A model for the joint distribution of income and wealth

Markus Jäntti, Eva Sierminska and Philippe Van Kerm

Abstract 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 four 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.

Key words: income, wealth, inequality, copula, multivariate Gini

1 Introduction

Relatively little is known about the dependence between income and wealth, especially outside the United States.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. 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.

This paper considers a parametric model for studying the joint bivariate distribution of income and wealth. The model handles specificities of the distribution of wealth, in particular the presence of zero and negative data, as well as the long tails of the distribution. A copula-based approach is adopted to handle the depen-

Markus Jäntti
SOFI, University of Stockholm, e-mail: markus.jantti@sofi.su.se

Eva Sierminska
CEPS/INSTEAD, Luxembourg, e-mail: eva.sierminska@ceps.lu

Philippe Van Kerm
CEPS/INSTEAD, Luxembourg, e-mail: philippe.vankerm@ceps.lu

dence betwen the two covariates. The model is estimated on data from four OECD countries using comparable household-level survey data.

2 A model for the joint distribution of income and wealth

Multivariate distributions can be expressed as a function of a copula and univariate marginal distributions, $F(y,w) = C(F_Y(y), F_W(w))$ where F is the joint distribution of income and wealth, F_Y and F_W denote the marginal distribution 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.

Specifying the marginal distribution of income is reatively unproblematic. We rely on a Singh-Maddala (SM) specification: $F_Y(y) = \mathrm{SM}(y;a,b,q) = 1 - \left[1 + \left(\frac{w}{b}\right)^a\right]^{-q}$. The SM distribution is a flexible model for unimodal distributions allowing varying degrees of skewness and kurtosis and dealing with the heavy tails typical of income and earnings distributions.

Specifying the marginal distribution of wealth is more difficult. The literature on wealth inequality is focusing on the concept of *net worth*, defined as the value of total assets (financial and non-financial) minus total debts. It is consequently relatively common to observe data with zero or negative net worth. This rules out virtually all specifications typically used for modelling income distributions, since these size distributions have positive density only over R_+. To accommodate zero and negative data, we use the specification proposed by Dagum (1990) and more recently advocated by Jenkins and Jäntti (2005). The specification is a finite mixture model where negatives, zeroes and positive data are modelled separately with an exponential distribution (negatives), a point-mass at zero and a Dagum-type I distribution (positives):

$$F_W(w) = \begin{cases} \pi_1 \ \exp(\theta w) & \text{if } w < 0 \\ \pi_1 + \pi_2 & \text{if } w = 0 \\ \pi_1 + \pi_2 + (1 - \pi_1 - \pi_2) \ \left(1 + \left(\frac{\beta}{w}\right)^{\alpha}\right)^{-\gamma} & \text{if } w > 0 \end{cases} \tag{1}$$

where parameters α, β, γ, θ are positive. π_1 and π_2 are the shares of negatives and zeroes.

For the copula, we rely on a Plackett specification:

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

The Plackett copula exhibits symmetric upper-tail and lower-tail dependence.

We estimate the parameters of the three components F_Y, F_W and C separately. All parameters for F_Y and F_W are first estimated by conventional maximum likelihood using the built-in Newton-Raphson optimizer of Stata$^{\text{TM}}$.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 estimation results

We use data for the United States, Germany, Italy, and Luxembourg. The data for US 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. The data contain information on multiple income sources and detailed information on financial, non-financial assets and debts. On the basis of this detailed information, we use the conceptual framework developed by the Luxembourg Wealth Study (Sierminska et al., 2006) for creating harmonized variables of net worth (total assets minus liabilities) and disposable income.

Parameter estimates for the proposed model estimated from each of the four samples are reported in Table 1. The bottom lines report Gini coefficient estimates for the (univariate) distributions of income and wealth, as well as the bi-dimensional Gini coefficient of Koshevoy and Mosler (1995). The latter index summarizes inequality in the joint distribution of income and wealth. It is determined by the inequality in the marginal distributions as well as by the association among the two covariates. To separate out the effect of covariate association from marginal distributions in cross-country differences in inequality, the last line of the table reports a counterfactual estimate of the Gini coefficient obtained by applying the US copula parameter to all countries. All indices are derived from model parameter estimates.

The similarity of the income distributions across countries is reflected in the coefficients of the Singh-Maddala distribution. Wealth distribution parameters vary substantially. To start with, the proportions of negatives (π_1) and zeroes differ. This is certainly, in part, reflecting differences in data collection (the four surveys do not collect exactly the same sets of assets and debts). For data with positive wealth, the scale parameter (β) also shows tremendous variations with much higher levels of wealth in Luxembourg.

Our copula function parameter –the measure of key interest in this application– splits our four countries in two groups: the USA and Italy on the on hand which exhibit strong dependence between income and wealth, and Germany and Luxembourg on the other hand with much smaller levels of dependence.

Gini coefficients on income are much lower than Gini coefficients on wealth, but the ranking of countries is the same on both covariates, with inequality being higher in the US sample. The bi-dimensional Gini takes intermediate values, again exhibiting greater inequality in the US sample. The similarity of the dimensional Ginis with the ‘common copula’ counterfactual Ginis suggest that relatively little of the cross-country variations in bi-dimensional inequality is due to differences in association parameters – the difference in marginal distributions is key.

Acknowledgements 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).

Table 1 Parameter estimates and Gini coefficients

	USA	Germany	Italy	Luxembourg
Income distribution				
a	2.00	1.94	2.53	2.77
b	40,824	47,249	34,022	58,380
q	1.01	2.36	1.22	1.32
Wealth distribution				
π_1	0.067	0.126	0.038	0.003
π_2	0.020	0.205	0.070	0.116
θ ($\times 10^{-7}$)	537	361	1193	69
α	1.36	2.27	2.60	3.29
β	419,981	363,369	522,802	1,071,562
γ	0.44	0.24	0.25	0.23
Copula distribution				
τ	**7.65**	**3.72**	**8.08**	**4.56**
Gini coefficients (derived from model parameter estimates)				
Gini wealth	0.804	0.790	0.609	0.566
Gini income	0.495	0.370	0.357	0.317
Bi-dimensional Gini	**0.787**	**0.688**	**0.550**	**0.506**
Counterfactual Gini coefficient (with US copula parameter)				
Bi-dimensional Gini	**0.787**	**0.678**	**0.551**	**0.500**

References

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.

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.

Koshevoy, G. A. and Mosler, K. (1995), 'Multivariate Gini indices', *Journal of Multivariate Analysis* **60**, 252–276.

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.