Capital income inequality: evidences from ECHP data

Francesca Greselin, Leo Pasquazzi, Ričardas Zitikis

Abstract Micro-data of European Union (EU) countries show that capital incomes account for a large part of disparity in populations and follow heavy-tailed distributions. Measuring and comparing the disparity requires incorporating the relative nature of 'small' and 'large,' and for the latter reason we employ the newly developed Zenga index of economic inequality. Its non-parametric plug-in estimator does not fall into any well known class of statistics, hence inferential theory has to be developed *ad hoc* by first constructing an appropriate heavy-tailed Zenga estimator, then establishing its asymptotic distribution, and finally deriving confidence intervals. The performance of the new asymptotic results is shown by measuring and comparing capital inequality in the ECHP data.

Key words: Capital income, inequality, heavy-tails, extreme-value theory

1 Motivation

The present research is motivated by the need for a better understanding of the distribution of capital incomes, which in many cases appears to be heavy tailed, hence requires the development of specific inferential methods. We shall show that this is the case by analyzing the 2001 wave of the ECHP survey conducted by EuroStat, which contains data coming from 121,122 individuals spread over the fifteen EU countries (see Table 1 for useful synthesis on the sample data). We focus on personal capital incomes which, according to the definition in this survey, means

Francesca Greselin, Leo Pasquazzi
Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali, Università di Milano Bicocca, Milan, Italy, e-mail: francesca.greselin@unimib.it, leo.pasquazzi@unimib.it

Ričardas Zitikis
Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada, e-mail: zitikis@stats.uwo.ca

income flows from financial assets actually received during the reference year. We restrict our attention to individuals with positive net capital incomes.

Table 1 General ECHP sample informations, with n_T denoting the total sample size and n_P the number of positive incomes in the sample.

	n_T	n_P	median	mean
Germany (D)	10,624	4,861	186.62	948.37
Denmark (DK)	3,789	1,135	231.42	1,071.06
Netherlands (NL)	8,608	2,863	214.18	660.74
Belgium (B)	4,299	690	1,374.81	5,309.17
Luxembourg (LUX)	4,916	769	1,214.68	1,982.62
France (F)	10,119	4,347	359.93	716.68
United Kingdom (UK)	8,521	3,477	368.83	1,522.18
Ireland (IRL)	4,023	949	99.04	604.58
Italy (I)	13,392	1,111	480.30	1,762.38
Greece (GR)	9,419	335	1,232.58	2,256.55
Spain (E)	11,964	6,541	137.43	240.84
Portugal (P)	10,915	600	116.26	1,232.67
Austria (A)	5,605	2,834	133.50	323.82
Finland (FIN)	5,637	1,509	180.63	3,662.57
Sweden (S)	9,291	5,637	84.50	601.53

2 Parameters and their estimators

Zenga's inequality index is based on the ratio of the lower and upper conditional expectations, thus taking into account the relative nature of poor and rich. Details can be found in the seminal paper (Zenga, 2007) and some motivations for employing this measure w.r.t. well known indices, such as the Gini, can be found in (Greselin *et al.*, 2009 and references wherein). Here we recall the definition of the Zenga index, given in terms of the lower and upper tail-value-at-risk, $TVAR_F^*(t)$ and $TVAR_F(t)$, evaluated with respect to the cumulative cdf F, which we suppose to be continuous, of the underlying non negative r.v. X

$$Z_F = \int_0^1 z_F(t)dt = \int_0^1 \left(1 - \frac{\mathrm{TVaR}_F^*(t)}{\mathrm{TVaR}_F(t)}\right) dt.$$

The curve z_F measures the inequality between the 'poorer' $t \times 100\%$ part of the population and the remaining 'richer' $(1-t) \times 100\%$ part. The index Z_F is a synthetic measure of the overall inequality in the population. With the empirical quantile function denoted by Q_n, we estimate $\mathrm{TVaR}_F(t)$ by

$$\mathrm{TVaR}_n(t) = \begin{cases} \dfrac{1}{1-t}\left(\displaystyle\int_t^{1-k/n} Q_n(s)ds + \frac{kX_{n-k:n}}{n(1-\widehat{\gamma})}\right) & \text{when} \quad 0 \le t < 1-k/n, \\ \left(\dfrac{k}{n}\right)^{\widehat{\gamma}} \dfrac{X_{n-k:n}}{(1-t)^{\widehat{\gamma}}(1-\widehat{\gamma})} & \text{when} \quad 1-k/n \le t < 1. \end{cases} \tag{1}$$

Notice that it is based on Weissman's (1978) estimator of high-quantiles and Hill's (1975) estimator of the tail index γ.

Having this estimator of TVaR_F, we next introduce an estimator of TVaR^*_F via the equation

$$\mathrm{TVaR}^*_n(t) = \frac{1}{t}\big(\mathrm{TVaR}_n(0) - (1-t)\mathrm{TVaR}_n(t)\big) \tag{2}$$

for all $0 < t \leq 1$. With these estimators of the upper and lower tail-values-at-risk, the empirical Zenga index is defined by

$$Z_n = 1 - \int_0^1 \frac{\mathrm{TVaR}^*_n(t)}{\mathrm{TVaR}_n(t)} dt.$$

Theorem 1. *Assume that the cdf F is regularly varying at infinity with tail index $\gamma \in (1/2, 1)$ and satisfies generalized second-order regular-variation condition with second-order parameter $\rho \leq 0$, and let $k = k_n \to \infty$ when $n \to \infty$ be such that $k/n \to 0$ and $\sqrt{k}\ \alpha(Q(1-k/n)) \to 0$. Then we have that*

$$\frac{\sqrt{n}\,(Z_n - Z_F)}{\sqrt{k/n} X_{n-k:n}\,\sigma_Z} \xrightarrow{d} N(0,1), \tag{3}$$

where

$$\sigma^2_Z = \frac{\gamma^4}{(1-\gamma)^4(2\gamma-1)} v^2(1).$$

To make the asymptotic result in (3) directly applicable in practice, we need to replace the standard deviation σ_Z by its estimator, which we do as follows:

$$\widetilde{\sigma}^2_Z = \frac{1}{n}\sum_{i=1}^{n} h^2_{F_n}(X_i).$$

where by $h_{F_n}(X_i)$ we denote the empirical influence values. Proof, which crucially relies on recent results of Necir *et al.* (2010) and the Vervaat process (see Zitikis (1998)), is given with details in Greselin *et al.* (2011).

3 An analysis of European capital incomes

Here we use the developed methodology to analyze capital incomes in EU countries. Examining the corresponding Hill plots, we saw that capital incomes are heavy-tailed, and so we constructed confidence intervals for the Zenga index based on the heavy-tailed estimator. In addition, we also satisfy our curiosity by looking at the performance of the light-tailed (i.e., plug-in) estimator. Figure 1 visualizes the point estimates and the 95% confidence bounds for the Zenga index.

We see that inequality is very high in all of the considered countries, with Austria having the lowest point-estimate of the Zenga index, where both light-tailed and heavy-tailed estimates are almost equal to 0.9. The sample from Germany yields an estimate of the tail-index larger than 1, and we could therefore not compute

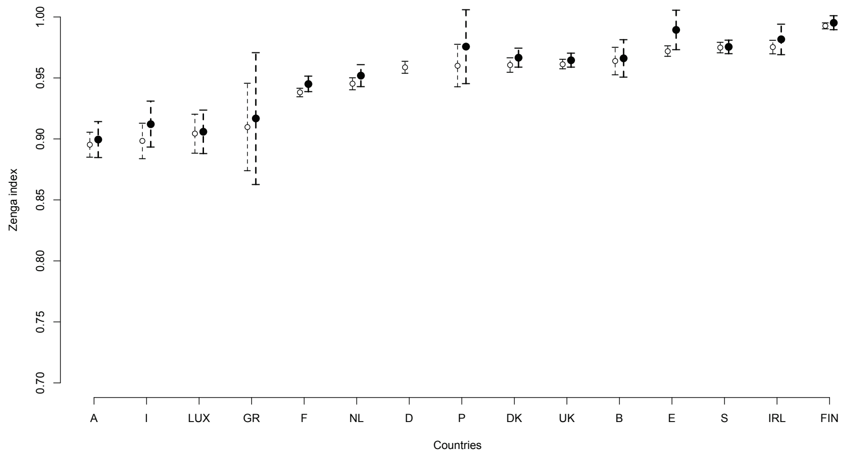

Fig. 1 95% confidence intervals based on the plug-in (empty circles) and the new estimator (solid circles).

the estimate of the Zenga index using the heavy-tailed estimator. Comparing the confidence intervals, we may finally say that in some countries, like Austria, Italy and Luxembourg, inequality is significantly lower than in others.

Acknowledgements We are indebted to the International Council for Canadian Studies and the University of Milano Bicocca for grants supporting our research.

References

Greselin, F., Pasquazzi, L. and Zitikis, R. (2011). European capital incomes are heavy tailed: statistical inferential theory and ECHP data analysis, *submitted*.

Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3 (5), 1163–1174.

Greselin F., Pasquazzi L., Zitikis R. (2010). Zenga's new index of economic inequality, its estimation, and an analysis of incomes in Italy. *Journal of Probability and Statistics*, 26 pp., DOI:10.1155/2010/718905.

Necir, A., Rassoul, A., and Zitikis, R. (2010). Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics*, doi:10.1155/2010/596839

Weissman, I. (1978). Estimation of parameters and larger quantiles based on the k largest observations. *Journal of the American Statistical Society*, 73 (364), 812–815.

Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni* 5, 3–27.

Zitikis, R. (1998). The Vervaat process. In: *Asymptotic Methods in Probability and Statistics*, pp. 667–694. North-Holland, Amsterdam.