Weighting elementary prices in Consumer Price Index (CPI) construction using spatial autocorrelation

Beatriz Larraz[(1)], Jose Manuel Pavía[(2)] and Guido Ferrari[(3)]

[(1)]University of Castilla-La Mancha, Spain

[(2)]University of Valencia, Spain

[(3)]University of Florence, Italy & Renmin University of China, PR China

Abstract

The importance of inflation (deflation) in current economic systems calls for it to be measured as accurately as possible. The Consumer Price Index (CPI) has been used to measure inflation. Both in the United States and the European Union, the CPI is a Laspeyres-type index which suffers, *inter alia*, from the very well-known drawbacks of having a fixed-base weighting system and therefore failing to capture the substitution effect or quality modifications. They both however suffer from an original sin that has been systematically overlooked. When constructing CPIs, official institutions have ignored the spatial dimension of elementary prices. Prices are collected on spatial or geographical locations, i.e, they are georeferenced data. Therefore, ignoring this fact implies consider prices as spatially independent, when in fact they are not, as is the case with many other socioeconomic variables. As a result, the first step of the procedure to estimate inflation is biased. In order to solve this problem, this paper proposes a new approach to elaborate price relatives based on kriging methodology. The idea is to retain the spatial coordinates where prices are surveyed, weighting the basic price data by taking into account the spatial correlation they display. The weighted geometric and arithmetic averages proposed generalize and improve the simple geometric and arithmetic averages currently in use.

Keywords: Price indexes; geostatistics; kriging; spatial analysis; elementary indexes.

Acknowledgments: This research has been supported by the Spanish MICINN through the project CSO2009-11246. The authors wish to thank Tony Little for revising the English of the manuscript.

1. Introduction

Practically all citizens in developed countries are affected by changes in the Consumer Price Index (CPI). In general, the CPI is mainly used as an economic indicator, as a means of adjusting income payments, as a way of preventing inflation-induced tax changes and as a deflation instrument to express other economic series at constant prices. In particular, in the United States (US) it is used to adjust billions of dollars in Federal payments and programs including annual adjustments to Social Security payments and Federal income tax brackets. Meanwhile, in the European Union (EU) the Harmonized Indices of Consumer Prices (HICP) provide the official measure of consumer price inflation in the euro-zone. Monetary policy and the assessment of inflation convergence in the euro zone are both affected by these measures.

The importance of CPIs in current economies therefore requires as precise and up-to-date measurement as possible of inflation and calls for the introduction of as many technical improvements as can be reasonably achieved. In fact, CPIs have undergone a large number of changes and avatars over time in order to perfect it. Changes have been made right from the outset to enhance the surveys and methods and to update the sample and weights in an attempt to provide a better way of reflecting consumer substitution behaviour.

Historically, the first attempts to measure inflation date back to the 19th century, although it was not until after World War I when the Bureau of Labor Statistics (BLS) began to publish a US national CPI on a regular basis. Later, in the 20th century, Eurostat set up a homogenising methodology across European countries (Reinsdorf and Triplett, 2009). At present, a Laspeyres price index is used in both the US and the EU to aggregate elementary price indexes using expenditure information on each of the elementary aggregates as weights. As regards calculating the basic index [the so-called elementary aggregate index in European terminology; European Communities, 2001, p. 209], which shows the average price change in each item in each geographical area (BLS, 2008), the geometric mean is, as a rule, preferred to the arithmetic mean. In addition to the superiority that geometric mean displays when dealing with percentages and rates of change and the well-known sensitivity of arithmetic averages to anomalous data (e.g., Fenwick, 1998), it seems that geometric means also eliminate the so-called functional form bias problem (BLS, 1998), which appears when the arithmetic mean is employed. In fact, according to Boskin et al. (1996), the use of the arithmetic mean implicitly implies that the quantity purchased of an item does not change when prices change. Hence, following the advice from what is commonly known as the Boskin Commission (Advisory Commission to Study the Consumer Price Index; see Boskin et al. 1996), the European Commission recommended in September 1996, coinciding with the implementation of HICP measures, the use of ratios of geometric or arithmetic means of elementary prices (European Communities, 2001, p. 217). Furthermore, in January 1999 the BLS introduced the geometric mean formula to average price change within most of its categories (covering more than 60 percent of total consumer spending).

Anyway, despite the continuous improvements, a weak point still persists in the base of CPI elaboration: item elementary prices are handled as geographically independent entities, omitting the fact that, according to Tobler's first law of geography (Tobler, 1970), prices (of one same item) drawn in closer outlets are more alike than prices collected further apart. Fortunately, this shortcoming could be easily solved by merely slightly modifying the way data are collected and by applying the appropriate methodologies. On the one hand, it would be sufficient for the spatial coordinates of the outlets (or stores) where prices come from to be saved along with prices; while, on the other hand, employing the recorded coordinates, kriging techniques could be used to geographically weight prices at elementary level.

Spatial Statistics deals with the importance of space in statistics and comprises a series of techniques (including geostatistics) that take into account the topological, geometric or geographic characteristics of the entities under study. Kriging stands out among these techniques. Developed in the 1950's by Daniel Gerhardus Krige (a South African mining engineer) to solve ore reserve estimation problems, kriging was expanded in the sixties by Georges Matheron (1962) from the Commissariat de l'Energie Atomique in France. Nowadays, apart from the well-established earth sciences and environmental applications where it is widespread (e.g., Lasslet et al., 1987; Englund, 1993; Montero, Chasco and Larraz, 2010), kriging is being used a great deal in almost all social sciences, from economics (e.g., Nagle, 2010) to political science (Pavia, Larraz and Montero, 2008), including medicine (e.g., Goovaerts, 2006), real estate (Chica-Olmo, 2007; Montero, Larraz and Paez, 2009) or epidemiology (e.g., Lawson, 2001). In contrast with the classic random sampling approach, where data are commonly assumed to be spatially independent, kriging incorporates the spatial correlation of sample data in the inference process. This topic comes under a spatial stochastic process framework, in which each phenomenon is considered as a set of random variables that can be spatially dependent.

Under this framework, the general theory of index numbers can be redirected to geostatistics and the information provided by the geographical location where outlets are placed easily incorporated. This article suggests moving forward in the process of improving inflation measurement by showing how to include the structure of the spatial correlation of prices into the number index theory. In particular, we propose weighting prices at elementary level taking advantage of the information provided by the spatial location where prices are collected. The main aim of the paper is to offer better alternatives to estimate the mean prices of elementary aggregates (in European terminology) and basic indexes (in American terminology). The method leads to a better weighted estimator of the geometric and arithmetic means than the way they are currently used in the US and the EU when computing the first simple index of the elemental aggregate in CPI methodology.

The remainder of the paper is structured as follows. Section 2 briefly describes the parts of US and EU CPI methodologies that are relevant to this paper. Section 3 presents the technical details of kriging in a CPI environment and the tools we will use to estimate the structure of the spatial correlation that prices display. Section 4 weights basic prices by way of a kriging approach and presents the specific formulae. Section 5 summarises and concludes.

2. Consumer Price Index Methodologies: European Union and United States

In all countries, CPIs are elaborated following a multi-stage process, the first stage of which involves estimating elementary price indexes for elementary expenditure aggregates. Subsequently, in successive stages of aggregation, these elementary price indexes are combined to obtain higher-level indexes using information on the expenditures on each of the elementary aggregates as weights (Turvey, 2004). Focusing on the basic concepts, procedures and formulae used to construct the so-called price relatives for each area-item combination, the first step of European and American methodologies are succinctly described in this section. The aim is to become familiar with the terminologies involved at this first stage, at elementary level, when the prices of each item are averaged. Discussion on the way the basic index of each item is aggregated to other basic indexes in the second and further stages is beyond the scope of this paper. More detailed information on this topic is available in BLS (2008) and European Communities (2001).

2.1. European Union

Since March 1997, Eurostat publishes, on a common reference base 1996=100, the comparable indexes of consumer prices (HICP) produced by each EU member state, as required by European Communities (1995). The aim is to measure inflation by means of CPIs on an equivalent basis, removing differences in national definitions. The harmonization project is based on detailed regulations, each one establishing specific implementing measures determining the computation of the HICP. In particular, the first Commission Regulation (European Communities, 1996) deals with the initial coverage, recently significant goods and services, elementary aggregates (the object of this paper) and minimum standards required for quality adjustments, sampling and prices. The HICP covers those goods and services which are included in household final monetary consumption expenditure classified according to the four digit categories and subcategories of the Classification of Individual Consumption by Purpose (United Nations, 2000) adapted to the needs of HICPs (COICOP/HICP).

According to European Communities (1996, Article 7), HICP regulations establish the formulae to be used for computing the price indexes for elementary aggregates, which refers to the "expenditure or consumption covered by the most detailed level of stratification of the HICP and within which reliable expenditure information is not available for weighting purposes" (European Communities, 1996, Article 2, Definition j), the elementary aggregate index being the "price index for an elementary aggregate comprising only price data" (European Communities, 1996, Article 2, Definition i). In particular, Annex II of the regulation document (European Communities, 1996) states that either the ratio of arithmetic mean prices or the ratio of geometric mean prices shall be used and recommends that the arithmetic mean of price relatives should not normally be used. Table 1 reports such definitions. Note that none of the methods used to compile elementary aggregates are weighted. In fact, not even estimated quantities of the items purchased are used as weights. This implicitly implies the assumption of giving the same importance to the quotations of one same item in different outlets, despite this not being the case, as expenditure is not the same in each store.

Table 1: Methods for compiling elementary aggregates in the EU

Formulae to be used		Formula not to be used
Ratio of geometric mean prices	Ratio of arithmetic mean prices	Arithmetic mean of price relatives
$\frac{\left[\prod p^t\right]^{1/n}}{\left[\prod p^b\right]^{1/n}}$	$\frac{\frac{1}{n}\sum p^t}{\frac{1}{n}\sum p^b}$	$\frac{1}{n}\sum \frac{p^t}{p^b}$

p^t is the current price

p^b is the reference price

n is the number of prices in the elementary aggregate

Source: Own elaboration following European Communities (1996).

Following the stepwise approach, firstly the elementary aggregates (which are defined on a regional level for most EU member states) are calculated. Secondly, these regional prices are compiled to the lowest level of aggregation within which reliable expenditure information is available for weighting purposes. Finally, these indexes are combined with expenditure groups at country level.

2.2. United States

Chapter 17 on *Handbook of Methods* (BLS, 2008) describes the methodology followed to construct the CPI in the US. The origin of the CPI dates back to 1919 in large cities, although regular publication of a US national index began in 1921, indexes being estimated back to 1913. The major outputs are the so-called American price indexes, which comprise three CPI series—(i) CPI for All Urban Consumers (CPI-U); (ii) Chained CPI for All Urban Consumers (C-CPI-U) and (iii) CPI for Urban Wage Earners and Clerical Workers (CPI-W)—as well as selected average prices.

The sample, weights, coverage and methodology have been enhanced and updated several times, with the most significant developments over the past 30 years being: (i) the introduction of the rental equivalence concept in 1983; (ii) the sixth comprehensive revision carried out in 1998, when timelier consumer spending weights were introduced and geographic and housing samples were updated, among other changes; (iii) the use of the geometric mean formula to average price change within most item categories in 1999 and (iv) the implementation of the biennial weight updates and the inclusion of Chained CPI-Us, among other changes, in 2002.

At present, the basket of goods and services that people buy for day-to-day living is divided into 211 categories (called 'item strata') and the urban part of the US CPI into 38 geographic areas (called 'index areas'). The calculation of price indexes consists of two stages. First, basic indexes are obtained for the average change in each item within each of the 8,018 (38x211) possible area-item combinations in order to, secondly, compute aggregate indexes by averaging across different subsets of the 8,018 item-area combinations.

Focusing on the first stage, from January 1999 the price relative calculation for most of the item strata (around 61% of total consumer spending) uses the geometric mean index formula, which entails the use of a geometric weighting of price ratios with the expenditures on each item in the corresponding sampling period as weights. The rest of the items continue applying a modified Laspeyres index number formula. Table 2 shows both procedures clarifying their meanings.

Table 2: Formulae used in US to calculate price relatives for area-item combination, *a,i*, from previous period *t-1* to current month *t*.

Geometric Formula	Laspeyres Formula
${}_{a,i}R^{G}_{[t;t-1]} = \prod_{j\in a,i}\left(\frac{P_{j,t}}{P_{j,t-1}}\right)^{\left(W_{j,POPS}/\sum_{k\in a,i}W_{k,POPS}\right)}$	${}_{a,i}R^{L}_{[t;t-1]} = \frac{\sum_{j\in a,i}\left(W_{j,POPS}/P_{j,POPS}\right)P_{j,t}}{\sum_{j\in a,i}\left(W_{j,POPS}/P_{j,POPS}\right)P_{j,t-1}}$

$P_{j,t}$ is the price of the *j*th observed item in month *t* for area-item combination *a,i*

$P_{j,t-1}$ is the price of the *j*th observed item in month *t*-1 for area-item combination *a,i*

$P_{j,POPS}$ is an estimate of the item's *j*'s price in the sampling period when its POPS[(a)] was conducted

$W_{j,POPS}$ is item's *j*'s weight in the POPS[(b)]

Source: Own elaboration following BLS (2008).

(a) Point of Purchase Survey; (b) Details could be found in BLS (2008, p. 22).

3. Basic Spatial Methodology adapted to Consumer Price Index Theory

Classical sampling statistics looks at the observations drawn in a sample as realizations of independent random variables identically distributed, which share the same probability distribution as the population where they come from. When the data are geographically located, however, this interpretation can hardly be accepted on account of cluster effects and spatial interdependence. As a result, they have to be regarded as realizations of spatially varying phenomena. From a mathematical point of view, geostatistics builds its inferences on the stochastic process or random function theory assuming that there is a family of random variables $X = \{X(\mathbf{s}), \mathbf{s} \in S\}$ whose distributions vary as a function of a parameter **s** (space) belonging to S (the domain). This characterises a random field that in the multivariate case is defined by a vector of stochastic processes represented by a set of *p* random functions $\mathbf{X} = \{X_1, X_2, \ldots, X_p\}$,

Under this framework, a set of collected prices $\{x_i(\mathbf{s}_j)\}$, the data, could seem an array of realizations of a finite set of random variables $\{X_i(\mathbf{s}_j)\}$ $(i = 1, \ldots, p,\ j = 1, \ldots, n)$; where $X_i(\mathbf{s}_j)$ represents the variable 'price of good or service *i*' collected in the outlet placed at the geographical location $\mathbf{s}_j = (x_j, y_j)$, and $x_i(\mathbf{s}_j)$ denotes the price of the good or service *i* observed in the outlet *j* (located in $\mathbf{s}_j$). Thus, taking into account that the prices of closer outlets tend to be more similar than the prices of outlets that are far away from each other, in this geostatistical context the prices of the good or service *i* in a particular geographical area no longer have to be considered as independent. What is more, given that they will usually be spatially autocorrelated their structure of spatial correlation should be explicitly considered in order to make proper inferences. For example, to attain the best linear unbiased estimate of, say, the price mean of the good or service *i* discounting the spatial dependence of the data, a two-step procedure is required. First, we must estimate the structure of the spatial correlation to secondly obtain the weights of the linear estimator that minimizes the variance of prediction under that structure of spatial autocorrelation.

The structure of spatial correlation can be represented by two functions: the covariance function (or covariogram) and the so-called variogram—the main tool in geostatistics. The covariance function of the stochastic process is a non random function defined by:

$$C\left[X_i(\mathbf{s}_j), X_i(\mathbf{s}_k)\right] = E\left[\left(X_i(\mathbf{s}_j) - \mu_i(\mathbf{s}_j)\right)\left(X_i(\mathbf{s}_k) - \mu_i(\mathbf{s}_k)\right)\right], \forall \mathbf{s}_j, \mathbf{s}_k \in S \subset \mathbb{R}^2, i = 1, \ldots, p$$

which, under the usual second-order stationary assumptions (constant mean and variance), collapses in a covariance function that only depends on the distance between outlets, **h**, being $\mathbf{h} = \mathbf{s}_j - \mathbf{s}_k$. This new function shows the evolution of the correlation with distance:

$$C_i(\mathbf{h}) = E\left[X_i(\mathbf{s})X_i(\mathbf{s}+\mathbf{h})\right] - \mu_i^2, \ \forall \mathbf{s}, \mathbf{s}+\mathbf{h} \in S \subset \mathbb{R}^2, \ i = 1, \ldots, p.$$

Nevertheless, due to the fact that in the intrinsically stationary random function case (when the first-order increments are second-order stationary) covariance cannot exist at the origin, the variogram is the most widely used tool to represent the structure of spatial correlation. It is defined by half of the variance of the increments:

$$\gamma_i(\mathbf{h}) = \frac{1}{2} V\left[X_i(\mathbf{s}+\mathbf{h}) - X_i(\mathbf{s})\right], \ \forall \mathbf{s}, \mathbf{s}+\mathbf{h} \in S \subset \mathbb{R}^2, i = 1, \ldots, p$$

and leads to $\gamma_i(\mathbf{h}) = \frac{1}{2} E\left[\left(X_i(\mathbf{s}+\mathbf{h}) - X_i(\mathbf{s})\right)^2\right]$ in the intrinsically stationary random function case. Among other properties, the covariance function is a positive-definite function, whereas the variogram is a conditionally negative-definite function, both in order to assure that the variance of any linear combination of sample variables is non-negative (more details can be found in Wackernagel, 2003, p.35-40). Furthermore, if the variogram is bounded, it is not difficult to prove the relationship:

$$\gamma_i(\mathbf{h}) = C_i(\mathbf{0}) - C_i(\mathbf{h})$$

Fortunately, these two theoretical functions, covariance and the variogram, could be estimated from the available data (the prices of the good or service *i* collected in each outlet *j*, $j = 1, \ldots, n$).

The classic empirical variogram is computed using the estimator based on the method of moments proposed in Matheron (1962), through:

$$\gamma_i^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left(X_i(\mathbf{s}_j + \mathbf{h}) - X_i(\mathbf{s}_j)\right)^2,$$

where $X_i(\mathbf{s}_j)$ denotes the price of the good or service *i* in an outlet located at $\mathbf{s}_j$ when the price of the same good is also available in another outlet sited at $\mathbf{s}_j + \mathbf{h}$, and $N(\mathbf{h})$ represents the number of pairs of outlets at distance **h**.

Likewise, adapting the time series covariogram to the current context, Cressie and Glonek (1984) suggest estimating the covariogram by:

$$C_i^*(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left(X_i(\mathbf{s}_j + \mathbf{h}) - Med(X_i)\right)\left(X_i(\mathbf{s}_j) - Med(X_i)\right)$$

where $Med(X_i)$ is the median of the set of observed prices $\{ x_i(\mathbf{s}_j) \}$.

This median-based covariogram estimator is preferable, due to its minor bias (Cressie and Glonek, 1984), to the straightforward and frequently used mean-based estimator (e.g. Genton and Gorsich, 2002):

$$C_i^*(\mathbf{h}) = \frac{1}{N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}\left(X_i(\mathbf{s}_j+\mathbf{h})-\bar{X}\right)\left(X_i(\mathbf{s}_j)-\bar{X}\right)$$

where $\bar{X}_i = \left(1/n\right)\sum_{j=1}^{n} X_i(\mathbf{s}_j)$.

In practice, however, there are seldom sufficient (if any) observations linked by each vector **h** and so the data must be grouped for each "approximate distance" **h** (using a certain tolerance) in order to reach a discrete (punctual) estimation of the variogram (covariogram). These punctual estimates help to identify the form of the spatial correlation function of the data and are used to estimate a valid theoretical variogram (covariogram) by fitting the empirical estimates to a proper theoretical function following the linear model of regionalisation.

Once the structure of spatial dependence is available, kriging is the best linear alternative, taking into account the spatial correlation that the prices show, to make inferences about the random field (including the area mean price). Kriging is a linear minimum mean squared error statistical procedure for spatial estimation that assigns different weights to each observation to construct estimates. The weights $\lambda_j,\ j=1,\ldots,n$ are obtained by minimising the variance of the estimation error under the condition of unbiasedness. This estimator provides the best linear unbiased punctual estimate. In this sense, kriging estimates are more precise than any estimates obtained by any other linear estimator and have minimal variance in the Gaussian case.

4. Weighting basic prices through kriging

As defined by the BLS (2008), the CPI is "a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services." In general, a CPI sums up the evolution of the set of prices of goods and services purchased by the households of the population that resides in a particular area and is constructed using the prices of a sample of representative items drawn periodically at fixed points in time (*m*: month, *t*: year) from a set of outlets in the target area $r,\ r=1,\ldots,R$. In the initial stage of CPI construction, basic indexes (or elemental aggregate simple indexes in European terminology) for each product *i* are estimated using the prices recorded in the sample. It is at this stage when space should be explicitly taken into account to consider the information derived from it, as once one begins to aggregate basic indexes, the geo-reference of the information becomes increasingly diffuse and unavailable.

4.1. European Union case

As stated in Table 1, the arithmetic mean of price relatives was considered inadequate by the Commission of the European Communities (European Communities, 1996, Annex II, p. 217) to compile elemental aggregate simple indexes. As alternatives, either the ratio of arithmetic mean prices or the ratio of geometric mean prices were proposed. Observing both expressions (see Table 1), it is easy to note that both are constructed as a quotient of means (either geometric or arithmetic) in which all the prices receive the same weight: $\frac{1}{n}$. That is, they are constructed as if the set of drawn prices was a simple random sample of the price population, omitting the spatial dependence that exists among them. In this section, taking account the spatial autocorrelation of prices, we propose more accurate alternative estimators, within the same family. In particular, (i) the *Ratio of Kriged Means* is proposed as a substitute of the ratio of arithmetic mean prices and (ii) the so-called G*eometric Weighted Mean Prices* is suggested as an alternative to the ratio of geometric mean prices.

4.1.1. Arithmetic Weighted Mean Prices: Ratio of Kriged Means

It is a well-known fact that price indexes based on arithmetic averages perform worse, as a rule, than indexes based on geometric means and consequently warrant less interest. However, as they must be considered, we had decided to present them first because, out of all the statistics proposed to construct elementary indexes by EU and US CPI authorities, they are the easiest to adjust in order to incorporate spatial autocorrelation, thereby representing a good way to begin familiarising the reader with the topic. Indeed, in this case spatial dependence can be included in straightforward fashion by using the well-known (in geostatistics literature) *kriging the mean* statistic (e.g., Wackernagel, 2003).

In particular, given a particular area, let n_i^{mt} be the number of commercial establishments where prices of the good or service *i* have been collected and let *p* be the size of the set of consumer goods and services (elemental aggregates) drawn. Then, instead of using a quotient of arithmetic means, the proposal is to use, for each item *i* (with $i = 1,\ldots, p$), a quotient of spatial weighted arithmetic means ${}_{AW}\overline{P}_i^{mt}$ and ${}_{AW}\overline{P}_i^{Dec(t-1)}$—

${}_{AW}\overline{P}_i^{mt}$ being the spatial weighted mean price of the elemental aggregate *i* in month *m* of year *t*, ${}_{AW}\overline{P}_i^{Dec(t-1)}$ the same value in December of the previous year and *p* the size of the set of consumer goods and services (elemental aggregates) drawn—, where a generic ${}_{AW}\overline{P}_i^{mt}$ is obtained from $P_i^{mt}(\mathbf{s}_j)$ using the kriged mean formula:

$$ {}_{AW}\overline{P}_i^{mt} = \sum_{j=1}^{n_i^{mt}} w_j P_i^{mt}(\mathbf{s}_j), $$

with $P_i^{mt}(\mathbf{s}_j)$ denoting the price of the item *i* collected from the outlet at $\mathbf{s}_j$ in month *m* of year *t*, $\left\{\mathbf{s}_j, j=1,\ldots,n_i^{mt}\right\}$ representing the set of spatial locations where prices are collected, and the spatial weights, w_j, being obtained after imposing the conditions of unbiasedness and minimum error variance on the estimator.

Unbiasedness condition: Let μ_i^{mt} be the expected value of the prices of the elementary aggregate *i* in month *m* of year *t:* $E\left(P_i^{mt}(\mathbf{s}_j)\right) = \mu_i^{mt}$.Then, given that the expected value is supposed to be constant under second-order stationarity, it follows that the unbiasedness condition is equivalent to $E\left({}_{AW}\overline{P}_i^{mt} - \mu_i^{mt}\right) = 0$, and hence:

$$ E\left({}_{AW}\overline{P}_i^{mt}\right) = E\left(\sum_{j=1}^{n_i^{mt}} w_j P_i^{mt}(\mathbf{s}_j)\right) = \sum_{j=1}^{n_i^{mt}} w_j E\left(P_i^{mt}(\mathbf{s}_j)\right) = $$
$$ = \sum_{j=1}^{n_i^{mt}} w_j \mu_i^{mt} = \mu_i^{mt} \sum_{j=1}^{n_i^{mt}} w_j = \mu_i^{mt} \Leftrightarrow \sum_{j=1}^{n_i^{mt}} w_j = 1 $$

That is, from the unbiasedness condition it follows (i) that the set of weights, w_j, of the proposed estimator must sum up to one and, furthermore, (ii) that any linear estimator with weights that sum up to one will be unbiased. The best, therefore, will be that which displays minimum variance.

Minimum variance: The variance of the estimator can be expressed by:

$$ V\left({}_{AW}\overline{P}_i^{mt}\right) = E\left[\left({}_{AW}\overline{P}_i^{mt}\right)^2\right] - E\left[{}_{AW}\overline{P}_i^{mt}\right]^2 = E\left[\left(\sum_{j=1}^{n_i^{mt}} w_j P_i^{mt}(\mathbf{s}_j)\right)^2\right] - \left(\mu_i^{mt}\right)^2 = $$
$$ = \sum_{j=1}^{n_i^{mt}} \sum_{k=1}^{n_i^{mt}} w_j w_k E\left(P_i^{mt}(\mathbf{s}_j) P_i^{mt}(\mathbf{s}_k)\right) - \left(\mu_i^{mt}\right)^2 , $$

or, equivalently, denoting the covariance function of the stochastic process P_i^{mt} by $C\left(P_i^{mt}(\mathbf{s}_j), P_i^{mt}(\mathbf{s}_k)\right) = E\left[\left(P_i^{mt}(\mathbf{s}_j) - E\left(P_i^{mt}(\mathbf{s}_j)\right)\right)\cdot\left(P_i^{mt}(\mathbf{s}_k) - E\left(P_i^{mt}(\mathbf{s}_k)\right)\right)\right]$, it follows that the expected value of the product of variables can be written in covariance terms as:

$$V\left({}_{AW}\overline{P}_i^{mt}\right) = \sum_{j=1}^{n_i^{mt}}\sum_{k=1}^{n_i^{mt}} w_j w_k \left[C\left(P_i^{mt}(\mathbf{s}_j), P_i^{mt}(\mathbf{s}_k)\right) + \left(\mu_i^{mt}\right)^2\right] - \left(\mu_i^{mt}\right)^2$$

an expression that, taking into account that the weights sum up to one, simplifies to:

$$V\left({}_{AW}\overline{P}_i^{mt}\right) = \sum_{j=1}^{n_i^{mt}}\sum_{k=1}^{n_i^{mt}} w_j w_k C\left(P_i^{mt}(\mathbf{s}_j), P_i^{mt}(\mathbf{s}_k)\right)$$

Thus, in order to minimise the variance of the estimator under the condition of unbiasedness, we can define the following Lagrange function:

$$L\left(w_j, \vartheta\right) = \sum_{j=1}^{n_i^{mt}}\sum_{k=1}^{n_i^{mt}} w_j w_k C\left(P_i^{mt}(\mathbf{s}_j), P_i^{mt}(\mathbf{s}_k)\right) - 2\vartheta\left(\sum_{j=1}^{n_i^{mt}} w_j - 1\right),\ j = 1,\ldots,n_i^{mt},$$

which, after setting partial derivatives to zero, yields the following $n_i^{mt} + 1$ equation system:

$$\begin{cases} \sum_{k=1}^{n_i^{mt}} w_j C\left(P_i^{mt}(\mathbf{s}_j), P_i^{mt}(\mathbf{s}_k)\right) - \vartheta = 0,\ j = 1,\ldots,n_i^{mt} \\ \sum_{j=1}^{n_i^{mt}} w_j = 1 \end{cases}$$

This system can be easily expressed in matrix form through: $\mathbf{C}\cdot\mathbf{W} = \mathbf{A}$, with $\mathbf{C}$, $\mathbf{W}$ and $\mathbf{A}$ given respectively by:

$$\mathbf{C} = \begin{pmatrix} C(\mathbf{0}) & C(\mathbf{s}_1-\mathbf{s}_2) & C(\mathbf{s}_1-\mathbf{s}_3) & \cdots & C(\mathbf{s}_1-\mathbf{s}_{n_i^{mt}}) & -1 \\ C(\mathbf{s}_2-\mathbf{s}_1) & C(\mathbf{0}) & C(\mathbf{s}_2-\mathbf{s}_3) & \cdots & C(\mathbf{s}_2-\mathbf{s}_{n_i^{mt}}) & -1 \\ C(\mathbf{s}_3-\mathbf{s}_1) & C(\mathbf{s}_3-\mathbf{s}_2) & C(\mathbf{0}) & \cdots & C(\mathbf{s}_3-\mathbf{s}_{n_i^{mt}}) & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C(\mathbf{s}_{n_i^{mt}}-\mathbf{s}_1) & C(\mathbf{s}_{n_i^{mt}}-\mathbf{s}_2) & C(\mathbf{s}_{n_i^{mt}}-\mathbf{s}_3) & \cdots & C(\mathbf{0}) & -1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{n_i^{mt}} \\ \vartheta \end{pmatrix} \text{ and } \mathbf{A} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

where, to shorten the notation, $C\left(P_i^{mt}(\mathbf{s}_j), P_i^{mt}(\mathbf{s}_k)\right)$ has been replaced by $C\left(\mathbf{s}_j - \mathbf{s}_k\right)$.

Thus, provided that $\mathbf{C}$ is invertible, it is easy to obtain the set of weights $w_j,\ j = 1,\ldots,n_i^{mt}$ through $\mathbf{W} = \mathbf{C}^{-1}\cdot\mathbf{A}$, and from them the best linear unbiased estimator of the mean of the prices of item *i* in month *m* of year *t* and, subsequently, the desired ratio. Note that this solution is a generalisation of the EU expression. Indeed, it is not difficult to prove (e.g., Montero and Larraz, 2010) that in the case of spatial independence this process leads to the usual (non-weighted) arithmetic mean.

Finally, standard errors of the estimators can be obtained through the Lagrange multiplier as:

$$V\left({}_{AW}\overline{P}_i^{mt}\right)=\sum_{j=1}^{n_i^{mt}}\sum_{k=1}^{n_i^{mt}}w_j w_k C\left(P_i^{mt}(\mathbf{s}_j),P_i^{mt}(\mathbf{s}_k)\right)=\sum_{j=1}^{n_i^{mt}}w_j\sum_{k=1}^{n_i^{mt}}w_k C\left(P_i^{mt}(\mathbf{s}_j),P_i^{mt}(\mathbf{s}_k)\right)=$$
$$=\sum_{j=1}^{n_i^{mt}}w_j\vartheta=\vartheta$$

4.1.2. Geometric Weighted Mean Prices

Proceeding in similar fashion to when we generalized the ratio of the arithmetic mean statistic, the spatial-based estimator suggested for the geometric price index of an elementary aggregate is in this case defined straightforwardly by:

$${}_{Dec(t-1)}I_i^{mt}=\frac{{}_{GW}\overline{P}_i^{mt}}{{}_{GW}\overline{P}_i^{Dec(t-1)}}\ ,\ \forall i=1,\ldots,p$$

where ${}_{GW}\overline{P}_i^{mt}$ denotes the spatially weighted geometric mean price of the elemental aggregate *i* in month *m* of year *t* and ${}_{GW}\overline{P}_i^{Dec(t-1)}$ the corresponding value of December of the previous year, and a generic ${}_{GW}\overline{P}_i^{mt}$ obtained through the expression:

$${}_{GW}\overline{P}_i^{mt}=\sqrt[\sum_{j=1}^{n_i^{mt}}\omega_j]{\prod_{j=1}^{n_i^{mt}}\left(P_i^{mt}(\mathbf{s}_j)\right)^{\omega_j}}=\prod_{j=1}^{n_i^{mt}}\left(P_i^{mt}(\mathbf{s}_j)\right)^{w_j}\text{, where } w_j=\frac{\omega_j}{\sum_{j=1}^{n_i^{mt}}\omega_j}$$

Compared to the previous case, however, it is a bit more complicated to obtain the spatial weights, w_j, that characterise the proposed geometric mean of the elementary aggregates. In this case, based on the (approximate) relationship $E\left(Ln(X)\right)\approx Ln\left(E(X)\right)$, the spatial weights are obtained by linearising the estimator through a natural logarithm (which maps multiplication into addition) before imposing the minimum mean squared error condition. That is, the weights are those associated to the best spatial linear unbiased estimator of the log-transformed prices. Indeed, taking logs in ${}_{GW}\overline{P}_i^{mt}$, we obtain:

$$ {}_{w}\bar{Y}_i^{mt} = Ln\left({}_{GW}\bar{P}_i^{mt} \right) = \frac{1}{\sum_{j=1}^{n_i^{mt}} \omega_j} \sum_{j=1}^{n_i^{mt}} \omega_j LnP_i^{mt}(\mathbf{s}_j) = \sum_{j=1}^{n_i^{mt}} w_j LnP_i^{mt}(\mathbf{s}_j), $$

that defining $Y_i^{mt}(\mathbf{s}_j) = LnP_i^{mt}(\mathbf{s}_j)$ collapses in an expression of the same type as that addressed in the previous section. In particular, η_i^{mt} representing the expected value of the natural logarithm of the prices of the elementary aggregate *i* in month *m* of year *t:* $E\left(Y_i^{mt}(\mathbf{s}_j)\right) = \eta_i^{mt}$, which is supposed to be constant under second-order stationarity, the condition of unbiasedness once again leads to:

$$ E\left({}_{w}\bar{Y}_i^{mt} \right) = E\left(\sum_{j=1}^{n_i^{mt}} w_j Y_i^{mt}(\mathbf{s}_j) \right) = \sum_{j=1}^{n_i^{mt}} w_j E\left(Y_i^{mt}(\mathbf{s}_j) \right) = $$
$$ = \sum_{j=1}^{n_i^{mt}} w_j \eta_i^{mt} = \eta_i^{mt} \sum_{j=1}^{n_i^{mt}} w_j = \eta_i^{mt} \Leftrightarrow \sum_{j=1}^{n_i^{mt}} w_j = 1 $$

And, in the same way as before, denoting the covariance function of the log-transformed stochastic process Y_i^{mt} with $C\left(Y_i^{mt}(\mathbf{s}_j), Y_i^{mt}(\mathbf{s}_k)\right) = E\left[\left(Y_i^{mt}(\mathbf{s}_j) - E\left(Y_i^{mt}(\mathbf{s}_j)\right) \right) \cdot \left(Y_i^{mt}(\mathbf{s}_k) - E\left(Y_i^{mt}(\mathbf{s}_k)\right) \right) \right]$, we have that the variance of ${}_{w}\bar{Y}_i^{mt}$ could be expressed by:

$$ V\left({}_{w}\bar{Y}_i^{mt} \right) = \sum_{j=1}^{n_i^{mt}} \sum_{k=1}^{n_i^{mt}} w_j w_k \left[C\left(Y_i^{mt}(\mathbf{s}_j), Y_i^{mt}(\mathbf{s}_k) \right) + \left(\eta_i^{mt} \right)^2 \right] - \left(\eta_i^{mt} \right)^2, $$

which, considering that the weights must sum up to one, simplifies to:

$$ V\left({}_{w}\bar{Y}_i^{mt} \right) = \sum_{j=1}^{n_i^{mt}} \sum_{k=1}^{n_i^{mt}} w_j w_k C\left(Y_i^{mt}(\mathbf{s}_j), Y_i^{mt}(\mathbf{s}_k) \right) $$

Thus, defining the corresponding Lagrange function by:

$$ L\left(W_j, \vartheta \right) = \sum_{j=1}^{n_i^{mt}} \sum_{k=1}^{n_i^{mt}} w_j w_k C\left(Y_i^{mt}(\mathbf{s}_j), Y_i^{mt}(\mathbf{s}_k) \right) - 2\vartheta \left(\sum_{j=1}^{n_i^{mt}} w_j - 1 \right), \; j = 1, \ldots, n_i^{mt} $$

and setting partial derivatives to zero (in order to minimise estimator variance), the following equation system is obtained:

$$ \begin{cases} \sum_{k=1}^{n_i^{mt}} w_j C\left(Y_i^{mt}(\mathbf{s}_j), Y_i^{mt}(\mathbf{s}_k) \right) - \vartheta = 0, \; j = 1, \ldots, n_i^{mt} \\ \sum_{j=1}^{n_i^{mt}} w_j = 1 \end{cases}, $$

and from it, the weights $w_j, \; j = 1, \ldots, n_i^{mt}$.

Once the weights have been obtained, the log-transformed estimator can be computed and, taking antilogarithms, the desired spatial weighted geometric mean: ${}_{GW}\bar{P}_i^{mt} = e^{{}_w\bar{Y}_i^{mt}}$. We should note again that the proposed estimator is a generalisation of the EU ratio of non-weighted geometric means. If prices were spatially independent they would both coincide.

Finally, using the fact that the variance of ${}_w\bar{Y}_i^{mt}$ is the Lagrange multiplier:

$$V\left({}_w\bar{Y}_i^{mt}\right) = \sum_{j=1}^{n_i^{mt}} \sum_{k=1}^{n_i^{mt}} w_j w_k C\left(Y_i^{mt}(\mathbf{s}_j), Y_i^{mt}(\mathbf{s}_k)\right) = \sum_{j=1}^{n_i^{mt}} w_j \sum_{k=1}^{n_i^{mt}} w_k C\left(Y_i^{mt}(\mathbf{s}_j), Y_i^{mt}(\mathbf{s}_k)\right) =$$
$$= \sum_{j=1}^{n_i^{mt}} w_j \vartheta = \vartheta$$

and a well-know result of the variance of the natural logarithm of a random variable X, $V\left(Ln(X)\right) \approx \left[E\left(X\right)\right]^{-2} V(X)$, it follows that:

$$V\left({}_{GW}\bar{P}_i^{mt}\right) \approx \left[E({}_{GW}\bar{P}_i^{mt})\right]^2 \vartheta \approx \left[e^{E({}_w\bar{Y}_i^{mt})}\right]^2 \vartheta = \vartheta e^{2\eta_i^{mt}},$$

from which it is straightforward to obtain an estimator for the variance of the spatial weighted geometric mean by replacing the unknown parameter η_i^{mt} with its best available estimator ${}_w\bar{Y}_i^{mt}$.

4.2. United States case

As mentioned in subsection 2.2., since January 1999 the BLS has implemented two different statistics to calculate its basic indexes: a geometric weighted mean of price ratios and a modified Laspeyres index number formula. The geometric mean formula is, nevertheless, recommended ahead of the Laspeyres formula and is broadly used in all item strata. In fact, the expression of the expenditure-share-weighted geometric average, ${}_{a,i}R_{[t;t-1]}^G$, is applied, except for some selected items (some shelter services, some utilities and government charges and some medical services), for most of all usable quotes in each of the 8,018 area-item combination *a,i*.

As in the EU case, alternative formulae, which take into account the information provided by the spatial coordinates of the stores where prices are collected, are proposed in this section. In particular, the *New Spatial-Expenditure Weighted Geometric Average* is suggested as a substitute for the US geometric formula and the *New Spatial-Expenditure Weighted Arithmetic Average* is provided as an option instead of the Laspeyres expression.

4.2.1. New Spatial-Expenditure Weighted Geometric Average

Adapting the notation introduced previously to the US BLS case, let us assume that the price relative measuring the short-term price change between any two consecutive (either monthly or bimonthly) periods (*t* -1, *t*) of an item *j* of item-strata *i* can be built in each outlet drawn in an area *a*:

$$I_{j,t-1}^{t}(\mathbf{s}_j) = \frac{P_{j,t}(\mathbf{s}_j)}{P_{j,t-1}(\mathbf{s}_j)}, \text{ for } j = 1,\ldots,n_{a,i},$$

where $I_{j,t-1}^{t}(\mathbf{s}_j)$ represents the price relative of item *j* in an outlet located at $\mathbf{s}_j$ in area *a*, $P_{j,t}(\mathbf{s}_j)$ the price of the *j*th observed item in month *t* in outlet $\mathbf{s}_j$, $P_{j,t-1}(\mathbf{s}_j)$ the price of the same item in period *t* -1 and $n_{a,i}$ denotes the number of items observed for area-item combination *a,i*.

In order to construct basic indexes, the current US methodology already weights these price relatives, $I_{j,t-1}^{t}(\mathbf{s}_j)$, albeit using only the proportion of estimated expenditures of the item, $W_{j,POPS} / \sum_{k\in a,i} W_{k,POPS}$, as weights in the POPS of the corresponding sampling period (BLS, 2008). That is, in order to construct basic indexes, the BLS takes into account expenditure structure but omits the spatial correlation of price relatives.

Fortunately, however, without breaking away from the current methodology, the basic indexes can also be spatially weighted using, after some algebra, an approach similar to that followed in subsection 4.1.2. In particular, given that $\sqrt[\sum_{k\in a,i} W_{k,POPS}]{\prod_{j\in a,i}\left(I_{j,t-1}^{t}(\mathbf{s}_j)\right)^{W_{j,POPS}}} = \sqrt[n_{a,i}]{\prod_{j\in a,i}\left(I_{j,t-1}^{t}(\mathbf{s}_j)\right)^{n_{a,i}(W_{j,POPS}/\sum_{k\in a,i} W_{k,POPS})}}$ and the fact that the expected value of $n_{a,i}(W_{j,POPS} / \sum_{k\in a,i} W_{k,POPS})$ tends to one as $n_{a,i}$ tends to infinity, it would not be difficult to observe the expression of ${}_{a,i}R_{[t,t-1]}^{G}$ as a (un-weighted) geometric mean of the "weighted" price relatives, $\left(I_{j,t-1}^{t}(\mathbf{s}_j)\right)^{n_{a,i}(W_{j,POPS}/\sum_{k\in a,i} W_{k,POPS})}$, and therefore making it possible to use the procedures in the previous subsection.

Thus, in order to introduce the spatial structure of the stochastic process of price relatives into the expenditure-share-weighted geometric average statistic, we propose defining the new spatial-expenditure weighted geometric mean by

$$ {}_{w,a,i}R^{G}_{[t,t-1]} = \sqrt[\sum_{j\in a,i}\lambda_j]{\prod_{j\in a,i}\left[\left(I^{t}_{j,t-1}(\mathbf{s}_j)\right)^{n_{a,i}(W_{j,POPS}/\sum_{k\in a,i}W_{k,POPS})}\right]^{\lambda_j}} = \prod_{j\in a,i}\left[\left(I^{t}_{j,t-1}(\mathbf{s}_j)\right)^{n_{a,i}(W_{j,POPS}/\sum_{k\in a,i}W_{k,POPS})}\right]^{\rho_j} $$

with the spatial weights, $\rho_j = \dfrac{\lambda_j}{\sum_{j\in a,i}\lambda_j}$, being obtained by imposing the conditions of unbiasedness and minimum variance on the log-transformation of the "weighted" price relatives. These weights, combined with the expenditure weights, $\rho_j n_{a,i}(W_{j,POPS}/\sum_{k\in a,i}W_{k,POPS})$, make it possible to weight price relatives taking into account both the spatial dependence and POPS expenditure structure of drawn items.

Repeating the steps followed previously, it is straightforward to obtain the spatial weights. First, taking natural logarithms, the estimator is linearised yielding:

$$ {}_{w,a,i}Y^{G}_{[t,t-1]} = Ln\left({}_{w,a,i}R^{G}_{[t,t-1]}\right) = \sum_{j\in a,i}\rho_j Ln\left(\left(I^{t}_{j,t-1}(\mathbf{s}_j)\right)^{n_{a,i}(W_{j,POPS}/\sum_{k\in a,i}W_{k,POPS})}\right) = \sum_{j\in a,i}\rho_j Y^{t}_{j,t-1}(\mathbf{s}_j), $$

with $Y^{t}_{j,t-1}(\mathbf{s}_j) = Ln\left(\left(I^{t}_{j,t-1}(\mathbf{s}_j)\right)^{n_{a,i}(W_{j,POPS}/\sum_{k\in a,i}W_{k,POPS})}\right)$.

Second, the condition of unbiasedness yields $\sum_{j\in a,i}\rho_j = 1$.

Third, imposing the minimum error variance condition leads to the Lagrange Function $L\left(\rho_j,\vartheta_j\right) = \sum_{j\in a,i}\sum_{k\in a,i}\rho_j\rho_k C\left(Y^{t}_{j,t-1}(\mathbf{s}_j), Y^{t}_{k,t-1}(\mathbf{s}_k)\right) - 2\vartheta\left(\sum_{j\in a,i}\rho_j - 1\right)$, which after deriving and setting to zero produces the system:

$$ \begin{cases} \sum_{k\in a,i}\rho_k C\left(Y^{t}_{j,t-1}(\mathbf{s}_j), Y^{t}_{k,t-1}(\mathbf{s}_k)\right) - \vartheta = 0, \; j\in a,i \\ \sum_{j\in a,i}\rho_j = 1 \end{cases}, $$

Fourth, solving the system the ρ_j spatial weights are obtained and from them the linear estimator ${}_{w,a,i}R^{G}_{[t,t-1]}$ is computed. Finally, replacing $E({}_{w,a,i}Y^{G}_{[t;t-1]})$ by ${}_{w,a,i}R^{G}_{[t,t-1]}$ in

$V\left({}_{w,a,i}R^{G}_{[t;t-1]}\right) \approx \left[E({}_{w,a,i}R^{G}_{[t;t-1]})\right]^2 V\left({}_{w,a,i}Y^{G}_{[t;t-1]}\right) = \left[e^{E({}_{w,a,i}Y^{G}_{[t;t-1]})}\right]^2 \vartheta$, an estimate for the variance of the new spatial-expenditure weighted geometric mean is obtained.

4.2.2. New Spatial-Expenditure Weighted Arithmetic Mean

The modified Laspeyres index number formula shown in Table 2 is currently used to estimate basic indexes in 13 item-strata —rent of primary residence; owners' equivalent rent of primary residence; housing at school, excluding board; electricity; residential water and sewerage maintenance; landline telephone services, local charges; utility (piped) gas service; State and local registration, license, and motor vehicle property tax; physicians' services; hospital services; dental services; services by other medical professionals; and nursing homes and adult daycare (BLS, 2008, Table 4)— which represent around 39% of total consumer spending. This section proposes to modify this Laspeyres formula, which is a base-period-quantity-weighted arithmetic average, using the new Spatial-Expenditure Weighted Arithmetic Mean, which accounts for the spatial autocorrelation of price relatives.

Expressing the Laspeyres formula of Table 2 in terms of price relatives:

$$ {}_{a,i}R^{L}_{[t;t-1]} = \frac{\sum_{j\in a,i}\left(W_{j,POPS}/P_{j,POPS}\right)P_{j,t}}{\sum_{j\in a,i}\left(W_{j,POPS}/P_{j,POPS}\right)P_{j,t-1}} = \frac{\sum_{j\in a,i}\frac{P_{j,t}}{P_{j,t-1}}\left(W_{j,POPS}/P_{j,POPS}\right)P_{j,t-1}}{\sum_{j\in a,i}\left(W_{j,POPS}/P_{j,POPS}\right)P_{j,t-1}} = \frac{\sum_{j\in a,i} I^{t}_{j,t-1} w_j}{\sum_{j\in a,i} w_j}, $$

with weights $w_j = \frac{W_{j,POPS}}{P_{j,POPS}} P_{j,t-1}$ quote depending; it is not difficult to write it as a "non-weighted" arithmetic mean:

$$ {}_{a,i}R^{L}_{[t;t-1]} = \frac{1}{n_{a,i}} \sum_{j\in a,i} I^{t}_{j,t-1} \left(\frac{w_j}{\sum_{j\in a,i} w_j} \right) n_{a,i} $$

which using the notation introduced in subsection 4.2.1 could be represented by:

$$ {}_{a,i}R^{L}_{[t;t-1]} = \frac{1}{n_{a,i}} \sum_{j\in a,i} I^{t}_{j,t-1}(\mathbf{s}_j) \left(\frac{w_j}{\sum_{j\in a,i} w_j} \right) n_{a,i} $$

Thus, in the same line as before, the proposal to incorporate spatial auto correlation into the new spatial-expenditure weighted arithmetic mean would lead this time to the statistic:

$$ {}_{w,a,i}R^{L}_{[t;t-1]} = \frac{1}{\sum_{j\in a,i}\lambda_j}\sum_{j\in a,i}\lambda_j I^t_{j,t-1}(\mathbf{s}_j)\left(\frac{n_{a,i}w_j}{\sum_{j\in a,i}w_j}\right) = \sum_{j\in a,i}\rho_j I^t_{j,t-1}(\mathbf{s}_j)\left(\frac{n_{a,i}w_j}{\sum_{j\in a,i}w_j}\right), $$

with the spatial weights, $\rho_j = \frac{\lambda_j}{\sum_{j\in a,i}\lambda_j}$, obtained (imposing again the conditions of unbiasedness and minimum error variance) using the spatial process defined by $Y^t_{j,t-1}(\mathbf{s}_j) = \frac{n_{a,i}w_j}{\sum_{j\in a,i}w_j} I^t_{j,t-1}(\mathbf{s}_j)$ and the variance of the estimator coinciding this time with the Lagrange multiplier of the corresponding Lagrange function.

5. Conclusions and Final Remarks

Inflation (deflation) is a topic of great concern to economic authorities and citizens in developed countries, where almost every macro and microeconomic decision is markedly affected by it. It is therefore vitally important for inflation measures to be as accurate and up-to-date as possible. This explains why Consumer Price Indexes, the tools used to measure inflation, have been repeatedly modified over the years in an attempt to improve them. Despite frequent modifications, however, the spatial dimension of the data collected to elaborate them has been only superficially exploited. Although the weights used to aggregate basic indexes usually depend on the region (or area), the fact of the matter is that the spatial dependence shown by elementary prices has not as yet been taken into account in CPI construction, despite the prices drawn from stores closer together tending to be more similar than prices collected in stores that are further apart, as is the case with many other socioeconomic variables. In order to correct this situation, this paper proposes to explicitly consider the geographical coordinates where items are collected and to use the structure of the spatial correlation of prices to weight price relatives in the construction of basic indexes. This will only entail a slight modification of the way the survey is conducted, retaining the locations of the stores where prices are collected and adapting the formulae of the indexes currently in use accordingly. In particular, this paper shows how EU elementary aggregate and US basic index formulae could be appropriately altered to include the spatial dependence of prices.

Firstly, in order to familiarize readers with the terminology used in price index construction, EU and US CPI methodologies are succinctly analyzed. In both cases, the paper focuses on the most detailed level of stratification: the elementary aggregate indexes in European terminology, within which reliable expenditure information is not available for weighting purposes, and basic indexes, in American terminology, where weights are equal to expenditures on the items in their sampling periods. Secondly, geostatistical notation is adapted to CPI theory and we show how the structure of the spatial dependence of prices could be estimated using a covariogram or variogram function. Finally, we propose weighting elementary prices in CPI construction using kriging methodology. In each of the cases, the information provided by the location of the stores where sample prices have been collected is considered at the very beginning of CPI elaboration, before aggregation.

In the European Union framework, two spatial alternatives have been developed: the *Ratio of Kriged Means* and the *Geometric Weighted Mean Prices*. Both of them provide the best estimator of price indexes of elementary aggregates in the presence of spatially autocorrelated prices for each of the formulae allowed in European regulations, both being unbiased and with minimum variance. In the US framework, the BLS uses a geometric formula in all item strata except for some selected items, which still use a Laspeyres formula average, both being expenditure-weighted means. In order to incorporate the information from the spatial location of the outlets where prices have been collected, thereby improving the weighted geometric and arithmetic means currently in use, this paper suggests a *New Spatial-Expenditure Weighted Geometric Average* and a *New Spatial-Expenditure Weighted Arithmetic Average* for each area-item combination. In all cases, in the presence of spatial independence, the proposed weighted statistics reduce to the estimators recommended in either the European Commission or the BLS technical documents. The proposals could be therefore observed as more accurate generalizations of the CPIs currently in use.

References

BLS (1998) Planned Change in the Consumer Price Index Formula. Bureau of Labor Statistics. April 16.

BLS (2008) Handbook of Methods. Chapter 17. The Consumer Price Index. Available online: http://www.bls.gov/opub/hom/pdf/homch17.pdf. Access date: June, 2010.

Boskin, M.J., Dulberger, E. and Griliches, Z. (1996) Toward a more accurate measure of the cost of living. Final report of the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. Washington, U.S., Government Printing Office, December. http://www.ssa.gov/history/reports/boskinrpt.html

Chica-Olmo, J. (2007) Prediction of Housing Location Price by a Multivariate Spatial Method: Cokriging. Journal of Real Estate Research, 2007, 29(1), 91-114.

Cressie, N. and Glonek, G. (1984) Median based covariogram estimators reduce bias. Statistics and Probability Letters, 2, 299-304.

Englund E. J. (1993) *Spatial simulation: environmental applications*. In Goodchild M F, Parks B O, Steyaert L T (eds) Environmental modeling with GIS. New York, Oxford University Press: 432–437

European Communities (1995) Council regulation (EC) No 2494/95 of 23 October 1995 concerning harmonized indices of consumer prices. Official Journal No L257, 27.10.1995, p.1.

European Communities (1996) Commission regulation (EC) No 1749/96 of 9 September 1996 on initial implementing measures for Council Regulation (EC) No 2495/95 concerning harmonized indices of consumer prices. Official Journal No L229, 10.9.1993, p.3.

European Communities (2001) Compendium of HICP. Reference documents. Working Documents (2/2001/B/5). Office for Official Publications of the European Communities.

Fenwick, D. (1998). The impact of sample design on the performance of the sample geometric mean and related issues. Fourth Meeting of the International Working Group on Price Indexes. Washington, US.

http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+4/$file/1998%204th%20Meeting%20-%20Fenwick%20-%20The%20Impact%20of%20Sample%20Design%20on%20the%20Performance%20of%20the%20Sample%20Geometric%20Mean%20and%20Related%20Issues.pdf

Genton H.G. and Gorsich, D.J. (2002) Nonparametrics variogram and covariogram estimation with Fourier-Bessel matrices. Computational Statistics and Data Analysis, 41, 47-57.

Goovaerts, P. (2006) Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *International Journal of Health Geographics* 5: 52.

Laslett G M, McBratney A B, Pahl P J, Hutchinson M F 1987 Comparison of several spatial prediction methods for soil pH. Journal of Soils Science 38: 32–50

Lawson, A.B. (2001) Statistical methods in spatial epidemiology. New York: Wiley.

Matheron, G. (1962). Traité de Géostatistique Appliquée. Technip, Paris.

Montero, JM and Larraz, B. (2010) Estimating housing prices: a proposal with spatially correlated data. International Advances in Economic Research, 16(1), pp39-51

Montero, J.M., Chasco, C. and Larraz, B. (2010) Building an Environmental Quality Index for a big city: a spatial interpolation approach combined with a distance indicator. Journal of Geographical Systems. 435-459.

Montero, J.M., Larraz, B. and Paez, A. (2009) Estimating Commercial Property Prices: An Application of Cokriging with Housing Prices as Anciliary Information. Journal of Geographical Systems, 11(4), 407-425

Nagle, N. (2010) Geostatistical Smoothing of Areal Data: Mapping Employment Density with Factorial Kriging Geographical Analysis, 42, 99-117.

Pavía, J.M., Larraz, B. and Montero, J.M. (2008) Election forecasts using spatio-temporal models. Journal of the American Statistical Association, 103 (483), 1050-1059

Reinsdorf, M and Triplett, J.E. (2009) A review of reviews: ninety years of professional thinking about the consumer price index. In Price Index Concepts and Measurement, Studies in Income and Wealth Volume 70, Edited by W. Erwin Diewert, Jon Greenlees and Charles Hulten

Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234-240.

Turvey, R. (2004). *Consumer Price Index Manual: Theory and Practice*. International Labour Office.

United Nations (2000) Classifications of expenditure according to purpose, *Statistical papers*, Series M, No. 84. Department of economic and social affairs. Statistics division: New York.

Wackernagel, H. (2003) *Multivariate Geostatistics*. Springer-Verlag, Berlin Heidelberg.