SMALL AREA METHODS FOR MONITORING OF POVERTY AND LIVING CONDITIONS IN EU

Jean Monnet Chair
European Commission
Università di Pisa

Workshop on
Small Area Methods and living conditions indicators in European poverty studies in the era of data deluge and Big data

The computation of subnational Spatial Price Indexes for the study of inequalities at a local level in monitoring

Luigi Biggeri
University of Florence and ASED, Dagum Centre

Pisa, May 10, 2018

This provisional presentation benefits of the works and discussions made with Tiziana Laureti and Prasada Rao

Outline

1. Introduction
2. Methodological general framework to compute Spatial Price Indexes (SPIs) across areas
3. Evidence from researches and experiments for the computation of SPIs
4. Conceptual framework for the computation of SPIs by using Consumer Price Index (CPI) data
5. Challenges of computing local SPIs by using big data
6. Concluding remarks

1

Introduction

Introduction -1-

- **Spatial price Indexes (SPIs)** are measures of differences in price levels across areas, essential to compare economic well-being indicators/aggregates, also for defining and implementing the various policy interventions at a local level (ECINEQ- Society for the Study of Economic Inequality, USA, 2017)
- For the comparisons across countries there is the **International Comparison Program (ICP)** managed by World Bank, that gained prominence over the last forty to fifty years (World Bank, 2013)
- The ICP computes the **Purchasing Power Parities (PPPs),** within the framework of **National Accounts (N.A.) expenditure side** (for GDP and its components)
- In the most recently completed 2011 round, the ICP **cover 177 country**, and now the 2017 round is under implementation

Introduction -2-

- What about the **computation of Sub-national Spatial Price Indexes**?
- The importance of constructing these indexes within a country has been recognized in literature and during the last three decades many researches, experiments and debates have been conducted on the field
- However, up to now, **no systematic attempts have been made by the National Statistical Offices** (NSOs) with the exception of the **USA** (Biggeri et al., 2010 and 2017; Laureti and Rao, 2018)
- Deaton said that "National and international statistical systems are strangely reticent on differences in price levels within countries……The ICP...publishing noting on the within country differences…." (Deaton and Dupriez, 2011) and surmise that

Introduction -3-

"the lack of these [spatial price] indexes more likely reflects the difficulty and cost of producing them"

In my opinion, may be that the situation depend also on the difficulties of interpretation by the users and on "political" reasons and possible discussions among different policy makers.

- However, within the ICP, during the 2011 ICP Round, the TAG devoted to the topic a specific agenda item (Biggeri et A., February 2010)
- Finally, now in the 2017 Round, the Governance Bodies of the ICP included the Sub-National PPPs topic on the research agenda items as an activity of a specific Task Force

Introduction -4-

- The aim of my presentation is to provide an overview of the main methods and of the practical organization process for computing the **sub-national spatial price indexes**, to discuss the main issues and the possible solutions

✓ The presentation refers to general **Spatial Price Indexes** in order to avoid the possible confusion with the Purchasing Power Parities (PPPs) computed by ICP that are computed for the components of the N.A.

- In general my main references are:

✓ the description of the process of compiling PPPs within the ICP (it is quite complex; a complete description can be found in World Bank, 2013 and in particular in the Chapters 1 and 4 written by Prasada Rao)

✓ the evidence from the researches and experiments for compiling sub-national spatial price indexes (Biggeri et al., 2017; Laureti and Rao, 2018)

- The challenges of using new big data, are also discussed in order to produce specific sub-national spatial price indexes

2

Methodological general framework to compute Spatial Price Indexes (SPIs) across areas

The requests (properties) of Economic Theory that the tool should satisfy to be correctly applied in an economic context

Methodological general framework to compute Spatial Price Indexes across areas -the problem-

- Let p_{ij} and q_{ij} represent price and quantities, respectively, of i-th product (and or service) in j-th area, where i= 1,2, …,N and j= 1,2,…,j,k,…M; (j can be any territorial area of interest)
- An economic nominal aggregate for area j can be written as

$$\mathrm{EAj} = \sum_{ij} p_{ij} q_{ij}$$

and for the area k the same aggregate is

$$\mathrm{EA_k} = \sum_{ik} p_{ik}\, q_{ik}$$

- The comparison between the aggregates EA_j and EA_k is affected both by the level of prices and quantities in the two areas
- To do the comparison in **real term** we have to eliminate the difference dues to the different level of prices in the two areas, by computing a spatial price index between the areas, which is the **binary spatial price index** (considering as base the area j or k)

Methodological general framework to compute Spatial Price Indexes across areas –elementary binary price indexes-

- Let us suppose that we have two correctly defined price vectors, (p_{1j} , …, p_{ij} ,…, p_{Nj}) and (p_{1k} ,…, p_{ik} ,…, p_{Nk}), of N product items for area j and area k and that the two vectors are technically comparable*
- Having the two price vectors or the elementary price indices (P_{1jk} ,…, P_{ijk} ,…, P_{Njk}), where $P_{ijk} = p_{ik}/p_{ij}$. , is the elementary price index of the i-th product for area k compared with the base or reference area j.
- The problem of determining a **synthetic binary index** P_{jk} (the **SPI**) can be solved either by utilising some «aggregation» function of the prices, or of the elementary indices in the two areas, in the following ways:

* The p_{ij} can be only the price observed for each product i, or an average price derived by a set of prices for each product i (that for the time being we consider not available)

Methodological general framework to compute Spatial Price Indexes across areas- synthetic binary price indexes-

- $P_{jk} = F(p_{1k}, \ldots, p_{ik}, \ldots, p_{Nk}) / G(p_{1j}, \ldots, p_{ij}, \ldots, p_{Nj})$
- $P_{jk} = H(P_{1jk}, \ldots, P_{ijk}, \ldots, P_{Njk})$
- The problem of computing the spatial price binary index P_{jk} comparing the two areas, is then solved by the **choice of the aggregation functions** *F* and *G* or *H* and by the derived synthetic index expressions (formulae).
- Such a choice is conditioned by the characteristics of the price vectors and by the purposes we have in mind in constructing the synthetic index, since we must be able to give it a **precise economic meaning**

Methodological general framework to compute Spatial Price Indexes across areas –properties for multilateral comparisons

- Obviously we can do the comparisons between all pair of areas and obtain the following binary price index matrix

$$\mathrm{P} = \begin{bmatrix} P_{11} & P_{12} & P_{1k} & P_{1M} \\ P_{21} & P_{22} & P_{2k} & P_{2M} \\ P_{j1} & P_{j2} & P_{jk} & P_{jM} \\ P_{M1} & P_{M2} & P_{mk} & P_{MM} \end{bmatrix}$$

where the binary comparisons have a **clear meaning** with the reference to the index formula chosen and the properties that the index satisfy.

- However, comparisons among all areas are usually considered important in order to do **multilateral comparisons**. At this end the multilateral P_{jk} are required to **satisfy** other two basic properties: ***transitivity*** and ***base invariance***.

Methodological general framework to compute Spatial Price Indexes across areas -transitivity and base invariance-

- ***Transitivity*** stipulates that a direct comparisons between any two areas must equal any indirect comparison between the two areas obtained through a third areas. Transitivity for three areas j, k, l requires that the method used in computing P_{jk} should be such as that

$$P_{jk} = P_{jl} \cdot P_{lk}$$

It is useful to note that some of the commonly used price index numbers such as the Laspeyres, Paasche, Fisher and Tornqvist indexes do not satisfy transitivity

- The requirement of ***base invariance*** *(*or *area symmetry)* simply requires that all the areas are treated symmetrically in making price comparisons (the use of only one area as reference area does not allow for a direct comparison between two different areas; **Pjk and Pjl are not directly comparable**)

Methodological general framework to compute Spatial Price Indexes across areas -Different situation about data-aggregation by Jevons index

- **Hypothesis that all the product items are priced in all areas, and also information on the quantity are available**

Among the many index formulae proposed to compute all the binary P_{jk} included in the previous matrix, the most recommended formulae are the Jevons-Tornqvist index and the Fisher index

- The binary **Jevons index** is the simple **unweighted geometric average** of price ratios in the areas, that is

$$P_{jk}^{Jevons} = \prod_{i=1}^{N} \left[\frac{p_{ik}}{p_{ij}} \right]^{1/N}$$

The index satisfies transitivity and base-invariance properties, but if the index is computed as a **weighted geometric average** of the price relatives (as for example in **Tornqvist index**) or when not all products are priced in all areas, the indexes became **not transitive**

Methodological general framework to compute Spatial Price Indexes across areas - Fisher index – Need for GEKS

- **Hypothesis that all the product items are priced in all areas, and also information on the quantity are available**

- The binary **Fisher index** is the geometric average of the so called Laspeyres and Paasche (weighted) indexes

$$P_{jk}^{L} = \frac{\sum_{i \in N_{jk}} p_{ik} q_{ij}}{\sum_{i \in N_{jk}} p_{ij} q_{ij}} \qquad P_{jk}^{P} = \frac{\sum_{i \in N_{jk}} p_{ik} q_{ik}}{\sum_{i \in N_{jk}} p_{ij} q_{ik}} \qquad P_{jk}^{F} = \sqrt{P_{jk}^{L} \cdot P_{jk}^{P}}$$

Also this index is **not transitive**.

Interpretation more difficult!
Biggeri, 2016

- Therefore, having computed the binary price index matrix, it is necessary to use the **GEKS** (Gini, Elteto, Koves, Szulc) **procedure** to derive the **transitive indexes**, by **minimizing**

$$\sum_{j=1}^{M} \sum_{k=1}^{M} \left[\ln P_{jk}^{GEKS} - \ln P_{jk} \right]^2$$

subject to $P_{jk}^{GEKS} = P_{jl}^{GEKS} \cdot P_{lk}^{GEKS} \quad for\ all\ \ j,k \text{ and } l$

Methodological general framework to compute Spatial Price Indexes across areas -the aggregation: GEKS procedure

- **Application of GEKS procedure to Jevons and Fisher binary indexes**

- To the **Jevons weighted binary index**, we obtain GEKS-Jevons

$$P_{jk}^{GEKS-Jevons} = \prod_{l=1}^{M} \left[P_{jl}^{Jevons} \cdot P_{lk}^{Jevons} \right]^{1/M}$$

To the **Fisher binary index**, we obtain GEKS-Fisher index

$$P_{jk}^{GEKS-Fisher} = \prod_{l=1}^{M} \left[P_{jl}^{F} \cdot P_{lk}^{F} \right]^{1/M}$$

- The GEKS approach has been generalized and a weighted GEKS (WGEKS) has been proposed in Rao (2009) and the properties are discussed further in Hajargasht et al. (2017)

- The satisfaction of the transitivity property makes the **interpretation** of the differences a little more difficult

Methodological general framework to compute Spatial Price Indexes across areas -the aggregation by CPD methods-1

- **Hypothesis that not all the product items are priced in all areas**

- In this case **specific version of the GEKS-Jevons index** can be used (see World Bank, 2013 pp. 101-105), and this is the approach followed by **OECD-Eurostat** computation of the PPPs

- Anyway, the approach for the aggregation of prices now followed by the **ICP** is the **Country Product Dummy (CPD) method**, that considering the areas instead of the countries we can called and write as **APD method**

- The method was firstly introduced by Summers (1973) as a simple **regression-based method** to fill missing price data, and then many researchers (in particular Prasada Rao) have discussed and showed its important properties.

- The model can be written also as a **standard hedonic regression** model in which the characteristic used are the **area** and the **product** specifications.

Methodological general framework to compute Spatial Price Indexes across areas -the aggregation by APD methods-2

- The **APD method** suggests that price levels can be estimated by regressing the logarithms of prices on area and product dummy variable and the model is given by

$$\ln p_{ij} = \ln P_j + \ln P_i + \ln u_{ij}$$
$$= \pi_j + \eta_i + v_{ij}$$
$$= \sum_{k=1}^{M} \pi_k D^k + \sum_{i=1}^{N} \eta_i D^i + v_{ij}$$

where v_{ij} are random disturbance terms (assumed to satisfy the standard assumption of a multiple regression model), D^k is a area-dummy variable that takes value equal to 1 if the price observation is from k-th area, and D^i is a product-dummy variable that takes value equal to 1 if the price observation is for the i-th product.

- The **parameter π_j is interpreted as the general price level in area j** relative to prices in other areas included in the comparison. It is possible to express π_j to a reference area, and the price comparisons are given by:

$$P_{jk}^{RPD} = \frac{\exp(\hat{\pi}_k)}{\exp(\hat{\pi}_j)} \quad for\ all\ \ j,k = 1,2,\ldots,M$$

Methodological general framework to compute Spatial Price Indexes across areas -the aggregation by CPD methods-3

Passing over the method used to estimate the parameters, that requires transitive price comparisons, we can observe that the APD method:

- provides price comparisons that are **transitive and base-invariant**
- allows to compute the **standard errors** of the estimates of π_j
- is general enough to accommodate the **different cases about the availability and characteristics of price data** (extended APD model to incorporate additional characteristics associated with each price quotation like type of outlet etc.)
- if average price for each product are used and if we have information on the associated standard errors, these can be incorporated into the model
- if **all the product items are priced** in all the areas, then the price comparisons are **identical to the Jevons-based comparisons**.

Methodological general framework to compute Spatial Price Indexes across areas -practical situation: aggregation above sub-groups-a

- Moving from a theoretical to a practical situation, the first topic refers to the SPIs objectives. Usually, it is of interest to compute SPIs for **sub-group of products** (i.e. for COICOP classification of the consumption products and services) and/or for **sub-groups of areas** (different level of territorial areas, i.e. Nuts 3 and Nuts 3)
- In these cases the methods already presented can be used for the computation of the SPIs at the level of each sub-group, and then we have to consider the **methods to be used for aggregation above the sub-groups** (for higher level aggregates)
- The main index number formulae used at international level for that are:

- The **Fisher-based GEKS** index (already presented; actually used in ICP)
- The **Geary-Khamis** (GK) method (used in ICP until 2005)
- The **Weighted Area-Product-Dummy** (WAPD) Method and **Stochastic approach**

Methodological general framework to compute Spatial Price Indexes across areas -practical situation: aggregation above sub-groups-b

- The **Geary-Kamis method** is defined by an inter-related system of equations regarding P_j and P_i . It compares prices in each area j with the average prices of the items obtained by averaging observed prices in different areas after adjusting for differences in the price level in different areas. SPIs for each area j is obtained by using quantities of the products in area at observed prices p^j_i.

✓ Despite some criticisms, the method is still used when the additive consistency property are required in the comparisons (Aten and Figueroa, 2015; Aten, 2017)

- The **Weighted Area-Product-Dummy** (WAPD) method can be applied for the purpose of aggregating price data also in the absence of data on the weights. It also referred to as the stochastic approach. Rao (2010) showed how most of the multilateral index numbers can be derived using weighted APD model

➢ **Spatial chaining of price indexes**: in **multilateral** spatial price comparisons we need to pay attention to the **order to follow in the subsequent binary areas comparisons that is not "a priori" defined**. Following **different paths**, **different results** are obtained depending on the **structure of relative prices and quantities in the two following areas under comparisons**. Need to chose the **best path** by using a **similarity** or **dissimilarity measure**. There many papers on this topic issue and in particular on the Minimum Spanning Trees (MST). See for example (Hill, 1999, 2009; Diewert, 2009, 2013; Rao et al. 2017)

Methodological general framework to compute Spatial Price Indexes across areas -practical situation: different types of data

- The use of the various methods for the estimation of the SPIs depends also by the **type and characteristics of data available** on prices and weights (quantities or values like expenditure). Therefore the decision depend on the specific cases. Many experiment have been conducted in order to chose the most adequate methods

- For example Biggeri et al. 2017 used the following CPD models:

✓ CPD model CPD model based on individual price data: the hedonic approach

✓ CPD model using average prices: unweighted *vs* weighted approach

✓ CPD model with spatially auto-correlated error structure

- The issue of the **spatial autocorrelation among prices of the different areas** is important, because its presence usually affect the estimations of the SPIs. Certainly, this issue is more relevant among areas within a country (particularly for small areas) than across countries. In any case, it is necessary to take into account of that autocorrelation with adequate methods. Laureati et al. (2018) devoted accurate attention to this issue providing also interesting experiments.

3

Evidence from researches and experiments for the computation of SPIs

Evidence from researches and experiments for the computation of SPIs – attempts to compile PPPs and SPIs

- **Research projects and studies** have been conducted by NSOs and individual researchers **in various countries** (see a detailed description in Biggeri et al., 2017; Laureti and Rao,2018; the specific researches are quoted in references)

➢ There are **three principal approaches** to compute the SPIs:

✓ The **first** and foremost is the approach whereby **data collected for the Consumer Price Indexes (CPIs)** are used in constructing SPIs

✓ The **second** approach is to use unit-value prices, derived from **Household Expenditure or Budget Sample Survey (HBS)**, and applying to them a **demand system model** to obtain SPIs

✓ The **third approach** is to use alternative sources of data and sometime to use just **data on specific product or service**, like the cost for the housing service.

➢ **First approach** (the framework and principles of ICP are followed)

✓ in a few cases, a complete set sub-national GDP-PPPs referred to all the aggregates of GDP, have been computed, usually implementing a new system of data collection of prices adequate both for CPI and ICP computations. Usually the experiments have been conducted under the "supervision" of experts of the ICP Global Office and/or Regional ICP agencies (see for example Dikhanov et al.,2011; Capilit and Dikhanov, 2017; Skaini, 2017)

Evidence from researches and experiments for the computation of SPIs – attempts by using CPIs data

- ✓ In most cases, on the base of CPI data, sub-national **SPIs for Household Consumptions** have been computed (sometimes called sub-national household consumption PPPs). In this case, additional data are sometimes collected by using ad hoc survey when suitable CPI data are not available (for example for: clothing, furniture, etc.), as happened in Italy and UK.

There is an important difference:

- in the first case the computation have been done by using the SPD ICP classification and integrating it in the COICOP classification used for CPI. The work is much more but in this case a real integration and synergies between CPI and ICP collection of data are reached.
- in the second case the NSOs are using easily the national COICOP classification of products without more work and any tentative to link it at the SPD ICP classification of products

Evidence from researches and experiments. USA RPPs and Oecd-Eurostat countries computation of SAFs

- The best and complete researches and applications within this approach have been conducted in **USA starting from 1940s**. After a lot of research and experiments conducted by the Bureau of Economic Analysis in collaboration with agencies (Aten, 2017), finally in 2000s regular **compilation of spatial price differences** through the compilation of Regional Price Parities (**RPPs**) for 50 States and District of Columbia and 366 Metropolitan areas. For the computation of the RPPs also data on rent of the houses have been used. Then, since 2014, the RPPs and the price-adjusted estimates of regional personal income have **become official statistics.**

- An other important work is currently done by some countries within the Oecd-Eurostat program. The PPPs are computed on the data collected for the **capital cities** of each country. In order to have a better picture of the PPPs for the country, the computation of **Spatial Adjustment Factors (SAFs)** are requested every 6 years. This SAFs are a special kind of sub-national SPIs. Some countries publish this data as SPIs giving interesting overview at regional level (see in particular ONS, 2018; D' Silva and Bucknall, 2018)

Evidence from researches and experiments. The use of data coming from HBS and estimations based on a demand system model

- **Second approach that use data from HBS**

- ✓ This approach was used for the first time by Coondoo et al. (2004) who reported the results obtained for India. Other researchers continued the work with the aim of proposing a **unified framework for estimating intra and intercountry PPPs for household consumption** (Majumder and Ray, 2016)
- ✓ The approach is based on the anchored on the **Engel's Curve** and/or a demand system model, and in particular to **Quadratic Almost Ideal Demand System (QAIDS) model**, by using the **unit-value** coming form the HBS. The methodology used in the papers is then related to some of the known variants of CPD system
- ✓ However, the approach has the **limit** due to the fact that much of their works focuses purely on food price index numbers as the household expenditure surveys provide reliable unit value measures for food items.
- ✓ The approach is surely of interest for the topic of sub-national SPIs estimations and should be extended to other countries to check its validity and to compare the results with the findings obtained with other methods of computation

Evidence from researches and experiments. The use of other sources of data and the use of house rent data

- Several studies for computing sub-national price level differences have been carried out by various researchers in particular for the **Chinese provinces** (Biggeri et al., 2017a)
- ✓ The first most important estimation was conducted **using the cost basket method** (Brandt and Holtz, 2006).
- For sub-national cost-of-living adjustments to compare poverty, also spatial indexes based on the **cost of housing are used**. In particular, these indexes are used in the USA because their variation across areas can be significant (Renwick, 2009; Renwick et al., 2014)
- ✓ The most used index refers to the **average monthly rent for different type of houses**. The hypothesis is that renting is the most important issue faced by poor, representing 40%-50% of their total consumption expenditures, that is approximately the share of expenditure devoted to the cost of housing by the poor households
- ✓ Recently, a **Paasche index** price for the household consumptions and **rental data** have been used to spatially impute the value for Romanian Municipalities (Wagner De Azevedo and Corral Rodas, 2018)

Evidence from researches and experiments. Potential and current uses of the SPIs

Potential uses (of the national and sub-national policy makers, economists, etc.)

- ✓Development and competitiveness analysis
- ✓Cross sub-national area comparisons of economic data computing the real per capita GDP and for all the sub-National Accounts components
- ➢ Spatial price level analysis, also to check the urban and rural differences
- ➢ Well being, Inequality and poverty analysis
- ➢ Operational policy purposes, i.e. allocation of funds policy intervention to reduce poverty

and so on

Current uses (Most are in inequality and poverty analysis)

➢ Now days the poverty analysis are extensively developed at local level by using the **Small Area Estimation (SAE)** methods. See previous speakers at this Workshop and the World Bank-EU project (Simler, 2016)

Evidence from researches and experiments. Potential and current uses of the SPIs: possible effects on the organization of data collection

- Two main issues are coming from the poverty analysis at local level:

✓ need for spatial price indices at very detailed territorial level (or availability of a substitute indicator)

✓ need for poverty specific sub-national spatial price indices (considering both their consumer basket and prices payed in the outlets and/or markets where the poor do the purchases)

➢ More experiments in these two fields are necessary and may be possible that ask to **revise the organization of the collection** of the CPI data and of the HBS

➢ The **purposes** and **uses** of the interested sub-national SPIs surely affect the organization and implementation of the **data collection** and the choice of the **methods of computations**

4

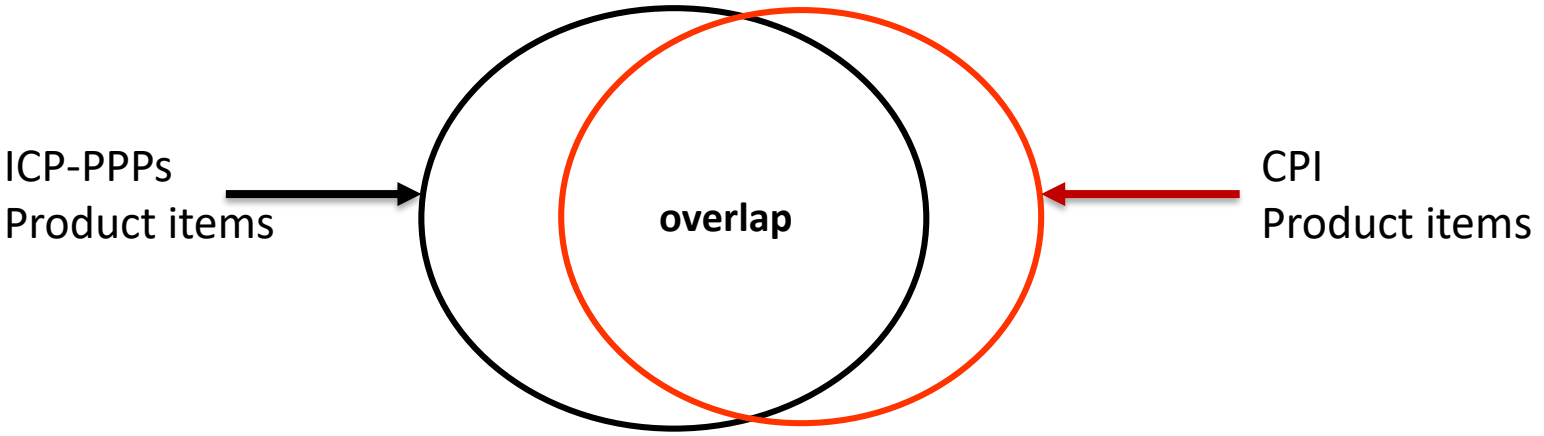
Conceptual framework for the computation of SPIs by using Consumer Price Index (CPI) data

We have the tool kit, but all the results depend on which data we use and on the quality of the data!!

Computation of SPIs by using CPI data: conceptual framework

- To compute **sub-national SPIs and/or sub-national PPPs**, the most promising approach is to **use the CPI data**, taking into account the conceptual framework provided by the ICP.
- Most (all) **countries are computing CPIs** collecting a lot of prices for each product (for example, for the computation of CPIs, in Italy **Istat** up to 2017 collected prices for about 1,500 Products, in 40,000 outlets, 87 capital city provinces, and collect about **700,000 quotations every month**).....are this Big Data?
- The NSOs could have the **objective** to satisfy the request of the ICP, but, above all, to produce **SPIs data more appealing** for research objectives and for national and local policy makers.
- It is important that the NSOs compute the **Sub-national household consumption SPIs** **(Spatial CPIs)**
- May be that for the NSOs it is better to compile the spatial CPIs, **using their classifications**, that are already very well knowledge by the users

Computation of SPIs by using CPI data: intersection between ICP and CPI product items classification



- the **integration of the CPI and ICP** classifications of products could be done, but the **resources necessary** to do it depend on the overlapping between the two classifications (see Biggeri and Laureti, 2010)

- In any case , the NSOs have **to reengineer the collection of data**

Computation of SPIs by using CPI data: property requirements

- Need to satisfy the requirements of **the representativity** and **comparability** *for each product or groups of products* (likely **less serious** issues in the context of sub-national spatial indices in comparison with ICP)
- It seems more **important** to consider the representativity and the importance of each item (a principle generally followed by the NSOs) and this must be satisfied also at the level of sub-areas

➢ One **important question**: could the exact **comparability** of the products in all the area be **relaxed**?

- **May be** this could be done considering comparable the products mostly purchased in the different areas that satisfy the same consumer need (also if of the different brands, quality, etc.), that give the same level utility for the consumer of each specific area. It could be a **kind of Economic** Spatial Price Index, like this

$$P_{jk}^{E} = C\,(p_{1k},\ldots,p_{2k},\ldots,p_{Nk};\,U_i)/\,C\,(p_{ij},\ldots,p_{ij},\ldots,p_{Nj};\,U_i)$$

where *C* is a cost function and *U* the **level of utility** for each product i.

Example for **different brands and quality of Rice** in different Italian Region!

Computation of SPIs by using CPI data: organization of the collection of data

- For the compilation of the sub-national spatial consumer price indices, it is necessary to define their structure, usually done through a **pyramid approach**, to building up the indices at various levels for which the process of CPIs production collects prices and estimates the system of expenditure weights

- This is very important because the application of CPD methods and other methods of aggregation depends of the characteristics of data available

- The framework for the construction of the CPI **must** refers to a kind of ideal *multistage stratified sampling design*.

- The population of items should be considered as structured by different hierarchical levels.

➢ **Price data**: to be collected $P_{i^*, o, j, r, \ldots}$

$_{i^*}$ refers to the different characteristics of the product; o is the type of outlet; $_j$ is the elementary area, $_r$ is referred to aggregation of j

Computation of SPIs by using CPI data: need for a multistage stratified sampling design

- At the first stage, of hierarchical structure, the territory of the country may be partitioned into *geographical areas* ***r***, as regions and provinces, and into *local areas j* - i.e. municipalities- which may be grouped into different geographic regions, while the *outlets* are the elements of the second level of the hierarchical structure.

- In the **product dimension**, the elementary aggregates (considered as product strata in the ideal sampling) are aggregates at different levels, following the COICOP hierarchical classification.

- The list of products to compute the spatial indices should be as large as possible considering the detail of CPI data collected, so many different groups of population can be used to compute the indices

- In general terms, the **overall SPIs based on CPI data** may be obtained by **successive aggregations** of the elementary indices following **different 'paths'.**

Computation of SPIs by using CPI data: the list of products

- The list of products to compute the spatial indices should be **as large as possible** considering the detail of CPI data collected, so many different groups of population can be used to compute the indices

- To satisfy the needs of national and local policy makers that want to use the SPIs, three items are important:

✓ Consider the areas as small as possible (**local areas**)

✓Compile the PPPs separately for **urban and for rural areas** (in I*taly as in many other countries only data in the capital provincial cities!)*

✓Collect data on prices and expenditures in order to compile also the **poverty-specific PPPSPIs**.

➢ Deaton proposal

Computation of SPIs by using CPI data: the matrix of elementary indexes

- After having decided the **Data requirements**, it is important to define the main objectives of the **data preparation (**indicating also the different sources of data)
- At the end of the data preparation phase, two matrixes must be prepared:

✓ A **matrix of prices**

✓ A **matrix of expenditure weights** (or other kind of weights)

- Than the compilation of the **matrix of the elementary binary spatial indexes** must be done

➢ Usually there is a phase of data cleaning, adjustment and editing (use of Quaranta table)

➢ Need to propose a specific evaluation of the **error profile** for each data collection and final data production

5

Challenges of computing local SPIs by using big data

The use of scanner data to compute SPIs

Istat Project

- Within the European Multipurpose Price Statistics project, Istat has been exploring the possibility of using big data (retail scanner data) for compiling sub-national PPPs.
  - 16 modern distribution chains (food, beverages, personal care products, etc.) **4 millions of quotation every month! But not high coverage**

- However, scanner datasets provide both opportunities and challenges for price statisticians

❑ **Advantages**:
  - It is recorded what is actually sold and universe of all transactions is considered
  - Prices available every day of the month and for 1,781 outlets
  - Detailed information on turnover and quantities for each item code (GTINs)
  - Territorial coverage is high (all Italian provinces)
  - A wide range of methods may be used due to availability of quantities and expenditure data

The use of scanner data to compute SPIs

- **Issues**:
  - Hard discount are not included.
  - Several sub-classes of food products cannot be considered since these products are sold at price per quantity and are not pre-packaged with GTIN codes (i.e. vegetables, fruit, meat and fresh fish)
  - High variability of products sold among cities (chaining methods?)

- **Open questions:**

- How to combine scanner data with other sources to compile household consumption sub-national PPPs?
  - ✓ traditional CPI data (already representative and in some cases also comparable),
  - ✓ ad hoc collected data for certain groups of products (clothing, furniture, etc.)
  - ✓ data collected on the web also through web scraping techniques.

The use of data from the archives of the Revenue and Tax Agency to compute SPIs for the housing rent cost

- For sub-national cost-of-living adjustments to compare poverty, also spatial indexes **based on the cost of housing are used**.
- In particular, these indexes are used in the USA because their variation across areas can be significant.
- The most used index refers to the average monthly rent for different type of houses. The hypothesis is that **renting is the most important issue faced by poor**, representing 40%-50% of their total consumption expenditure.
- Moreover, it is important to consider that **many policies to combat poverty** are **based on rental housing subsidies** or on providing no housing cost to the poor.

EXAMPLE

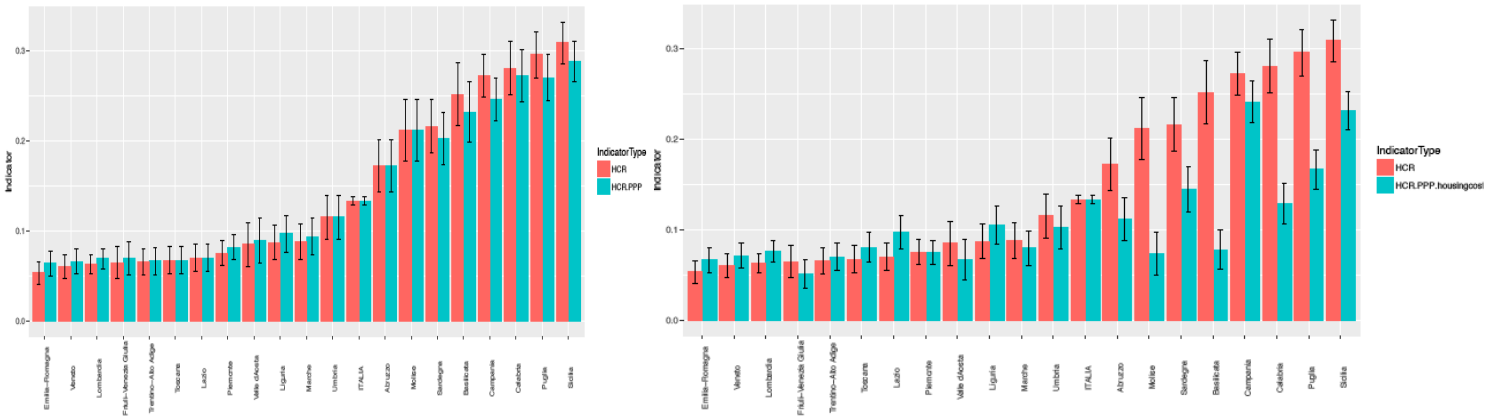


Fig. 3 Household HCR for Italian regions computed with NPL adjusted by PPPs (left) and RMHRs (right) – 2012

As we can see, the use of the conversion factors reshuffles in some way the territorial distribution of the HCRs values. More reshuffling effect is evident when the MHR is used. The range of the two spatial indexes is quite different: 13.2 percent points for PPPs and 68.8 percentage points for MHR.

The use of data from the archives of the Fiscal Agency to compute SPIs for the housing rent cost

- The Italian **Revenue and Tax Agency** has a special directorate called «**Estimate and Observatory on the Real Estate**»
- The observatory estimates and publish (every six months), the **real estate quotations** of the estate and of the **rents for their location.**
- The estimates refer to the minimum and maximum **market values** in Euro for square meter for each **typology of the estates** and for **each territorial zone** (area) within all the Italian Municipalities (more than 8,000).
- The value are obtained not as a punctual estimation, but as a **process of subjective assessments (**Min and Max)
- Possibility to compute **SPIs for 5 zone**, within the municipality and for **5 typology of houses**. The aggregation could be for **different level size of the territorial area**: Local Labor System (group of Municipality, more than 600), Provinces (110), Regions (20)
- Possible issue: the synthesis of the subjective assessments

6 Concluding remarks

- Use of SPIs: need to know

- Statistical methods, but also Economic theory and the assumption made
- The process of production of price and weight data and their errors profile

- In Italy, Istat is investing in the future production of the SPIs. The process is difficult therefore Istat should invest more on this field of research, with the support of the policy makers

- Dagum Centre and their members are working on that within Maxwell Project (together with Istat and other European institute of research

I am an optimist person

So, **I am confident that we will reach very good results**

THANKS
FOR
YOUR
ATTENTION

References quoted and other publications in the field

References quoted and other publications in the field

- Aten B.H. (2005), Report on Interarea Price Levels, Working Paper No. 2005-11, Bureau of Economic Analysis, Washington D.C.
- Aten B.H. (2006) Interarea Price Levels: An Experimental Methodology, Monthly Labor Review, 129, 9, 47–61.
- Aten B.H (2009), Cities in Brazil: An interarea price comparison. In International and interarea comparisons of income, output, and prices (eds A. Heston and R. Lipsey), pp.211–229, Chicago: University Chicago Press.
- Aten B.H. (2017), Regional Price Parities and Real Regional Income for the United States, Proceedings of the 60th World Statistics Congress of the International Statistical Institute, 2017.
- Aten B.H and Figueroa E.B. (2015), Real personal income and regional price parites for state and metropolitan areas,2009-2013, Survey of Current Business, ,95(7)
- Biggeri L., De Carli R. and Laureti T. (2008). The interpretation of the PPPs: A method for measuring the factors that affect the comparisons and the integration with the CPI work at regional level. In Proc. Joint UNECE/ILO meeting on Consumer Price Indices, May 8–9, Geneva.
- Biggeri L. and Laureti T. (2010) Are integration and comparison between CPIs and PPPs feasible?, in Price Indices in Time and Space, L. Biggeri and G. Ferrari (eds.), Springer, 2010
- Biggeri L., Laureti T. and Rao D.S.P. (2010), Sub-national PPPs based on integration with CPIs-Research project, draft proposal, 2011 ICP Round, 2nd Technical Advisory Group, Document 10.01
- Biggeri L., Laureti T. and Polidoro F. (2017a). Computing Sub-national PPPs with CPI Data: An Empirical Analysis on Italian Data Using Country Product Dummy Models. Special Issue, Social Indicators Research

Biggeri, L., Ferrari, G., Zhao, Y. (2017b), Estimating Cross Province and Municipal City Price Level Differences in China: Some Experiments and Results, *Social Indicators Research*

Biggeri L. and Pratesi M. (2017), Monetary poverty indicators at local level: definitions, methods of estimations and comparisons in real terms, Proceedings of the 60th World Statistics Congress of the International Statistical Institute, 2017.

Bishop J.A., J. M. Lee and L.A. Zeager, Incorporating spatial price adjustments in U.S. public policy analysis, ECINEQ. Society for the Study of Economic Inequality, WP 438, USA

Capilit E. and Dikhanov Y. (2017), Subnational Purchasing Power Parities in Asia: Prospects and Feasibilities, Proceedings of the 60th World Statistics Congress of the International Statistical Institute, 2017.

Coondoo, D., Majumder, A., & Ray, R. (2004). A method of calculating regional consumer price differentials with illustrative evidence from India. Review of Income and Wealth, 50(1), 51-68.

D' Silva E. and Bucknall R. (2018), Spatial Adjustment Factors for the United Kingdom, ONS, UK

De Azevedo J.P.W. and Rodas P.A.C. (2018), Adjustment for Sub-National Prices: A Spatial and Small Areas Approach to Adjust for Subnational Cost of Living, Global Solutions Group on Welfare Measurement and Statistical Capacity, World Bank, ppt provided by authors.

Deaton, A. (2003) Prices and poverty in India, 1987–2000. Economic and Political Weekly, 38(4), 362–368

Deaton A. and Dupriez O. (2011), Using unit-values to assess spatial price differences: evidence from India and Brazil, 5th Technical Advisory Group Meeting, Document 09.01.

Diewert W.E. (2009), Similarity Indexes and Criteria for Spatial Linking, in: D.S.P Rao (ed.)

Giusti C., Masserini L. and Pratesi M. (2017), Local comparisons of small area estimates of poverty: an application within Tuscany region in Italy, Special Issue, Social Indicators Research
Hamadeh N., M. Mouyelo-Katoula, P. Konijn and F. Koechlin (2017), Purchasing Power Parities of Currencies and Real Expenditures from the International Comparison Program: Recent Results and Uses, Special Issue, Social Indicators Research
Hill R.J. (1999), Comparing Price Levels across Countries Using Minimum Spanning Trees, Review of Economics and Statistics (81)
Hill R.J. (2009), Comparing Per Capita Income Levels across Countries Using Spanning Trees: Robustness, Prior Restrictions , Hybrids and Hierarchies, in D.S.P. Rao (ed.)
Istat (2008), Le differenze nel livello dei prezzi tra i capoluoghi delle regioni italiane per alcune tipologie di beni Anno 2006, Roma (in Italian language).
Istat (2010), La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regionii taliane, Anno 2009, Roma (in Italian language).
Istat (2015), La misura dell'inflazione per classi di spesa delle famiglie, Flash, Luglio 2015 (in Italian language)
Jolliffe D. (2006), The Cost of Living and the Geographic Distribution of Poverty, USDA**,** Economic Research Report Number 26
Kelly P. and Ranoto L. (2016), Partnership, processes and possibilities: the South African experience of integrating PPP and CPI work, Paper presented at the ILO/UNECE meeting of CPI experts, Geneva, 2-4 May 2016
Kokoski M. (1991) New Research on Interarea Consumer Price Differences, Monthly Labor Review, Vol. 114, 31.
Kokoski, M., Moulton, B., and Zieschang, K. (1999). Interarea price comparisons for heterogeneous goods and several levels of commodity aggregation. In International and interarea comparisons of income, output and prices (eds. A. Heston and R. Lipsey), pp. 123–166. Chicago: University of Chicago Press.
Koo J., Phillip, K. R., and Sigalla F. D. (2000), Measuring regional cost of living. Journal of Business and Economic Statistics, 18(1), 127–136
Laureti T. and Polidoro F. (2017), Testing the use of scanner data for computing sub-national Purchasing Power Parities in Italy, Proceedings of 61st ISI World Statistics Congress, Marrakesh
Laureti T., Ferrante C, and Dramis B. (2017), Using scanner and CPI data to estimate Italian sb-national PPPs, Procedeeings of the 49th Scientific Meeting of the Italian Statistical Society, Florence, pp. 581-588
Laureti T., D.S.P, Rao (2018), Measuring Spatial Price Level Differences within a Country: Current Status and Future Developments, Estudos de Economia Aplicada, vol. 36-1, pags.119-148
Laureti T., Montero JM., Minguez R., Fernàndez-Aviles G. (2018), A Stochastic Model with Penalized Coefficients for Spatial Price Comparisons: An Application to Regional Price Indexes in Italy, Review of Income and Wealth, forthcoming
Moulton, B. R. (1995). Interarea indexes of the cost of shelter using hedonic quality adjustment techniques. Journal of Econometrics, 68(1), 181-204.
McCarthy P. (2010), Asia and Pacific Region: Sub-National Purchasing Power Parities. Case study for the Philippines, 2011 ICP Round, 2nd Technical Advisory Group, Document 10.02
Majumder et and 2014; Mishra and Ray, 2014, and Majumder and Ray, 2016
Majumder A., Ray R. and Sinha K. (2013), Spatial Comparisons of Prices and Expenditure in a Heterogeneous Country:

Rao D.S,P. ed. (2009), Purchasing Power Parities:: Recent Advances in Methods and Applications, Edward Elgar Publishing Company
Rao DSP and Hajargasht (2016), Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP), Journal of Econometrics, 191(2), pp.414-425
Rao D.S.P., R. Hill, S. Shankar and R. Hajargasht 2017), Spatial Chaining of Price Indexes to Improve International Comparisons of Prices and Real Incomes, Rao ppt presentation at a Seminar in Pisa , Italy, 15th June, provided by the author
Renwick, T. (2009), Alternative Geographic Adjustments of U.S. Poverty Thresholds: Impact on State Poverty Rates. Working paper 2009-11, Washington DC: U.S. Department of Commerce, Census Bureau.
Renwick T., Aten B., Figueroa E., and Martin T. (2014), Supplemental Poverty Measure: A Comparison of Geographic Adjustments with Regional Price Parities vs. Median Rents from the American Community Survey, SEHSD Working Paper No. 2014-2, U.S. Census Bureau
Roos, M. (2006). Regional price levels in Germany. Applied Economics, 38(13), 1553–1566
Secondi L. and Marchetti S. (2016), Estimates of household consumption expenditure at provincial level in Italy by using small area estimation methods: "real" comparisons using purchasing power parities, Special Issue, Social Indicators Research
Simler K. (2016), Pinpointing Poverty in Europe: New Evidence for Policy Making. World Bank, Washington D.C.
Skaini M. and Samara A. (2017), Synergies between CPIs and PPPs and Integration of Survey Activity. UNSVWA, paper presented at the meeting of ICP Task Force held in October 2017
Walschots J. (2013), Workshop multipurpose price statistics – results and next steps, Item 4.3 of the Agenda, Meeting of the Price Statistics Working Group, Eurostat, Luxembourg 22-23 October 2013
Waschka A. M., W. Khoo, J. Quirey T. and Zhao S. (2003) Comparing Living Costs in Australian Capital Cities. A Progress Report on Developing Experimental Spatial Price Indexes for Australia, Australian Bureau of Statistics.
Wingfield, D., Fenwick, D., and Smith, K. (2005). Relative regional consumer price levels in 2004. Economic Trends (Office for National Statistics, UK), 615, 36–46.
World Bank (2013). *Measuring the Real Size of the World Economy: The Framework, Methodology and Results of the International Comparison Program.* Washington DC: World Bank.
World Bank (2014). *Purchasing Power Parities and the Real Size of World Economies: A Comprehensive Report of the 2011 International Comparison Program.* Washington DC: World Bank

Aten B. (2005), Report on Interarea Price Levels, Working Paper No. 2005-11, Bureau of Economic Analysis, Washington D.C.
Aten, B. (2006) Interarea Price Levels: An Experimental Methodology, Monthly Labor Review, 129, 9, 47–61.
Aten B. (2009), Cities in Brazil: An interarea price comparison. In International and interarea comparisons of income, output, and prices (eds A. Heston and R. Lipsey), pp.211–229, Chicago: University Chicago Press.
Aten B. (2017), Regional Price Parities and Real Regional Income for the United States, Proceedings of the 60th World Statistics Congress of the International Statistical Institute, 2017.
Biggeri L., De Carli R. and Laureti T. (2008). The interpretation of the PPPs: A method for measuring the factors that affect the comparisons and the integration with the CPI work at regional level. In Proc. Joint UNECE/ILO meeting on Consumer Price Indices, May 8–9, Geneva.
Biggeri L., Laureti T. and Rao D.S.P. (2010), Sub-national PPPs based on integration with CPIs-Research project, draft proposal, 2011 ICP Round, 2nd Technical Advisory Group, Document 10.01
Biggeri L., Laureti T. and Polidoro F. (2017a). Computing Sub-national PPPs with CPI Data: An Empirical Analysis on Italian Data Using Country Product Dummy Models. Special Issue, Social Indicators Research
Biggeri, L., Ferrari, G., Zhao, Y. (2017b), Estimating Cross Province and Municipal City Price Level Differences in China: Some Experiments and Results, *Social Indicators Research*
Biggeri L. and Pratesi M. (2017), Monetary poverty indicators at local level: definitions, methods of estimations and comparisons in real terms, Proceedings of the 60th World Statistics Congress of the International Statistical Institute, 2017.
Capilit E. and Dikhanov Y. (2017), Subnational Purchasing Power Parities in Asia: Prospects and Feasibilities, Proceedings of the 60th World Statistics Congress of the International Statistical Institute, 2017.
Coondoo, D., Majumder, A., & Ray, R. (2004). A method of calculating regional consumer price differentials with illustrative evidence from India. Review of Income and Wealth, 50(1), 51-68.
De Azevedo J.P.W. and Rodas P.A.C. (2018), Adjustment for Sub-National Prices: A Spatial and Small Areas Approach to Adjust for Subnational Cost of Living, Global Solutions Group on Welfare Measurement and Statistical Capacity, World Bank, ppt provided by authors.
Deaton, A. (2003) Prices and poverty in India, 1987–2000. Economic and Political Weekly, 38(4), 362–368
Deaton A. and Dupriez O. (2011), Using unit-values to assess spatial price differences: evidence from India and Brazil, 5th Technical Advisory Group Meeting, Document 09.01.
Dikhanov Y., Palanyandy C., and Capilit E. (2011) Sub-national purchasing power parities toward integration of international comparison program and the consumer price index: The case of Philippines. ADB Economics Working Paper Series, No. 290, Asian Development Bank.
Fenwick, D., & O'Donoghue, J. (2003). Developing estimates of relative regional consumer price levels. Economic Trends, 599, 72-83
Giusti C., Masserini L. and Pratesi M. (2017), Local comparisons of small area estimates of poverty: an application within Tuscany region in Italy, Special Issue, Social Indicators Research
Hamadeh N., M. Mouyelo-Katoula, P. Konijn and F. Koechlin (2017), Purchasing Power Parities of Currencies and Real