	0.913	0.670	0.892	0.882	0.944
NW=0	0.020	0.205	0.070	0.115	0.009
NW<0	0.067	0.124	0.038	0.003	0.047
Mean					
Net worth	572,015	136,472	284,394	578,364	339,744
Income	68,542	33,101	37,368	59,424	34,348
Median					
Net worth	133,900	21,739	175,976	407,088	235,330
Income	40,522	27,159	30,107	50,415	29,096
Proportion in quantile groups					
Q1Q1	9,30	11,93	10,16	9,57	5,16
Q5Q5	11,60	8,98	11,64	8,83	7,17

and Spain).[6] Mean wealth holdings are particularly large in Luxembourg and the US. Median wealth holdings in Luxembourg are particularly striking in comparison to other countries.

To gain insight on the degree of dependence between wealth and income in our samples, Table 1 reports the proportion of observations found in the bottom and top quintile groups of both the income and wealth distributions. If there was perfect correlation between ranks in income and wealth, we would observe 20% of the samples in each of the groups. If, on the other hand, there was no correlation in ranks, we would observe about 4% of the samples in each of the Q1Q1 and Q5Q5 cells. Negative correlation could lead at the extreme to observing no data in these cells. In most countries, 8 to 11 percent of the sample can be found in each of the cells with the exception of Spain. There we find 5% of the sample in the bottom quantiles of the two groups and 7% in the top two quantiles.

Table 2 shows additional summary indices of the association between income and net worth in our samples. The Pearson correlation coefficient gives us an indication of the linear relationship between income and wealth. The others –Spearman's ρ and Kendall's

[6] Note that unlike for the other countries for Spain we observe gross income only, thereby inflating values of income compared to other countries.

Table 2: Association measures

	US	Germany	Italy	Luxembourg	Spain
Correlations (not weighted)					
Pearson	0.666	0.556	0.468	0.147	0.108
Spearman	0.833	0.471	0.630	0.494	0.341
Tau-a	0.652	0.326	0.452	0.343	0.232
Tau-b	0.652	0.330	0.453	0.349	0.233

tau– are rank correlation indicators. They give indication of the non-parametric relationship between income and wealth. We observe the highest linear correlation between income and wealth in the US (0.67), Germany (0.56), Italy (0.47), and then Luxembourg (0.15) and Spain (0.11), where the correlation is surprisingly low. When it comes to rank correlations, the country rankings change slightly. The highest Spearman correlation is still in the US (0.83), then Italy (0.63), but then Luxembourg (0.49), Germany (0.47) and Spain (0.34). Note that this ordering is also slightly different from what comes out from the Q1Q1 and Q5Q5 statistics of Table 1 in which Italy tends to show greater dependence. Overall this suggests positive dependence between income and net worth but a potentially complex dependence that is not univocally captured by summary indices.

Finally, a more detailed description of the association pattern between income and net worth in our samples is provided in Figure 3 which shows 'transition probability color plots'. In a spirit similar to the Q1Q1 and Q5Q5 statistics, the plots show conditional wealth quantile groups conditioning on income quantile groups.[7] Each fine horizontal stripe stacked in the plots represents an income quantile group (from low income at the top to high income at the bottom) and each stripe is colored according to the location of sample observations from this income quantile group in the wealth distribution. Observations located at the bottom of the wealth distribution are colored in light blue, observations located at the top are colored in dark blue. So, the size of light colors represents the share of people in the lowest wealth percentile and the darkest in the highest conditional on income. This is a convenient way of showing details of the distribution of wealth across the income distribution.

In Germany and Spain, we see that the low values of wealth (the large share of zero

[7]Transition probability plots are originally used for visualizing income mobility (Van Kerm, 2011).

wealth in Germany) are distributed throughout the income distribution. This is less so in the other countries. High wealth tends to show more concentration toward high income than does low income and low wealth. This is particularly strong in the US and in Italy. In the other three countries, high wealth is much more dispersed over the income groups, in particular in Luxembourg and Spain (notice however the extreme top income group in Spain which consists exclusively of high wealth people). Overall, the similarity in patterns is clear between Italy and the US and the low association in Spain is also obvious.

4 Estimation results

We now return to our parametric model for the joint distribution of income and wealth. We first briefly report the detailed parameter estimates and later present inequality indicators derived from the models, along with counterfactual constructs to quantify the role played by cross-national differences in association parameters on differences in bi-dimensional inequality.

Parameter estimates for the proposed model estimated from each of the five samples are reported in Table 3. The top panel shows parameter estimated for the Singh-Maddala distribution of income; the middle panel shows parameters for the mixture model of net worth, namely the share of negative and zero data, and the four distribution parameters; the bottom panel shows estimates of the various copula function parameters considered. Let us first discuss estimation issues. First we could not estimate reliably the model for negative wealth in Luxembourg given the small number of observations reporting negative net worth – we therefore discarded these few observations and worked with a model with just two components (positives and zeros) in this country. Second, the copula parameters could not be estimated by our likelihood optimizer on a number of occasions, in particular for the rotated Clayton model and for the more complex mixture Clayton model.

The similarity of the income distributions across countries is reflected in the relative similarity of coefficients of the Singh-Maddala distribution. Wealth distribution parameters on the other hand vary substantially, in line with differences in the shape of the

Table 3: Parameter estimates

	USA	Germany	Italy	Lux'g	Spain
Income distribution					
a	1.99	1.89	2.53	2.77	1.96
b	41,059	49,173	34,022	58,197	46,377
q	1.02	2.50	1.22	1.32	2.15
Wealth distribution					
π_1	0.067	0.123	0.038	0.000	0.047
π_2	0.020	0.208	0.070	0.116	0.009
θ ($\times 10^7$)	544	327	1193		531
α	0.73	0.73	0.84	0.98	0.97
β	756,007	3,519,022	5,431,515	4,999,554	3,358,732
γ	2.63	10.38	12.94	8.69	10.04
Copula parameters					
Plackett τ	7.65	3.76	7.93	4.44	2.18
Clayton τ	0.69	0.21	0.87	0.59	
Rotated Clayton τ	1.35	0.67			
Mixture Clayton τ_1	2.46				
Mixture Clayton τ_2	1.39				
Mixture Clayton p	0.114				

wealth distributions across countries. For example, for data with positive wealth, the scale parameter (β) shows large variations with much higher levels of wealth in Luxembourg and shape parameters suggesting relatively low inequality in Luxembourg.

In line with the rank correlation statistics reported in Table 2, our copula function parameter estimates splits our five countries in two groups: the USA and Italy on the one hand which exhibit strong dependence between income and wealth (high copula parameters), and Luxembourg, Germany and Spain on the other hand with much smaller levels of (positive) dependence.

The overall goodness of fit of our model can be gauged by comparing summary statistics derived from the model parameters to the statsitics computed from the raw data. Estimates reported in Table 5 in the Appendix suggest that our model approximates the distribution of income and wealth reasonably well. Except for some small underestimation of both levels and dispersion of income and wealth in the US, the marginal distributions seem satisfactorily captured. Correlation coefficients are however much less precisely approximated. Perhaps unsurprisingly given the complex empirical dependence

Table 4: Gini coefficient estimates (model-based)

	USA	Germany	Italy	Lux'g	Spain
Gini coefficients					
Income	0.485	0.371	0.354	0.315	0.373
Wealth	0.774	0.809	0.628	0.591	0.565
Joint (Plackett)	0.747	0.703	0.560	0.523	0.540
Joint (Clayton)	0.671	0.645	0.527	0.485	
Joint (Rot.Clayton)	0.731	0.662			
Counterfactual joint Gini coefficients (fix US copula)					
Joint (Plackett)	0.747	0.692	0.561	0.516	0.526
Joint (Clayton)	0.671	0.654	0.524	0.487	0.492
Joint (Rot.Clayton)	0.731	0.680	0.549	0.506	0.515
Counterfactual joint Gini coefficients (fix US wealth margin)					
Joint (Plackett)	0.747	0.693	0.675	0.663	0.700
Joint (Clayton)	0.671	0.636	0.635	0.619	
Joint (Rot.Clayton)	0.731	0.652			

structure shown in Section 3, no single model clearly outperforms the others and there is no systematic under- or over-estimation of the empirical correlation measures. However, overall, the comparison suggests that estimates based on the Plackett copula generally lead to more acceptable results. The inability to fit the (rotated) Clayton model in three countries is also suggestive that the Plackett provides a closer approximation to the sample data distributions.

We now exploit our models to separate out contributions by the dependence parameter and by marginal distributions in accounting for cross-national differences in inequality.

Table 4 first reports Gini coefficient estimates for the (univariate) distributions of income and wealth derived from our model parameters.[8] It is directly clear that wealth is substantially more unequally distributed than income. Countries with the highest inequality in income are not necessarily those with highest wealth inequality. Spain has comparatively large income inequality but the lowest level of wealth inequality –although this may be due to the fact that income are before taxes in our Spanish sample. Germany on the contrary has median income inequality but the highest wealth inequality among the five countries.

[8]No closed-form expression exist for deriving the various inequality measures we consider. Estimation is therefore based on Monte Carlo sampling on the basis of the models and estimated parameters.

To capture inequality in the joint distribution of income inequality, we rely on the bi-dimensional Gini coefficient of Koshevoy and Mosler (1997) which summarizes inequality in the joint distribution of income and wealth.[9] It is determined by the degree of inequality in the marginal distributions as well as by the association among the two covariates. Table 4 reports three different estimates of the bi-dimensional Gini, each based on an alternative specification for the copula function. While levels of bivariate inequality differ according to the specification used, the ordering of countries is preserved. In all cases, it appears that inequality in wealth remains a key determinant of overall inequality: bivariate Ginis are close to univariate Ginis for wealth. Notice however that because the association between income and wealth is not perfectly positive, the bivariate takes on a value between the two marginal Ginis. The joint Gini has a lower value when the Clayton copula (which exhibits a stronger association at the bottom of the distribution) is used and larger when stronger upper-tail dependence is assumed via the rotated Clayton copula.

Table 4 finally reports two sets of counterfactual estimates of the bivariate Ginis. The first set is obtained by fixing the US dependence parameter and applying it to all countries marginal distributions. The indices obtained therefore give us the value of the bivariate Gini which would be observed in the different countries if the association between income and wealth were as high as in the US. Comparison of the counterfactual indices with the previous estimates gives us indication of the impact of cross-national differences in the association parameter on overall inequality differences.

Our estimates suggest that swapping the dependence parameter would only have a small impact on bivariate inequality. To benchmark this effect, the second set of coun-

[9]The Koshevoy and Mosler (1997) bidimensional Gini index is an extension of the univariate Gini based on a bivariate, Euclidian distance-based extension of the Gini Mean Deviation:

$$G_2 = \frac{1}{4N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\left(\frac{y_i}{\mu_y} - \frac{y_j}{\mu_y} \right)^2 + \left(\frac{w_i}{\mu_w} - \frac{w_j}{\mu_w} \right)^2 \right)^{\frac{1}{2}}.$$

Compare this with the univariate version:

$$G_1 = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \frac{y_i}{\mu_y} - \frac{y_j}{\mu_y} \right|.$$

terfactuals shows what would be the bivariate Ginis if the parameters of the marginal *wealth* distribution were fixed at the level of the US. For Germany –another country with high wealth inequality– the impact of the swap is very small, but now Spain, Luxembourg and Italy would see bivariate inequality grow by about twenty percent if their distribution of wealth was as in the United States (keeping their income distribution and their dependence parameter). The impact of cross-national differences in dependence is comparatively very small.

5 Concluding remarks

In this paper, we set out to consider a parametric model for the joint distribution of income and wealth. We use this model to examine whether joint income and wealth inequality provides us with a different pattern of social inequality than the traditional income only approach. Our parameter of focus is the dependence parameter between income and wealth. We use our model to disentangle the relative contribution of cross-national differences in this dependence parameter from cross-national differences in marginal distributions in shaping differences in overall, bi-dimensional inequality. This is done by exploiting the simple structure of our model to generate counterfactual distributions of interest.

The framework offered by the Sklar theorem to build a model for the bivariate distribution is attractive in this context since the marginal distributions of income and wealth have specificities that require differentiated treatment –in particular the presence of zero and negative net worth. It would therefore be difficult to identify a relevant joint distribution directly. Specification of the copula function for capturing the rank-order association is a key step of the process. We have considered here some simple, tractable functions, yet we find that none of these functions fully capture the complex dependence between income and wealth observed in our samples.

Deriving inequality indicators from the estimated models, we observe that Gini coefficients on income are much lower than Gini coefficients on wealth. A bi-dimensional Gini coefficient summarizing inequality in the joint distribution of income and wealth returns intermediate values –with the highest value found in the US. The bi-dimensional Gini appears largely driven by inequality in the wealth distribution and differences across countries in the dependence between income and wealth does not appear to be a key driver in international differences in the bi-dimensional inequality.

The parametric methods presented in this paper could be extended in a number of directions. Firstly, perhaps the most promising avenue is to exploit recent advances in robust statistics to estimate marginal distribution parameters robust to the presence of outliers and data contamination. Wealth distributions in particular have typically long tails. This potentially leads to extreme sample data which, in turn, can exert strong influence on the estimation of distribution parameters and inequality indices (see, e.g. Cowell and Flachaire, 2007, Van Kerm, 2007). Robust estimation techniques used in place of classical maximum likelihood would be relevant to keep the impact of extreme data under control at the estimation stage.[10]

Secondly, our modelling of the dependence parameter primarily relies on simple, conventional copula functions. Yet, detailed inspection of the bivariate data suggests that the dependence structure may be relatively complex and may therefore not be entirely satisfactorily captured by these simple specifications. The search for more sophisticated specifications –perhaps of different specifications for different countries or samples– would therefore be of obvious interest.[11]

Thirdly, we have restricted ourselves at this stage to a largely illustrative analysis based on the relatively simple Koshevoy and Mosler's (1997) bi-dimensional Gini index. More sophisticated measures of multi-dimensional inequality or poverty could be considered within the same framework for more in-depth analysis (Lugo, 2005).[12]

References

Biewen, M. and Jenkins, S. P. (2005), 'Accounting for differences in poverty between the USA, Britain and Germany', *Empirical Economics* **30**(2), 331–358.

[10] We conducted preliminary analysis with the optimal B-robust estimators (OBRE) for distribution parameters proposed by Victoria-Feser (2000). While income distribution parameters could be estimated robustly, our OBRE algorithms failed to achieve convergence for the wealth distributions. Alternative estimators may therefore need to be considered to handle such distributions. We leave this issue open for further research.

[11] Note however that our approach based on a Clayton copula mixture did not reveal successful so far.

[12] The presence of negative net worth would however remain a critical constraint on the choice of relevant multidimensional inequality measures, as it is in the univariate case (Jenkins and Jäntti, 2005).

Bonhomme, S. and Robin, J.-M. (2009), 'Assessing the equalizing force of mobility using short panels: France, 1990–2000', *Review of Economic Studies* **76**(1), 63–92.

Brachmann, K., Stich, A. and Trede, M. (1996), 'Evaluating parametric income distribution models', *Allgemeines Statistisches Archiv* **80**, 285–298.

Chau, T. W. (2010), Essays on earnings mobility within and across generations using copula, PhD thesis, University of Rochester, Dept. of Economics.
URL: *http://hdl.handle.net/1802/11125*

Cowell, F. A. and Flachaire, E. (2007), 'Income distribution and inequality measurement: The problem of extreme values', *Journal of Econometrics* **141**(2), 1044–1072.

Dagum, C. (1990), A model of net wealth distribution specified for negative, null and positive wealth. A case study: Italy, *in* C. Dagum and M. Zenga, eds, 'Income and Wealth Distribution, Inequality and Poverty', Springer, Berlin and Heidelberg, pp. 42–56.

Genest, C. and McKay, J. (1986), 'The joy of copulas: Bivariate distributions with uniform marginals', *American Statistician* **40**(4), 280–3.

Jäntti, M., Sierminska, E. and Smeeding, T. (2008), The joint distribution of household income and wealth: Evidence from the Luxembourg Wealth Study, OECD Social Employment and Migration Working Paper 65, OECD, Directorate for Employment, Labour and Social Affairs.

Jenkins, S. P. and Jäntti, M. (2005), Methods for summarizing and comparing wealth distributions, ISER Working Paper 2005-05, Institute for Social and Economic Research, University of Essex, Colchester, UK.

Jenkins, S. P. and Van Kerm, P. (2009), The measurement of economic inequality, *in* W. Salverda, B. Nolan and T. M. Smeeding, eds, 'Oxford Handbook of Economic Inequality', Oxford University Press, chapter 3.

Kennickell, A. B. (2009), Ponds and streams: wealth and income in the U.S., 1989 to 2007, Finance and Economics Discussion Series 2009-13, Board of Governors of the Federal Reserve System (U.S.).

Kleiber, C. and Kotz, S. (2003), *Statistical size distributions in Economics and Actuarial Sciences*, John Wiley and Sons, Inc., New Jersey.

Koshevoy, G. A. and Mosler, K. (1997), 'Multivariate Gini indices', *Journal of Multivariate Analysis* **60**, 252–276.

Lugo, M. A. (2005), Comparing multidimensional indices of inequality: Methods and application, ECINEQ Working Paper 14, Society for the Study of Economic Inequality.

McDonald, J. B. (1984), 'Some generalized functions for the size distribution of income', *Econometrica* **52**(3), 647–663.

Plackett, R. L. (1965), 'A class of bivariate distributions', *Journal of the American Statistical Association* **60**, 516–522.

Sierminska, E., Brandolini, A. and Smeeding, T. (2006), 'The Luxembourg Wealth Study: A cross-country comparable database for household wealth research', *Journal of Economic Inequality* **4**(3), 375–383.

Singh, S. K. and Maddala, G. S. (1976), 'A function for size distribution of incomes', *Econometrica* **44**(5), 963–970.

Sklar, A. (1959), 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229–231.

StataCorp (2011), *Stata Statistical Software: Release 12*, StataCorp LP, College Station.

Trivedi, P. K. and Zimmer, D. M. (2007), 'Copula modeling: An introduction for practitioners', *Foundations and Trends in Econometrics* **1**(1), 1–111.

Van Kerm, P. (2007), Extreme incomes and the estimation of poverty and inequality indicators from EU-SILC, IRISS Working Paper 2007-01, CEPS/INSTEAD, Differdange, Luxembourg.

Van Kerm, P. (2011), Picturing mobility: Transition probability color plots, United Kingdom Stata Users' Group Meetings 2011 18, Stata Users Group.
URL: *http://ideas.repec.org/p/boc/usug11/18.html*

Victoria-Feser, M.-P. (2000), ‘Robust methods for the analysis of income distribution, inequality and poverty’, *International Statistical Review* **68**(3), 277–93.

Figure 1. Kernel density estimates of the marginal distributions of income and wealth (x-axis scaled by an inverse hyperbolic sine transformation)

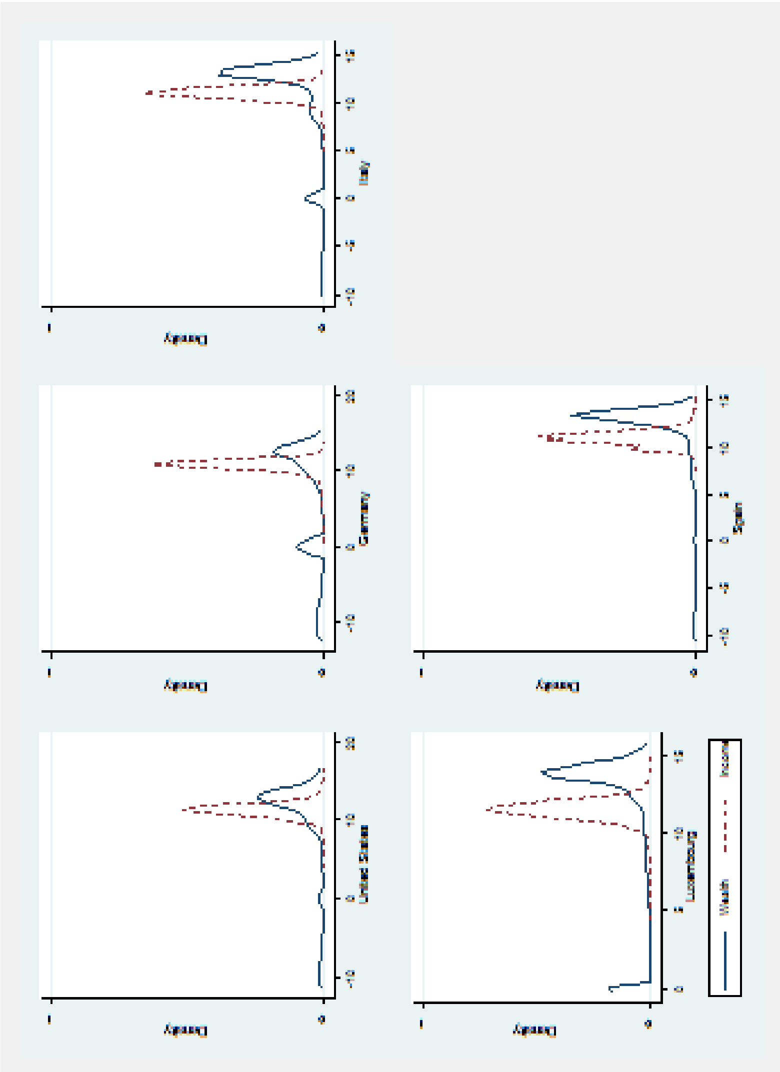

Figure 2. Relative quantile difference plot [(Wealth-Income)/Income].

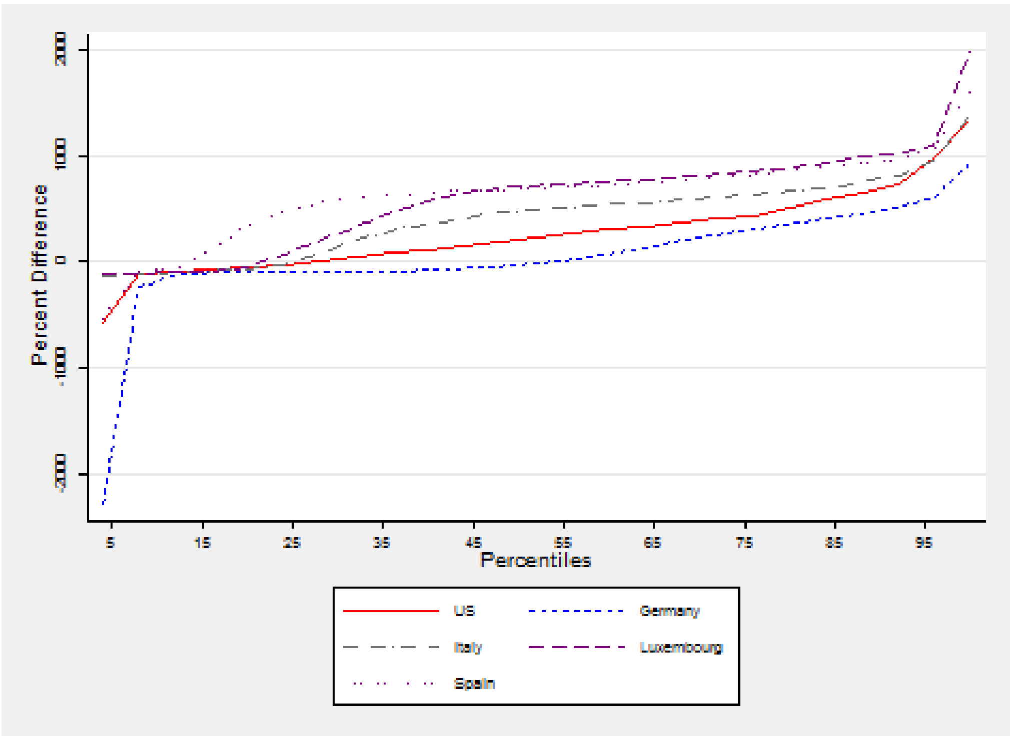

Figure 3. Probability color plot of income (y-axis) and wealth (x-axis).

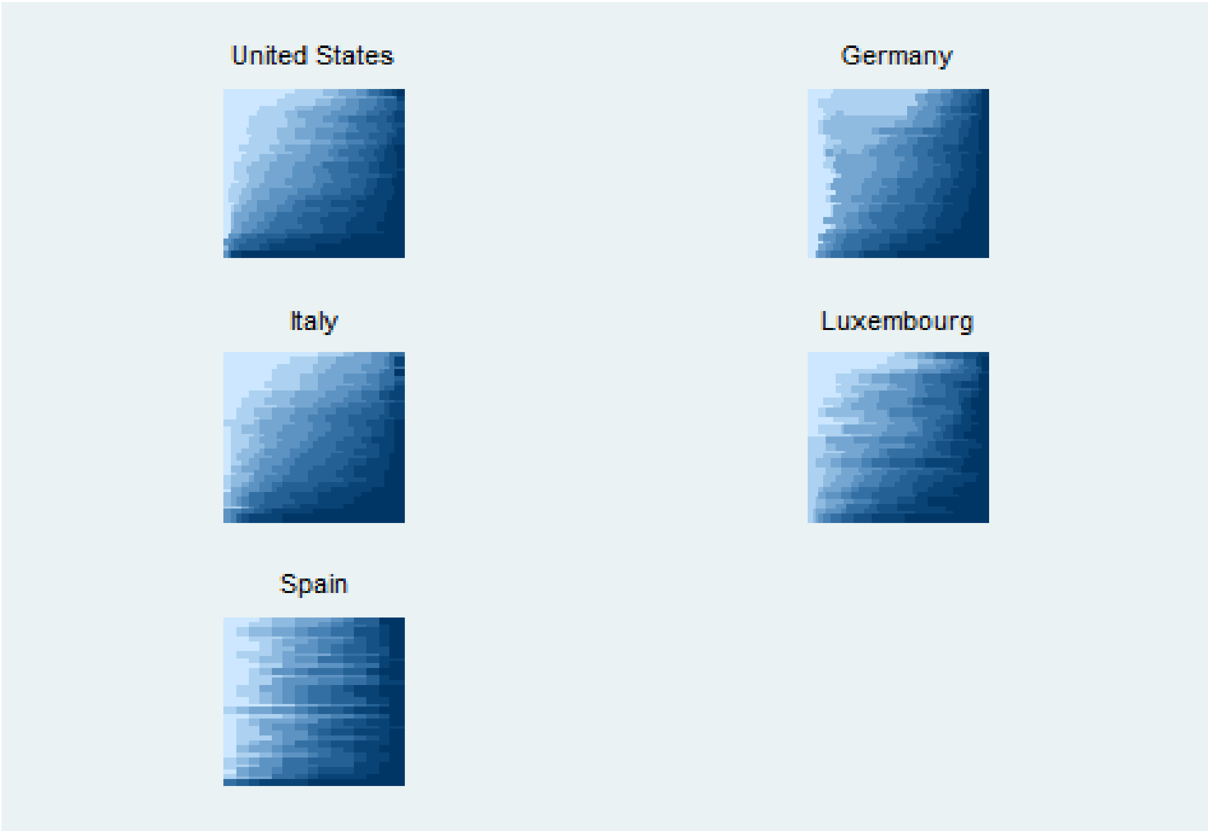

Table 5: Empirical vs. model-based estimates

	USA	Germany	Italy	Lux'g	Spain
Empirical estimates:					
Mean income	68,752	32,833	37,368	59,266	34,348
Mean wealth	572,845	135,664	284,394	575,854	339,744
Gini income	0.538	0.374	0.354	0.316	0.376
Gini wealth	0.809	0.807	0.611	0.578	0.544
Pearson correlation	0.641	0.578	0.520	0.175	0.169
Pearson correlation (unw.)	0.679	0.563	0.468	0.148	0.108
Spearman correlation (unw.)	0.831	0.469	0.630	0.494	0.341
Kendall's τ (unw.)	0.650	0.324	0.452	0.343	0.232
Model-based estimates:					
Mean income	61,497	32,323	37,298	59,106	34,241
Mean wealth	482,151	130,372	278,726	546,841	329,575
Gini income	0.485	0.371	0.354	0.315	0.373
Gini wealth	0.774	0.809	0.628	0.591	0.565
Pearson corr. (Plackett)	0.597	0.414	0.604	0.459	0.253
Pearson corr. (Clayton)	0.895	0.980	0.859	0.916	
Pearson corr. (Rot.Clayton)	0.767	0.894			
Spearman corr. (Plackett)	0.428	0.283	0.434	0.318	0.170
Spearman corr. (Clayton)	0.742	0.880	0.695	0.769	
Spearman corr. (Rot.Clayton)	0.597	0.728			
Kendall's τ (Plackett)	0.428	0.290	0.435	0.321	0.170
Kendall's τ (Clayton)	0.742	0.900	0.697	0.774	
Kendall's τ (Rot.Clayton)	0.597	0.744			

Note: Empirical estimates are derived directly from the raw samples. Model-based estimates are derived from the model parameter estimates (by simulation). Rank correlation estimates from the raw samples ignore sampling weights and are therefore not directly comparable to the model-based estimates. All estimates are based on a single set of multiply imputed data and may therefore differ marginally from estimates reported in Section 3 (Table 2).