Economic Modelling 27 (2010) 1473–1483

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

Bivariate income distributions for assessing inequality and poverty under dependent samples

Andrea Vinh [a], William E. Griffiths [b,*], Duangkamon Chotikapanich [a]

[a] Department of Econometrics and Business Statistics, Monash University, Australia
[b] Department of Economics, University of Melbourne, Australia

ARTICLE INFO

Keywords:
C23
C46
Copulas
Income distributions
Inequality
Panel data
Dependent sample

ABSTRACT

As indicators of social welfare, the incidence of inequality and poverty is of ongoing concern to policy makers and researchers alike. Of particular interest are the changes in inequality and poverty over time, which are typically assessed through the estimation of income distributions. From this, income inequality and poverty measures, along with their differences and standard errors, can be derived and compared. With panel data becoming more frequently used to make such comparisons, traditional methods which treat income distributions from different years independently and estimate them on a univariate basis, fail to capture the dependence inherent in a sample taken from a panel study. Consequently, parameter estimates are likely to be less efficient, and the standard errors for between-year differences in various inequality and poverty measures will be incorrect. This paper addresses the issue of sample dependence by suggesting a number of bivariate distributions, with Singh–Maddala or Dagum marginals, for a partially dependent sample of household income for two years. Specifically, the distributions considered are the bivariate Singh–Maddala distribution, proposed by Takahasi (1965), and bivariate distributions belonging to the copula class of multivariate distributions, which are an increasingly popular approach to modelling joint distributions. Each bivariate income distribution is estimated via full information maximum likelihood using data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey for 2001 and 2005. Parameter estimates for each bivariate income distribution are used to obtain values for mean income and modal income, the Gini inequality coefficient and the headcount ratio poverty measure, along with their differences, enabling the assessment of changes in such measures over time. In addition, the standard errors of each summary measure and their differences, which are of particular interest in this analysis, are calculated using the delta method.

1. Introduction

The incidence of inequality and poverty is of ongoing concern not only in developing regions but also amongst some of the world's economic leaders. As indicators of social welfare, much of the literature has been drawn towards the accurate specification and estimation of various inequality and poverty measures. In particular, policy makers and researchers alike are often interested in assessing the changes in inequality and poverty over time. That is, one would ideally expect to observe a reduction in inequality and poverty from one year to another in order to determine whether policies implemented for that purpose have been effective. Such assessments are typically performed through the estimation of income distributions, from which income inequality and poverty measures, along with their differences and standard errors, can be derived and compared.

Traditionally, comparisons of inequality and poverty over time have been made with income distributions for different years being treated as independent. Various univariate functional forms have been suggested in the literature and income distributions have been estimated accordingly using conventional inference techniques. Initially, the gamma, lognormal and Pareto distributions were commonly used, with the gamma distribution generally found to fit better than the lognormal distribution (Salem and Mount, 1974; McDonald and Ransom, 1979; McDonald, 1984). Other two-parameter distributions that have been considered include the beta, Fisk and Weibull distributions. In addition, a number of three-parameter distributions have been proposed in the literature, including the Singh–Maddala distribution, which contains the Pareto, Fisk and Weibull distributions as special cases, and the Dagum distribution, which, although not as widely applied as the Singh–Maddala distribution, has shown to provide a better fit (Kleiber, 1996; Kleiber and Kotz, 2003).

A major issue in taking the univariate approach, however, is that panel data are becoming more frequently used to make comparisons of inequality and poverty over time. Consequently, as some members of

* Corresponding author.
E-mail address: wegrif@unimelb.edu.au (W.E. Griffiths).

0264-9993/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.econmod.2010.07.014

the panel will be common between years, recorded incomes are likely to be correlated, resulting in a dependent sample. Therefore, treating one income distribution for any given year independently of another does not take into account that those who earned a high income in one year are also likely to earn a high income in a subsequent year and vice versa. This is of concern particularly in regions which exhibit low income mobility, as it coincides with a high degree of correlation.

There are two main consequences of estimating separate univariate distributions for different years of a panel. The first is that the parameter estimates are likely to be less efficient than a bivariate or multivariate approach that recognises correlation in incomes from year to year. The second is that the standard errors for between-year differences in various inequality and poverty measures will be incorrect. This is due to the estimated differences being functions of the parameter estimates from both marginal distributions. These parameter estimates will be correlated, and separate estimation does not provide a covariance term for computing the standard error of the difference. These features provide motivation for the use of multivariate techniques for the distribution of income when using panel data, in order to account for possible correlation between years.

The presence of dependence in a sample of income taken from a panel study has been left largely unaddressed in the literature. In a paper by Kmietowicz (1984), a bivariate lognormal distribution is suggested for the joint distribution of household size and income, rather than income over time, which is then used to derive estimates of the Gini inequality measure. Sarabia et al. (2005) adapt this model by deriving extensions of the bivariate lognormal distribution and applying each to data from the European Community Household Panel. In both papers, the proposed models have marginal income distributions which follow the univariate lognormal distribution. However, it has been historically found at the univariate case that although the lognormal distribution performs well at lower income levels, it fits poorly at higher income levels (Singh and Maddala, 1976). In addition, the Singh–Maddala and Dagum distributions have been subsequently shown to provide a better fit than the lognormal distribution (Singh and Maddala, 1976; McDonald and Ransom, 1979; McDonald, 1984). Therefore, a multivariate distribution which has either Singh–Maddala or Dagum marginals would be better suited to approximating the joint distribution of income rather than one with lognormal marginal distributions.

Other studies which have recognised the issue of dependent samples often ignore the problem by selecting a subsample of the data to create either an independent sample or a completely dependent sample, with both Kmietowicz (1984) and Sarabia et al. (2005) guilty of the latter. This is of concern as the disregard of large proportions of available data creates the potential for the marginal distributions of income to be estimated inaccurately. Intuitively, a solution would be the use of a partially dependent sample, which contains both the dependent observations within a panel as well as the independent observations.

This paper seeks to address the issue of sample dependence by applying various bivariate distributions, with Singh–Maddala or Dagum marginals, to a partially dependent sample of household income for two (non-consecutive) years, with a view to assessing the changes in inequality and poverty over that period. For ease of analysis only the bivariate case is being considered in this paper. One of the distributions suggested is the bivariate Singh–Maddala distribution proposed by Takahasi (1965). The appeal of this distribution is that both the marginal and conditional distributions follow a univariate Singh–Maddala specification (Kleiber and Kotz, 2003). Other bivariate distributions being considered belong to the copula class of multivariate distributions. Using copulas to model multivariate distributions is extremely popular in the finance and actuarial context, particularly for capturing dependence amongst stocks. This approach is appealing as copulas are easily estimated using maximum likelihood techniques, and there are many variants available in the literature which capture a wide range of dependence structures beyond simple correlation. In addition, copulas are flexible in that they can be applied to any specification of the marginal distribution, including allowing for the marginal distributions to have different specifications. This provides an attractive method for capturing the dependence structure contained in the joint distribution of income under partially dependent samples.

Each of the above bivariate income distributions is estimated via full information maximum likelihood using income data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey for 2001 and 2005. Once the parameters for each bivariate income distribution have been estimated, values for various measures of inequality and poverty are obtained for each marginal distribution along with their differences, enabling the assessment of changes in such measures over time. More specifically, the summary measures to be considered in this paper include mean income and modal income, the Gini inequality coefficient and the headcount ratio poverty measure. In addition, the standard errors of each of the differences, which are of particular interest in this analysis, are calculated using the delta method. For comparative purposes, estimates of each measure will also be obtained for the Singh–Maddala and Dagum marginal distributions which are estimated as univariate distributions under independence.

The remainder of this paper is organised as follows. Section 2 discusses and defines the concept of a partially dependent sample. Specifications for each of the bivariate distributions proposed for the joint distribution of income, along with the inequality and poverty measures considered in this analysis are provided in Section 3. Section 4 defines a likelihood function for partially dependent samples with particular emphasis on the likelihood function for a copula. Section 5 summarises the characteristics of the HILDA panel data used in the analysis. Empirical results of the analysis as applied to the data, including parameter estimates, tests for independence, and estimates for the inequality and poverty measures are presented and discussed in Section 6; conclusions appear in Section 7.

2. Partially dependent samples

Given the nature of panel studies, it is difficult to maintain a completely dependent sample over an extensive period of time, as members will typically enter and exit the panel from year to year. Therefore, if a sample of income is taken from any two years of a panel, there will be members of the panel common to both years as well as members who have only participated in one year. This feature makes it impossible to obtain a completely dependent sample without discarding valuable data, inhibiting the accurate approximation of the marginal income distributions for each year. In order to include all of the available data when estimating dependent income distributions, a partially dependent sample should be used, where those members of the panel which remain in both years are paired and treated as dependent, but those members who have only recorded data in one year are retained in the sample and treated as independent. The concept of a partially dependent sample is further defined as follows.

Let y_1 be a sample of income for one particular year of a panel data set, with a sample size denoted by n_1, and y_2 a sample of income from a subsequent year of the panel, with a sample size, n_2. Note that n_1 and n_2 need not be equal. Suppose there are k members of the panel which have recorded incomes in both years, where $k \leq \min(n_1, n_2)$. These observations can be matched (and ordered for ease of notation), giving $\{(y_{1,1}, y_{2,1}), \ldots (y_{1,k}, y_{2,k})\}$ paired observations which follow some bivariate income distribution, $f(y_1, y_2)$. With respect to the remaining observations in the sample which are not matched with another observation from the other year, $(y_{1,k+1}, \ldots, y_{1,n_1})$ is independent of y_2 and $(y_{2,k+1}, \ldots, y_{2,n_2})$ is independent of y_1. In addition, it should be noted that $(y_{1,1}, \ldots, y_{1,k})$ and $(y_{1,k+1}, \ldots, y_{1,n_1})$ are observations from the same marginal income density, $f_1(y_1)$. Similarly, $(y_{2,1}, \ldots, y_{2,k})$ and $(y_{2,k+1}, \ldots, y_{2,n_2})$ come from the same marginal income density, given by $f_2(y_2)$. The main objective of this