Soc Indic Res (2017) 131:93–121
DOI 10.1007/s11205-015-1217-x

Computing Sub-national PPPs with CPI Data: An Empirical Analysis on Italian Data Using Country Product Dummy Models

Luigi Biggeri[1] · Tiziana Laureti[2] · Federico Polidoro[3]

Accepted: 22 December 2015 / Published online: 4 January 2016

Abstract The Italian National Statistical Office is implementing a new project for computing sub-national Purchasing Power Parities (PPPs) on a regular basis, which is based on an appropriate use of existing Consumer Price Index (CPI) data and new sources of data. Concerning the use of CPI data, in this paper the role of the Country Product Dummy (CPD) method for compiling sub-national PPPs at Basic–Heading (BH) level is analysed together with the specific issues that arise in this context, such as the need to take spatial autocorrelation among price relatives into account. The results of various experiments based on CPI data concerning 7 BHs and 19 Italian regional chief towns are presented and discussed with the aim of exploring the performance of various CPD models and analysing to what extent the type and characteristics of the data affect the estimates obtained. The statistical uncertainty associated with the ranks derived from the regional PPPs was then calculated by means of a simulation procedure. Our findings prove to be interesting and confirm that methods and CPI data for spatial comparisons are reciprocally influenced.

Keywords Sub-national PPPs · Hedonic CPD · Spatial correlation · Spatial error models · Multipurpose price statistics

✉ Tiziana Laureti
laureti@unitus.it

[1] Department of Statistics, University of Florence, Florence, Italy

[2] Department of Economics and Management, University of Tuscia, Via del Paradiso 47, 01100 Viterbo, Italy

[3] Consumer Price Division, Istat, Rome, Italy

1 Introduction

Over the last three decades there has been much debate about constructing sub-national Purchasing Power Parities (PPPs) for carrying out analyses on inter-area price levels, standards of living and real income comparisons on topics such as poverty, rural–urban and regional (local) differences. When reviewing international practices, it is important to note that more and more academics are carrying out interesting experimental estimations of sub-national PPPs yet up to now few countries have produced official indexes of spatial prices or have carried out experiments to this aim. In addition, the Technical Advisory Committee (TAG) of the International Comparison Program (ICP) at the World Bank during its meeting on February 2010 discussed and stressed the importance of the computation of sub-national PPPs (ICP-TAG 2010).

As yet, the most promising approach is to compute sub-national household consumption PPPs using the data collected by the National Statistical Offices (NSOs) for compiling Consumer Price Indexes (CPIs) and Country Product Dummy (CPD) methods. In this context, the Italian National Statistical Office (Istat) is implementing a project with the aim of computing regional PPPs on a regular basis. There are two main objectives to this project: to develop a large data warehouse containing the elementary information required for compiling sub-national PPPs by integrating the data obtained from new sources with the data already currently collected for CPIs; to identify the best method to adapt this elementary information for constructing sub-national PPPs.

Within this framework, this paper aims at exploring various CPD models for estimating sub-national PPPs, at Basic–Heading[1] (BH) level, using Istat CPI data. Therefore, after reviewing studies on the construction of sub-national spatial price indexes in various countries, we focus on the role of the CPD methods for compiling sub-national PPPs at BH level since it is essential to obtain reliable PPPs at this level of aggregation as they are the foundations of overall comparisons (Hill and Hill 2009; Hill and Syed 2014). To this aim the performance of various CPD models based on different kind of data (average price or individual price data) are analysed and compared together with specific issues that arise in this context, such as the need to take spatial autocorrelation among prices into account that will certainly be more relevant among areas within a country than across nations (Aten 1996; Rao 2004). With the aim of making multiple comparisons thus determining if the purchasing power of a country's currency unit in a region can be significantly distinguished from the others in that country we apply a simulation method, based on ranks, suggested by Marshall and Spiegelhalter (1998) and previously used in other fields (Leckie and Goldstein 2011). To analyse whether and to what extent the performance of the various CPD models for estimating the PPPs at BH level are influenced by the type and characteristics of the data we carried out various experiments based on Istat CPI data regarding 7 BHs belonging to the Food and non-alcoholic beverage group and concerning the 19 Italian regional chief towns included in the 2014 CPI survey.

The remainder of this paper is structured as follows. Section 2 reviews the proposals and the experiments carried out for constructing sub-national PPPs with the aim of obtaining some guidelines for our research study. Section 3 deals with the methodological and empirical issues related to the estimation of sub-national PPPs using CPD methods at

[1] In the ICP, the Basic Heading (BH) is defined as the lowest level of aggregation within the National Accounts at which expenditure and expenditure share data are available. In the 2011 ICP Round the PPPs have been computed for 155 BHs.

BH level. In Sect. 4 the results obtained from various CPD models by using different sets of information are presented and discussed. Some concluding remarks are drawn in Sect. 5.

2 Review of Proposals and Analyses Carried Out for Computing Sub-National PPPs

2.1 Computation of Sub-national PPPs Carried Out by NSOs

Although the first official measure of inter-area differences in the cost of living was developed in the 1940s in the US (Sherwood 1975), up to now few countries have produced official indexes of spatial prices or have carried out experiments to this aim (US, Australia, UK and Italy).

In the US, the inter-area price level comparisons led to interesting developments in the mid-1990s as a result of the studies carried out by various researchers, mostly employed by the Bureau of Labor Statistics and the Bureau of Economic Analysis, see for example Kokoski (1991) and Kokoski et al. (1999). Further studies were later carried out by Aten (see for example 2005 and 2006), who has continued to broaden research in this field concerning both products and areas considered, thus obtaining Regional Price Parities (RPPs) for 50 states and the District of Columbia and 366 Metropolitan areas (Aten et al. 2014). These estimations, which compare the average price level of a specific area with the national average price level, were computed by using detailed price data collected for the CPI calculation and both the Gini–Elteto–Koves–Szulc formula (GEKS) and various CPD models. The analysis of the characteristics of the data and especially the methods used for computing PPPs at product level, basic headings and aggregated commodity group level have been improved over the last decade, thus including the measures in an economic approach framework (Aten and Reinsdorf 2010).

In the early 2000s the UK Office for National Statistics (ONS) carried out various exercises to produce indicative figures concerning the variation in prices between regions. These first analyses were carried out as a by-product of a survey conducted to provide data for the Eurostat's PPP programme (Fenwick and O'Donoghue 2003). The ONS produced relative regional consumer price levels based on CPI price data (mainly food items, tobacco and drinks) supplemented with administrative data sources and, more importantly, with a purpose-designed regional price level survey for items of expenditure where suitable data was not available (mainly clothing, furniture, electrical goods and travel), as reported by Wingfield et al. (2005). The ONS then produced relative regional consumer price levels for 2010 following a more suitable procedure (ONS 2011) which involved using data on price observations from the existing monthly CPI collection and regional price surveys conducted for computing the Spatial Adjustment Factors (SAFs) required by the Eurostat-OECD PPP program.

The Australian Bureau of Statistics (ABS) compiled and disseminated experimental indexes on the cost of living in the 8 capital cities using existing price data collected for CPIs and calculated spatial price indexes using GEKS (Waschka et al. 2003). ABS continued to compute and disseminate data on the average retail prices of selected items in eight capital cities, advising a careful interpretation of the results; however in 2011, the ABS discontinued the publication, as established in the June CPI release.

The Italian Statistical Office (Istat) carried out two experiments for computing regional PPPs in cooperation with *Unioncamere* and *Istituto Tagliacarne*, and disseminated the results in two occasions: in 2008 with reference to 2006 data and in 2010 with reference to 2009 data. In order to calculate sub-national PPPs for consumer prices for 20 Italian cities (the regional chief towns), the procedure for international comparison was adopted. Therefore a complicated analysis of CPI data was carried out with the aim of checking whether the characteristics of products were the same across the various cities. Moreover, *ad-hoc surveys* were designed and carried out for the product groups "Clothing and Footwear" and "Furniture". In the first experiment, sub-national PPPs, obtained by means of the GEKS formula, were computed for three expenditure divisions (Food and Beverages, Clothing and Footwear, Furniture), which represented approximately 34 % of the total consumer expenditures (Istat 2008). In the second experiment, the sub-national PPPs were compiled for all the COICOP expenditure divisions. The sources of data and methodology remained the same with the exception regarding actual rents for which the spatial comparison was carried out using CPD models and Household budget survey data that includes some detailed information about the characteristics of the dwellings (Istat 2010).

It is worth noting that Statistics New Zealand (SNZ) has been looking in the possibility of carrying out spatial comparisons of prices since 2005 and it has appointed two experts to develop a methodology for constructing Spatial Cost of Living Indexes. The experts' report (Melser and Hill 2005) deals with all the issues and offers advice on constructing sub-national PPPs, in particular on the use of CPD methods, thus providing a reference text for implementing PPPs in other countries. However, SNZ has never disseminated data on sub-national PPPs to the authors' knowledge.

2.2 Computation of Sub-national PPPs Carried Out by Researchers

Many researchers have carried out studies and experiments with the aim of estimating sub-national PPPs. Most studies have been developed through the cooperation with the national or international statistical agencies, which have provided CPI data for the analyses.[2] One of the most important paper is the study carried out by Aten (1999) who used the CPD method for estimating the 1987 price levels on food products and the general price levels in 10 Brazilian cities. At that time regional price statistics were not widely available therefore the practices for price collection were discussed and three alternative index number methods for constructing spatial regional indexes, such as Geary, GEKS and Fisher average methods, were examined.

Some studies for computing spatial indexes (SPIs) have been carried out also for China's provinces. The most important works have been conducted by Brandt and Holz (2006), who estimated the SPIs for the base year 1990 by using the cost basket method, and by Li et al. (2005) who estimated PPPs for the year 2002.

The World Bank and the Asian Development Bank carried out a specific experiment in the Philippines (ICP-TAG 2010; Dikhanov et al. 2011) aimed at analyzing the plausibility

[2] In order to estimate consumer sub-national PPPs a different approach have been proposed which is based on the Engel's Curve and/or a demand system model applied to data collected with sample surveys on household consumption expenditure. To the authors' knowledge, this approach was used for the first time by Coondoo et al. (2004), who reported results obtained for India. The Authors explained that their procedure follows the CPD methodology in a generic sense as price equations essentially share the hedonic feature. Majumder, Ray and Sinha have continued to work on this approach in order to compute regional PPPs for India, Indonesia and Vietnam, with the aim of proposing a unified framework for estimating intra and inter-country PPPs, and obtained interesting results (Majumder et al. 2013a, b, 2014; Mishra and Ray 2014). This approach proved to be a useful tool and should be extended to other countries in order to check the validity and compare the results with the findings obtained from analyses carried out with other methods of computation.

of integrating the ICP with the Philippine CPIs by computing sub-national PPPs using regional prices and expenditure weights from the CPI. The purpose of the study was to establish whether the prices collected for the CPI could be used to provide reliable estimates of price levels for a range of products in each region and to verify if the results are consistent with the information obtained from the ICP process. This paper describes all the necessary phases for constructing regional PPPs underlining the main issues and solutions for overcoming them, both concerning data preparation and methods of index computation and aggregation. The CPD method was used in elementary aggregation.

Moreover, the topic of estimating sub-national PPPs has been discussed in several meetings of Eurostat working groups. Indeed, Eurostat has supported various countries as already mentioned for the NSO in UK, which broadened regional price surveys carried out for computing SAFs as required by the Eurostat-OECD PPP program.

2.3 Evidence from the Review of Literature and Experiments

Bearing in mind the various studies and analyses of the mentioned works and the current ICP strategy, it is evident that more in-depth analyses are required for computing sub-national PPPs using CPI data. On the other hand, many advantages can be obtained by using CPIs data and CPD methods for computing sub-national PPPs:

(i) Availability—in most countries price data collected for computing CPIs can be easily obtained on a regular basis;
(ii) Money and time saving—the ICP surveys could reduce the collection of specific data in the countries that participate to the Program and obtain the requested integration between PPP and CPI construction (Rao 2001a; Biggeri et al. 2008; Biggeri and Laureti 2010);
(iii) Unified framework—price data collections may be carried out contemporarily for computing CPIs, sub-national and international PPPs.

3 Exploring CPD Models for Estimating Sub-national PPPs Based on CPI Data

3.1 Why Should We Use CPD Methods?

Various CPD models can be specified with the aim of constructing inter-area PPPs. The choice of the most suitable model depends on the type and characteristics of the data which in turn may affect the performance of the methods.

The requirements established for international price comparisons should also be met for price data to use for sub-national PPP estimation, that is, they should be representative of consumers' expenditures and comparable across areas within a country. By definition, the products priced to be used for constructing CPIs within a country are representative of the expenditures in that country, otherwise they would not have been included in the basket of goods and services being priced. Therefore, representativeness should not be an issue when comparing price levels across areas within a country with CPI data. However, comparability can cause some problems. The main issue concerning comparability is the large range of products that the staff collecting price products for computing CPIs may select by following broad parameters that distinguish the type of product to be selected. Apart from satisfying the broad specifications, the main criterion when selecting a product to be priced for computing CPIs is that it must be representative of local consumption behavior (for example the criterion of the most sold elementary item) and must be monitored month by month so that the prices observed are not affected by changes in quality. Consequently, the actual products priced for the CPIs are not necessarily identical across the various areas.

Therefore, for sub-national PPPs, it is essential to match the prices for those items that not only meet the CPI specifications but which must also be comparable, or, in other words, have the same brand and model if they are price-determining characteristics. When identifying the products to be compared, the problem of missing price observations will be accentuated and the risk of a poorer coverage of the spatial index will be increased if a "tight" product specification is adopted. On the other hand, "loose" product specifications increase the possibility of matching items, but may result in price comparisons of dissimilar items. Therefore, there is a trade-off between having tight specifications and poor coverage of items and loose specifications with price differences contaminated by quality differences (Silver and Heravi 2005; Silver 2009).

In order to deal with these data problems arising from the quality variation of items across areas and gaps in the available price data, the Country Product Dummy (CPD) regression methodology may be used (Summers 1973).

The CPD, which is essentially an implementation of the *hedonic* approach accounting for the quality variations in price data, provides a regression analysis-based econometric methodology for constructing multilateral price index numbers that accounts for the quality variations in the cross-area price data (Kokoski et al. 1999). Over the years the CPD methodology has undergone immense theoretical improvements. Although the initial application of the CPD method was for filling gaps in the price tableau prior to the computation of PPPs at BH level (Summers,1973), several Authors have demonstrated that, due to its econometric nature, the CPD method could be extended and generalized in order to provide a comprehensive framework for carrying out both international comparisons[3] (Rao 2001b, 2002, 2005, 2013; Diewert 2005, 2010) and intra-national comparisons (Kokoski et al. 1999; Aten 2005, 2006; Dikhanov et al. 2011).

Since the CPD method is based on a stochastic formulation, it has the additional advantage that it enable us to use a range of econometric tools and techniques that are not generally used for computing PPPs. For example, the basic CPD method can be extended to account for spatial autocorrelation. Moreover, recent studies by Hill and Hill (2009) and Hill and Syed (2014) have underlined that the most pressing concern in spatial price comparisons is to obtain unbiased price indexes at basic heading level as they are the foundations of overall comparison. In this framework the CPD methods can help to reduce errors that are likely to arise when calculating basic heading price indexes. The choice of the CPD model to be used for estimating intra-national PPPs is strongly influenced by the type and characteristics of the CPI data. The structure and characteristics of the price database may vary across countries as national statistical offices adopt procedures that best suit their own needs.

[3] In the 2005 ICP round the CPD approach has been the recommended method of aggregation below the basic heading level. Hill and Syed (2010) used detailed price data provided by the Word Bank with the aim of analysing how the CPD could improve PPP estimates in the context of the ICP.

However, it is reasonable to assume that a micro-level data matrix with information on annual average prices[4] for each product from the various outlets (price quotes) is available from all areas. In this case, intra-national PPPs can be estimated using CPD models for ungrouped data (Diewert 2004, Silver 2009). Moreover, quality adjustments for not strictly comparable products can be incorporated by using a hedonic framework (Kokoski et al. 1999; Silver 2009).

On the other hand, it may be interesting to analyse to what extent the type and characteristics of the data affect the estimates obtained in order to understand which data should be used and whether the differences between the methods depend on data. To this aim a data matrix composed of average prices (across outlets) for each product in all areas should be constructed by calculating an unweighted average of all the prices collected for that product. The average is often an arithmetic mean, although the geometric mean is becoming more and more popular thanks to its useful properties.

3.2 The CPD Model Based on Individual Price Data: The Hedonic Approach

Let us assume that we are attempting to make a sub-national comparison of prices between R areas at BH level (no expenditure weights are available for the price comparisons) and for each BH p_{knr} denotes the annual price in outlet k of product n in area r ($n = 1, 2, \ldots N$; $r = 1, 2, \ldots, R$; $k = 1, \ldots K_{nr}$). It is possible to incorporate available individual price quotes into an unweighted CPD model as observations along the lines suggested by Diewert (2004). Assuming that $Z_1, Z_2, \ldots, Z_J$ represent a set of quality characteristics, associated with each quotation, including information on the type of outlet and product brand, the hedonic CPD model estimates the following regression equation separately for each basic heading:

$$\ln p_{knr} = \sum_{r=1}^{R} a_r D_r + \sum_{n=1}^{N} b_n D_n^* + \sum_{j=1}^{J} \lambda_j Z_j + v_{knr} \tag{1}$$

where D_r and D_n^* are respectively area and commodity dummy variables while v_{knr} denote random error terms which are independently and identically normally distributed with zero mean and variance σ^2. The intra-national PPP for area r is given by: $PPP_r = \exp(a_r)$.

This model generalizes the standard formulation developed by Summers (1973) which was employed in the calculations for the initial studies of the World Bank's International Comparisons Program and in which missing "unmatched" prices are assumed to be randomly distributed. By using CPD formulation (1) the price variability is modeled right down to the level of the individual prices that are collected by the areas involved in the comparison (Diewert 2004). Moreover, it is possible to include strictly non-comparable items when estimating sub-national PPPs. On the other hand, incorporating quality characteristics will improve the efficiency of the estimates and remove potential bias[5] also in the case of comparable products. Indeed, as a normal part of CPI calculations, each product price will include the detailed specifications which define the type of product (size, brand) and the type of outlet from which the product is purchased. Preferences concerning brands and type of outlet may vary across areas due to different consumption habits. Therefore, it

[4] Usually, monthly observations are averaged by using an arithmetic mean. Recently, Dikhanov et al. (2011) suggested a combined spatial–temporal model, called Country-Time-Product-Dummy method, to use quarterly CPI data.

[5] Silver (2009) suggested that it is possible to improve on the specification of (1) by also introducing quality and product interaction terms.

is essential to include this information into the CPD model specification in order to improve PPP estimations.

3.3 The CPD Model Using Average Prices: Unweighted Versus Weighted Approach

When constructing inter-area PPPs it is reasonable to assume that several individual price quotes for each product are available and therefore incorporated in the CPD model (1). On the other hand, an average price may be computed on price quotes collected in various outlets to be used in inter-area PPP construction.

In this case, the CPD model can be written as:

$$\ln \overline{p}_{nr} = \sum_{r=1}^{R} a_r D_r + \sum_{n=1}^{N} b_n D_n^* + v_{nr} \tag{2}$$

where D_r ($r = 1, 2, \ldots, R$) and D_n^* ($n = 1, \ldots, N$) are respectively area and product dummy variables and $\overline{p}_{nr}$ is the average price of product n in area r.

However, if average prices for each area are used, it is important to include information on the number of quotations and the standard deviations per product for estimating a Weighted Least Squares (WLS) version of CPD (2) with the aim of improving the efficiency of the estimated BH price indexes. Indeed, when CPD formulation (2) is estimated with Ordinary Least Estimator (OLS) equal weights are given to the average prices of all areas. As underlined by Silver (2009) the use of the CPD method as formulated in Eq. (2) is negatively affected by the use of grouped data and the inability to include quality variations. In other words, grouping individual price quotes can result in efficiency loss due to heteroskedasticity that may be mitigated by using a WLS estimator.

Both arithmetic and geometric means may be used for computing the average price for product n in each area r. However, although the arithmetic mean has been used for computing average prices in ICP process up to now, various Authors have argued that geometric means are better as they rely on weaker assumptions (Rao 2004, Hill and Syed 2014). Therefore, we considered as dependent variable in Eq. (2) the logarithm of the geometric mean of price quotes for product n in area r, which can be expressed as the arithmetic mean of the logarithms of price quotes:

$$\ln \overline{p}_{nr} = \ln \left[\left(\prod_{k=1}^{K_{nr}} p_{nrk} \right)^{1/K_{nr}} \right] = \frac{1}{K_{nr}} \sum_{k=1}^{K_{nr}} \ln p_{nrk} \tag{3}$$

Hill and Syed (2014) suggested that for measuring reliability of the various average price observations when geometric means are used it is necessary to have information on the variances of the log price quotes $\mathrm{var}(\ln p_{nrk})$ and the number of price quotes K_{nr}. In this case the following expression can be calculated:

$$w_{nr} = \frac{1}{\overline{\mathrm{var}}_n + \mathrm{var}(\ln \overline{p}_{nr})} \tag{4}$$

where the variance of the log average prices[6] is estimated using $\mathrm{var}(\ln \overline{p}_{nr}) = \frac{\mathrm{var}(\ln \mathrm{p_{nrk}})}{K_{nr}}$ and $\overline{\mathrm{var}}_n$ is the average variance of the log average prices across the whole set of areas and is included to ensure that w_{nr} is still defined whenever $\mathrm{var}(\ln \overline{p}_{nr}) = 0$ for a particular area r.

[6] Hill and Syed (2014) underlined that their derivation assumes only that the price quotes are identically and independently distributed.

3.4 The CPD Model with Spatially Auto-Correlated Error Structure

In this section we focus on some spatial econometric aspects related to hedonic CPD models. Indeed, an issue which can frequently arise, when estimating sub-national PPPs, is that of spatial correlation among disturbances across areas for a specific commodity. However, this issue has not yet been investigated in the context of sub-national PPP construction by means of CPD models.

Regarding international comparisons, evidence of spatial autocorrelation for a number of commodity groups among countries was observed by Aten (1996) and Rao (2001b). Using household consumption data for 64 countries and 23 aggregate headings in 1985, Aten (1996) tested for spatial autocorrelation among price relatives and found that all headings were significantly and positively autocorrelated by at least one of the weight matrices used. Focusing on methodological issues, Rao (2004) drew attention to the effects of spatial autocorrelation and the adjustments required for the estimates.

It is worth noting that the inclusion of area dummies in CPD models (1), (2) leads to the so-called spatial fixed effects, which relates to a cross-sectional setting (Anselin and Arribas-Bel 2013). However, this specification may not be sufficient to correct for the presence of spatial correlation. In other words, the removal of spatial autocorrelation by spatial fixed effects may be spurious when the true data generating process (DGP) takes the form of a spatial lag or spatial error dependence.

Spatial patterns in consumer prices may arise from a combination of spatial heterogeneity and spatial dependence[7] (Anselin 2006). It is reasonable to assume that spatial heterogeneity may derive from spatially differentiated characteristics of demand, supply and retailers. Particular attention should be paid to this systematic variation in the behavior of economic agents across areas, since any model that imposes homogeneity will be misspecified.

Positive spatial autocorrelation may appear when either the prices or characteristics of products that are purchased in adjacent areas are more similar to one another than from products that are purchased in more distant areas. Alternatively, spatial autocorrelation may also be due to measurement problems in explanatory variables, omitted variables and other forms of model misspecification (Baumont 2004).

In order to test for the presence of spatial autocorrelation the well-known Moran's I statistic can be used. Rao (2004) suggested using residuals from an OLS CPD regression in order to test for spatially autocorrelated errors since the disturbance term for the *n-th* commodity in area r is the logarithm of price of commodity n, p_{nr}, converted by using PPPr and expressed in relation to the national average price of the same commodity.

Measures of spatial autocorrelation for residuals take into account the dependence among observations by means of a spatial weight matrix **W**, which defines the structure of spatial relationships. For a set of R areas the spatial matrix **W** is a $R \times R$ matrix with the diagonal elements equal to zero; the other elements w_{rs} represent the intensity of the effects of the area r on area s (Anselin and Bera 1998). Moran's I is essentially a cross product

[7] As outlined in Anselin (2006) "spatial dependence or spatial autocorrelation is a special case of cross-sectional dependence in which the structure of the covariation between observations at different locations is subject to a spatial ordering. This ordering is related to the relative positioning, distance or spatial arrangement of the observations in geographic space, or, more generally, in (social) network space. Spatial heterogeneity is a special instance of structural instability, which can be observed or unobserved. The spatial aspect of this issue is that spatial structure provides the basis for the specification of the heterogeneity. This may inform models for spatial structural change (referred to as spatial regimes), heteroskedasticity, or spatially varying and random coefficients".

correlation measure that incorporates "space" by means of **W**. This statistic can be expressed as follows:

$$I = \frac{R\sum_{r=1}^{R}\sum_{s=1}^{S} w_{rs}\left(\hat{v}_{nr} - \overline{\hat{v}}\right)\left(\hat{v}_{ns} - \overline{\hat{v}}\right)}{S_0\sum_{r=1}^{R}\left(\hat{v}_{nr} - \overline{\hat{v}}\right)^2} \quad r \neq s \tag{5}$$

where $S_0 = \sum_{r=1}^{R}\sum_{s=1}^{S} w_{rs}$. Significance can be based on analytical derivations, or, more commonly, on a comparison to a reference distribution obtained by randomly permuting the observed values.

From a theoretical viewpoint, a spatial error specification is a more natural way to include spatial effects in a hedonic model. Unobserved neighborhood effects will be shared by products and services in the same area and naturally lead to spatially correlated error terms. Elhorst (2014) underlined that "interaction effects among the error terms are consistent with a situation where determinants of the dependent variable omitted from the model are spatially autocorrelated, or with a situation where unobserved shocks follow a spatial pattern". This results in a non-diagonal error variance–covariance matrix: $Var(vv') = E(vv') = \Sigma$ where $\mathbf{\Sigma} \neq \mathbf{I}$, with $\mathbf{I}$ as the identity matrix. Typically, $\mathbf{\Sigma}$ contains "nuisance" parameters that need to be estimated consistently.[8] Consequently, OLS remains unbiased, but it is no longer efficient and the classical estimators for standard errors will be biased.

In practice, the most commonly used specification assumes a spatial first order autoregressive process for the error terms. By referring to the CPD model based on individual price quotes and stacking all the R observations for each commodity into a vector, the model containing spatial dependence in the disturbances can then be expressed as follows:

$$\ln \mathbf{p}_n = \mathbf{x}_n\beta + \mathbf{v}_n \tag{6}$$

$$\mathbf{v_n} = \lambda\mathbf{W}\mathbf{v_n} + \boldsymbol{\varepsilon}_\mathbf{n}$$

with the data generating process (LeSage and Pace 2009):

$$\ln \mathbf{p_n} = \mathbf{x_n}\beta + (\mathbf{I_R} - \lambda\mathbf{W})^{-1}\varepsilon_\mathbf{n}$$
$$\varepsilon_n \sim N\left(0, \sigma^2 I_R\right)$$

where λ is the spatial autoregressive coefficient. The resulting error variance–covariance matrix is as follows: $E(\mathbf{vv}') = \sigma^2\left[(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})\right]^{-1}$. If there is no spatial correlation between the errors for neighboring or connected area i and j, the spatial error parameter λ will be 0, and the model reduces to the standard linear regression model, thus the CPD model may be estimated by OLS in the conventional manner. However, if the spatial error parameter $\lambda \neq 0$, then a pattern of spatial dependence exists between the errors for connected area. In particular, the error terms are positively correlated if $\lambda > 0$ and negatively correlated if $\lambda < 0$. In order to obtain estimates of the unknown parameters a maximum likelihood (ML) estimator may be used which is based on the assumption of normal error terms (Ord 1975; Anselin 1988) and is expressed as follows:

[8] The interpretation of the nuisance parameters differs greatly from the spatial autoregressive coefficient in the spatial lag model, in that there is no particular relation to a substantive model of spatial interaction.

$$\ln L(\beta, \sigma, \lambda) = \ln|\mathrm{I} - \lambda\mathbf{W}| - N/2\ln(2\pi) - N/2\ln\left(\sigma^2\right) - \frac{(y - X\beta)'(y - \lambda\mathbf{W})'(y - X\beta)}{2\sigma^2}$$

By relying on Ord's (1975) result, the determinant can be written as a function of the product of the eigenvalues ϖ_i of the spatial matrix, i.e. $|\mathrm{I} - \lambda\mathbf{W}| = \prod_i (1 - \lambda\varpi_i)$. Since the estimates for β and σ can be expressed in function of λ after substituting these expressions into the likelihood, a concentrated likelihood can be found, which is solely a function of the autoregressive parameter λ. A ML estimate for λ can thus be found by means of a search over the acceptable interval ($1/\omega_{\min}$, $1/\omega_{\max}$).

It is worth noting that the ML estimators explicitly incorporate a weight matrix **W** which needs to be specified in advance. By means of the spatial weights matrix **W**, a neighbour set is specified for each area. However, many different definitions of the neighbour relation are possible, and there is little formal guidance in the choice of the "correct" spatial weights (Anselin 2002). Various measures of proximity may be used to specify **W**, such as contiguity, cardinal distance (e.g. kilometers) and ordinal distance (e.g., the k closest neighbors). However more and more research studies focus on the "economic" distance with the aim of defining spatial weights matrices. In this context various measures have been suggested in order to model spatial interactions (see for example, Case et al. 1993; Fingleton 2001; Corrado and Fingleton 2012).

In this paper we base the specification of the spatial weights matrix **W** on the economic distance suggested by Fingleton and Le Gallo (2008), hypothesizing that shocks (spillover) impact areas (towns) with similar economic and employment structures in a similar way, but also depend on inter-area distance. This measure is based on a negative exponential function of straight line distance d_{ij} (in kilometers) between areas i and j, and on the size of each area's economy measured in terms of the total employment level in 2014. The resulting matrix is then row standardized. A similar economic distance measure, based on the volume of trade between countries, has been previously used by Aten (1996).

4 Empirical Application Using Italian CPI Data

4.1 The Istat's Project and the Selection of Data for Experimental Analyses

The sub-national PPP results obtained in the two analyses mentioned in Sect. 2.1 have encouraged Istat to confirm the project of producing spatial indices of consumer prices at regional level on a regular basis.

As CPIs and PPPs have different purposes, the price data available from the CPI sources are not usually in a suitable form for carrying out spatial price comparisons. Although the issues regarding *comparability* and *representativeness* are likely to be less serious in the context of subnational PPPs, it is essential to consider them carefully when using CPI data: products collected for CPI may not be comparable or representative across different areas especially in the case of diverse countries in terms of climate, tastes and preferences. Moreover, if the items are not comparable within all the areas the current ICP methodology cannot use CPI price data efficiently. Therefore, two steps are required for constructing sub-national PPPs. Firstly, it is essential to build up the data base and then to test the methodological approaches, such as the various CPD models which may differ from that used in the ICP exercise.

Istat is attempting to build up a database suitable for constructing sub-national PPPs taking into account the results of previous experiments as well as the possibilities coming from the Multipurpose Price Statistics (MPS) project, which is an important Eurostat project regarding consumer price statistics to which the NSOs are contributing. Two of the pillars of the project are the integration of the spatial and temporal dimensions of price statistics and the modernization of data collection by widening the use of electronic devices and scanner and web data (scraped by automatic procedures and robots). This framework can be useful for dealing with the issue of the availability of data that fulfill the requirements of representativeness and comparability that emerge when compiling regional spatial consumer price indices.

Within a data warehouse approach which integrates various sources of data, the compilation of spatial indices of consumer prices should be based on three main blocks of information: use of CPI data (already representative and in some cases also comparable), scanner data, *ad hoc* collected data for certain groups of products by means of electronic devices and a dedicated software that has already been tested in the field.

The final objective of the Istat's project is to construct and regularly update a matrix of prices based on a unique data warehouse containing data obtained from various sources, defined by elementary products and regions where:

- The Italian regions (NUTS 2) are represented by the regional chief towns[9] as shown in Fig. 1.
- A certain amount of elementary quotes will be available in the cells of the matrix. Analyses must be carried out in order to decide whether to use elementary quotes or average prices obtained by averaging individual price quotes.

The end of 2015 and 2016 are essential for laying the foundations, in terms of availability of suitable data and of methodological choices, for producing sub-national PPPs on a regular basis in 2017.

Therefore, with the aim of exploring the available CPI data and understanding whether and to what extent data characteristics affect the selection of the method for computing intra-national PPPs which in turn influences the estimates obtained, Istat has started a research project in collaboration with the Florence and Tuscia Universities.

Within the frame of this research project, on the basis of a preliminary analysis of the basket, 7 BHs (groups of products) have been selected[10] for carrying out experiments with the aim of understanding whether to use elementary quotes or average prices obtained by averaging individual price quotes and exploring the performance of various CPD models based on data characterized by different levels of aggregation.

As can be seen in Table 1, these BHs refer to the most important CPI product group, namely Food and non-alcoholic beverages, which account for 16.5 % of household final monetary consumption expenditure. In particular we considered fresh meat, all fresh fish species, all the different varieties of fresh fruit and fresh vegetables, weighing approximately 30.3 % of the Food and non-alcoholic beverage group and 5.2 % of the entire

[9] In order to assess the statistical foundation of this choice, an analysis should be carried out to determine the presence of spatial autocorrelation among all the capital towns of the provinces belonging to the same region. It is clear that this kind of analysis requires data which may not be necessarily be available for all the products, even if the new sources of data (first of all scanner data) could enhance data availability for some products.

[10] The data set was provided by the Italian Statistical Institute (Istat) and the elementary price quotes were treated in order to respect the statistical confidentiality and the authors worked at Istat in order test the various hypotheses of the CPD approach.

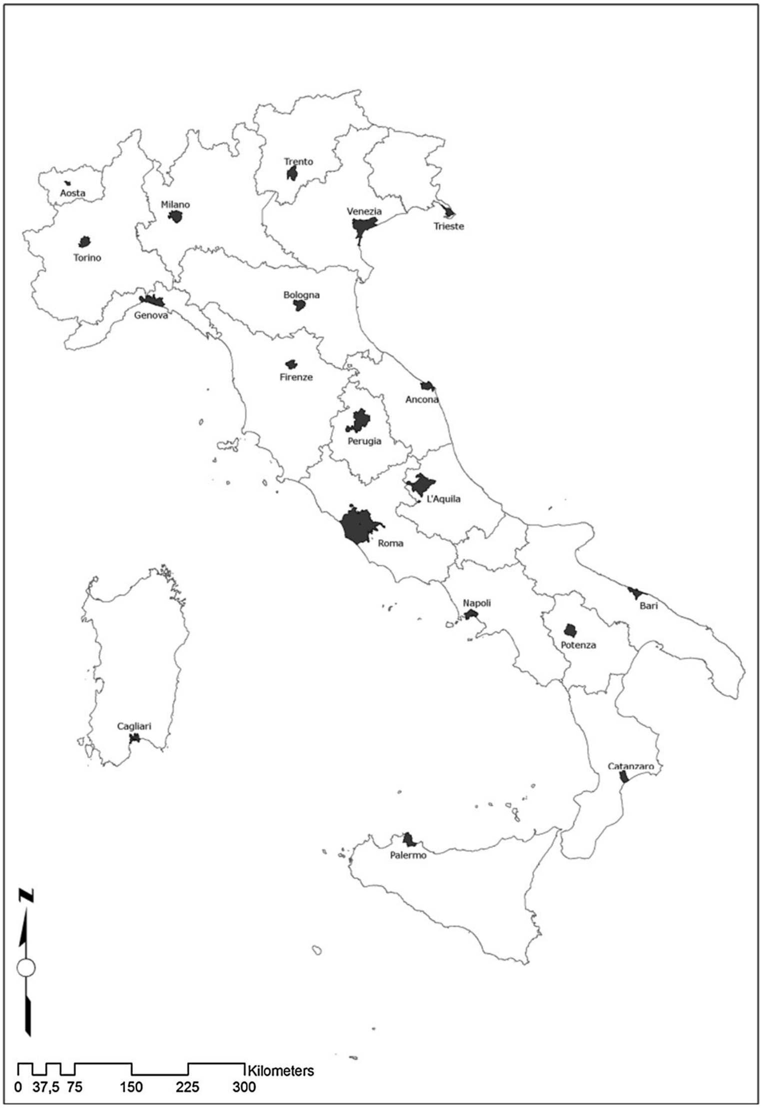

Fig. 1 Italian regional chief towns. *Note* Campobasso, which is the regional chief town of Molise, is not included in the 2014 CPI Survey

basket, as they are comparable by definition and do not require further specifications in addition to those already present in the basket. In this way the spatial price indexes of the products compared should not be biased by dissimilarities among products, thus allowing for the analysis of the various CPD model specifications based on individual price quotes and average prices. Only the group of products "Beef" include different varieties that cannot be considered in the CPD estimation process because they are not coded a priori and the data collectors usually select specific elementary items specifying the variety selected. Therefore in this case we have a "loose" product description and a weaker comparability.

Table 1 List of basic headings and number of monthly price quotes

BH	Description	Num. of price quotes
11.01.13.1	Fresh, chilled or frozen fish and seafood	55,276
11.01.16.1	Fresh or chilled fruit	64,655
11.01.17.1	Fresh or chilled vegetables other than potatoes	63,917
11.01.12.1	Beef and veal	16,884
11.01.12.2	Pork	8424
11.01.12.3	Lamb, mutton and goat	3552
11.01.12.5	Other meats and edible offal	5520
	Total	218,228

Total number of individual price quotes for each basic heading is reported

The dataset used consists in 218,228 monthly price quotes from the 19 regional chief towns[11] considered in CPI computations for 2014 which are: Ancona, Aosta, L'Aquila, Bari, Bologna, Cagliari, Catanzaro, Firenze, Genova, Milano, Napoli, Palermo, Perugia, Potenza, Roma, Torino, Trento, Trieste, Venezia (Fig. 1).

Firstly, the items for which price data were collected in a single region only were eliminated in order to ensure that the incomplete price tableau is connected and therefore the CPD method is feasible. Then, the monthly individual price quotes, uniquely identified by a set of characteristics, were converted into annual average prices for each product for each outlet in each regional chief town by using the arithmetic mean. When computing annual prices for specific product varieties in each BH we considered also the type of outlet in which the product is sold without taking into account the location of the outlet within the regional chief town. As there are multiple quotes for all the observations and considering the loose specification for "Beef" products, the annual data set contains approximately 6000 unique individual price observations, each identified by outlet type, item code, and geographic area. Finally, by using these annual prices, the geometric means for each product in the various regional chief towns were calculated, thus obtaining a synthetic value of the price of the goods sold in the various outlets. Therefore two different price matrices referring to the year 2014 were obtained. The first includes individual price quotes (ungrouped data) while the second considers average prices (grouped data).

With the aim of identifying the most suitable methods for constructing sub-national PPPs using two different kinds of information, i.e. individual price quotes and average prices, firstly various CPD models are estimated and then a synthetic measure of the differences between estimation results is provided. The issue of spatial correlation is subsequently explored by estimating a CPD model which uses individual price quotes and includes spatial autocorrelated errors. Finally, in order to determine the reliability of PPP estimations we constructed confidence intervals for the ranks for each Italian regional chief town in four CPD models: unweighted CPD, weighted CPD, hedonic CPD and spatial CPD both based on individual price quotes.

[11] Campobasso was not included in our analysis because it was the only regional chief town which did not participate in the CPI survey, from which it was eliminated at the beginning of 2014 for inconsistencies in the data collected and unavailability of data collectors.

4.2 Assessing Differences Between CPD Approaches: Average Prices Versus Individual Price Quotes

In order to analyse to what extent the type of the data affect the estimates obtained we compared PPP estimation results obtained from CPD models based on individual price quotes with the results obtained from CPD models based on average prices (across outlets) estimated by using unweighted and weighted least squares. Therefore the basic CPD using average price without weights are compared with the weighted CPD using the weight formula (4) suggested by Hill and Syed (2014) and with the CPD model using individual price quotes.

In order to summarize the differences between the results obtained with the CPD models using average prices, with and without weights, and the individual price quote CPD model, we used a synthetic measure as suggested by Hill and Syed (2014).

By following this approach we compared the various CPD methods in pairs. Firstly we compared the unweighted CPD (CPD_gm) with the weighted CPD (CPD_wgm), then the unweighted CPD (CPD_gm) with the hedonic CPD based on individual price quote model (CPD_ipq), finally the weighted CPD (CPD_wgm) with the CPD_ipq (Table 2). For each BH, the comparison is based on the results of 19 models estimated using each of the other regional chief towns in turn as the base. This means that for comparing each couple of methods 133 regressions are estimated from which the spatial price indexes, expressed

Table 2 Average differences in the basic heading price indexes between pairs of methods

	Unweighted CPD (CPDgm)/ Weighted CPD (CPD_wgm)	Unweighted CPD (CPDgm)//CPD_ipq	Weighted CPD (CPD_wgm)/CPD_ipq
Ancona	3.753	6.792	3.089
Aosta	4.408	6.614	2.576
L'Aquila	8.608	13.666	3.639
Bari	4.195	8.367	2.141
Bologna	4.942	8.543	3.343
Cagliari	3.699	7.659	2.155
Catanzaro	9.263	9.036	2.890
Firenze	4.028	7.728	2.709
Genova	3.940	7.723	2.614
Milano	3.846	16.920	3.863
Napoli	4.161	9.012	3.144
Palermo	3.825	6.567	2.172
Perugia	4.289	8.217	3.354
Potenza	3.900	8.021	3.201
Roma	3.770	8.519	2.463
Torino	3.888	10.825	2.539
Trento	4.596	7.337	3.507
Trieste	3.749	7.717	2.461
Venezia	3.915	7.484	2.553
Average	4.567	8.776	2.864

Average percentage differences between couples of method are reported relative to the base region *b,* are obtained. The differences observed are summarized by averaging over regions and BHs.

Considering, for example, the comparison between the unweighted CPD (CPD_gm) and the weighted CPD (CPD_wgm), the metric used can be expressed as follows:

$$A_r(CPDgm, CPDwgm) = \frac{100}{H(R-1)} \sum_{b \neq r}^{R} \sum_{h=1}^{H} \left[\max \left(PPP_{bh,rh}^{CPDgm} \Big/ PPP_{bh,rh}^{CPDwgm}, PPP_{bh,rh}^{CPDwgm} \Big/ PPP_{bh,rh}^{CPDgm} \right) - 1 \right] \tag{7}$$

where H is the number of basic headings considered which in our case is equal to 7, R denotes the number of regions (regional chief towns) included in the comparison which in our case is equal to 19 and $PPP_{bh,rh}^{CPDgm}$ denotes the spatial price index of region r for basic heading h calculated using the unweighted CPD (CPD_gm) while $PPP_{bh,rh}^{CPDwgm}$ indicates the spatial price index of region r for basic heading h calculated using the weighted CPD (CPD_wgm). Therefore, the metric (7) measures the average percentage difference between the basic heading price indexes generated by the two methods considered for region r.

Table 2 shows that the average difference between the unweighted CPD and the weighted CPD, using weights expressed by (4), is 4.57 %. However, the greatest difference, equal to 8.78 %, can be observed when setting the unweighted CPD against the individual price quote CPD. There is a relatively small difference (2.86 %) between using ungrouped data and the WLS estimation method. These findings are similar to those obtained by Hill and Syed (2014) and suggest that the WLS approach may provide some

Table 3 Differences between the weighted CPD based on average prices (CPD_wgm) and hedonic CPD based on individual price quotes (CPD_ipq)

	Beef	Fish	Fruit	Vegetables	Total
Ancona	6.524	2.532	2.412	0.956	3.089
Aosta	4.139	2.794	2.523	0.864	2.576
L'Aquila	6.710	1.987	4.397	0.781	3.639
Bari	3.516	2.221	2.905	0.793	2.141
Bologna	7.276	2.623	4.755	0.897	3.343
Cagliari	3.564	2.277	2.056	0.841	2.155
Catanzaro	3.502	3.206	5.951	0.785	2.890
Firenze	3.716	2.929	4.788	0.872	2.709
Genova	6.829	1.792	2.689	1.112	2.614
Milano	5.061	7.001	4.478	0.926	3.863
Napoli	3.867	4.996	5.657	0.868	3.144
Palermo	3.505	1.607	2.780	1.248	2.172
Perugia	5.913	4.649	4.202	1.044	3.354
Potenza	3.566	6.832	3.867	1.010	3.201
Roma	3.617	3.476	2.657	1.068	2.463
Torino	4.104	1.563	3.010	0.894	2.539
Trento	4.652	2.406	4.274	1.328	3.507
Trieste	4.316	1.697	2.694	1.584	2.461
Venezia	3.575	2.479	2.468	1.388	2.553
Average	4.629	3.109	3.609	1.014	2.864

Average percentage differences between CPD_wgm and CPD_ipq are reported advantages of the CPD based on individual price quotes which proved to be the best method as illustrated in Sect. 3.2. When the available data are average prices the WLS approach should be used instead of the OLS. This means that if the price matrix to be used for constructing sub-national PPPs is composed by geometric average prices instead of individual price quotes, information available on the number of quotations and the standard deviations should be used in order to reduce the impact of switching from using individual price quotes to geometric average price.

Although the average difference between the results obtained from the WLS and the CPD based on individual price quotes is small (2.86 %), there is high heterogeneity among BHs which may be due to the characteristics of the data. Table 3 shows the percentage difference between spatial price indexes from these two CPD methods for the most important BHs considered in our analysis and for the various regional chief towns. On average "Beef", "Fruit" and "Fish" show higher impacts while only slight differences are observed for "Vegetables". One possible explanation is the influence on prices of the different distribution channels (modern vs traditional), whose effects are accounted for when estimating CPD based on individual price quotes; this influence is relevant for the "Beef", "Fruit" and "Fish" BHs, but not for the "Vegetables" BH. The reason why "Vegetables" do not behave in the same way as "Fruit", even if both BH product groups are distributed through similar outlets (above all open markets), may be due to a better specification of the varieties for the latter compared to the former which increases the "density" of information in the "Fruit" microdata that are able to explain some of the regional differences that are smoothed in the models using average prices.

4.3 Exploring the Issue of Spatial Autocorrelation

In order to carry out an in-depth analysis of the sub-national PPPs compiled for the selected BHs using the CPD models based on individual price quotes, we tested for the presence of autocorrelation among disturbances across regional chief towns for each BH by estimating spatial error models.

Bearing in mind the aim of this paper and the fact that our database is composed of the Italian regional chief towns that do not share land borders (Fig. 1) we constructed the **W** matrix based on the economic distance measure suggested by Fingleton and Le Gallo (2008) which takes into account both the geographical distance between regional chief towns and the size of each town's economy measured by the total employment level[12] in 2014.

In order to test for the presence of spatial autocorrelation we refer to the Moran's I statistic using residuals from a CPD model based on individual price quotes, denoted $\widehat{v}_{knr}$. Since the lowest level of spatial information available in our study is the location of the outlet in one of the 19 regional chief town, following previous studies in the literature on hedonic spatial models we assumed that all observations in the same regional chief town are uniformly distributed, such that their ''mean location'' can be properly approximated by the town centroid (Kim et al. 2003). Therefore, the weights matrix for all individual price quotes are obtained by considering all prices in the same regional chief town as neighbors.

As expected, an autocorrelation among disturbances was observed for all the BHs under analysis even if Moran's I is quite low in some cases. With the aim of illustrating the

[12] It is worth noting that we specified the economic distance by considering other measures of the size of the towns' economy, such as the unemployment rate and the GDP per capita, obtaining similar results.

results obtained, let us focus on two BHs, namely "Beef" and "Fruit", which include products with different kinds of information.[13] Figure 2 shows the Moran scatter plots[14] which is a bivariate scatter plot with values $\widehat{v}_{knr}$ on the horizontal axis and the spatial lag residuals $\mathbf{W}\widehat{v}_{knr}$ on the vertical axis. The Moran's I statistics based on the residuals obtained from the individual price quote CPD for these two BHs is also reported in Fig. 2. The vertical and horizontal lines that cross at the average values of $\widehat{v}_{knr}$ and $\mathbf{W}\widehat{v}_{knr}$ divide the scatter plot into 4 quadrants that correspond to the following 4 different types of spatial associations (clockwise from top right): high–high (HH), low–high (LH), low–low (LL) and high-low (HL).

It is clear that in both cases we found negative spatial autocorrelation among error terms and that the spatial correlation among residuals for "Beef" is high while it is quite low in the case of "Fruit" even if significant. Indeed, Fig. 2 portrays a scatter of points that aligns along an axis sloping from the upper left-hand to the lower right-hand quadrant which refers to a geographic distribution of values in which the neighbours of towns with large values have small values, the neighbors of towns with intermediate values have intermediate values, and the neighbors of towns with small values have large values. Although in empirical analyses negative spatial autocorrelation appears far less frequently than positive spatial autocorrelation, more and more researchers are focusing on this issue in various fields (see for example, Blonigen et al. 2007; Griffith and Arbia 2010; Elhorst and Zigova 2014).

Following Griffith and Arbia (2010) we believe that the presence of negative spatial autocorrelation among disturbances may derive from a geographic competition for market areas among regional chief towns. Moreover, the comparisons of price levels among regional chief towns may be affected by various factors including differences in market structure, in transport costs and in the characteristics of demand, supply and retailers. Our results seem to be in contrast to those found by the two previous empirical studies carried out on this issue by Aten (1996) and Rao (2001b). However it is really hard to compare our results to the positive autocorrelation observed by these Authors since they used different levels of aggregation of price data and considered different geographical areas. Indeed, Aten (1996) found positive values of Moran's I-statistics calculated using normalized price relatives for household expenditures in 64 countries at the aggregate level for 23 headings, ranging from food to expenditures on restaurants and hotels. Rao (2001b) tested the presence of spatial autocorrelation by means of Moran's I-statistics using residuals from ordinary least squares CPD based on the 1985 global comparison results from the ICP for 56 countries with eight aggregated expenditure categories, such as Food, Clothing, Education, etc. It is worth noting that, as shown in Anselin (2002), the parameters of spatial models estimated at an aggregate level do not correspond to those at the individual level and the essential aspect determining the extent of this problem is the intra-unit heterogeneity. Moreover as stressed by Griffith and Arbia (2010) "…negative spatial autocorrelation may go undetected mostly because geographic situations in which it exists also contain counterbalancing positive spatial autocorrelation".

[13] For example, "Ground beef meat, first cut" is one of the products representative of "Beef" in the CPI basket. Within this representative position, data collectors may select different varieties that are to be specified without being coded a priori. In contrast in the case of "Fruit", the variety of the representative position "Apple" presents an "a priori" specification (for example "Red delicious apple") that drives the data collection, is coded and easily manageable in the CPD models with hedonic adjustments.

[14] The MI scatter plots are obtained using STATA and the *splagvar* command.

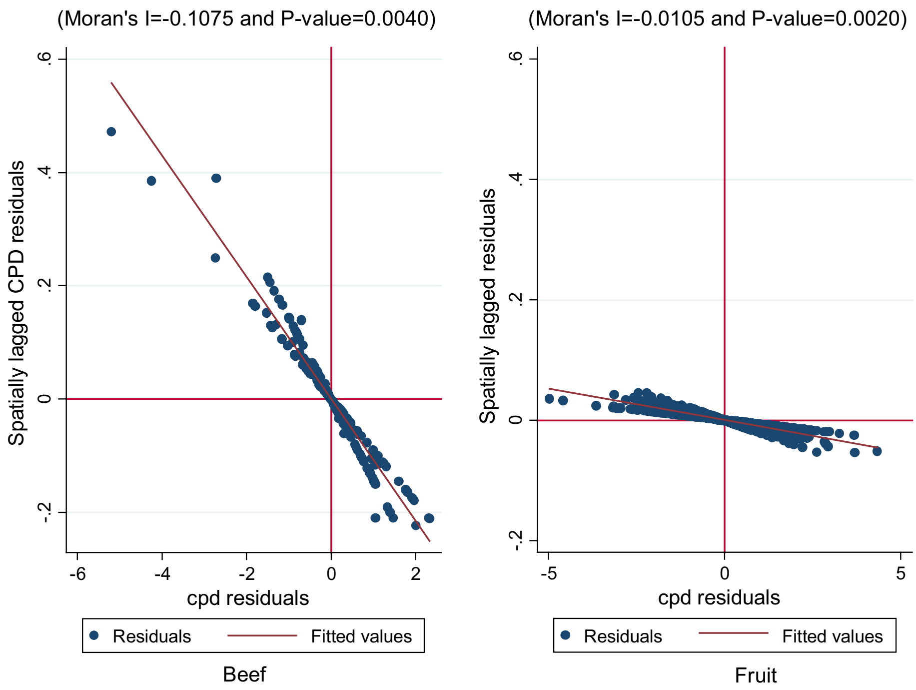

Fig. 2 Moran scatter plot and Moran's I statistics. *Note* The Moran's I *p* value is calculated using a random permutation procedure

Table 4 shows estimation results for the hedonic CPD model and the spatial error CPD model both based on individual price quotes. As can be seen, ignoring spatial correlation among residuals when present will yield biased estimates of standard errors in the traditional ordinary least squares CPD if the proper adjustments are not carried out. This will yield biased t-tests and misleading indications of precision of the resulting sub-national PPPs.

It is interesting to note the contrasting behaviour of the products of these two BHs which could be due to the differences in the description of the products belonging to the BHs that are available to the data collectors in the Italian basket of consumer prices. Indeed, the product description is "loose" for "Beef" and "tight" for "Fruit" and consequently data collectors may select elementary items that are more heterogeneous by following broader parameters for "Beef" than for "Fruit". Therefore it is not possible to take into account the heterogeneity of "Beef" elementary items in the CPD model, due to the lack of specific classifications that may lead to price comparisons of dissimilar items.

Although this evidence is partial, as it only refers to two BHs, it is important since it poses the issue of how to use CPI data for sub-national PPPs aims. If the elementary data (which the elementary quotes refer to) are well specified in terms of quality characteristics, the use of the CPD model with hedonic adjustment does not necessarily need to take spatial autocorrelation into account, even if it improves the estimates. Otherwise if we deal with elementary items with less detailed quality characteristics, it may be necessary to estimate sub-national PPPs by using a CPD spatial error model.

Table 4 Estimation results for CPD models based on individual price quotes: hedonic versus spatial error model

	"Beef"						"Fruit"					
	CPD_ipq		Sign.	CPD_spatial		Sign.	CPD_ipq		Sign.	CPD_spatial		Sign.
	Coef.	SE		Coef.	SE		Coef.	SE		Coef.	SE	
Ancona	0.071	0.039	**	0.070	0.017	***	0.204	0.023	***	0.204	0.008	***
Aosta	0.059	0.049		0.059	0.013	***	0.224	0.032	***	0.225	0.007	***
L'Aquila	0.007	0.055		0.006	0.017		−0.082	0.023	***	−0.082	0.005	***
Bari	−0.115	0.049	**	−0.115	0.016	***	−0.114	0.025	***	−0.114	0.007	***
Bologna	−0.001	0.079		−0.001	0.013		0.237	0.027	***	0.237	0.006	***
Cagliari	−0.149	0.048	***	−0.149	0.013	***	0.041	0.024	*	0.042	0.007	***
Catanzaro	−0.239	0.058	***	−0.239	0.013	***	−0.205	0.023	***	-0.205	0.007	***
Firenze	−0.131	0.054	**	−0.131	0.013	***	0.085	0.025	***	0.085	0.006	***
Genova	0.095	0.038	**	0.095	0.017	***	0.129	0.022	***	0.129	0.006	***
Milano	−0.137	0.107		−0.138	0.017	***	0.409	0.027	***	0.409	0.004	***
Napoli	−0.208	0.057	***	−0.208	0.013	***	−0.041	0.024	*	−0.041	0.007	***
Palermo	−0.138	0.060	**	−0.138	0.013	***	0.020	0.022		0.019	0.006	***
Perugia	0.016	0.052		0.016	0.016		0.105	0.028	***	0.105	0.008	***
Potenza	−0.249	0.042	***	−0.249	0.013	***	−0.007	0.029		−0.008	0.009	
Torino	0.121	0.048	**	0.120	0.015	***	0.023	0.022		0.023	0.005	***
Trento	0.053	0.092		0.053	0.017	**	0.224	0.025	***	0.225	0.006	***
Trieste	0.025	0.043		0.026	0.015	*	0.195	0.023	***	0.195	0.006	***
Venezia	0.044	0.030		0.044	0.013	***	0.209	0.021	***	0.209	0.007	***
constant	−0.242	0.056	***	−0.288	0.039		0.826	0.031		0.826	0.028	
outlet type1	0.044	0.025	*	−0.051	0.029		0.040	0.017		0.041	0.014	
outlet type3	2.836	0.034	***	2.889	0.029		0.141	0.009		0.140	0.009	
lambda				−3.138	0.448	***				−7.596	1.611	***
R^2 and PsuedoR^2	0.829			0.829			0.777			0.776		

Table 4 continued

	"Beef"						"Fruit"					
	CPD_ipq		Sign.	CPD_spatial		Sign.	CPD_ipq		Sign.	CPD_spatial		Sign.
	Coef.	SE		Coef.	SE		Coef.	SE		Coef.	SE	
Root MSE	0.1811						0.1764					
Obs.	1673			1673			177					

CPD_ipq (hedonic CPD based on individual price quotes) CPD_spatial (spatial error hedonic CPD model based on individual price quotes)

* 10 %

** 5 %

*** 1 %

Table 5 Sub-national PPPs for "Fruit" and "Beef"-Rome = 100 using various CPD specification

	"Beef"				"Fruit"			
	CPD_gm	CPD_wgm	CPD_ipq	CPD_spatial	CPD_gm	CPD_wgm	CPD_ipq	CPD_spatial
Ancona	98.3	98.4	107.3**	107.3***	123.0***	122.2***	122.6***	122.7***
Aosta	106.1	106.8	106.1	106.1***	125.9***	124***	125.2***	125.3***
L'Aquila	92.2**	92.1**	100.7	100.6	92.5***	91.7***	92.1***	92.1***
Bari	89.2***	88.8***	89.1**	89.1***	90.4***	89.4***	89.2***	89.2***
Bologna	99.9	103.8	99.9	99.9	127.1***	126.3***	126.8***	126.8***
Cagliari	86.1**	86.0**	86.1***	86.1***	104.5*	103.8*	104.2*	104.3***
Catanzaro	78.8***	78.4***	78.8***	78.8***	81.9***	81.1***	81.5***	81.5***
Firenze	87.8**	87.9**	87.8***	87.8***	109.8***	109.3***	108.9***	108.9***
Genova	100.7	100.5	110.0**	109.9***	114.5***	113.4***	113.7***	113.8***
Milano	79.8***	81.6***	87.2	87.1***	149.7***	147.9***	150.6***	150.6***
Napoli	81.2***	81.5***	81.2***	81.2***	96.4	95.8**	96*	96***
Palermo	87.1**	86.5**	87.1**	87.1**	101.3	101	102	101.9***
Perugia	95	93.9	101.6	101.7	110.8***	108.4**	111.1***	111.1***
Potenza	77.9***	77.8***	77.9***	77.9***	97.9	96.8	99.3	99.2
Torino	109.6	108.1	112.9**	112.7***	103.1	102.7	102.4	102.3***
Trento	96.6	99.5	105.4	105.4***	124.9***	122.1***	125.2***	125.2***
Trieste	97.7	97.5	102.5	102.6*	121.9***	120.2***	121.5***	121.6***
Venezia	104.5	104.4	104.5	104.5***	123.8***	122.6***	123.2***	123.2***

CPD_gm (unweighted CPD model), CPD_wgm (weighted CPD) CPD_ipq (hedonic CPD based on individual price quotes) CPD_spatial (spatial error hedonic CPD model based on individual price quotes)

* 10 %

** 5 %

*** 1 %

In order to carry out an in-depth analysis of the statistical performance of CPD models (and in particular of using CPD with spatial error models in order to improve the estimates), Table 5 presents PPP results for the "Fruit" and "Beef" BHs. This table compares the sub-national PPPs estimated from the weighted and unweighted CPD methods based on average prices with the models based on individual price quotes, that is the hedonic CPD and hedonic spatial CPD which takes into account the correlation across CPD disturbances.

Generally speaking, the use of CPD_spatial increases the number of regional chief towns for which the coefficients estimated are significant compared to the hedonic CPD, thus improving the quality of the estimations obtained.

The main evidences from this table are the followings:

- As expected, the adoption of CPD_ipq and the spatial error CPD model produces point estimates of sub-national PPPs that are almost the same but with different standard errors which are lower when we account for spatial correlation among residuals.
- The position of each regional chief town with respect to Rome changes in a significant number of cases if we compare CPD based on individual price quotations and the other CPD models only in the case of "Beef". In fact sub-national PPPs for "Beef", compiled with CPD_ipq, are over 100 (Rome) for half of the regional chief towns considered, while in the case of CPD models applied to average prices, only five regional chief towns are over 100. Moreover, in the case of CPDipq models, most of the towns located in Centre – North of Italy (with the exception of Firenze and Milano) are "more expensive" than Rome concerning "Beef" consumption, which is a result more coherent with our expectations.

Therefore, by using the individual price quote CPD and the spatial CPD, we can improve the efficiency of PPP estimations also when the specifications of the product are not tight and the detailed information on the characteristics of the elementary items are not taken into account since they are not codified (as in the case of "Beef").

4.4 Evaluating the Reliability of the Sub-national PPPs

Besides the analysis of the various PPP estimates, it is important to verify their accuracy in order to determine if the purchasing power of a country's currency unit in a region can be significantly distinguished from the others in that country (i.e. the purchasing power of euro across the various Italian regions represented by the regional chief towns). This issue is particularly important considering the possibility of using PPPs for converting household consumption expenditure or income levels. As already underlined, the 95 % confidence intervals should be used to communicate how precisely each BH sub-national PPPs is estimated. However, the main use of confidence intervals is to judge whether each regional chief town is significantly more or less expensive than the base regional chief town, which in our case is Rome. Therefore, with the aim of making multiple comparisons we propose to use a simulation method, based on ranks, suggested by Marshall and Spiegelhalter (1998) and previously used in other fields (Leckie and Goldstein 2011). In this way the statistical uncertainty in regional chief towns' ranks is displayed by plotting these ranks together with their associated 95 % confidence intervals, where the lower and upper bounds for these 95 % confidence intervals are given by the 2.5th and 97.5th percentiles from their respective distributions of M simulated ranks.

Therefore, we compared the estimated regional PPPs for the "Beef" and "Fruit" BHs obtained from the four CPD models graphically, by plotting the regional chief towns in rank order where the rankings are derived as the median ranking (based on the estimated PPPs) from 1000 bootstrap replications of the various CPD models. For each of the four PPPs, the regional chief town with the highest value of PPP is denoted as having the top ranked (1) spatial price index. On the basis of the 1000 bootstrap replications we derived 95 % confidence intervals for the ranks for each regional chief town in the various CPD models. In particular, we carried out 1000 iterations and for each regional chief town ordered the 1000 simulated ranks. We then identified the median rank (that is, position 500) and the 95 % confidence interval (the distance between those ranks in position 25 and 975). Confidence intervals that do not overlap, enable us to consider the purchasing power of a country's currency unit in a regional chief town significantly different from the others, while a considerable overlapping in the confidence intervals suggests that there is a low level of accuracy in the estimates, thus emphasizing a high level of uncertainty in the rankings.

The results shown in Figs. 3 and 4 provide another important perspective on the results obtained in the previous subsections concerning the extent to which the characteristics of the data influence the performance of the CPD methods. In particular, Fig. 4 shows a considerable uncertainty in the rank positions of the regional chief towns. Although a small number of regional chief town PPPs at the extremes of the ranking have small confidence intervals, most of the other regional chief towns have very wide overlapping confidence intervals, thus indicating considerable uncertainty in the rankings derived.

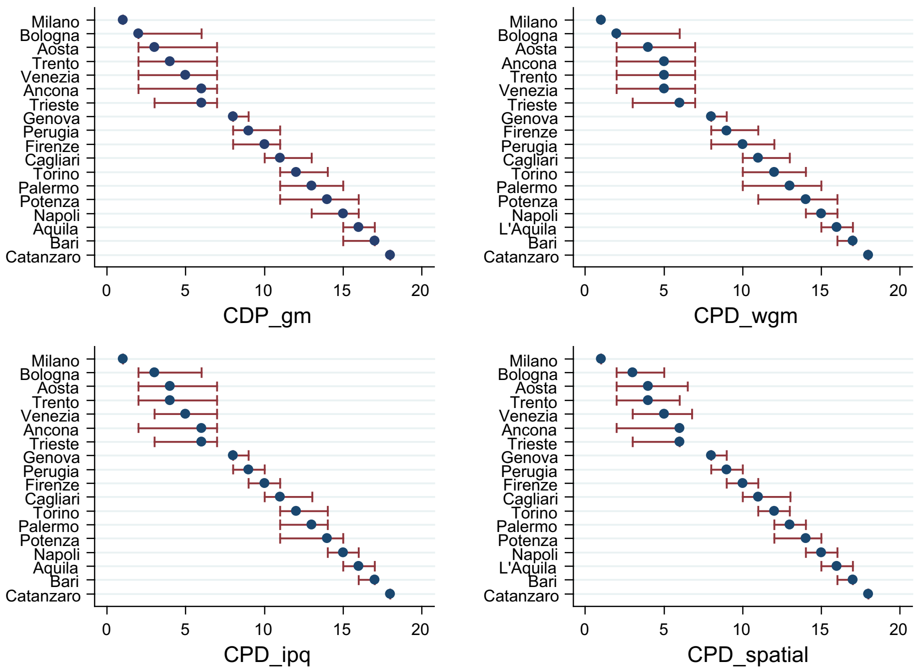

Fig. 3 PPP ranks with simulated 95 % confidence intervals: comparison between four CPD methods for the BH "Fruit". *Legend* PPP ranks with the simulated 95 confidence intervals are reported on the x-axis while regional chief towns are reported on the y-axis. CPD_gm (unweighted CPD model), CPD_wgm (weighted CPD), CPD_ipq (hedonic CPD based on individual price quotes), CPD_spatial (spatial error hedonic CPD model based on individual price quotes)

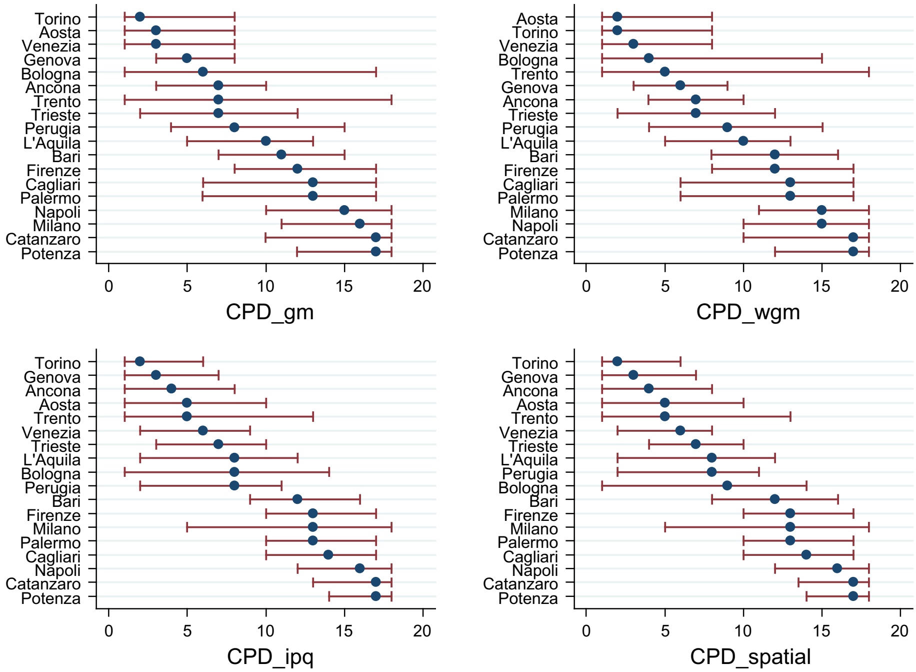

Fig. 4 PPP ranks with simulated 95 % confidence intervals: comparison between four CPD methods for the BH "Beef". *Legend* PPP ranks with the simulated 95 confidence intervals are reported on the x-axis while regional chief towns are reported on the y-axis. CPD_gm (unweighted CPD model), CPD_wgm (weighted CPD), CPD_ipq (hedonic CPD based on individual price quotes), CPD_spatial (spatial error hedonic CPD model based on individual price quotes)

Therefore, for "Beef", for which little information on elementary items is available, even if the use of CPD_ipq reduces the degree of overlapping of the various regional chief towns, the accuracy of the estimates remains low. There is a higher level of accuracy for "Fruit" rather than "Beef" which is further increased by using the CPD_ipq and in particular the spatial CPD.

5 Concluding Remarks

Spatial price indexes measuring differences in price levels across regions within a country are essential for comparing real income, standards of living and consumer expenditure patterns. The relevance of compiling sub-national PPPs for a country like Italy is shared by several players of the economic and social debate. This relevance is due to the socio-economic heterogeneity across the various geographical areas which is confirmed by the results released by Istat in 2008 and 2010 as well as by the analyses carried out in the study. The differences of the levels of consumer prices across the various geographical areas do not seem to be negligible and even high in some cases and therefore it is essential to measure them with the aim of helping policy makers to define economic strategies and make decisions.

The main purpose of this study was to find a solution to the methodological and data availability issues with the aim of guaranteeing a compilation and dissemination of sub-national PPPs for Italy on a regular basis.

One of the most important aspects of the Istat's project, carried out within the European MPS project, is to use CPI data appropriately. On this topic much attention has been paid by estimating various CPD models and using two different kinds of products from the Italian CPI basket:

(a) Products characterized by a tight definition in the Italian basket at national level, thus allowing for the comparison of the data collected across the various geographical areas without a specific analysis and cleaning of the data, which is the case of fresh fruit, fresh vegetables and fish products.
(b) Products characterized by a broader definition in the Italian basket such as all meat products.

In particular tests and analyses were carried out on seven BHs using data from the Italian CPI data base in order to compare sub-national PPP estimations obtained by using different CPD models. Although preliminary and not definitive, the results obtained proved to be interesting thus confirming that methods and data are reciprocally influenced.

From the results illustrated in the previous sections, it is worthwhile drawing some conclusions in order to highlight the following research steps to be taken on this topic:

(i) In general the application of CPD models on ungrouped data seems to be preferable compared to the basic approach on average data;
(ii) Even if the results in terms of PPP values are similar, it is essential to take spatial autocorrelation into account whenever present in order to improve the reliability of the PPP estimates by reducing the standard errors;
(iii) However, a detailed specification and coding of the qualitative characteristics of the elementary items is required in order to estimate CPD models on individual price quotations, as they allow for hedonic adjustments. This adds to the research required on the issue of the availability of CPI data suitable for spatial comparison as it necessary to improve the basic information concerning brand and varieties as well as the retail trade channels;
(iv) The reliability of the PPP estimates obtained by constructing the bootstrap confidence intervals of ranks proved to be useful for evaluating whether the comparison between pairs of regional PPPs are valid or not.

Our findings, which mainly concern the methods for improving the use of CPI data for carrying out spatial comparisons, may be useful for further developing the research studies implemented by Istat in cooperation with the University of Florence and the University of Tuscia.

To this aim it is essential to acquire and use scanner data as well as to collect price quotes in the field by means of software initially developed for inter country PPPs.

Therefore, the tests and the experiments should be continued, both for various kinds of BHs and data deriving from different sources within the MPS logic, by focusing on the analysis of elementary item characteristics and spatial autocorrelation issues.

The final objective is to produce and disseminate regional PPPs by 2017, thus filling a gap in the statistical information concerning consumer prices in Italy.

Acknowledgments The authors wish to thank Stefania Occhiobello (Istat) for her support and assistance.

References

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27*(3), 247–267.

Anselin, L. (2006). Spatial econometrics. In T. C. Mills & K. D. Patterson (Eds.), *Palgrave handbook of econometrics*. London: Palgrave Macmillan.

Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science, 92*(1), 3–17.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237–289). New York: Marcel Dekker.

Aten, B. H. (1996). E*vidence of spatial autocorrelation in international prices, review of income and wealth*, June, 42, 2, pp. 149–163.

Aten, B. H. (1999). Cities in Brazil: An interarea price comparison. In *International and interarea comparisons of income, output, and prices* (pp. 211–229). Chicago: University of Chicago Press.

Aten, B. H. (2005). *Report on interarea price levels, orking paper no. 2005-11*. Bureau of Economic Analysis, May.

Aten, B. H. (2006). Interarea price levels: An experimental methodology. *Monthly Labor Review, 129*(9), 47–61.

Aten, H. B., Figueroa, E. B., & Martin, T. M. (2014). *Regional price parities for states and metropolitan ares, 2006–2010*. Washington, US: Survey of Current Business, Bureau of Economic Analisis.

Aten, B. H., & Reinsdorf, M. (2010). *Comparing the consistency of price parities for regions on the US in an economic approach framework, paper presented at the 31st general conference of IARIW*. St. Gallen Switzerland.

Baumont, C. (2004) *Spatial effects in housing price models: Do housing prices capitalize urban development policies in the agglomeration of Dijon (1999)*? Working Paper, Université de Bourgogne.

Biggeri, L., De Carli, R., & 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, paper presented at the Joint UNECE/ILO meeting on Consumer Price Indices*, May 8–9, Geneva.

Biggeri, L., & Laureti, T. (2010). Are integration and comparison between CPIs and PPPs feasible? In L. Biggeri & G. Ferrari (Eds.), *Price indices in time and space*. Springer, 2009, forthcoming.

Blonigen, B. A., Davies, R. B., Waddell, G. R., & Naughton, H. T. (2007). FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review, 51*(5), 1303–1325.

Brandt, L., & Holz, C. A. (2006). Spatial price differences in China: Estimate and implications. *Economic Development and Cultural Change, 55*(1), 43–86.

Case, A., Rosen, H. S., & Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the States. *Journal of Public Economics, 52*, 285–307.

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.

Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of Regional Science, 52*(2), 210–239.

Diewert, W. E. (2004). *On the stochastic approach to linking regions in the ICP*. Paper prepared for Development Economics Data Group (DECDG), World Bank, Washington, DC.

Diewert, W. E. (2005). *Weighted country product dummy variable regressions and index number formulae, review of income and wealth*, 51, 4, December, pp. 561–570.

Diewert, E. (2010). *New methodological developments for the international comparison program Review of Income and Wealth*, 56, Special Issue 1, S11–S31.

Dikhanov, Y., Palanyandy, C., & Capilit, E. (2011). *Subnational 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.

Elhorst, J. P. (2014). *Spatial econometrics from cross-sectional data to spatial panels*. New York: Springer.

Elhorst, J. P., & Zigova, K. (2014). Competition in research activity among economic departments: Evidence by negative spatial autocorrelation. *Geographical Analysis, 46*(2), 104–125.

Fenwick, D., & O'Donoghue, J. (2003). Developing estimates of relative regional consumer price levels. *Economic Trends, 599*, 72–83.

Fingleton, B. (2001). Equilibrium and economic growth: Spatial econometric models and simulations. *Journal of Regional Science, 41*(1), 117–147.

Fingleton, B., & Le Gallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science, 87*(3), 319–339.

Griffith, D. A., & Arbia, G. (2010). Detecting negative spatial autocorrelation in georeferenced random variables. *International Journal of Geographical Information Science, 24*(3), 417–437.

Hill, R. J., & Hill, T. P. (2009). Recent developments in the international comparison of prices and real output. *Macroeconomic Dynamics, 13*(Supplement 2), 194–217.

Hill, R. J., & Syed, I. A. (2010). *The international comparisons program and the country-product-dummy method: An application to the Asia-Pacific Region*, paper received by the authors.

Hill, R. J., & Syed, I. A. (2014). Improving international comparisons of prices at basic heading level: An application to the Asia-Pacific Region. *Review of Income and Wealth,*. doi:10.1111/roiw.12116.

ICP-TAG. (2010). *Papers on sub-national PPPs presented at the 2° meeting held on February 2010*. http://siteresources.worldbank.org/ICPEXT/Resources/ICP_2011.html, Meetings, ICP Technical Advisory Group.

Istat. (2008). *Le differenze nel livello dei prezzi tra i capoluoghi delle regioni italiane per alcune tipologie di beni Anno 2006*, Roma.

Istat. (2010). *La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane, Anno 2009*, Roma.

Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of environmental economics and management, 45*(1), 24–39.

Kokoski, M. (1991). *New research on interarea consumer price differences, monthly labor review*, Vol. 114, p. 31.

Kokoski, M., Moulton, B., & Zieschang, K. (1999). Interarea price comparisons for heterogenous goods and several levels of commodity aggregation. In A. Heston & R. Lipsey (Eds.), *International and interarea comparisons of income, output and prices* (pp. 123–166). Chicago: University of Chicago Press.

Leckie, G., & Goldstein, H. (2011). Understanding uncertainty in school league tables. *Fiscal Studies, 32*(2), 207–224.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. New York: Chapman & Hall.

Li, X., Zhang, L., & Du, Y. (2005). Study on the method of regional purchasing power parity in China. *China-USA Business Review, 4*(7), 41–48.

Majumder, A., Ray, R., & Sinha, K. (2013a). *Spatial comparisons of prices and expenditure in a heterogeneous country: Methodology with application ti India*. Manuscript provided by authors.

Majumder, A., Ray, R., & Sinha, K. (2013b). *Estimating purchasing power parities from household expenditure data using complete demand systems with application to living standards comparisons: India and Vietnam*. Review of income and wealth.

Majumder, A., Ray, R., & Sinha, K. (2014). *A unified framework for the estimation of intra and inter country food purchasing power parities with application to cross country comparisons of food expenditure: India, Indonesia and Vietnam*. Manuscript provided by the authors.

Marshall, E. C., & Spiegelhalter, D. J. (1998). Reliability of league tables in in vitro fertilisation clinics: Retrospective analysis of live birth rates. *British Medical Journal, 316*, 1701–1704.

Melser, D., & Hill, R. (2005). *Developing a methodology for constructing spatial cost of living indexes*. Sidney, Australia: Economic Department, University of New South Wales.

Mishra, A., & Ray, R. (2014). Spatial variation in prices and expenditure inequalities in Australia. *Economic Record, 90*(289), 137–159.

ONS. (2011). *UK relative regional consumer price levels for goods and services for 2010*, UK, pp. 1–15.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association, 70*(349), 120–126.

Rao, D. S. P. (2001a). *Integration of CPI and PPP: Methodological issues, feasibility and recommendations*. Washington, DC: Joint World Bank-OECD Seminar on Purchasing Power Parities.

Rao, D. S. P. (2001b). *Weighted EKS and Generalised CPD methods for aggregation at basic heading level and above basic heading level Joint World Bank—OECD seminar on purchasing power parities*. Washington, DC: Recent Advances in Methods and Applications.

Rao, D. S. P. (2002). *On the equivalence of weighted country-product-dummy (CPD) method and the rao system for multilateral price comparisons*. Armidale, Australia: School of Economics, University of New England.

Rao, D. S. P. (2004). *The country-product-dummy method: A stochastic approach to the computation of purchasing power parities in the ICP*. Paper presented at the SSHRC conference on index numbers and productivity measurement, Vancouver, June 30–July 3, 2004.

Rao, D. S. P. (2005). On the equivalence of weighted country-product-dummy (CPD) method and the Rao-system for multilateral price comparisons. *Review of Income and Wealth, 51*(4), 571–580.

Rao, D. S. P. (2013). Computation of basic heading purchasing power parities (PPPs) for comparisons within and between regions, chapter 4. In World Bank (Ed.), *Measuring the real size of the world economy* (pp. 93–119). World Bank, Washington DC.

Sherwood, M. K. (1975). *Family budgets and geographic differences in price levels*. Montly Labor Review, April 1075, pp. 8–15.

Silver, M. (2009). *The hedonic country product dummy method and quality adjustments for purchasing power parity calculations*. IMF working paper WP/09/271.

Silver, M., & Heravi, S. (2005). Purchasing power parity measurement and bias from loose item specifications in matched samples: An analytical model and empirical study. *Journal of Official Statistics, 21*(3), 463.

Summers, R. (1973). International comparisons with incomplete data. *Review of Income and Wealth, 29*(1), 1–16.

Waschka, A., Milne, W., Khoo, J., Quirey, T., & 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., Feniwick, D., & Smith, K. (2005). *Relative regional consumer price levels in 2004*. Office for National Statistics, Economic Trends, p. 615.