Sarabia, José María; University of Cantabria
sarabiaj@unican.es

Authors: Jose Maria Sarabia, Department of Economics, University of Cantabria
Markus Jantti, Department of Political and Economic Studies, University of Helsinki

Title: Estimation of income inequality from grouped data

Over the past decades, there has been a growing interest in the distributional patterns of income in both, the economic literature and the international policy arena. The introduction of the Sustainable Development Goals has highlighted the relevance of this topic since Goal 10 calls for reducing inequalities in income, thus positioning disparities as a key concern. Although addressing inequality trends has become essential, individual data on income or consumption is not often available. Instead, group data from nationally representative surveys are, in most cases, used to assess the evolution of inequality levels. Hence, most prior research on economic inequality relies on lower bounds of inequality measures, which neglect income differences within shares, introducing, therefore, a potential source of measurement error. The aim of this paper is to explore a nuanced alternative to estimate income inequality, which leads to a reliable representation of the income distribution within shares. We examine the performance of a flexible parametric family of income distributions, the generalized beta of the second kind (GB2), to estimate different inequality measures. The accuracy of these estimates is compared with the nonparametric lower bound in more than 5000 datasets covering 182 countries over the period 1867-2015. We deploy two different econometric strategies to estimate the parametric distributions, non-linear least squares and generalised method of moments, both implemented in R and conveniently available in the package GB2group. Despite its popularity, even the simplest two-parameter models outperform the nonparametric approach. Our results confirm the excellent performance of the GB2 distribution to represent income data for a heterogeneous sample of countries, which provides highly reliable estimates of several inequality measures. This strong result and the access to an easy tool to implement the estimation of this family of distributions, we believe, will incentivize its use, thus contributing to the development of reliable estimates of inequality trends.