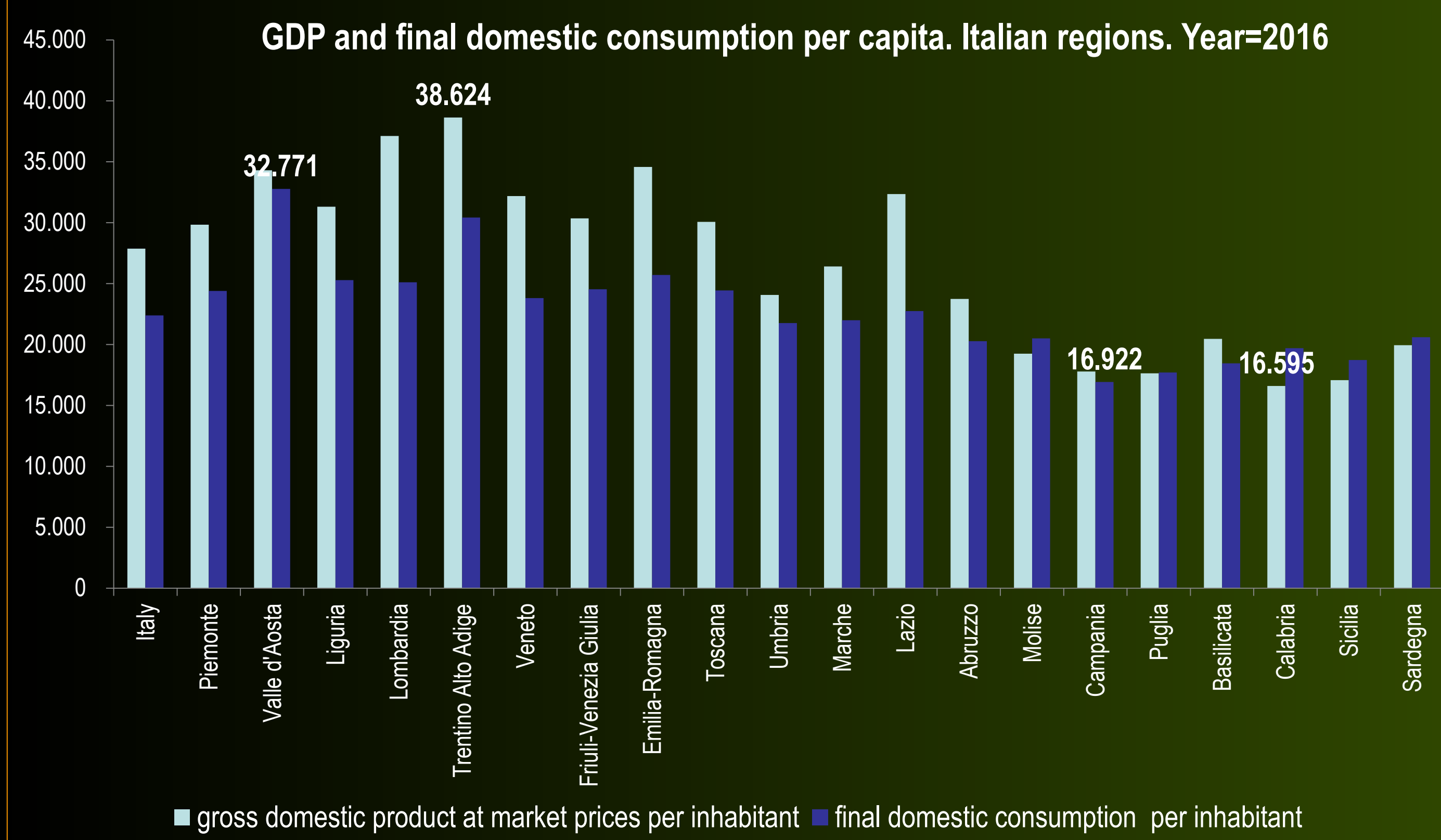


### Introduction

The main aim of the paper is to investigate how to deal, from the methodological and computational point of view, with the availability of different sources of elementary data of consumer prices (data coming from traditional data collection, scanner data, administrative data, other sources) to compile sub-national, that is **Regional Purchasing Power Parities (RPPPs)** adopting a multi-sources approach (already on the way for consumer price survey). Multi-sources approach means specific treatments and specific methodological solutions to allow each data source to contribute to the compilation of RPPPs.

### 1. Socio-economic differences across Italian regions



### 2. Results obtained in previous experiments of RPPPs

Italian National Statistical Institute (Istat) computed regional consumer price level indexes, with specific reference to household consumption, in two experiments, in 2008 and 2010.

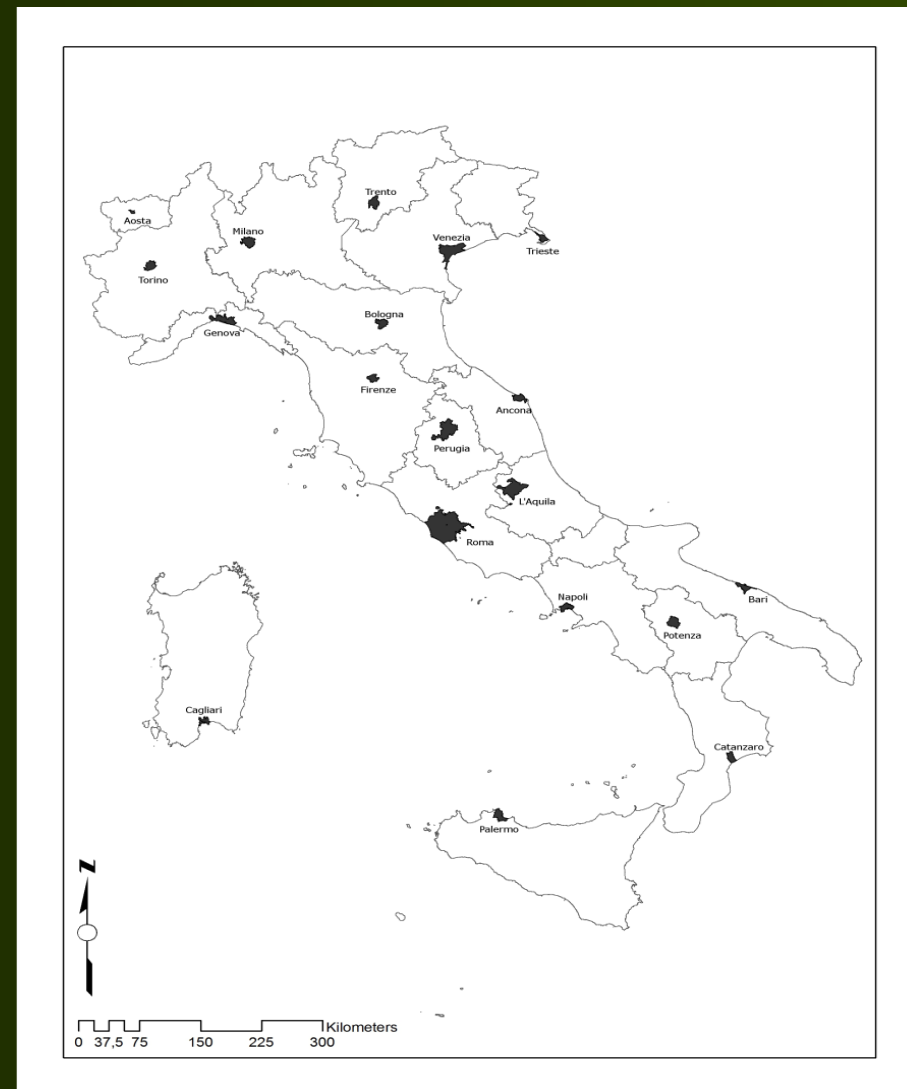
In 2010 (with reference to 2009 data) RPPPs were compiled for all COICOP expenditure divisions. GEKS formula and CPD model for actual rents were used. CPI data and ad hoc surveys referred only to regional chief towns.

#### Main results obtained

- significant differences in the level of consumer prices
- consumer price levels in northern cities generally higher than in the Centre and especially in the South
- Bolzano (105.6) and Milano (104.7) showed the highest prices compared with the Italian average (100.0) while the less expensive city proved to be Naples (93.8)

#### Main limits and drawbacks

- time consuming extensive data editing was necessary for using CPI data
- high costs for carrying out ad hoc surveys
- due to costs and labor-intensive preliminary analyses, spatial comparisons involved only the 20 Italian regional chief towns as representative for the entire region

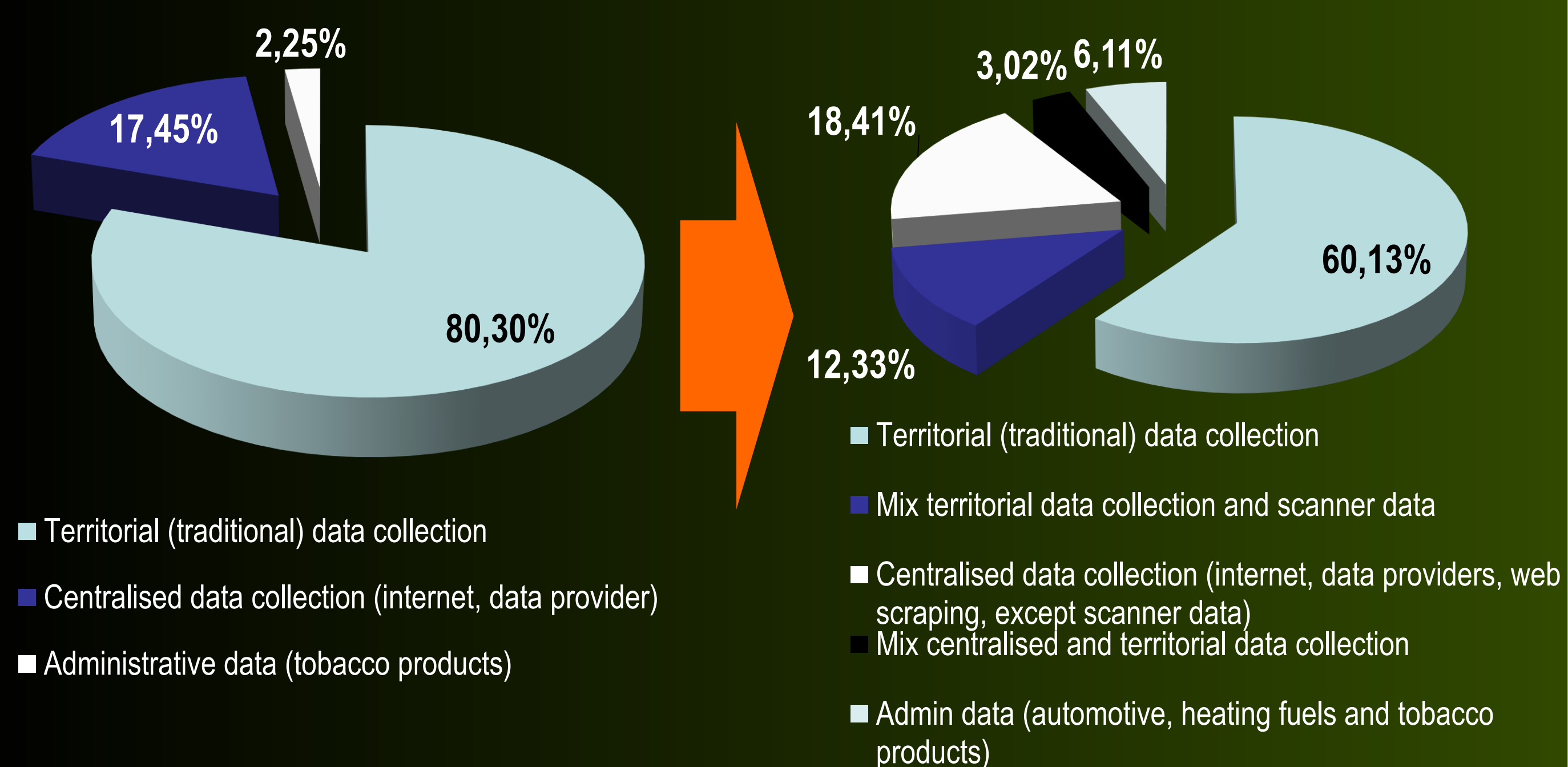


Starting from these experimental activities, Istat has planned to go towards the **regular production of RPPPs by including a specific project in the Italian National Statistical Plan.**

### 3. Evolution towards multiplicity of data sources for Italian consumer price statistics

Italian HICP data sources and data collection techniques. 2009 weights

Italian HICP data sources and data collection techniques. 2019 weights



### 4. Issues, solutions, potentialities of different CPI data sources for compiling RPPPs

#### Data coming from territorial data collection

- Identifying comparable products and enhancing the overlap degree
  - Implementing a string matching algorithm for identifying comparable products
  - Widening the use of electronic data collection besides Rome that is involved in OECD-Eurostat exercise
  - Carrying out an additional survey (every 2 years) for collecting prices for computing RPPPs
- Geographical coverage (for CPI it regards chief towns of the Italian provinces participating in the CPI survey that in 2019 are 79 out of 107 for the whole product aggregates included in the basket)
  - Widening the territorial coverage by motivating Municipal Statistical Offices using financial supports
  - Widening the use of alternative sources (admin data above all) as it the case of actual rents

#### Data coming from centralised data collection

- Localization of data collected. In some cases not relevant (mobile phone services) in other cases relevant (some goods), taking into account the growing role of e-commerce that tend to reduce price differences between different geographical areas

#### Administrative data

- Implementing the use of prices of fuels for RPPPs aims and starting the use of data of actual rents (coming from Italian Tax Office) for both temporal and spatial comparison aims

#### Scanner data

- Several advantages of this data source: granularity, comparability of elementary items through GTINs, availability of turnover and quantity information (weights), territorial coverage
- Expanding besides Hyper and Supermarkets to other retail trade channel (discounts, drug specialists, small surfaces of selling) the use of scanner data for both CPIs and RPPPs (2020)

### 5. The experimental use of scanner data to compile Italian RPPPs

- First tests in 2016-2017 using 2015 data in order to explore potential advantages of the use of scanner data compiling RPPPs (suitability of scanner data for making spatial comparisons) and to deal with the empirical issues deriving from the use of this new data source
- In 2018 (data of 2017) further analyses aimed at exploring the feasibility of implementing various aggregation methods at BH level and assess their performance to estimate RPPPs for product aggregates. Data and survey framework (the same for CPI) were:

#### OUTLETS

- Stratified random sample (888 strata identified by province, retail trade chains and kind of outlets, covering the entire national territory) of 1,781 outlets (510 hypermarkets and 1,271 supermarkets) selected from a universe of 9,000 retailers belonging to the 16 most important retail chains (94% of modern retail chain distribution), with probabilities proportional to 2016 turnover

#### ITEMS

- 487,094 different products (GTINs) belonging to food, beverages and personal and home care products: five divisions of the ECOICOP (01, 02, 05, 09, 12), 56 Basic Headings with scanner data cover 55.4% of the total retail trade for this category of products
- Items were selected with probabilities proportional to the 2016 turnover for each product aggregate (at 60% cut-off line). Chain structure in overlapping products

#### PRICE CONCEPT

- Annual averages of weekly prices (average of prices paid by consumers) for each item and outlet using turnover as weights
- Provincial averages using sampling weights for each outlet

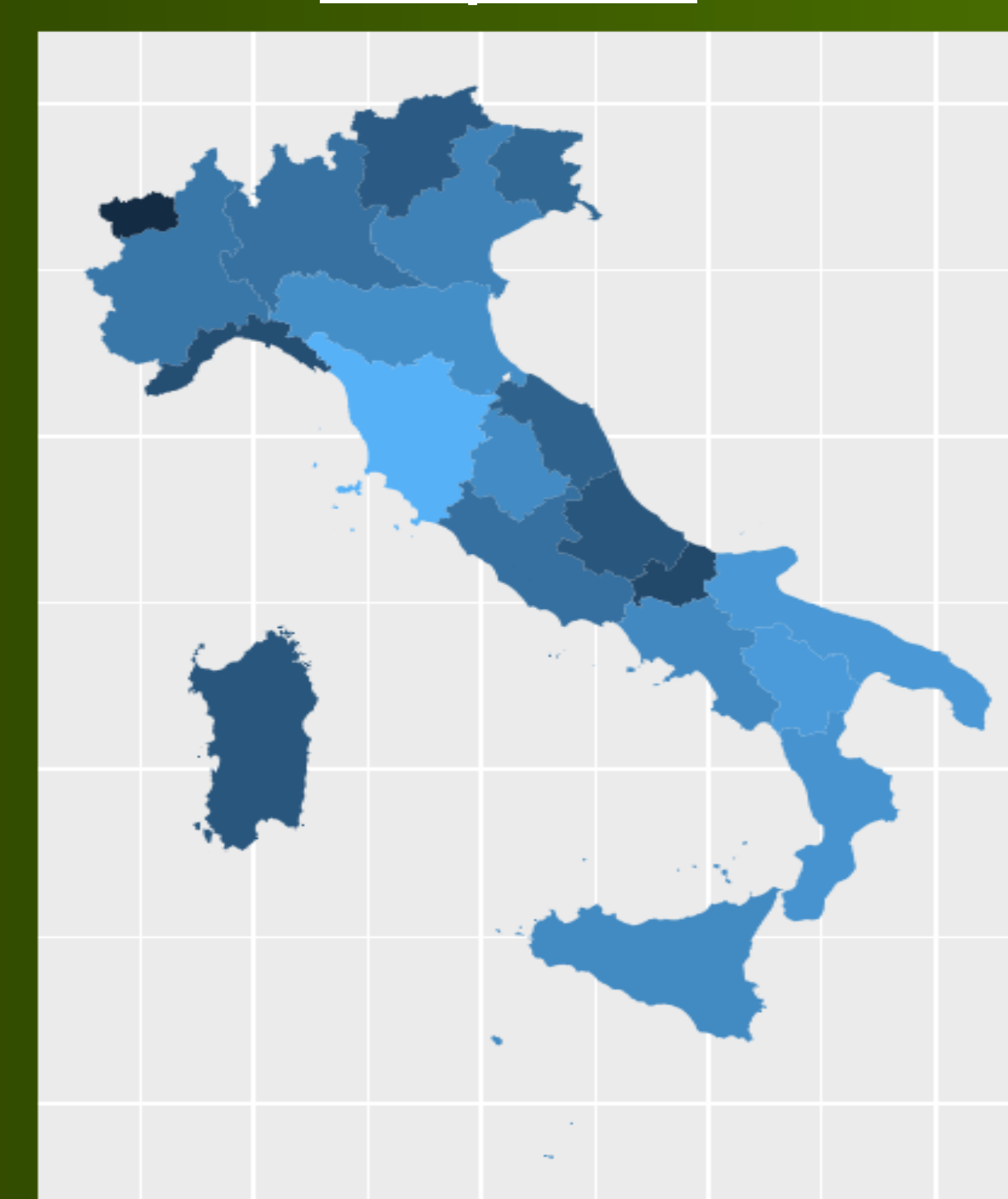
#### METHODS

- Weighted regional CPD using expenditure weights at item and BH level and then GEKS to aggregate above BH level

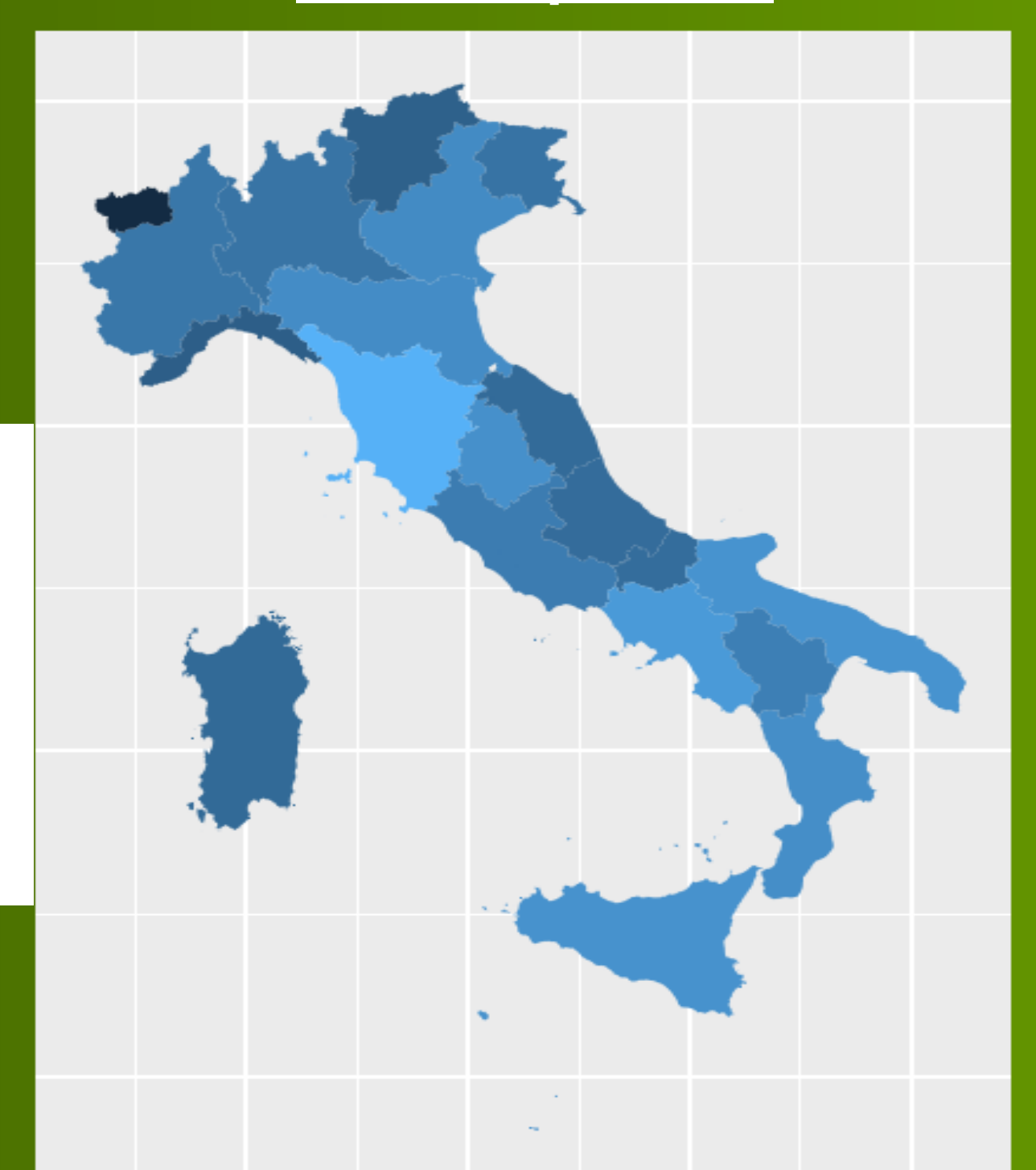
### 6. The results of 2017 analyses of RPPPs using scanner data

#### RPPPs FOR FOOD AND NON FOOD PRODUCTS (ITALY=100)

##### Food products



##### Non food products



- Southern regions with price levels that are below the national average both for Food and Non-Food products, with the exception of Abruzzo (101.90 and 101.33, respectively), Molise (102.90 and 101.24) and Sardegna (101.93 and 101.57)
- Some Northern regions with price levels lower than the national average, such as Emilia-Romagna (98.31 and 98.40), Veneto (99.09 and 98.48) and Piemonte for Food products (99.80). Toscana proved to be the less expensive region for both product aggregates (96.24 and 95.17)
- Considering the retail modern distribution, the expected relationship in terms of price levels between the North and South of Italy partially changes

### 7. Aggregating different data sources for different retail trade channels for RPPPs

- A mix of different retail trade channels, data sources and groups of product have to be taken into account (for RPPPs, but also for CPI)
- In particular for grocery products (and in the future for unprocessed food and other groups of products) the weight of each retail trade channel and also (if not always) of its own data source has to enter the RPPPs compilation
- The example of grocery products with elementary prices collected by municipal data collectors for the traditional retail trade channel, through scanner data for the modern distribution or the case of clothing and footwear, where traditional and modern distribution are increasingly joined by e-commerce channel
- The mix of two sources of data has to be carried out at the lowest level of aggregation that is BH level through a two-step procedure similar to the one used in the ICP whereby provinces are grouped into regions:
  - Provincial sub-national PPPs (or within-regional) are computed by comparing products sold in the various provinces within each region and using different methods according to the data sources:
    - Unweighted country product dummy models (in the case of price data from the territorial data collection for the traditional distribution)
    - Weighted country product dummy models with expenditure share weights for scanner data/modern distribution
  - RPPPs (or between-region) are computed by using regional prices adjusted for differences among provinces for each region (obtained by dividing provincial prices by the within-regional PPPs) thus taking into account the regional heterogeneity in consumer prices. Weighted CPD models are used with a different specification of weights:
    - Unweighted country product dummy models (in the case of price data from the territorial data collection for the traditional distribution)
    - Weighted country product dummy models with expenditure share weights for scanner data/modern distribution

Therefore, after the two previous steps, two set of RPPPs are obtained for the two retail trade channels (traditional and modern). They are aggregated using:

#### As Aggregation method

- GEKS- Fisher (ICP and Eurostat-OECD)
- Standardized the GEKS-Fisher based PPPs (S-GEKS)

#### As Weights

- Expenditure share weights defined using information from the HBS at regional level and taking into account market share of traditional retailers

### 8. Conclusive remarks and perspectives

Istat is committed in the activity to achieve the objective of compiling reliable RPPPs to be produced on regular basis (every two years) solving the main issues emerged in the last years of experimentation. The first step is the definition of a "matrix" where for each BH the map of the retail trade channels, of service providers and of data sources available is accurately traced together with the methodologies to compile the PPPs at the lowest level of aggregation and the structure of weights. Then an intensive work on the CPI data base of elementary prices traditionally collected will be carried out in the coming months to make feasible their correct use for RPPPs. Moreover it is on the table the hypothesis to launch a cycle of data collection at territorial level in the next year (May or June) to integrate further CPI information with prices collected on the basis of basket specified for spatial comparison aims using the software recently developed. Afterwards experimental RPPPs will be compiled covering the entire basket of products and as far as possible the different retail trade channels and they will be disseminated via Istat web site dedicated to the experimental statistics.