Confidence sets for dynamic poverty indexes

Guglielmo D’Amico a and Riccardo De Blasis b

aDepartment of Economics, Università ‘G. D’Annunzio’, Chieti, Italy; bDepartment of Management, Finance, and Technology, LUM University, Casamassima (BA), Italy

ABSTRACT
In this study, we consider different poverty indexes in a dynamic framework where individuals change their rate of income randomly in time. The primary objective of this paper is to assess the accuracy of the approximation of the indexes that can be obtained by applying the strong law of large numbers to an economic system composed of an infinite number of agents. The main result is a multivariate central limit theorem for dynamic poverty measures, which is obtained applying the theory of U-statistics. We also show how to get the confidence sets for the considered dynamic indexes, which show the appropriateness of the model. An application to the Italian income data from 1998 to 2012 confirms the effectiveness of the considered approach and the possibility to determine the evolution of poverty and inequality in real economies.

1. Introduction
Economists and econometricians have dedicated a lot of effort to the investigation of poverty. The literature is vast and comprehends theoretically oriented contributions as well as applied researches. The advancement to more powerful indexes of poverty is always of interest and it aims at capturing specific peculiarities of the phenomenon that were ignored by previous indicators (see, e.g.[15]). In particular, multidimensional measures of poverty are relevant in this context. They include nonmonetary sources of deprivation that affect the well-being of individuals and households such as disability, exposure to environmental hazards, and limited availability of healthcare services (see, e.g.[1,24,25]). Besides, other authors based their measures on methodologies derived from data analysis. One example is the use of the clustering techniques proposed by Kana Zeumo et al. [18]. This research stream has its origin in the seventies of the last century when the stately contribution by Professor Sen appeared [28]. Since then, continuous improvements and generalizations have been made (see, e.g.[14,29,30]) included poverty measures based on a holistic and system modeling approach as proposed in [19] and the adoption of a multiple classification method able to recognize spatial properties of poverty [26].
Almost at the same time, it was recognized that poverty is not a static notion and that this characteristic should be investigated in relation to time. The main idea is to statistically assess the frequencies of poverty conditions and their variations through time. These statistical properties of poverty mobility were determined and confirmed in several studies (see, e.g. [2, 11, 31, 32]). In general, understanding the features of poverty over time needs the adoption of a stochastic model of income evolution. Therefore, Markov chain models were frequently used. Some examples are available in McCall [23], Breen and Moisio [5], Cappellari and Jenkins [6], Formby et al. [13], and Langheine and Pol [20]. In addition, poverty rates and transition probabilities have been estimated in relation to noisy data by Lee et al. [21]. A recent contribution by Hobza et al. [16] has focused on the change between the poverty proportions of two consecutive years by applying a unit-level temporal binomial-logit mixed model.

In contrast to previous research, the idea of advancing dynamic indexes of poverty and income inequality is relatively new and limited to a few contributions. Precisely, to the authors’ knowledge, two approaches can be identified. They share the same starting idea, namely, the extension of static indicators into a dynamic framework, but then move away from each other in relation to methods and results to answer different questions. On the one hand, Ewald and Yor [12] consider a sequence of distributions parameterized in time and they look for conditions under which the corresponding sequence of the indicator (in the specific case, the Gini index) increases over time. On the other hand, starting from the work by D’Amico and Di Biase [7], D’Amico et al. [8] and D’Amico et al. [9], an economic system is built up by advancing a set of assumptions on the time evolution of income for every agent who is member of the economic system. This approach was also adopted in D’Amico and Regnault [10] in relation to dynamic measures of poverty where the dynamic indexes were evaluated both for finite and infinite size economic systems. Specifically, the infinite size system (i.e. an economy with an infinite number of agents) is revealed to be particularly interesting. In fact, for each index, using probabilistic arguments (i.e. strong law of large numbers), it is possible to shape a deterministic function (of the parameters of the model) to which the index of the real economy converges to.

In this paper, we move further steps in this direction. First, we extend the computation of the dynamic Gini index including the inequality within each class of poverty where people are classified according to their income. This extension also impacts the Sen index that is a function of the Gini index. Second, we establish a multivariate central limit theorem to whom the dynamic indexes obey. This result is obtained by using the fundamental limit theorems for U-statistics presented in Hoeffding [17] and successively applied by Xu [33] for static poverty indexes. Thus, we generalize the contribution by Xu showing how the theory of U-statistics allows the study of poverty indexes also in a dynamic framework. The multivariate central limit theorem gives us the possibility to determine confidence sets, i.e. bounds that at a fixed probability level expressing the goodness of the approximations based on the strong law of large numbers. These results derive from the specific assumptions defining the model which are based on the probabilistic equality of the incomes of people belonging to the same income class and that may migrate in time from one class to another according to a continuous-time Markov process. The confidence sets are estimated consistently by using empirical estimators of the model’s parameters. Finally, we present an application of the aforementioned probabilistic approximations on the Italian income data provided by the Italian Central Bank from 1998 to 2012, which contains information
about family net disposable incomes and household members. The results of the application suggest the effectiveness of the considered approach and confirm the possibility to apply it for the determination of the evolution of poverty and inequality in real economies.

The remainder of the paper is organized as follows. Section 2 sets out the assumptions that define the model and presents the main theoretical results of the paper including probabilistic approximations of the indexes and their confidence sets. Section 3 illustrates the result of the application to real data and demonstrates the adequacy of the proposed approach to the investigation of the evolution in time of dynamic indexes of poverty in real economic systems. Section 4 summarizes our contributions and results. All proofs are deferred to the Supplemental Material.

2. The stochastic model and confidence sets for dynamic poverty indexes

In this section, we present the mathematical model. First, we introduce the stochastic model of income evolution and the dynamic version of four poverty indexes in the case of infinite size economic systems. Then, we derive the confidence sets for the poverty indexes. They are obtained by proving the central limit theorem for the stochastic processes expressing the dynamic poverty indexes.

2.1. Dynamic poverty indexes

Following D’Amico and Regnault [10], we consider an economic system composed of a set $H$ of $N$ individuals. The rate of income produced by each economic agent evolves randomly in time and can be described through a stochastic process $Y = (Y_h(t))_{t \in \mathbb{R}_+}$, where $h$ denotes the $h$th individual in the economic system and $t$ is the time variable. For our purposes, we classify individuals according to their rate of income in one of three exhaustive and exclusive income classes denoted by a random process $C_h(t)$ such that

$$
C_h(t) := \begin{cases} 
C_1 & \text{if } Y_h(t) \leq y_{ep}, \\
C_2 & \text{if } y_{ep} < Y_h(t) \leq y_p, \\
C_3 & \text{if } Y_h(t) > y_p,
\end{cases}
$$

where $y_p$ and $y_{ep}$ are the poverty and extreme poverty threshold rates, respectively. Clearly, the possibility to extend the model to multiple richness classes is straightforward.

In the remainder of the paper, the simplifying notation $\{1, 2, 3\}$ will be used to denote the set $\{C_1, C_2, C_3\}$.

The following assumptions advanced in D’Amico and Regnault [10] define the model:

A1: the number $N$ of individuals in the economic system is finite and constant in time;
A2: the income rate processes $(Y_h)_{h \in H}$ are independent and hence the class allocation processes $(C_h)_{h \in H}$;
A3: the processes $(C_h)_{h \in H}$ are identically distributed ergodic Markov processes taking values in the set $\{C_1, C_2, C_3\}$ with infinitesimal generator matrix $\Lambda$;
A4: for any time $t \in \mathbb{R}$ and any individual $h \in H$, the conditional distribution of the rate of income $Y_h(t)$ knowing that $C_h(t) = C_i$, with $C_i \in E$, does not depend on past income values, nor on $t$ or $h$. We denote it as $F_i$ and we assume that it admits pdf $f_i$. 


and possesses finite first and second order moments. In symbol,
\[ D(Y_h(t) \mid \sigma_{t^-}(Y_h), C_h(t) = C_i) = D(Y_h(t) \mid C_h(t) = C_i) =: F_i(\cdot) \quad \forall t \in \mathbb{R}, \forall h \in \mathcal{H}, \]
where \( \sigma_{t^-}(Y_h) := \lim_{s \to t^-} \sigma_s(Y_h) \) is the sigma-algebra generated by the income process of agent \( h \) up to time \( t \) but excluding it.

Now, according to D’Amico and Regnault [10] we present the stochastic extension of the poverty indexes. To this end, we denote by
\[ P(t) = \{ h \in \mathcal{H} : Y_h(t) \leq y_p \}, \]
the set of poor agents at time \( t \) and by
\[ n(t) = \{ n_1(t), n_2(t), n_3(t) \}, \quad t \in \mathbb{R}_+, \]
the multivariate counting process denoting the composition of the income classes in time. Precisely, \( n_i(t) \) is the number of individuals allocated in class \( C_i \) at time \( t \).

**Definition 2.1:** The Dynamic Headcount ratio, The Dynamic Income-gap ratio, the Dynamic Gini and the Dynamic Sen Index are defined as follows:
\[ H(t) := \frac{n_1(t) + n_2(t)}{N}, \]
\[ I(t) := 1 - \frac{\sum_{h \in P(t)} Y_h(t)}{y_p(n_1(t) + n_2(t))}, \]
\[ G(t) := \frac{\sum_{h \in P(t)} \sum_{l \in P(t)} | Y_h(t) - Y_l(t) |}{2(n_1(t) + n_2(t))}\sum_{h \in P(t)} Y_h(t)), \]
\[ S(t) = H(t) \cdot [I(t) + (1 - I(t)) \cdot G(t)]. \]

Specifically, the headcount index at time \( t \) is defined as the ratio between the number of the poor \( n_1(t) + n_2(t) \) and the population size \( N \). Formally, the income gap ratio is defined as the percentage of the income short-fall of people below the poverty line from the poverty level. Formula (2) can be represented also as
\[ I(t) = \frac{\sum_{h \in P(t)} (y_p - Y_h(t))}{y_p(n_1(t) + n_2(t))}, \]
where the numerator indicates the missing income to the poor agents to exit the poverty condition and the denominator is the poor agents income as if they had an income level equal to the poverty threshold. Finally, the Gini index among the poor is expressed algebraically as one half of the mean of the absolute value of the differences between all pairs of poor agents income relative to their mean income.

Although the previous indexes share the same functional form with their static counterparts, they are of different nature being stochastic processes due to the randomness of the counting process \( n(t) \) and of the rates of income \( \{ Y_h(\cdot) \}_{h \in P(t)} \). As a consequence, also the set of poor agents \( P(t) \) at every instant of time \( t \) is random.

This set of assumptions defines an economic system that describes the evolution of people according to their income. The study of this system is very complex and since the number of involved individuals \( N \) is very large, it requires a big computational effort.
An alternative strategy has been implemented in D’Amico and Regnault [10] where assumption A1 is relaxed in favor of a new assumption:

A1': (large-size population) the number of individuals \( N \) in the economy is large enough to be considered as infinity.

This new hypothesis allows us to use stochastic approximations based on fundamental limit theorems of probability theory.

Next proposition is the first result of this strategy:

**Proposition 2.2:** Under assumptions A1’–A4, we have that

\[
\mathcal{H}(t) \xrightarrow{a.s.} \mathcal{H}_\infty(t) = H(\mu, \Lambda, t) := \mu' (\mathbf{P}_1(t) + \mathbf{P}_2(t)),
\]

where \( \mu' \) is the transpose of the initial distribution and \( \mathbf{P}_1(t) \) and \( \mathbf{P}_2(t) \) are the first and second column of \( \mathbf{P}(t) = \exp(t\Lambda) \), respectively.

Similarly, the Dynamic Income-gap ratio \( \mathbb{I}(t) \), the Dynamic Gini index \( \mathcal{G}(t) \), and the Dynamic Sen index \( \mathcal{S}(t) \) converge almost surely to

\[
\mathbb{I}_\infty(t) := 1 - \frac{y_1 \mu' \mathbf{P}_1(t)}{y_p \mathcal{H}_\infty(t)} - \frac{y_2 \mu' \mathbf{P}_2(t)}{y_p \mathcal{H}_\infty(t)},
\]

\[
\mathcal{G}_\infty(t) := \frac{(\mu' \mathbf{P}_1(t))^2 z_1 + 2(y_2 - y_1)(\mu' \mathbf{P}_1(t))(\mu' \mathbf{P}_2(t)) + (\mu' \mathbf{P}_2(t))^2 z_2}{2\mathcal{H}_\infty(t)(y_1 \mu' \mathbf{P}_1(t) + y_2 \mu' \mathbf{P}_2(t))}
\]

\[
\mathcal{S}_\infty(t) := \mathcal{H}_\infty(t) \cdot [\mathbb{I}_\infty(t) + (1 - \mathbb{I}_\infty(t)) \cdot \mathcal{G}_\infty(t)],
\]

where \( y = (y_1, y_2) \) is the vector of mean incomes per poor classes and

\[
\bar{z}_1 := \int_{y_p}^{y_{cp}} \int_{y_p}^{y_{cp}} |y - x| \, dF_1(y) \, dF_1(x), \tag{5}
\]

\[
\bar{z}_2 := \int_{y_p}^{y_{cp}} \int_{y_p}^{y_{cp}} |y - x| \, dF_2(y) \, dF_2(x). \tag{6}
\]

**Proof:** See Supplemental Material.

**Remark 2.1:** Proposition 2.2 was already demonstrated in D’Amico and Regnault [10]. However, with respect to the Gini index, and in turn to the Sen index, the proof was limited to the case of equivalence of the incomes of people belonging to the same class while the hypotheses of the model advance only the equivalence of their probability distributions. The proof we provide in this paper overcomes this limitation with the addition of inequality within each class.

### 2.2. Central limit theorem for dynamic poverty indexes

The next step forward in the global understanding of the time evolution of the dynamic poverty indexes is the assessment of specific central limit theorems for each index and the subsequent derivation of the confidence sets. The confidence sets are centered on
the asymptotic values $\mathbb{H}_\infty(t), \mathbb{I}_\infty(t), \mathbb{G}_\infty(t), \mathbb{S}_\infty(t)$ obtained in Proposition 2.2 and have amplitudes proportional to their variances which we are going to compute. This finding represents the main result. However, first we anticipate two auxiliary lemmas which are useful tools for obtaining the proof of our main result.

**Lemma 2.3:** For any $h \in \mathcal{H}$ and for every $t \in \mathbb{R}$, let $F(t; x) := \mathbb{P}[Y_h(t) \leq x]$. The cumulative distribution function (cdf) $F(t; \cdot)$ is continuous everywhere and is given by

$$F(t; x) = F_1(x)\mu'(\mathbb{P}_1(t)) + F_2(x)\mu'(\mathbb{P}_2(t)) + F_3(x)\mu'(\mathbb{P}_3(t)).$$

Accordingly, for every $r \geq 1$ it results that

$$\mathbb{E}[Y_h(t) | Y_h(t) < y_p] = \frac{y_1^{(r)}\mu'(\mathbb{P}_1(t)) + y_2^{(r)}\mu'(\mathbb{P}_2(t))}{\mu'(\mathbb{P}_1(t)) + \mu'(\mathbb{P}_2(t))},$$

where $y_1^{(r)} := \mathbb{E}[(Y_1)^r]$ and $y_2^{(r)} := \mathbb{E}[(Y_2)^r]$.

**Proof:** See Supplemental Material.

The random variable $Y_h(t)$ is a mixture of the rates of income of the three considered classes. The weights are given by the probability to find the individual in a given class at time $t$.

Since the asymptotic distributions of poverty indexes in the static case have been determined by Xu [33] using the theory of U-statistics originally developed by Hoeffding [17], we are going to introduce now some statistical functionals of the distribution $F(t; \cdot)$ which will play a fundamental role in the determination of the central limit theorem of the dynamic indexes of poverty. Whenever possible, our notation will follow the one in [33].

Let us denote by

$$\Theta_1(t) = \int_0^{y_p} dF(t; y) = \mu'(\mathbb{P}_1(t) + \mathbb{P}_2(t)),$$

$$\Theta_2(t) = \int_0^{y_p} y dF(t; y) = y_1^{(1)}\mu'(\mathbb{P}_1(t)) + y_2^{(1)}\mu'(\mathbb{P}_2(t)),$$

and set

$$\zeta(\Theta_2(t)) = \Theta_2^{(2)}(t) - (\Theta_2(t))^2,$$

where $\Theta_2^{(2)}(t) = y_1^{(2)}\mu'(\mathbb{P}_1(t)) + y_2^{(2)}\mu'(\mathbb{P}_2(t))$.

Further statistical functional are needed to be considered and computed in terms of the model’s parameters. This is the content of the next lemma.
Lemma 2.4: For any \( h \in \mathcal{H} \) and for every \( t \in \mathbb{R} \), denote by

\[
\Theta_3(t) = \int_0^{y_p} \int_0^{y_p} |y - x| \, dF(t; y) \, dF(t; x),
\]

then

\[
\Theta_3(t) = z_1(\mu'_P.1(t))^2 + 2(y_2 - y_1)(\mu'_P.1(t))(\mu'_P.2(t)) + z_2(\mu'_P.2(t))^2.
\]

Moreover, the quantity

\[
\zeta(\Theta_3(t)) = \int_0^{y_p} \left[ \int_0^{y_p} |y - x| \, dF(t; x) \right]^2 dF(t; y) - \Theta_3^2(t),
\]

can be obtained using

\[
\int_0^{y_p} \left[ \int_0^{y_p} |y - x| \, dF(t; x) \right]^2 dF(t; y)
\]

\[
= (\mu'_P.1(t))^3 \int_0^{y_p} \left[ \int_0^{y_p} |y - x| \, dF_1(x) \right]^2 dF_1(y) + (\mu'_P.2(t))^2(\mu'_P.1(t))(y_1^{(2)} - 2y_1y_2 + y_2^2) + (\mu'_P.1(t))^2(\mu'_P.2(t))(y_2^{(2)} - 2y_1y_2 + y_1^2)
\]

\[
+ (\mu'_P.2(t))^3 \int_0^{y_p} \left[ \int_0^{y_p} |y - x| \, dF_2(x) \right]^2 dF_2(y).
\]

Finally, the quantity

\[
\zeta(\Theta_2(t), \Theta_3(t)) = \int_0^{y_p} \int_0^{y_p} y|y - x| \, dF(t; y) \, dF(t; x) - \Theta_2(t)\Theta_3(t),
\]

can be evaluated using

\[
\int_0^{y_p} \int_0^{y_p} y|y - x| \, dF(t; y) \, dF(t; x)
\]

\[
= (\mu'_P.1(t))^2 \int_0^{y_p} \int_0^{y_p} y|y - x| \, dF_1(y) \, dF_1(x) + (\mu'_P.2(t))^2 \int_0^{y_p} \int_0^{y_p} y|y - x| \, dF_2(y) \, dF_2(x)
\]

\[
+ (\mu'_P.1(t))(\mu'_P.2(t))(y_2^{(2)} - y_1^{(2)}).
\]

Proof: See Supplemental Material.

To obtain the confidence sets for the aforementioned poverty indexes, we are going to express them according to three U-statistics as done by Xu [33] for the static case. The notable difference here is that our U-statistics are time-dependent as the indexes are too.
We then prove that the results in [33] can be effectively extended to our more general framework and evaluated according to the model parameterization. Consequently, define by

$$U_1(t) = \frac{\sum_{h=1}^{N} 1_{[G_h(t) \in [C_1, C_2]]}}{N} = \frac{n_1(t) + n_2(t)}{N} = \mathbb{H}(t).$$

Introduce a second U-statistic

$$U_2(t) = \frac{\sum_{h=1}^{N} Y_h(t) 1_{[G_h(t) \in [C_1, C_2]]} - Y_l(t) 1_{[G_h(t) \in [C_1, C_2]]}}{N} \cdot \frac{\sum_{h=1}^{N} Y_h(t) 1_{[G_h(t) \in [C_1, C_2]]}}{y_p \mathbb{H}(t) N} = (1 - \mathbb{P}(t)) y_p U_1(t) \Rightarrow (1 - \mathbb{P}(t)) = \frac{U_2(t)}{U_1(t) y_p}.$$

Finally, introduce a third U-statistic

$$U_3(t) = \frac{\sum_{h=1}^{N} \sum_{l=1}^{N} Y_h(t) - Y_l(t) 1_{[G_h(t) \in [C_1, C_2]]} 1_{[G_l(t) \in [C_1, C_2]]}}{N(N - 1)} \cdot \frac{2|\mathcal{P}(t)|}{2|\mathcal{P}(t)|} \cdot \frac{\sum_{h=1}^{N} Y_h(t) 1_{[G_h(t) \in [C_1, C_2]]}}{N(N - 1)} = 2 \mathbb{G}(t) \mathbb{H}(t) \cdot \frac{N}{N - 1} \Rightarrow \mathbb{G}(t) = \frac{N - 1}{N} \cdot \frac{U_3(t)}{2U_1(t) U_2(t)}.$$

It is simple to realize that the asymptotic behavior of the dynamic Gini index is equivalent to that of the function $\frac{U_3(t)}{2U_1(t) U_2(t)}$. Thus, this index is hereafter approximated by

$$\tilde{\mathbb{G}}(t) = \frac{U_3(t)}{2U_1(t) U_2(t)}.$$

Accordingly, the Sen index has the same asymptotic distribution of

$$\mathbb{S}(t) \sim \mathbb{H}(t)[\mathbb{P}(t) + (1 - \mathbb{P}(t)) \tilde{\mathbb{G}}(t)] = U_1(t) \left[ 1 - \frac{U_2(t)}{y_p U_1(t)} + \frac{U_2(t)}{y_p U_1(t)} \frac{U_3(t)}{2U_1(t) U_2(t)} \right] = U_1(t) \left[ \frac{2y_p U_1^2(t) - 2U_1(t) U_2(t) + U_3(t)}{2y_p U_1^2(t)} \right] = U_1(t) - \frac{U_2(t)}{y_p} + \frac{U_3(t)}{2y_p U_1(t)} = f(U_1(t), U_2(t), U_3(t)).$$

Now, first observe that the random variables $\{Y_h(t)\}_{h \in \mathcal{H}}$ share the same distribution for each fixed time $t$. Moreover, the kernels $\Phi_1$, $\Phi_2$ and $\Phi_3$ which correspond to the U-statistics $U_1$, $U_2$ and $U_3$ are independent of the number of agents $N$ being equal to

$$\Phi_1(a_1) = a_1, \quad \Phi_2(a_1, a_2) = \frac{1}{2}(a_1 - a_2), \quad \Phi_3(a_1, a_2) = |a_1 - a_2|.$$
These conditions, jointly with the continuity of the cdf \( F(t; \cdot) \), satisfy the hypotheses of Theorem 7.1 in [17] providing

\[
\sqrt{N}(U_1(t) - \Theta_1(t)), \sqrt{N}(U_2(t) - \Theta_2(t)), \sqrt{N}(U_3(t) - \Theta_3(t)) \sim N[0, \Sigma(t)],
\]

where

\[\Sigma(t) = \begin{pmatrix}
\Theta_1(t)(1 - \Theta_1(t)) & \Theta_2(t)(1 - \Theta_1(t)) & 2\Theta_3(t)(1 - \Theta_1(t)) \\
\Theta_2(t)(1 - \Theta_1(t)) & \zeta(\Theta_2(t)) & 2\zeta(\Theta_2(t), \Theta_3(t)) \\
2\Theta_3(t)(1 - \Theta_1(t)) & 2\zeta(\Theta_2(t), \Theta_3(t)) & 4\zeta(\Theta_3(t))
\end{pmatrix}. \tag{9}
\]

Similarly to the work of Xu [33] applied to static poverty indicators, we define a vector-valued function, for each time \( t \), \( g : \mathbb{R}^3 \to \mathbb{R}^4 \) as follows:

\[
g(x_1, x_2, x_3) = \begin{bmatrix}
x_1 \\
1 - \frac{x_2}{x_3} \cdot \frac{1}{y_p} \\
x_1 \\
x_2 \\
x_3 \\
x_1 = \frac{x_2}{y_p} + \frac{x_3}{2y_p x_1}
\end{bmatrix}.
\]

The functions vector \( g \) is independent from \( N \), moreover, its Jacobian matrix evaluated in a neighborhood of \( \Theta(t) = (\Theta_1(t), \Theta_2(t), \Theta_3(t)) \) is

\[
J_g = \begin{bmatrix}
\nabla^T g_1 \\
\nabla^T g_2 \\
\nabla^T g_3 \\
\nabla^T g_4
\end{bmatrix} = \begin{bmatrix}
\frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \frac{\partial g_1}{\partial x_3} \\
\frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \frac{\partial g_2}{\partial x_3} \\
\frac{\partial g_3}{\partial x_1} & \frac{\partial g_3}{\partial x_2} & \frac{\partial g_3}{\partial x_3} \\
\frac{\partial g_4}{\partial x_1} & \frac{\partial g_4}{\partial x_2} & \frac{\partial g_4}{\partial x_3}
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & \\
\frac{x_2}{x_1} \cdot \frac{1}{y_p} & -\frac{1}{x_3} & 0 & \\
\frac{x_3}{2x_1 x_2} & -\frac{2x_1 x_2}{x_3} & 1 & \\
1 - \frac{2y_p x_1}{x_1} \cdot \frac{1}{y_p} & 0 & 2y_p x_1
\end{bmatrix},
\]

\[
\Rightarrow J_g(t) = J_g|_{x=\Theta(t)} = \begin{bmatrix}
\Theta_1^2(t) \cdot \frac{1}{\Theta_1(t)} & -\frac{1}{y_p \Theta_1(t)} & 0 & 0 \\
-\frac{2\Theta_1^2(t) \Theta_2(t)}{\Theta_3(t)} & -\frac{2\Theta_1(t) \Theta_2^2(t)}{\Theta_3(t)} & 2\Theta_1(t) \Theta_2(t) & 0 \\
1 - \frac{2y_p \Theta_1(t)}{\Theta_3(t)} & -\frac{1}{y_p} & 1 &
\end{bmatrix},
\]

from which it is simple to realize that \( g \) is continuous together with its second-order partial derivatives in some neighborhood of the point \( \Theta(t) = (\Theta_1(t), \Theta_2(t), \Theta_3(t)) \) for each time \( t \). Then, according to Theorem 7.5 in [17] we obtain that

\[
\sqrt{N}(g(U_1(t), U_2(t), U_3(t)) - g(\Theta_1(t), \Theta_2(t), \Theta_3(t))) \xrightarrow{L} N[0, J_g(t) \Sigma(t) J_g^T(t)],
\]

which is a central limit theorem for the considered three time-varying U-statistics.
2.3. Estimation of the covariance matrix

The final step before recovering the confidence sets of the dynamic poverty indexes is the estimation of the covariance matrix $\Gamma(t) = \mathbf{J}g(t)\Sigma_b(t)\mathbf{J}^T_g(t)$. Indeed, this matrix of functions possesses elements that are dependent on the model’s parameters and then should be estimated.

Due to the specification of our dynamic model, it is simple to realize that $\Gamma(t) = \Gamma(\mu, \Lambda, F_1, F_2)$, so that an estimation can be obtained by plugging in the estimators of the model’s parameters into $\Gamma$, i.e. $\hat{\Gamma} = \Gamma(\hat{\mu}, \hat{\Lambda}, \hat{F}_1, \hat{F}_2)$, where

- $\hat{\mu}$ is the MLE of $\mu$, given by
  \[
  \hat{\mu} = \frac{1}{N} \sum_{h=1}^{N} 1_{[C_h(0)=C_i]}, \quad i \in \{1, 2, 3\}
  \]

- $\hat{\Lambda} = (\hat{\lambda}_{ij})_{ij \in E}$ is the MLE of the generator of the Markov process where
  \[
  \hat{\lambda}_{ij} = \begin{cases} 
  \frac{K_{ij}}{R_i} & \text{if } i \neq j, \\
  \sum_{l \neq i} \hat{\lambda}_{il} & \text{if } j = i
  \end{cases}
  \]
  with $K_{ij}$ denoting the total number of jumps from class $C_i$ to $C_j$ and $R_i$ denotes the total time spent in the class $i$ along the $N$ trajectories of the agent’s income evaluation.

- $\hat{F}_i(\cdot)$ is the empirical estimator of the cdf $F_i(\cdot)$.

From these inputs, we have to estimate the statistical functionals $\Theta_1(t)$, $\Theta_2(t)$, $\Theta_3(t)$, $\zeta(\Theta_2(t))$, $\zeta(\Theta_3(t))$ and $\zeta(\Theta_2(t), \Theta_3(t))$. The estimation procedure depends on the adopted observational scheme. In D’Amico and Regnault [10], several estimation schemes for poverty indexes are discussed. However, for brevity reasons, we discuss here only the continuous sampling of class allocation processes. Under this scheme, having observed the evolution of the income process continuously for each agent $h \in \mathcal{H}$ over a time interval $[0, T]$, we can deduce that, as $N \to \infty$,

\[
\hat{\Theta}_1(t) \xrightarrow{a.s.} \Theta_1(t), \\
\hat{\Theta}_2(t) = \hat{y}_1(t) \hat{\mu}_1(t) + \hat{y}_2(t) \hat{\mu}_2(t) \xrightarrow{a.s.} \Theta_2(t),
\]

because $\hat{\mu}$, $\hat{P}_i(t)$ and $\hat{y}_i$ are strongly consistent estimators of the corresponding parameters $\mu$, $P_i(t)$ and $y_i$ (see [10]).

Consider now the term $\bar{z}_1$ in (5) and its corresponding empirical estimator

\[
\hat{\bar{z}}_1 = \int_0^{y_{ep}} \int_0^{y_{ep}} |y - x| d\hat{F}_1(y) d\hat{F}_1(x), \quad (10)
\]
provided an estimator of \( F_1(\cdot) \) is known. To this end, recall that 
\[ F_1(x) = \mathbb{P}[Y_h(t) \leq x | C_h(t) = 1], \forall t \in \mathbb{R}_+, \forall h \in \mathcal{H}. \]
Then we define
\[
A_1^{(h)}(x) = \begin{cases} 
0 & \text{if } x < 0, \\
\int_0^T \mathbf{1}_{\{Y_h(t) \leq x\}} \, dt & \text{if } x \in [0, y_{ep}], \\
R_1^{(h)} & \text{if } x > y_{ep},
\end{cases}
\]
where \( R_1^{(h)} \) is the total length of time the class \( C_1 \) is occupied by individual \( h \) in the sample path \((C_h(t), 0 \leq t \leq T)\). Define now the estimator
\[
\hat{F}_1(x) = \frac{\sum_{h=1}^N A_1^{(h)}(x)}{\sum_{h=1}^N R_1^{(h)}} = \frac{\sum_{h=1}^N A_1^{(h)}(x)}{\sum_{h=1}^N R_1^{(h)}}.
\]
But for \( N \to +\infty \), according to the fact that the \( A_1^{(h)}(x) \) are i.i.d., we have that
\[
\frac{\sum_{h=1}^N A_1^{(h)}(x)}{N} \xrightarrow{a.s.} \mu_1(x, T),
\]
where \( \mu_1(x, T) \) is the expected time an individual \( h \) has an income less than \( x \). Denote by
\[
\mu_1(x, T) = \int_0^T \mathbb{P}[Y_h(t) \leq x] \, dt
\]
\[
= \int_0^T (F_1(x) \mu' \mathbf{P}_1(t) + F_2(x) \mu' \mathbf{P}_2(t) + F_3(x) \mu' \mathbf{P}_3(t)) \, dt,
\]
and after observing that for \( x \in [0, y_{ep}] \) we have \( \hat{F}_2(x) = 0 \) and \( \hat{F}_3(x) = 0 \), we obtain
\[
\mu_1(x, T) = \int_0^T F_1(x) \mu' \mathbf{P}_1(t) \, dt = F_1(x) \int_0^T \mu' \mathbf{P}_1(t) \, dt.
\]
Furthermore, due to the fact that \( \{R_1^{(h)}\} \) are i.i.d., the strong law of large numbers gives
\[
\frac{\sum_{h=1}^N R_1^{(h)}}{N} \xrightarrow{a.s.} \mu_1(y_{ep}, T) = \int_0^T \mathbb{P}[Y_h(t) \leq y_{ep}] \, dt = \int_0^T \mu' \mathbf{P}_1(t) \, dt.
\]
Therefore, the present result suggests that
\[
\hat{F}_1(x) \xrightarrow{a.s.} \frac{\mu_1(x, T)}{\mu_1(y_{ep}, T)} = F_1(x).
\]
A similar reasoning leads to the definition of the estimator of \( F_2(\cdot) \). We define
\[
A_2^{(h)}(x) = \begin{cases} 
0 & \text{if } x < y_{ep}, \\
\int_0^T \mathbf{1}_{\{Y_h(t) \in [y_{ep}, x]\}} \, dt & \text{if } x \in [y_{ep}, y_p], \\
R_2^{(h)} & \text{if } x > y_p,
\end{cases}
\]
where \( R_2^{(h)} \) is the total length of time the class \( C_2 \) is occupied by individual \( h \) in the sample path \((C_h(t), 0 \leq t \leq T)\).
Using similar arguments, we have
\[
\hat{F}_2(x) = \frac{\sum_{h=1}^{N} A_2^h(x)}{\sum_{h=1}^{N} R_2^h} \xrightarrow{\text{a.s.}} F_2(x).
\]

Now, we can estimate \( \bar{z}_1 \) using the empirical estimators as plug-in estimators in formula (10). Note that
\[
|\hat{z}_1 - \bar{z}_1| \\
= |\int_0^{y_{ep}} \int_0^{y_{ep}} |y - x| \ d\hat{F}_1(y) \ d\hat{F}_1(x) - \int_0^{y_{ep}} \int_0^{y_{ep}} |y - x| \ dF_1(y) \ dF_1(x)| \\
\leq \int_0^{y_{ep}} \int_0^{y_{ep}} y_{ep} \left( |d\hat{F}_1(y) \ d\hat{F}_1(x) - dF_1(y) \ dF_1(x)| \right) \\
\leq y_{ep} \int_0^{y_{ep}} \int_0^{y_{ep}} \left( |d\hat{F}_1(x) d\hat{F}_1(y) - dF_1(y) dF_1(x)| + |dF_1(y) d\hat{F}_1(x) - dF_1(x) dF_1(x)| \right) \\
\leq y_{ep} \int_0^{y_{ep}} \int_0^{y_{ep}} d\hat{F}_1(x) |d\hat{F}_1(y) - dF_1(y)| + dF_1(y) |d\hat{F}_1(x) - dF_1(x)|.
\]

Observe now that by Glivenko–Cantelli Theorem, we have that
\[
|d\hat{F}_1(y) - dF_1(y)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } N \to +\infty,
\]
\[
|d\hat{F}_1(x) - dF_1(x)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } N \to +\infty,
\]
thus \( \hat{z}_1 \xrightarrow{\text{a.s.}} \bar{z}_1 \). The same holds true for \( \hat{z}_2 \xrightarrow{\text{a.s.}} \bar{z}_2 \). Therefore, the continuous mapping theorem gives
\[
\hat{\Theta}_3(t) \xrightarrow{\text{a.s.}} \Theta_3(t) \quad \forall t \in \mathbb{R}_+.
\]

Similar arguments can be used to prove that
\[
\zeta(\hat{\Theta}_2(t)) \xrightarrow{\text{a.s.}} \zeta(\Theta_2(t)),
\]
\[
\zeta(\hat{\Theta}_3(t)) \xrightarrow{\text{a.s.}} \zeta(\Theta_3(t)),
\]
\[
\zeta(\hat{\Theta}_2(t), \hat{\Theta}_3(t)) \xrightarrow{\text{a.s.}} \zeta(\Theta_2(t), \Theta_3(t)).
\]

Thus we have the following proposition.

**Proposition 2.5:** Under assumptions A1'-A4, given \( N \) trajectories of the income rate process sampled continuously, the plug-in estimator \( \hat{\Sigma}(t) \) of \( \Sigma(t) \) defined by (9) is strongly consistent as \( N \to \infty \).

**Remark 2.2:** The results of Proposition 2.5 could be adapted to more general sampling schemes such as periodic observations of the trajectories, or the i.i.d censoring times, or trajectories sampled according to a Poisson process.
2.4. Confidence sets for dynamic poverty indexes

Since the confidence sets for the dynamic poverty indexes are functions of the asymptotic covariance matrix, they are strongly and consistently estimated by using the plug-in estimator.

From a practical point of view, we may proceed by introducing the vectors $e_1 = (1, 0, 0, 0)$, $e_2 = (0, 1, 0, 0)$, $e_3 = (0, 0, 1, 0)$, $e_4 = (0, 0, 0, 1)$, $U = (U_1(t), U_2(t), U_3(t))$, and $\Theta = (\Theta_1(t), \Theta_2(t), \Theta_3(t))$. Having found that

$$\sqrt{N}(g(U) - g(\Theta)) \overset{\mathcal{L}}{\to} N(0, \Gamma(t))$$

and recalling that

$$g(U) = \begin{bmatrix}
H(t) = U_1 \\
\Pi(t) = 1 - \frac{U_2}{U_1} \\
G(t) = \frac{U_2}{U_1U_2} \\
S(t) = U_1 - \frac{U_2}{y_p} + \frac{U_3}{2y_pU_1}
\end{bmatrix}, \quad g(\Theta) = \begin{bmatrix}
H_\infty(t) = \Theta_1 \\
\Pi_\infty(t) = 1 - \frac{\Theta_2}{\Theta_1} \\
G_\infty(t) = \frac{\Theta_2}{\Theta_1\Theta_2} \\
S_\infty(t) = \Theta_1 - \frac{\Theta_2}{y_p} + \frac{\Theta_3}{2y_p\Theta_1}
\end{bmatrix},$$

where the time dependency of the elements $U_i$ and $\Theta_i$ with $i = 1, 2, 3$ has been omitted for clarity, we obtain, for the four indexes,

$$\sqrt{N}(H(t) - H_\infty(t)) = e_1 \sqrt{N}(g(U) - g(\Theta)) \overset{\mathcal{L}}{\to} N(0, e_1\Gamma(t)e_1^T),$$

$$\sqrt{N}(\Pi(t) - \Pi_\infty(t)) = e_2 \sqrt{N}(g(U) - g(\Theta)) \overset{\mathcal{L}}{\to} N(0, e_2\Gamma(t)e_2^T),$$

$$\sqrt{N}(G(t) - G_\infty(t)) = e_3 \sqrt{N}(g(U) - g(\Theta)) \overset{\mathcal{L}}{\to} N(0, e_3\Gamma(t)e_3^T),$$

$$\sqrt{N}(S(t) - S_\infty(t)) = e_4 \sqrt{N}(g(U) - g(\Theta)) \overset{\mathcal{L}}{\to} N(0, e_4\Gamma(t)e_4^T).$$

The confidence sets provide important information on the infinite size approximation strategy. Indeed, considering the headcount ratio as a matter of example, it is simple to verify that $e_1\Gamma(t)e_1^T = H_\infty(t)(1 - H_\infty(t))$.

Therefore, we can easily determine $\forall a, b \in \mathbb{R}$ an estimate of the probability $P(a \leq H_N(t) \leq b)$. Indeed, once we denote by $Z$ the standard normal distribution, the following approximation holds:

$$P(a \leq H_N(t) \leq b) \approx P\left(\frac{a - H_\infty(t)}{\sqrt{H_\infty(t)(1 - H_\infty(t))}} \leq Z \leq \frac{b - H_\infty(t)}{\sqrt{H_\infty(t)(1 - H_\infty(t))}}\right) = \Phi\left(\frac{b - H_\infty(t)}{\sqrt{H_\infty(t)(1 - H_\infty(t))}}\right) - \Phi\left(\frac{a - H_\infty(t)}{\sqrt{H_\infty(t)(1 - H_\infty(t))}}\right).$$

Similar arguments can be used to construct confidence sets for other indexes.
3. Empirical application

We test the model on the Italian income data provided by the Italian central bank, Banca d'Italia. The historical database is based on the survey of Italian household budgets from 1977 to 2012 and contains information about the characteristics of the individuals and their household members, along with the family net disposable incomes, which include financial assets. The poverty thresholds are reported by the Italian National Institute of Statistics (ISTAT) and are available from the year 1998 per the number of household components. However, as stated by ISTAT, the data from 1997 to 2012 are not directly comparable with the data from other years due to a substantial change in the design of the survey. Therefore, to have a clean and consistent dataset, we bind our analysis to the years 1997 to 2013. Within this range, household data are available on a biennial basis on even years.

The summary statistics for the net disposable income grouped by number of household components are reported in Table 1. It is worth to mention that the poverty thresholds are not computed based on the income data, to avoid any bias in the results. The first column of both tables indicates the dimension of the household. On average, the household income sharply increases sharply from 1- to 3-person household, with an approximate stability from dimension 3 to 6, followed by a final increase for households bigger than seven persons. A similar pattern is also observable in the income variability.

For the application of the model with two poverty classes, we need to define an extra poverty class. We identify a class of extremely poor households by setting its threshold, $y_{ep}$, at 60% of the poverty threshold, $y_p$. The choice of this threshold is arbitrary, however, results are equivalent for small variations of this threshold. Moreover, to make the incomes comparable between the number of household components and to account for the inflation during the years, we standardize the net disposable income. The standardization by the components is performed each year using the income for 1-component households as the base income. Conversely, the inflation adjustment is performed setting the first year income as the base income. Consequently, the poverty threshold $y_p$ is represented by the 1998 1-component value, i.e. 5479.50, and the extreme poverty threshold $y_{ep}$ is 60% of the previous threshold, i.e. 3287.70.

In addition, we require that all households show an income for each year, thus we exclude households with any missing data, reducing the total number of households to 914, and the number of observed incomes to 7312. Table 2 shows the evolution of the standardized income by the number of household components from 1998 to 2012 and Table 3 reports the summary statistics for the standardized income divided by classes.

| Table 1. Summary statistics of the net disposable income in euro of Italian households from 1998 to 2012 by number of family components. |
|---|---|---|---|---|---|---|---|
| Count | Mean | Std | Min | 25% | 50% | 75% | Max |
| 1 | 14374.0 | 18233.79 | 16525.48 | 0.0 | 10795.80 | 15449.50 | 21877.92 | 810218.64 |
| 2 | 18675.0 | 29414.22 | 22868.09 | 0.0 | 17281.59 | 24662.21 | 35372.21 | 587783.94 |
| 3 | 13279.0 | 36047.38 | 24476.71 | 0.0 | 21451.87 | 31813.74 | 44322.66 | 453843.73 |
| 4 | 12089.0 | 37108.13 | 26956.98 | 0.0 | 21200.00 | 32478.33 | 46372.60 | 1022616.85 |
| 5 | 3488.0 | 35108.13 | 26956.98 | 0.0 | 21200.00 | 32478.33 | 46372.60 | 1022616.85 |
| 6 | 8540.0 | 37331.73 | 28507.22 | 0.0 | 19551.39 | 31298.05 | 48646.07 | 368689.73 |
| 7+ | 211.0 | 41883.71 | 46678.48 | 0.0 | 17351.48 | 31000.00 | 53253.47 | 529872.81 |

Data sourced from the Italian households budgets survey from Banca d’Italia.
Table 2. Evolution of the average standardized income in euro of Italian households from 1998 to 2012 by number of household components.

| Year | 1     | 2     | 3     | 4     | 5     | 6     | 7+    |
|------|-------|-------|-------|-------|-------|-------|-------|
| 1998 | 14138.35 | 14637.67 | 11783.78 | 8676.17 | 8002.68 | 12215.24 |
| 2000 | 15142.21 | 14143.71 | 13850.06 | 9880.23 | 10295.98 | 11524.28 |
| 2002 | 16404.43 | 14963.10 | 15012.10 | 10739.47 | 10419.76 | 23570.36 |
| 2004 | 14711.59 | 14034.27 | 15034.27 | 11121.93 | 8896.92  | 16667.58 |
| 2006 | 15107.63 | 14193.89 | 14034.27 | 11121.93 | 12591.90 | 11667.71 |
| 2008 | 14865.08 | 12012.99 | 14761.12 | 12012.99 | 10948.79 | 12975.73 |
| 2010 | 16509.03 | 12192.64 | 15191.34 | 12192.64 | 11413.25 | 12975.73 |
| 2012 | 15383.99 | 13168.36 | 16472.31 | 13168.36 | 11239.69 | 10961.35 |

Table 3. Summary statistics of the standardized income in euro of Italian households from 1998 to 2012 by poverty class, where $C_1$ is the class of extreme poor households and $C_2$ is the class of poor households.

|       | Count | %    | Mean  | Std   | Min | 25% | 50%  | 75%  | Max  |
|-------|-------|------|-------|-------|-----|-----|------|------|------|
| $C_1$ | 182   | 2.5  | 2136.62 | 977.95 | 0.00 | 1509.42 | 2354.29 | 2941.39 | 3285.68 |
| $C_2$ | 416   | 5.7  | 4488.47 | 613.02 | 3293.41 | 4020.49 | 4529.34 | 5001.68 | 5467.46 |
| $C_3$ | 6714  | 91.8 | 14802.99 | 9023.69 | 5480.92 | 9360.72 | 12870.52 | 17698.43 | 222822.47 |

Now, considering that we do not observe each household income continuously but every two years, we can estimate the generator matrix using the periodic sampling of class allocation processes described in [10]. According to this methodology and with 914 independent trajectories of the class allocation processes $D_i$,  

$$D_1(1998), D_1(2000), \ldots, D_1(2012),$$  

$$\vdots$$  

$$D_{914}(1998), D_{914}(2000), \ldots, D_{914}(2012),$$  

where $D_i(t)$ denotes the income class occupied by household $i$ at year $t$, we can estimate the transition probability matrix $\hat{P} = \hat{p}_{ij}$ as  

$$\hat{p}_{ij} = \frac{K_{ij}}{K_i},$$  

where $K_{ij} = \sum_{k=1}^{914} \sum_{t=0}^{6} I(D_k(2t+1998) = i, D_k(2t+2000) = j)$ is the number of transitions from class $i$ to class $j$, and $K_i = \sum_{j=1}^{3} K_{ij}$ is the total number of times households have been allocated to class $i$.

Then, the maximum likelihood estimator $\hat{\Lambda}$ of the generator matrix $\Lambda$ satisfies the relation $\hat{P} = \exp(\eta \hat{\Lambda})$ and it can be obtained as the logarithm matrix,  

$$\hat{\Lambda} = \frac{\log(\hat{P})}{\eta},$$  

where $\eta$ is the period of observation, i.e. 2 years in our application. It should be remarked that the maximum likelihood estimator of $\Lambda$ under this observational scheme is not guaranteed to exist or to be unique (see, e.g. [4,27]) but as proved in [10], estimator (11) exists and is unique whenever the transition probability matrix is irreducible with positive eigenvalues.
The estimated transition probability matrix is

\[ \hat{P} = \begin{pmatrix}
0.37 & 0.38 & 0.25 \\
0.11 & 0.38 & 0.51 \\
0.01 & 0.03 & 0.96
\end{pmatrix}, \tag{12} \]

which can be readily recognized as an irreducible stochastic matrix. This matrix has positive eigenvalues, \([1, 0.542, 0.168]\). Thus, the generator matrix estimated through (11) is the following:

\[ \hat{\Lambda} = \begin{pmatrix}
-0.59 & 0.58 & 0.01 \\
0.17 & -0.59 & 0.42 \\
0.00 & 0.02 & -0.02
\end{pmatrix}. \]

Finally, given the estimated initial distribution in the year 1998,

\[ \hat{\mu}' = (0.050, 0.068, 0.882), \tag{13} \]

and the average income for the poverty classes as reported in Table 3, we calculate the four indexes, i.e. Headcount ratio, Income gap ratio, Gini Index, and Sen index, and their respective confidence intervals at 95% level of significance. Figure 1 shows the four indexes estimated from the model against the observed indexes computed with equations from (1) to (4). In all cases, the computed indexes follow the trajectories of the observed indexes.

**Figure 1.** Dynamic indexes of poverty for Italian households income from 1998 to 2012 computed with parameters given in (12) and (13).
Besides, it is important to notice that the observed indexes fall within the 95% confidence intervals with the only exception of the Gini index for the year 2010. In general, the plots show that the model has very good power in capturing the dynamic of the observed indexes.

Figure 1 also indicates that all indexes show a decreasing path in time. This means that the different aspects of poverty represented by them are moving towards better economic conditions of the given households, which include a reduction of the percentage of poor, a lower mean shortfall of people below the poverty line and a reduction of disparities among the poor. However, it is relevant to remark that at year 2012 all indexes are very close to their stationary levels which are obtained by letting $t$ tend to infinity. Thus, the stationary values express the scores to which the indexes converge in the long run. This implies that a further decrease of poverty must necessarily be accompanied by reinforcement of poverty containment policies or by the adoption of new ones because, if left in current conditions, the economic system cannot evolve towards a lower level of poverty.

As a robustness test for the model, we now proceed to estimate the generator matrix using a reduced set of data with only three years of observation, i.e. from 1998 to 2002. The number of households and poverty thresholds remains the same. However, the number of available incomes for the estimation reduces from 7312 to 2742. The objective of this test is to simulate a real-life application in which there might be limited availability of historical data.

Figure 2. Dynamic indexes of poverty for Italian households income from 1998 to 2012 computed with parameters estimated using only the initial three years of data, from 1998 to 2004. The vertical lines represent the separation between observed data on the left and forecast on the right.
data and the necessity to forecast the poverty indexes. In this new setting, the estimated transition probability matrix and the generator matrix become,

$$\hat{P} = \begin{pmatrix}
0.32 & 0.41 & 0.27 \\
0.12 & 0.37 & 0.51 \\
0.01 & 0.02 & 0.97
\end{pmatrix}, \quad \hat{\Lambda} = \begin{pmatrix}
-0.70 & 0.69 & 0.01 \\
0.20 & -0.63 & 0.43 \\
0.00 & 0.02 & -0.02
\end{pmatrix},$$

and the average incomes for the poverty classes become $y_1 = 2046.57$ and $y_2 = 4430.35$.

Figure 2 shows that, even with a smaller dataset for the estimation procedure, the model is capable of capturing the dynamic of the observed indexes during the forecast period. This may be particularly useful when the application of the model is required for the evaluation of the impact of general shocks to the economic system that can suddenly occur, causing effects that last several years.

4. Conclusion

The analysis of the literature on poverty has demonstrated the importance of a dynamic approach to the determination of poverty and inequality. With the advancements proposed in this paper, we aim at giving an additional tool to help the definition of the policies for the poverty in real economies.

In this study, we first proposed an extension of the dynamic Gini index, and consequently the Sen index, with the inclusion of the inequality within each class of poverty where people are classified according to their income. Then, we established the central limit theorem for each poverty index for the determination of their confidence sets. An application to the Italian income data from 1998 to 2012 confirmed the effectiveness of the considered approach and demonstrated that the model has very good power in capturing the dynamic of the observed indexes.

This study leaves some open possibilities for further research, which are based on the relaxation of some of the model’s assumptions. It is worth to mention the possibility to use income distributions for every class that are assumed to be time-independent but not necessarily identical distributed involving time-varying parameters. This requires the use of estimation techniques developed in [3]. On the application side, it would be interesting to assess the model in other real economies, especially in the condition of shocks, such as the recent Covid-19 disruption. The availability of a larger dataset would give the possibility to use a semi-Markov model, see e.g. [22] that allows the possibility to investigate duration effects of the income dynamic.

Notes

1. The data is available at https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/index.html.

Acknowledgments

The authors would like to thank Philippe Regnault for fruitful suggestions and discussions. Also, comments and helpful suggestions made by two anonymous Referees and by the Editor-in-Chief are gratefully acknowledged.
Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partially supported by Regione Puglia under grant REFIN [number 7E1D0BBB].

ORCID

Guglielmo D’Amico http://orcid.org/0000-0002-6948-2912
Riccardo De Blasis http://orcid.org/0000-0002-0383-6263

References

[1] P. Annoni, R. Bruggemann, and L. Carlsen, A multidimensional view on poverty in the European Union by partial order theory, J. Appl. Stat. 42 (2015), pp. 535–554.
[2] M.J. Bane and D.T. Ellwood, Slipping into and out of poverty: The dynamics of spells, J. Hum. Resour. 21 (1986), pp. 1–23.
[3] V.S. Barbu, A. Karagrigoriou, and A. Makrides, Statistical inference for a general class of distributions with time-varying parameters, J. Appl. Stat. 47 (2020), pp. 2354–2373.
[4] M. Bladt and M. Sørensen, Statistical inference for discretely observed Markov jump processes, J. R. Stat. Soc. Ser. B 67 (2005), pp. 395–410.
[5] R. Breen and P. Moisio, Poverty dynamics corrected for measurement error, J. Econ. Inequal. 2 (2004), pp. 171–191.
[6] L. Cappellari and S.P. Jenkins, Modelling low income transitions, J. Appl. Econom. 19 (2004), pp. 593–610.
[7] G. D’Amico and G. Di Biase, Generalized concentration/inequality indices of economic systems evolving in time, WSEAS Trans. Math. 9 (2010), pp. 140–149.
[8] G. D’Amico, G. Di Biase, and R. Manca, Income inequality dynamic measurement of Markov models: Application to some European countries, Econ. Model. 29 (2012), pp. 1598–1602.
[9] G. D’Amico, G. Di Biase, and R. Manca, Decomposition of the population dynamic Theil’s entropy and its application to four European countries, Hitotsubashi J. Econ. 55 (2014), pp. 229–239.
[10] G. D’Amico and P. Regnault, Dynamic measurement of poverty: Modeling and estimation, Sankhya B 80 (2018), pp. 305–340.
[11] G.J. Duncan, B. Gustafsson, R. Hauser, G. Schmauss, H. Messinger, R. Muffels, B. Nolan, and J.C. Ray, Poverty dynamics in eight countries, J. Popul. Econ. 6 (1993), pp. 215–234.
[12] C.O. Ewald and M. Yor, On increasing risk, inequality and poverty measures: Peacocks, lyrebirds and exotic options, J. Econ. Dyn. Control 59 (2015), pp. 22–36.
[13] J.P. Formby, W.J. Smith, and B. Zheng, Mobility measurement, transition matrices and statistical inference, J. Econom. 120 (2004), pp. 181–205.
[14] J.E. Foster, A class of chronic poverty measures, in Poverty Dynamics: Interdisciplinary Perspectives, T. Addison, D. Hulme, and R. Kanbur, eds., chap. Poverty Dynamics, Oxford University Press, 2009.
[15] D. Hirsch, M. Padley, J. Stone, and L. Valadez-Martinez, The low income gap: A new indicator based on a minimum income standard, Soc. Indic. Res. 149 (2020), pp. 67–85.
[16] T. Hobza, D. Morales, and L. Santamaria, Small area estimation of poverty proportions under unit-level temporal binomial-logit mixed models, Test 27 (2018), pp. 270–294.
[17] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Stat. 19 (1948), pp. 293–325.
[18] V. Kana Zeumo, A. Tsoukiàs, and B. Somé, A new methodology for multidimensional poverty measurement based on the capability approach, Socio-Econ. Plan. Sci. 48 (2014), pp. 273–289.
[19] G.K. Kanji and P.K. Chopra, *Poverty as a system: Human contestability approach to poverty measurement*, J. Appl. Stat. 34 (2007), pp. 1135–1158.
[20] R. Langheine and F.V.D. Pol, *A unifying framework for Markov modeling in discrete space and discrete time*, Sociol. Methods Res. 18 (2016), pp. 416–441.
[21] N. Lee, G. Ridder, and J. Strauss, *Estimation of poverty transition matrices with noisy data*, J. Appl. Econom. 32 (2017), pp. 37–55.
[22] G. Masala, *Earthquakes occurrences estimation through a parametric semi-Markov approach*, J. Appl. Stat. 39 (2012), pp. 81–96.
[23] J.J. McCall, *A Markovian model of income dynamics*, J. Am. Stat. Assoc. 66 (1971), pp. 439–447.
[24] E.Y. Park and S.J. Nam, *Multidimensional poverty status of householders with disabilities in South Korea*, Int. J. Soc. Welf. 29 (2020), pp. 41–50.
[25] G. Parodi and D. Sciulli, *Disability in Italian households: Income, poverty and labour market participation*, Appl. Econ. 40 (2008), pp. 2615–2630.
[26] R. Puurbalanta, *A clipped Gaussian geo-classification model for poverty mapping*, J. Appl. Stat. 48 (2020), pp. 1882–1895.
[27] P. Regnault, *Entropy estimation for M/M/1 queueing systems*, AIP Conf. Proc. 1443 (2012), pp. 330–337.
[28] A. Sen, *Poverty: An ordinal approach to measurement*, Econometrica 44 (1976), pp. 219–231.
[29] A.F. Shorrocks, *Revisiting the Sen poverty index*, Econometrica 63 (1995), pp. 1225–1230.
[30] N. Takayama, *Poverty, income inequality, and their measures: Professor Sen’s axiomatic approach reconsidered*, Econometrica 47 (1979), pp. 747–759.
[31] V. Verma and G. Betti, *Taylor linearization sampling errors and design effects for poverty measures and other complex statistics*, J. Appl. Stat. 38 (2011), pp. 1549–1576.
[32] C.T. Whelan, R. Layte, B. Maitre, and B. Nolan, *POVERTY DYNAMICS: An analysis of the 1994 and 1995 waves of the European Community Household Panel Survey*, Eur. Soc. 2 (2000), pp. 505–531.
[33] K. Xu, *U-statistics and their asymptotic results for some inequality and poverty measures*, Econom. Rev. 26 (2007), pp. 567–577.