Income inequality dynamic measurement of Markov models: Application to some European countries

Guglielmo D'Amico [a,*], Giuseppe Di Biase [a], Raimondo Manca [b]

[a] *Drug Sciences Department, "G. D'Annunzio", University of Chieti-Pescara, via dei Vestini, 31, 66013 Chieti Italy*
[b] *MEMOTEF Department, "La Sapienza" University of Rome, via del Castro Laurenziano, 9, 00161 Roma Italy*

ARTICLE INFO

ABSTRACT

In this paper we present a methodology for measuring income inequality dynamically within a Markov model of income evolution. The proposed methodology requires knowledge of the evolution of the population and the averages and medians of the incomes in a country and allows the computation of dynamic inequality indices. The methodology is supported with statistics from Eurostat data applied on France, Germany, Greece and Italy.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The themes of inequality and wealth concentration have been extensively studied by means of static inequality indices, see e.g. Athanasopoulos and Vahid (2003) and Kooleman and Van Doorslaer (2004).

One of the most common indices is Theil's Entropy (Theil, 1967).

The index is scale independent and invariant to replication of population, it satisfies the strong principle of transfers and it is additively decomposable, see e.g. Athanasopoulos and Vahid (2003) and Cowell (1995).

Many papers make use of Markov Chain modelling to describe income dynamics (Bickenbach and Bode, 2003; Quah, 1993, 1994). Nevertheless these contributions did not propose measuring inequality through dynamic indices. This drawback has been overcome in the recent paper (D'Amico and Di Biase, 2010), where the authors proposed dynamic inequality indices. This generalisation was made possible by considering a population that evolves over time according to a semi-Markov process and by considering the income of each economic agent as a reward process. Therefore it is possible to justify changes in indices when the population composition varies over time. The model was implemented in D'Amico et al. (2011) to simulate an artificial economic system with immigration; however critical

points that occur when handling with real-world applications were never faced.

The main problem of applying the model proposed in D'Amico and Di Biase (2010) is that of data availability. Indeed, the model requires microdata concerning income evolution of agents, which are very often unavailable to researchers. To overcome this problem we propose in this paper an application methodology that makes possible the application of the model when only the averages and medians evolution of the incomes in a country are available. The methodology is considered for a Markov Chain model of income evolution.

The results showed different types of temporal evolutions of the index in the considered countries, suggesting the strong necessity to implement an economic integration European policy.

The paper is organised as follows: Section 2 describes the stochastic model, the Dynamic Theil's Entropy and the computation of its expectation in the transient and in the asymptotic cases. Section 3 presents data and methodology of application. Section 4 provides the results. Finally, some conclusions and further developments can be found in Section 5.

2. The stochastic model

We briefly describe the model by following the research lines indicated in D'Amico and Di Biase (2010) and D'Amico et al. (2011), in this section, for the sake of completeness.

Let us assume a system of N economic agents. Suppose that each agent produces at time t I a quantity $y_i(t)$ of income.

We classify each agent by allocating it, at each time, in one of K mutually exclusive classes of income $E = \{C_1, C_2, ..., C_K\}$ through an

---

* Corresponding author.
 E-mail addresses: g.damico@unich.it (G. D'Amico), dibiase@unich.it (G. Di Biase),
raimondo.manca@uniroma1.it (R. Manca).

allocation map as in Section 3. This permits recovering N time series of states of E:

$$
\begin{matrix}
C^1(0) & C^1(1) & C^1(2) & \cdots & C^1(T) \\
C^2(0) & C^2(1) & C^2(2) & \cdots & C^2(T) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
C^N(0) & C^N(1) & C^N(2) & \cdots & C^N(T)
\end{matrix}
\tag{1}
$$

where $C^r(t)$ represents the income class of the rth agent at time t.

Suppose that such time series are realisations of a discrete time Markov Chain with transition probability matrix $\mathbf{P}$ whose element P_{ij} denotes the probability that an agent, now allocated in C_i, will enter the next allocation C_j.

We assume that each agent allocated in class $C_j \in E$ produces an income equal to y_{C_j}. A possible choice for the quantity y_{C_j} is illustrated in Section 3.

Let $t = 0$ be the initial observation time, $n_{C_i}(0)$ be the number of agents in class C_i at time zero and $\underline{n}(0) = \{n_{C_1}(0), n_{C_2}(0), \ldots, n_{C_K}(0)\}$ be the population structure at time zero. Moreover let $a_{C_j}(\underline{n}(0))$ be the initial share of income due to class C_j.

We define the process $\underline{a}(\underline{n}(t)) = (a_{C_1}(\underline{n}(t)), a_{C_2}(\underline{n}(t)), \ldots, a_{C_K}(\underline{n}(t)))$, which describes the time evolution of the shares of income among the classes of population:

$$
a_{C_j}(\underline{n}(t)) = \frac{n_{C_j}(t) y_{C_j}}{<\underline{n}(t), \underline{y}(t)>},
\tag{2}
$$

where $\underline{y}(t) = \left\{ y_{C_1}, y_{C_2}, \ldots, y_{C_K} \right\}$ is the vector of income production, $\underline{n}(t) = \left\{ n_{C_1}(t), n_{C_2}(t), \ldots, n_{C_K}(t) \right\}$ is the multivariate stochastic process describing the evolution of the population in time and $<\cdot, \cdot>$ is the usual scalar product.

Theil's Entropy is defined in Theil (1967) by

$$
T_e = \sum_{i=1}^{K} a_i (\log K a_i).
\tag{3}
$$

Given the population configuration $\underline{n}(0)$ and the vector of average incomes $\underline{y}(t)$, the Dynamic Theil's Entropy is the stochastic process:

$$
T_e(t) := \sum_{i=1}^{K} a_{C_i}(\underline{n}(t)) \left(\log K \, a_{C_i}(\underline{n}(t)) \right).
\tag{4}
$$

This index is defined as the sum of the products of the shares of the total income of each class multiplied by the logarithm of $K \, a_{C_i}(\underline{n}(t))$, where K is the number of the classes of the agents in the economic system. Notice that in Eq. (4) the shares of income are random processes rather than constant deterministic numbers. If the shares of production are known constant in time, then the classic Theil's index is recovered as a particular case from Eq. (4).

The range of values is between 0 and $1\text{n}(K)$. At a fixed time t, the index is 0 when the wealth at time t is equidistributed among the classes, whereas it reaches the value of $1\text{n}(K)$ when one class holds all the wealth.

We may summarize the process by computing the first order moment:

$$
\begin{aligned}
E[T_e(t)] &= \sum_{i=1}^{K} \sum_{\underline{n}' \in p.c.} Pr\left[\underline{n}(t) = \underline{n}' \big| \underline{n}(0) = \underline{n} \right] a_{C_i}(\underline{n}'(t)) \left(\log K \, a_{C_i}(\underline{n}'(t)) \right) \\
&= \sum_{i=1}^{K} \sum_{\underline{n}' \in p.c.} \frac{N!}{\prod_{h=1}^{K} n'_{C_h}!} \prod_{h=1}^{K} (P_h(t))^{n'_{C_h}} \left(\frac{\log K a_{C_i}(\underline{n}'(t))}{a_{C_i}^{-1}(\underline{n}'(t))} \right),
\end{aligned}
\tag{5}
$$

where p.c. is the set of all possible population configurations,

$$
P_i(t) = \sum_{h=1}^{K} \frac{n_h(0)}{N} p_{hi}^{(t)}
$$

and $p_{hi}^{(t)}$ are the t-step transition probabilities of the Markov Chain.

If the Markov Chain is ergodic with limit distribution, i.e.

$$
\Pi = (\pi_1, \pi_2, \ldots, \pi_K),
$$

then it is possible to determine the limit behaviour of the expected Dynamic Theil's Entropy as follows:

$$
\lim_{t \to \infty} E[T_e(t)] = \sum_{i=1}^{K} \sum_{\underline{n}' \in p.c.} \frac{N!}{\prod_{h=1}^{K} n'_{C_h}!} \prod_{h=1}^{K} (\pi_h)^{n'_{C_h}} \left(\frac{\log K a_{C_i}(\underline{n}')}{a_{C_i}^{-1}(\underline{n}')} \right).
\tag{6}
$$

The dynamic Theil's index is able to capture the randomness in the inequality evolution and, consequently, through its first moment Eq. (5), it provides a function that is an effective tool for forecasting wealth distribution for a given horizon time.

We would like to remark that the model considered in D'Amico and Di Biase (2010) relies on a semi-Markovian hypothesis. Here, with the aim of producing a real data application, due to the type of available data, the simpler Markovian assumption is done.

Finally, our model gives a complete description of the probabilistic evolution of the shares of income and it is able to study all kinds of functionals of the variables $a(n(t))$. For this reason the model can be easily extended to other classical inequality indices, like those by Gini (1912) and Hirschman (1964), since such indices also depend on the shares of income through specific functional relations different from Eq. (4).

3. Data and methodology of application

In order to apply the model we should retrieve time series of microdata concerning agent income evolution as in Eq. (1). Unfortunately, these data are rarely available to researchers. For this reason, we developed an application methodology of the proposed model that requires, for a given country, only the evolution of the total population, the averages and the medians of the incomes.

Data were downloaded from the Eurostat website at http://epp.eurostat.ec.europe.eu/. We executed two research paths. The former acquires data concerning incomes and the latter data concerning population. Both paths start by selecting the tree level *Statistics Database by Themes*.

We downloaded population, means and medians equivalised net income for France, Germany, Greece and Italy from the year 2005 to the year 2008 stratified in five age groups.

Table 1 summarizes data for the year 2005 for France, Germany, Greece and Italy.

In order to implement the model by means of our macrodata we propose the following step by step procedure.

3.1. Step 1: recovering the income distribution

Assume that incomes follow a lognormal distribution function. Indeed, incomes within countries typically follow a skewed distribution with a long high tail. Descriptive studies by offices for national statistics of different countries have found that national income distributions fit a lognormal distribution significantly. This hypothesis allows recovery of the parameters of the distribution once we know its Expectation and Median.