A Spatial Cost of Living Index for Colombia using a Microeconomic Approach and Censored Data

Atuesta, Laura and Paredes, Araya
University of Illinois, Universidad Catolica del Norte

29. April 2011

A Spatial Cost of Living Index for Colombia using a Microeconomic Approach and Censored Data

Laura Atuesta[a] * Dusan Paredes[a],[b]

[a] *Department of Agricultural and Consumer Economics,*
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[b] *Department of Economics,*
Universidad Catolica del Norte, Antofagasta, Chile

April 25, 2011

SUMMARY

This paper describes a methodology to calculate a spatial cost of living index using Colombian data for 2006 that takes into consideration the microeconomic behavior of households. Using the Almost Ideal Demand System and recovering the expenditure functions for the 23 main Colombian cities, the index proposed is compared to the traditional methodologies used to calculate the regional basket of goods in the country and to an alternative methodology proposed by Romero (2005). This comparison suggests that when the substitution effects are not considered, and the same basket of goods is evaluated in every city, the index is biased, and this bias increases when the difference between cities increases. For reducing the bias, we use a microeconomic approach that keeps the households' level of utility constant and allows substitution among different baskets of goods. According to our calculations, Bogota is still the most expensive city in the country followed by Armenia, Cali, Bucaramanga and Ibague.

*Corresponding author. E-mail: latuest2@illinois.edu

1. INTRODUCTION

The Spatial Cost of Living (SCOL) is crucial to compare monetary variables between different spatial units. For instance, nominal regional wages cannot be properly compared because the purchasing power in each region is affected by the level of local prices. Given this proposition, it is surprising to find that the SCOL is rarely available in official statistical agencies and Colombia no exception. This lack of information affects seriously any policy evaluation related to regional analysis of economic measures. In this paper, we cover this gap by estimating a SCOL for Colombia for the first time.

The estimation of a SCOL is highly conditioned by the availability of prices and quantities consumed in different spatial units. For the case of Colombia, the Integrated Household Survey of Income and Expenses 2006-2007 provides information about the expenditure and socioeconomic characteristics of the households for the main 23 Colombian cities. However, in terms of expenditures, the only consumption group with disaggregated prices and quantities is the food group. For the other groups (housing, transportation, clothing, etc), the price can be recovered, but Deaton (1987) argues that those magnitudes do not represent the unit price because the heterogeneity in quality is ignored. In order to obtain a consistent proxy of the SCOL, we compute it using only the information for food consumption which shows unit prices and quantities consumed per household. We are aware that our proxy does not represent the total cost variation, but the food basket is a crucial welfare measure that, together with housing would constitute a significant component of household expenditures.[1]

From the theoretical perspective, the SCOL is defined as the expenditure ratio generated by two price levels given a constant level of utility (Polak, 1971). For its estimation, the specification of the expenditure function must be consistent with the microeconomic foundations of the consumer theory. We follow this microeconomic approach estimating an Almost Ideal Demand System (AIDS) with a demographic component which has been widely accepted as a flexible and consistent theoretical framework (Ray, 1983, Cooper and McLaren, 1992). The demographic component allows to consider the different expen-

[1]Moreover, our methodology can be perfectly extended for the rest of the goods. However, the lack of information is a key constraint.

ditures patterns among the population, which could seriously affect the assumption of a homogeneous consumer or the non-homotheticity issue. While the literature has estimated the AIDS for recovering price and income elasticities in Colombia as a whole (see Barrientos, 2009; and Cortes and Perez, 2010), this is the first time that the AIDS is considered to estimate the SCOL in this country.

Although a microeconomic approach of the SCOL has not been estimated for Colombia, the National Department of Statistics (DANE) computes an alternative economic measure known as Regional Price Index (RPI).[2] The RPI is useful to evaluate how the prices change across time for a particular spatial unit, but it does not allow the comparison of differences in the cost of living across the space. In contrast, the SCOL reflects how much more expenditure an individual needs in region b, to maintain his same level of utility as in region a given the migration of the consumer from a to b. By keeping the utility constant, an estimation of the SCOL facilitates the comparison of the differences in the cost of living across cities (Koo et al., 2000). In summary, the SCOL provides a different perspective than the RPI, for interregional comparisons.

Notwithstanding the advantages derived from the microeconomic approach, the estimation is highly conditioned by the censored data. In consumption surveys, the censored data are represented by those households who declare a zero consumption. These observations play an important role on the estimation of food consumption, and a potential selection bias can appear if they are ignored. Shonkwiler and Yen (1999) propose a two-step procedure which is consistent with the estimation of the AIDS. We use this methodology to avoid the potential bias generated by zero-consumption observations.

Our estimated SCOL can be used in several applications where the RPI cannot provide a useful guide. For example, it can be used to understand the link between real wages and migration or commuting flows among spatial units. Both types of flows are affected by the real purchasing power, more than by the nominal wages. Our SCOL index helps to estimate this real measure, providing key information for regional analysis of migration. It can also be used for computing regional poverty rates. While the poverty line is generally constant across regions, we can use the SCOL for computing the minimum expenditure

[2] We refer like "microeconomic approach" every time when the assumption of fixed utility is maintained instead of fixed baskets.

to reach a level of utility which can be spatially variable. Strictly speaking, our SCOL measure can be used to make real comparisons of monetary variables among spatial units providing a better framework for the design of regional policies.

The paper proceeds as follows. Section 2 provides a review of the literature and the different methodologies used to calculate cost of living indexes. Section 3 explains the methodology to calculate our SCOL index using a microeconomic approach, and section 4 describes the data used for our analysis. The results and comparison with two different indexes are shown in section 5. Finally, section 6 gives some concluding remarks describing the limitations and providing some insights about further applications of this methodology.

2. LITERATURE REVIEW

The SCOL is an economic measure not available for most of the developing countries and the situation is not different even for developed regions such as Europe (Kosfeld et al., 2008). For the case of Colombia, the available information is grounded in the evolution of regional prices using the RPI perspective. Baron (2005) analyzes the inflation level for the main seven cities in Colombia. This approach represents the time evolution of local prices, but it does not allow comparison across units at one period of time. The author suggests a homogeneous inflation process for these cities and the convergence hypothesis is not rejected. Baron's work does not provide, however, a multilateral comparison of expenditure among different spatial units.

In the case of Colombia, a first approach to develop a SCOL is discussed by Romero (2005). The paper proposes a measure to evaluate the expenditure differentials among twelve cities using an axiomatic approach, and Bogota, Barranquilla, Cartagena and Medellin are identified as the group of cities with the highest costs of living. The axiomatic approach prices the same basket in two different regions, assuming a zero substitution effect even when the prices can be extremely different. The axiomatic approach has several problems when analyzing spatial units. For instance, it implies that every consumer does not substitute the consumed basket even when the prices can be strongly affected by local conditions such as climatic characteristics, culture or land availability. This problem is called substitution bias and it is not known how big it is for the case of Colombia.[3]

A second strand of literature estimates a consistent microeconomic demand system for Colombia, but without recovering the expenditure function. Cortes and Perez (2010) estimate three demand systems to recover the income and price elasticity for different groups of goods. The elasticities derived by the authors are relevant for policy and welfare analysis, but they are not helpful to compare the expenditure among spatial units. Barrientos (2009) uses a different approach to characterize the urban demand through the estimation of Engel curves, but the author does not make any inference about price differentials.

Although the literature has not estimated a SCOL for Colombia, some studies are available for Chile. Paredes and Aroca (2008) and Paredes (2011) estimate the cost differential among spatial units consuming the same housing basket. The authors highlight the contribution of "multilateral" comparison of regional expenditures. However, they also recognize the weakness of the axiomatic approach, showing that a homogeneous housing basket is a strong assumption of the consumer behavior. This critique also motivates the use of an AIDS for Colombia, where substitution is allowed among spatial units through the estimation of a demand system.

The substitution effect has not been considered in the estimation of the SCOL in Colombia. Romero (2005) contributes in this direction, but assumes a fixed basket structure which is highly restricted from the spatial perspective. Some additional estimations of demand systems have incorporated this economic consistency, but they do not explore the connections with a SCOL. Motivated by this lack of discussion, the present paper proposes a microeconomic theoretical background to carry out a better estimation of the SCOL.

3. MODEL AND METHODOLOGY

The estimation of a SCOL requires the definition of a expenditure function which must be consistent with microeconomic theory. We estimate our expenditure function

[3]A extensive literature has discussed this bias such as Braithwait (1980), Kokoski (1987), Boskin et al. (1998) and Polak (1998)

using an AIDS with equivalence scale (Deaton and Muellbauer, 1980, Ray, 1983). This approach has several advantages over related models such as the linear and the translog models which impose straight Engel curves for all different income levels of households. The first advantage is that it is more flexible, allowing as many free parameters as there are independent economic parameters (such as the cross-price and income elasticities of demand). Secondly, it considers non-homothetic preferences for each of the household income groups.

By using an equivalence scale estimation to represent the differences in expenditure functions according to socio-demographic variables, we avoid the problem of aggregation over consumers. This approach also represents the spatial heterogeneity that exists among different spatial units. The AIDS model assumes that the preferences of a rational consumer are represented by the following expenditure function:

$$c(p,u) = (1-u)\log(a(p)) + u\log(b(p)), \tag{1}$$

where

$$\log(a(p)) = \alpha_0 + \sum_{i=m}^{m} \alpha_i \log p_i + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \gamma_{ij} \log p_i \log p_j, \tag{2}$$

and

$$\log(b(p)) = \log(a(p)) + \beta_0 \prod_i p_i^{\beta_i}. \tag{3}$$

Both, namely $\log(b(a))$ and $\log(b(p))$, are homogeneous of degree one in prices satisfying the theoretical restrictions of the expenditure function. In the empirical estimation, the expenditure function is recovered from the estimated shares. According to Shephard´s lemma, the shares are just the derivatives of the expenditure function with respect to prices. Multiplying this derivative by $p_i/c(u,p)$, the estimable shares are defined as:

$$s_i = \alpha_i + \sum_{j=1}^{m} \gamma_{ij} \log p_j + \beta_i(\log w - (\alpha_0 + \log a)), \tag{4}$$

where α, β and γ are parameters of the model; s_i is the budget share of good i; p_i is the price of good i; and w is total expenditure. Following the methodology proposed by Ray (1983), we include a general equivalence scale to control for demographic differences between households and spatial units. The equivalence scale ($m_o(z, p, u)$) is disaggregated into two elements: a basic element $\overline{m}_0$ which is constant across price distributions and utility, and a price and utility-varying component φ:

$$m_o(z, p, u) = \overline{m}_0(z)\varphi(p, z, u). \tag{5}$$

The specification of both terms, $\overline{m}_0$ and φ, needs to be consistent with preference-consistent demand models without affecting the theoretical restrictions of the expenditure function. The form of ϕ which is consistent with the specification of the AIDS model is the following:

$$\varphi(z, p, u) = \exp\left(u \prod_j p_j^{\beta_j} \left\{ \prod_j p_j^{\theta_{1j}z_1 + \theta_{2j}z_2} - 1 \right\} \right). \tag{6}$$

By including the equivalence scale in the AIDS estimable shares, equation 4 is modified as follows:

$$\begin{aligned} s_i = \quad & \alpha_i + \textstyle\sum_{j=1}^m \gamma_{ij} \log p_j + (\beta_i + \theta_{i1}z_1 + \theta_{i2}z_2 + \theta i3 z_3)(\log w \\ & -(\alpha_0 + \log(1 + \rho_1 z_1 + \rho_2 z_2 + \rho_3 z_3) + \log a)) + \textstyle\sum_c d_c city_c, \end{aligned} \tag{7}$$

where θ and ρ are the new demographic parameters, and z_1, z_2 and z_3 are the demographic components of each household. Dummy variables ($city_c$) for the 23 cities are also included in the model in order to highlight geographical differences in expenditure levels. The β parameters give information about the characteristics of the goods with respect to income level. If $\beta_i > 0$, an increase in the expenditure would increase the budget share of i, then, the good i is a luxury. On the contrary, if $\beta_i < 0$, the good i is a necessity. γ parameters measure the changes in the budget shares following a change in the relative prices.

The AIDS model satisfies restrictions of adding-up, homogeneity and symmetry: it adds up to total expenditure (the sum of the budget shares is equal to the total expenditure), it is homogeneous of degree zero in prices and total expenditure, and the total expenditure satisfies the Slutsky symmetry. These theoretical restrictions above are imposed through the linearity of the parameters in the following way:

$$\sum_{i=1}^{n} \alpha_i = 1, \qquad \sum_{i=1}^{n} \gamma_{ij} = 0, \qquad \sum_{i=1}^{n} \beta_i = 0, \qquad \sum_{j=1}^{3} \theta_{ij} = 0, \tag{8}$$

$$\sum_{j} \gamma_{ij} = 0, \tag{9}$$

$$\gamma_{ij} = \gamma_{ji}. \tag{10}$$

The second consideration for building our SCOL is related to the bias produced by the households that reported zero consumption. According to Urzua (2010), zero consumption in the household surveys is due to two reasons. The first one is because of the existence of corner solutions, when the good is too expensive for the household to buy it. The second one is when the household only buys the good infrequently or does not buy it at all. Both reasons are approached by different techniques. In the first case, the bias is reduced by including censored data in the estimation of the budget shares (see Heien and Wessels, 1990; and Shonkwiler and Yen, 1999) while in the second one the problem is solved by using income as an instrumental variable of total spending (Keen, 1986).

Perales and Chavas (2000) show that, for the case of Colombian urban households, the reason why the survey reports zero consumption is because of corner solutions. The authors analyze the distribution of the zero expenditures by income class and within income groups and conclude that the zero shares are explained by non-consumption by some household groups. Following this suggestion, we follow the two-step method proposed by Shonkwiler and Yen (1999) to address this problem. In the first step, the binary variable of the decision of consumption is regressed as a function of demographic and socioeconomic characteristics of the household, and estimated using probit models for each consumption group. Then, this regression is used to estimate the cumulative (Φ) and the density (ϕ) probability functions. Finally, the cumulative probability function is included as a scalar multiplying the non-linear part of the estimable share in equation 7; and the density function enters as an extra variable in the estimation in a linear way. The modified estimable share with censored data is the following:

$$\begin{aligned} s_i = \quad & \Phi(x)[\alpha_i + \textstyle\sum_{j=1}^{m} \gamma_{ij} \log p_j + (\beta_i + \theta i 1 z_1 + \theta_{i2} z_2 + \theta i 3 z_3)(\log w \\ & -(\alpha_0 + \log(1 + \rho_1 z_1 + \rho_2 z_2 + \rho_3 z_3) + \log a))] + \textstyle\sum_c d_c city_c + \delta\phi(x), \end{aligned} \tag{11}$$

where δ is an extra parameter of the model with no restrictions. In order to maintain the additivity restriction of the shares, the system estimates $n-1$ equations, where n is the number of shares, and the last share is recovered as a residual of the $n-1$ shares. Once the demand system is estimated, the SCOL is calculated by recovering the expenditure function of a representative household for each city.[4] Finally, the SCOL between i and j is computed by:

$$SCOL_{ij} = \frac{c(\overline{p}_i, u)}{c(\overline{p}_j, u)}, \tag{12}$$

where $\overline{p}_i$ is the price paid by a representative consumer in the spatial unit i and j, respectively.

4. DESCRIPTION OF THE DATA

The data are obtained from the Integrated Household Survey of Income and Expenses (GEIH) of Colombia conducted at the national level in rural and urban areas in 2006-2007. The GEIH provides information at the household and individual level for consumption expenses, income, demographic and labor characteristics. However, the consumption data are available just at the household level with 41,118 observations. From this number of observations, 29,458 (72%) are urban households from the main 23 cities in the country: Bogota, Florencia, Neiva, Tunja, Villavicencio, Medellin, Monteria, Quibdo, Cali, Pasto, Popayan, Barranquilla, Cartagena, Rioacha, Santa Marta, Sincelejo, Valledupar, Bucaramanga, Cucuta, Manizales, Armenia, Ibague and Pereira. We select the urban group to focus on the price differential when the consumer can access to a similar set of goods,

[4] Generally, the literature of demand system uses the median consumer.

assumption that could not be mantained if we allow rural and urban households in our sample. The summary of statistics is shown in table I for the aggregated 23 cities, revealing the differences between the richest city, Bogota, and the poorest city, Quibdo.

The average stratification level is higher in Bogota than in the country, while it is lower in Quibdo. An important difference is also observed in the level of income. The mean monthly income for the total 23 cities is \$644,745 pesos, corresponding approximately to US\$336 dollars.[5] This figure increases to \$867,083 pesos in Bogota (US\$451 dollars), and decreases to \$643,494 pesos (US\$335 dollars) in Quibdo. If we compare the household head income with the total food expenditure per household, we observe that the food expenditure on average corresponds to 38.26% of the total household head income in the 23 cities; this proportion varies from 34.53% in Bogota to 27.60% in Quibdo. On average, the cities that spent the highest proportion of household head income in food consumption are Cartagena (62.01%), Sincelejo (58.87%) and Pasto (51.57%). The lowest proportions are observed in Armenia (28.28%), Medellin (27.70%) and Quibdo (27.60%).

The AIDS estimation requires the consumption and prices of goods, expenditure and socioeconomic characteristics of the household for the demographic component. Although the price is not asked in the questionnaires, the quantity bought and the total cost are used to derive the unit prices. The survey has 284 food items that we categorize in seven groups following the classification recommended by the International Standard Industrial Classification of Economic Activities (ISIC) provided by the U.N.: (1) breads and cereals, (2) vegetables and carbohydrates, (3) fruits, (4) meats, (5) eggs, milk and oils, (6) other foods, and (7) food consumed outside home (restaurants, precooked meals, fast food).[6] Following Urzua (2010), the weighting factors for each of the nine groups are calculated for each household as:

$$a_{jh} = \frac{w_{jh}}{W_{ih}}, \tag{13}$$

where w_{jh} is the expenditure of household h in the individual food item j (where $j = 1 \ldots 7$)

[5] One American dollar is approximately \$1,920 Colombian pesos at December, 2010.

[6] We further aggregate some of the groups to decrease the number of zero observations in some of the food items. Carbohydrates and vegetables are considered one group, and red meats are collapsed with seafood products.

that belongs to group i (where $i = 1...284$), and W_{ih} is the total expenditure of household h in group i. Using these weights and the unit prices derived for each food item, the composite price of group i is calculated as:

$$P_i = p_1^{a_1} p_2^{a_2} \dots p_n^{a_n}. \quad (14)$$

This is the price of group i used in the estimation of the AIDS model. The composite expenditure of group i is the sum of the expenditures of each of the goods j that belong to group i. The budget shares are easily estimated by dividing the expenditure of each of the groups over the total expenditure of the household. Finally, the demographic component includes three variables: education, income stratification level and number of persons.

5. RESULTS

Table II shows the results of the Probit estimation for the first step of the Shonkwiler and Yen (1999) methodology. In the estimation, the binary consumption decision is regressed as a function of income and demographic variables of the household head and his/her partner. The variables included are expenditure, household stratification level, number of persons in the household, number of employed persons, sex, skill, household head age, age of partner, household head income, education of household head and fixed effects by city. Seven different estimations are included, one for each of the consumption groups considered in the AIDS estimation.

The higher the income stratification of households, the lower the probability of food consumption of households for every food group. A negative relationship with food consumption is also observed in households with higher number of occupied persons, households with a male household head, skilled household head, and household head age and education. These results are significantly different from zero, and consistent with microeconomic theory. The results are used to calculate the cumulative and the density probability functions which are used later on the estimation of the AIDS (See $\Phi(x)$ and $\phi(x)$ in 11).

The AIDS estimates the budget shares for each of the consumption goods using prices, income, demographic characteristics, city fixed effects, and the cumulative and density functions as dependent variables. The results of the estimation are shown in table III. Most of the parameters are highly significant at the 1% level. The expenditure and the utility functions for each individual household can be recovered from the parameters as a function of prices and income. The literature suggests recovering the expenditure and utility functions of the median households and using them as representatives of the population. Then, the functions are recovered for the median household of each of the 23 cities considered.

The median expenditures for each city are used for the calculation of the SCOL. The results are shown in tables IV and V. Figure 1 is used to show the spatial distribution of the cities and the differences between the different SCOL. As expected, Bogota is the city with the highest expenditure and the highest cost of living, followed by Armenia, Cali, Bucaramanga and Ibague. Armenia shows a cost of living 16.5% lower than the cost of living of Bogota followed by Cali, Bucaramanga, Ibague and Tunja, all of them within the 70% of the Bogota food expenditure. A second group, with a cost of living between 50% and 70% of Bogota expenditures is composed by Villavicencio, Sincelejo, Cartagena, Medellin, Neiva, Santa Marta, Manizales, Cucuta, Popayan, Barranquilla and Florencia. Finally, cities with differences greater than 50% are Monteria, Pereira, Pasto, Valledupar, Rioacha and Quibdo. It is important to bear in mind that our SCOL measures only consumption of food. Then, even if some Colombian cities are very expensive in terms of real estate values or other expenditures, these differences are not considered here. However, our results are in the same line of the regional science literature, where the big cities tend to show higher inflationary pressures with higher levels of cost of living (Sudekum, 2009).

Our SCOL can be compared to other methodologies. The first one is the axiomatic methodology to calculate a fixed basket of goods. According to Polak (1971), the cost of living index is the ratio between two expenditure functions evaluated at two different level of prices. The quantities consumed are different because the households substitute consumption in order to maintain the same level of utility before the prices change. This economic approach is too difficult to estimate because it requires the specification of the expenditure and the utility function and the estimation of a demand system. Then, the axiomatic approach, which is commonly used for the estimation of the cost of living is preferred because of its simplicity. In this approach, the utility function is assumed to have a Leontief functional form, and substitution is not allowed even with changes in prices. The household maintains a fixed basket of good and the ratio between expenditures is calculated using the same basket.

The second methodology is proposed by Romero (2005) and compares the minimum expenditure in each city evaluated at prices of the other 13 main cities to calculate the regional differences. Romero (2005), as noted in section 2 also proposes an axiomatic approach with a fixed basket evaluated at different prices to calculate cost of living differences for 12 Colombian cities. His methodology is more complexed than the one explained above because it includes a labor participation component in which the individual optimizes time spent on working and leisure, and chooses its consumption patterns according to this rational behavior. However, the author does not solve the substitution bias because the same basket is used for different cities, and it does not take into consideration differences on regional consumption patterns.

Table VI shows the cost of living indexes using the three different methodologies. Our estimates are significantly different than the ones obtained using the axiomatic approaches suggesting that the geographical cost-of-living differentials among Colombian cities depend, not only on different price levels among cities, but also on different consumption patterns and substitution among different goods. When our SCOL (which is calculated using the economic approach) is compared to the axiomatic approach of the fixed basket of goods, the cost of living is similar for the most expensive cities. However, this difference increases when Bogota is compared to the poorest cities. In this sense, we can assume that the baskets of goods are very similar for the expensive cities, but when comparing expensive cities with poorer cities, the consumption patterns change and the bias in the index increases.

Figure 2 shows the bias generated by using the axiomatic approach for calculating the cost of living of different cities. The bias is very small when two expensive cities are compared, which is the case of Cali, Ibague, Sincelejo and Cartagena. But when Bogota is compared to poorer cities such as Pasto, Valledupar, Rioacha and Quibdo, this bias increases significantly, showing the difference in consumption patterns between different spatial units.

6. CONCLUSIONS

This paper proposes a Spatial Cost of Living Index for the case of Colombia using the food consumption. We propose the estimation of an Almost Ideal Demand System in order to keep constant the level of utility (economic approach) of the consumer instead of a fixed basket (axiomatic approach). Furthermore, the estimation of the AIDS considers the potential bias generated by the zero consumption of some of the households, a typical characteristic of the expenditure surveys. Using this methodology, the expenditure ratio is calculated for two representative consumers in two different spatial units. This approach allows capturing the substitution the consumer makes when facing different price levels, which is one of the most essential characteristic of consumer theory.

While most of the literature has computed cost of living measures across time, we discuss and put in evidence the complex scenario imposed by the geographical space. From the temporal perspective, the axiomatic approach, the most standard methodology for estimating price indexes, evaluates a fixed consumption basket between two different periods. In this sense, the literature recognizes that the prices change slowly across time assuming that the substitution effect should be minimal. Using this argument and the fact that this approach can be estimated easily, most of the official indexes are calculated using a fixed basket of goods.

However, the space can affect considerably the price of goods even for the same period of time. How can the price index be supported with a fixed basket of goods when the spatial structure imposes considerable price differentials? These differential can be explained by different transportation infrastructures, different climatic conditions or different cultures. Some regions require very particular set of goods, say high level of proteins for extreme cold places, versus those hot regions where fruits or vegetable can be more relevant. Considering these arguments, this paper opens the discussion about the significant relevance of the substitution effect when we are comparing expenditures across the space instead of time.

The empirical results suggest that Bogota is the most expensive city regarding food consumption. This result is a stylized fact. In general the spatial units with high level of population are characterized by higher level of prices (Sudekum, 2009). The city with the cheapest basket is Quibdo, showing a difference of 50% with respect to Bogota. The SCOL presents a high level of heterogeneity among cities, providing evidence of higher substitution processes among space than through time.

The future research agenda contains different areas of research. First, our SCOL is seriously limited by the availability of information. The incorporation of additional consumption groups would help to considerable improve the quality of the estimation. A crucial item is housing, which is considered by the literature like a good which represents properly the real measure of cost of living. According to Timmins (2006), the housing is a non-transable good and it absorbs the price pressure of each spatial unit. A second extension is related to the use of the SCOL for improving the inter-regional comparison of monetary variables. The spatial wage inequality of welfare analysis could suffer drastic variations if the estimation controls by spatial differences on the economic variables.

Table I: Summary of statistics at the national level and by two cities: Bogota and Quibdo.

Variable	Mean	Standard Deviation	Minimum	Maximum
Stratification level	2.32	1.07	0.00	6.00
Number of persons	3.92	1.97	1.00	22.00
Head age	48.25	15.16	13.00	97.00
Marital status	0.91	0.27	0	1.00
Partner's age	37.10	16.88	0	99.00
HH head income	644,745	896,939	0	16,700,000
Education	3.74	1.78	0	6.00
Sex	0.63	0.48	0	1.00
Food expenditure	246,680	205,747	433	3,751,945
Bogota				
Stratification level	2.65	0.95	1.00	6.00
Number of persons	3.56	1.71	1.00	13.00
Head age	47.40	15.55	18.00	91.00
Marital status	0.90	0.30	0	1.00
partner's age	37.63	16.97	0	92.00
HH head income	867,083	1,302,025	0	14,600,000
Education	4.09	1.67	0	6.00
Sex	0.66	0.47	0	1.00
Food expenditure	299,381	255,185	4,330	2,025,141
Quibdo				
Stratification level	1.39	0.69	0.00	4.00
Number of persons	4.24	2.19	1.00	13.00
Head age	46,14	15.28	17.00	87.00
Marital status	0.92	0.27	0	1.00
partner's age	31.70	16.83	0.00	95.00
HH head income	643,494	870,635	8.16	8,000,000
Education	3.79	2.01	0	6.00
Sex	0.47	0.49	0	1.00
Food expenditure	177,628	149,013	866	1,076,438

Table II: Probit estimation for censored consumption.

	Share 1	Share 2	Share 3	Share 4	Share 5	Share 6	Share 7
Expenditure	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
Stratification level	-0.142***	-0.212***	-0.053***	-0.174***	-0.099***	-0.136***	-0.012***
Number of persons	0.105***	0.101***	0.018***	0.077***	0.113***	0.057***	-0.035***
nocupados	-0.240***	-0.230***	-0.120***	-0.181***	-0.236***	-0.151***	0.280***
Male	-0.104***	-0.031***	-0.065***	0.027***	-0.105***	-0.072***	0.218***
Skill	-0.027***	-0.163***	-0.083***	-0.156***	-0.046***	-0.151***	0.005*
Husban age	0.001***	-0.003***	0.001***	-0.002***	0.001***	-0.003***	-0.002***
Head age	-0.002***	0.003***	0.001***	-0.001***	0.001***	-0.005***	-0.006***
Head income	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000*
Education	-0.065**	-0.109***	0.015***	-0.082***	-0.017***	-0.104***	0.091***
Bogota	-0.930***	-0.004	-0.128***	0.113***	-0.245***	0.007	0.583***
Florencia	-0.346***	-0.170***	0.188***	-0.098***	-0.191***	0.343***	0.601***
Neiva	0.492***	-0.221***	-0.283***	0.304***	0.395***	0.036***	0.385***
Villavicencio	-0.136***	-0.001	0.063***	0.019**	-0.228***	0.274***	0.092***
Medellin	0.156***	-0.133***	-0.291***	-0.134***	0.201***	0.201***	0.073***
Monteria	-0.104***	0.300***	0.082***	0.362***	-0.210***	0.682***	-0.456***
Quibdo	-0.731***	0.412***	0.147***	0.085***	-0.554***	0.358***	-0.105***
Cali	-0.450***	0.126***	0.226***	0.018***	-0.087***	0.413***	0.870***
Pasto	0.623***	0.258***	0.189***	-0.063***	0.372***	0.288***	-0.603***
Popayan	-0.197***	-0.021**	0.093***	0.031***	0.180***	0.463***	0.004
Barranquilla	-0.186***	0.029***	0.322***	-0.117***	-0.112***	0.387***	0.157***
Cartagena	0.357***	0.635***	0.577***	0.227***	0.377***	0.774***	0.168***
Riohacha	-0.492***	0.018	0.311***	0.167***	-0.298***	0.436***	0.095***
Santa Marta	0.016*	0.497***	0.411***	0.254***	0.098***	0.635***	0.321***
Sincelejo	0.238***	0.668***	0.774***	0.340***	0.319***	1.021***	0.333***
Valledupar	-0.282***	0.319***	0.153***	-0.215***	0.174***	0.557***	-0.074***
Bucaramanga	0.184***	-0.028***	0.201***	-0.098***	0.131***	0.344***	0.204***
Cucuta	-0.159***	0.050***	0.299***	-0.038***	-0.167***	0.489***	0.276***
Manizales	0.143***	0.018**	-0.475***	0.025***	0.161***	0.205***	0.043***
Armenia	0.397***	.286***	-0.053***	0.253***	0.118***	0.336***	0.385***
Ibague	0.067***	.097***	0.084***	0.094***	0.100***	0.253***	-0.094***
Pereira	-0.549***	.138***	-0.310***	-0.136***	-0.114***	0.126***	0.047***
Constant	1.564***	1.308***	-0.342***	1.174***	1.090***	0.866***	-0.455***

1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.

2) The number of observations is 3406666 (weighted).

Table III: Coefficients of the Almost Ideal Demand System.

City FE	Coeff.		Coeff.		Coeff.		Coeff.		Coeff.		Coeff.
$city_1$	0.005***	α_1	0.304***	γ_{11}	0.016***	θ_{11}	-0.001***	δ_1	0.007***	ρ_1	-0.005
$city_2$	0.019***	α_2	0.155***	γ_{12}	0.003***	θ_{21}	-0.002***	δ_2	0.164***	ρ_2	0.816***
$city_3$	0.006***	α_3	0.055***	γ_{13}	0.002***	θ_{31}	0.004***	δ_3	0.076***	ρ_3	-0.224***
$city_4$	0.012***	α_4	0.097***	γ_{14}	-0.004***	θ_{41}	-0.001***	δ_4	0.259***		
$city_5$	0.005***	α_5	0.297***	γ_{15}	0.001***	θ_{51}	0.002***	δ_5	0.026***		
$city_6$	0.014***	α_6	0.119***	γ_{16}	-0.005***	θ_{61}	-0.000***	δ_6	0.032***		
$city_7$	0.009***	β_1	-0.038***	γ_{22}	0.016***	θ_{12}	0.001***				
$city_8$	0.018***	β_2	0.003***	γ_{23}	-0.003***	θ_{22}	-0.001***				
$city_9$	0.016***	β_3	-0.003***	γ_{24}	-0.018***	θ_{32}	-0.002***				
$city_{10}$	0.017***	β_4	0.031***	γ_{25}	0.005***	θ_{42}	0.001***				
$city_{11}$	0.015***	β_5	-0.033***	γ_{26}	0.003***	θ_{52}	-0.002***				
$city_{12}$	0.011***	β_6	0.004***	γ_{33}	0.023***	θ_{62}	-0.002***				
$city_{13}$	0.007***			γ_{34}	-0.015***	θ_{13}	0.001***				
$city_{14}$	0.010***			γ_{35}	-.003***	θ_{23}	-.002***				
$city_{16}$	0.012***			γ_{36}	-0.001***	θ_{33}	0.000***				
$city_{17}$	0.013***			γ_{44}	0.055***	θ_{43}	-0.001***				
$city_{18}$	0.002***			γ_{45}	-0.009***	θ_{53}	0.002***				
$city_{19}$	0.008***			γ_{46}	-0.000**	θ_{63}	0.000***				
$city_{20}$	0.011***			γ_{55}	0.016***						
$city_{21}$	0.009***			γ_{56}	0.003***						
$city_{22}$	0.011***			γ_{66}	0.008***						

1) *, ** and *** represent the level of significance to 10%, 5% and 1%, respectively.

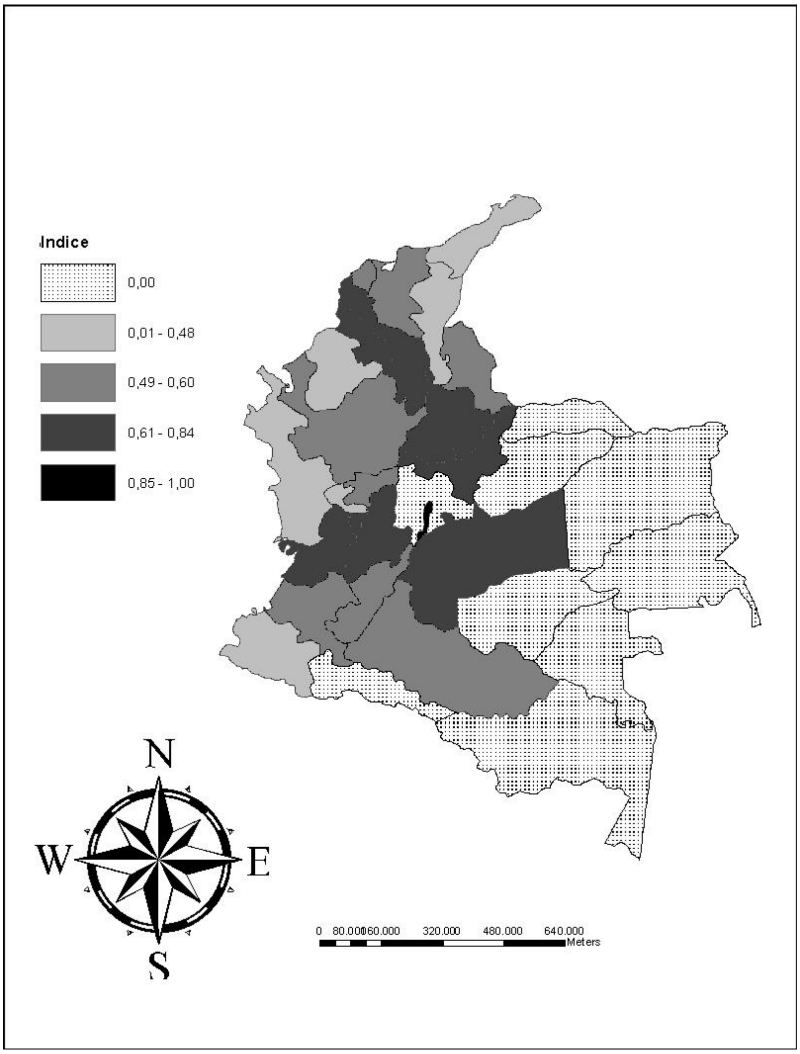

Figure 1: SCOL for the main Colombian 23 cities

Table IV: Food expenditure for the 23 main Colombian cities

	Bogota	Armenia	Cali	Bucaramanga	Ibague	Tunja	Villavicencio	Sincelejo	Cartagena	Medellin	Neiva
Bogota	1.00	0.84	0.80	0.77	0.75	0.71	0.69	0.67	0.66	0.60	0.58
Armenia	1.20	1.00	0.95	0.93	0.89	0.85	0.82	0.80	0.78	0.72	0.70
Cali	1.25	1.05	1.00	0.97	0.94	0.89	0.86	0.84	0.82	0.76	0.73
Bucaramanga	1.29	1.08	1.03	1.00	0.97	0.92	0.89	0.86	0.85	0.78	0.75
Ibague	1.34	1.12	1.07	1.04	1.00	0.95	0.92	0.90	0.88	0.81	0.78
Tunja	1.40	1.17	1.12	1.09	1.05	1.00	0.96	0.94	0.92	0.85	0.82
Villavicencio	1.46	1.22	1.16	1.13	1.09	1.04	1.00	0.98	0.96	0.88	0.85
Sincelejo	1.49	1.25	1.19	1.16	1.12	1.06	1.03	1.00	0.98	0.90	0.87
Cartagena	1.52	1.27	1.22	1.18	1.14	1.09	1.05	1.02	1.00	0.92	0.89
Medellin	1.66	1.39	1.32	1.29	1.24	1.18	1.14	1.11	1.09	1.00	0.97
Neiva	1.71	1.43	1.37	1.33	1.28	1.22	1.17	1.15	1.12	1.03	1.00
Santa Marta	1.74	1.45	1.39	1.35	1.30	1.24	1.19	1.16	1.14	1.05	1.02
Manizales	1.78	1.49	1.42	1.38	1.33	1.27	1.22	1.19	1.17	1.07	1.04
Cucuta	1.80	1.50	1.43	1.39	1.34	1.28	1.23	1.20	1.18	1.08	1.05
Popayan	1.83	1.53	1.46	1.41	1.37	1.30	1.25	1.22	1.20	1.10	1.07
Barranquilla	1.92	1.60	1.53	1.49	1.44	1.37	1.32	1.29	1.26	1.16	1.12
Florencia	1.93	1.62	1.54	1.50	1.45	1.38	1.33	1.29	1.27	1.17	1.13
Monteria	2.06	1.72	1.64	1.60	1.54	1.47	1.42	1.38	1.35	1.24	1.20
Pereira	2.21	1.85	1.76	1.71	1.65	1.57	1.52	1.48	1.45	1.33	1.29
Pasto	2.33	1.95	1.86	1.81	1.74	1.66	1.60	1.56	1.53	1.40	1.36
Valledupar	2.51	2.09	2.00	1.94	1.87	1.78	1.72	1.68	1.64	1.51	1.46
Riohacha	2.53	2.12	2.02	1.96	1.89	1.80	1.74	1.69	1.66	1.53	1.48
Quibdo	2.69	2.25	2.14	2.08	2.01	1.91	1.85	1.80	1.76	1.62	1.57

Table V: Food expenditure for the 23 main Colombian cities (Continuation)

	Santa Marta	Manizales	Cucuta	Popayan	Barranquilla	Florencia	Monteria	Pereira	Pasto	Valledupar	Riohacha	Quibdo
Bogota	0.58	0.56	0.56	0.55	0.52	0.52	0.48	0.45	0.43	0.40	0.39	0.37
Armenia	0.69	0.67	0.67	0.66	0.62	0.62	0.58	0.54	0.51	0.48	0.47	0.44
Cali	0.72	0.70	0.70	0.69	0.65	0.65	0.61	0.57	0.54	0.50	0.50	0.47
Bucaramanga	0.74	0.72	0.72	0.71	0.67	0.67	0.63	0.58	0.55	0.52	0.51	0.48
Ibague	0.77	0.75	0.74	0.73	0.70	0.69	0.65	0.61	0.57	0.53	0.53	0.50
Tunja	0.81	0.79	0.78	0.77	0.73	0.73	0.68	0.64	0.60	0.56	0.55	0.52
Villavicencio	0.84	0.82	0.81	0.80	0.76	0.75	0.71	0.66	0.63	0.58	0.58	0.54
Sincelejo	0.86	0.84	0.83	0.82	0.78	0.77	0.72	0.68	0.64	0.60	0.59	0.56
Cartagena	0.88	0.86	0.85	0.83	0.79	0.79	0.74	0.69	0.65	0.61	0.60	0.57
Medellin	0.95	0.93	0.92	0.91	0.86	0.86	0.80	0.75	0.71	0.66	0.66	0.62
Neiva	0.98	0.96	0.95	0.94	0.89	0.89	0.83	0.78	0.73	0.68	0.68	0.64
Santa Marta	1.00	0.98	0.97	0.95	0.91	0.90	0.84	0.79	0.75	0.69	0.69	0.65
Manizales	1.02	1.00	0.99	0.98	0.93	0.92	0.86	0.81	0.76	0.71	0.70	0.66
Cucuta	1.03	1.01	1.00	0.98	0.94	0.93	0.87	0.81	0.77	0.72	0.71	0.67
Popayan	1.05	1.03	1.02	1.00	0.95	0.94	0.89	0.83	0.78	0.73	0.72	0.68
Barranquilla	1.10	1.08	1.07	1.05	1.00	0.99	0.93	0.87	0.82	0.77	0.76	0.71
Florencia	1.11	1.09	1.08	1.06	1.01	1.00	0.94	0.88	0.83	0.77	0.76	0.72
Monteria	1.19	1.16	1.15	1.13	1.07	1.07	1.00	0.93	0.88	0.82	0.81	0.77
Pereira	1.27	1.24	1.23	1.21	1.15	1.14	1.07	1.00	0.95	0.88	0.87	0.82
Pasto	1.34	1.31	1.30	1.28	1.21	1.21	1.13	1.06	1.00	0.93	0.92	0.87
Valledupar	1.44	1.41	1.39	1.37	1.30	1.30	1.21	1.13	1.07	1.00	0.99	0.93
Riohacha	1.46	1.42	1.41	1.39	1.32	1.31	1.23	1.15	1.09	1.01	1.00	0.94
Quibdo	1.55	1.51	1.50	1.47	1.40	1.39	1.30	1.22	1.15	1.07	1.06	1.00

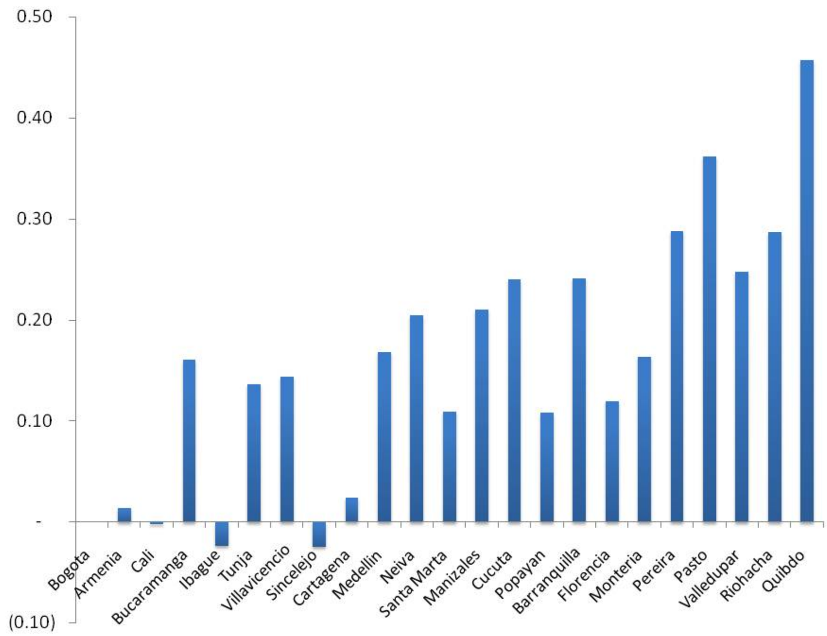

Figure 2: Spatial substiution bias.

REFERENCES

Baron J. 2005. *La inflacion en las ciudades de colombia: Una evaluación de la paridad del poder adquisitivo.* Centro de Estudios economicos regionales, Banco de la Republica de Colombia.: Documento de Trabajo No. 31.

Barrientos J. 2009. On the consumer behavior in urban Colombia: the case of Bogota. *Ensayos sobre politica economia* **27**: 46–82.

Boskin M, Dulberger E, Gordon R, Griliches Z, Jorgenson D. 1998. Consumer prices, the consumer price index, and the cost of living. *Journal of Economic Perspectives* **1**: 3–26.

Braithwait S. 1980. The substitution bias of the Laspeyres price index: an analysis using estimated cost-of-living indexes. *The American Economic Review* **70**: 64–77.

Cooper R, McLaren K. 1992. An empirically oriented demand system with improved regularity properties. *Canadian Journal of Economics* **25**: 652–668.

Cortes D, Perez J. 2010. *El consumo de los hogares colombianos, 2006-2007: estimacion de sistemas de demanda*. Universidad del Rosario.: Documento de Trabajo No. 86.

Deaton A. 1987. Estimation of own- and cross-price elasticities from household survey data. *Journal of Econometrics* **36**: 7–30.

Deaton A, Muellbauer J. 1980. An almost ideal demand system. *American Economic Association* **70**: 312–326.

Heien D, Wessels C. 1990. Demand systems estimation with microdata: A censored regression approach. *Journal of Business & Economic Statistics* **8**: 365–371.

Keen M. 1986. Zero expenditures and the estimation of engel curves. *Journal of Applied Econometrics* **1**: 277–286.

Kokoski M. 1987. Problems in the measurement of consumer cost-of-living indexes. *Journal of Business and Economic Statistics* **5**: 39–46.

Koo J, Phillips K, Sigalla F. 2000. Measuring regional cost of living. *Journal of Business and Economic Statistics* **18**: 127–136.

Kosfeld R, Eckey H, Turck M. 2008. New economic geography and regional price level. *Annals of Regional Science* **28**: 43–60.

Paredes D. 2011. A methodology to compute regional housing price index using matching estimator methods. *Annals of Regional Science* **46**: 139–157.

Paredes D, Aroca P. 2008. Metodología para estimar un indice regional de costo de vivienda en Chile. *Latin American Journal of Economics* **45**: 129–143.

Perales F, Chavas J. 2000. Estimation of censored demand equations from large cross-section data. *American Journal of Agricultural Economics* **82**: 1022–1037.

Polak R. 1971. *The Theory of the Cost of Living Index*. Research Division, Office of Prices and Living Conditions, U.S. Bureau of Labor Statistics: Research Discussion Paper No. 11.

Polak R. 1998. The consumer price index: A research agenda and three proposals. *The Journal of Economic Perspectives* **12**: 69–78.

Ray R. 1983. Measuring the costs of children: An alternative approach. *The Journal of Public Economics* **22**: 89–102.

Romero J. 2005. *¿Cuánto cuesta vivir en las principales ciudades colombianas? Índice de Costo de Vida Comparativo*. Centro de estudios economicos regionales, Banco de la Republica de Colombia.: Documento de Trabajo No. 57.

Shonkwiler J, Yen S. 1999. Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics* **81**: 972–982.

Sudekum J. 2009. Regional costs-of-living wih congestion and amenity differences: an economic geography perspective. *Annals of Regional Science* **43**: 49–69.

Timmins C. 2006. Estimating spatial differences in the Brazilian cost of living with household location choices. *Journal of Development Economics* **80**: 59–83.

Urzua C. 2010. *Notes on the estimation of demand systems*. URL http://economiccluster-lac.org/images/pdf/eventos/Fiscalidad/Mexico23y240310/Notes_on_Demand_Systems.pdf

Table VI: Comparison of cost of living indexes using different methodologies

	SCOL (economic)	**Laspeyres (axiomatic)**	**Romero (2005)**
Bogota	1.000	1.000	1.000
Armenia	0.836	0.849	
Cali	0.797	0.795	0.911
Bucaramanga	0.774	0.935	0.840
Ibague	0.748	0.724	
Tunja	0.712	0.848	
Villavicencio	0.686	0.830	0.911
Sincelejo	0.669	0.645	
Cartagena	0.656	0.680	0.976
Medellin	0.602	0.770	0.936
Neiva	0.584	0.789	
Santa Marta	0.575	0.685	
Manizales	0.561	0.772	0.767
Cucuta	0.556	0.796	
Popayan	0.548	0.655	
Barranquilla	0.521	0.762	0.991
Florencia	0.517	0.637	
Monteria	0.485	0.648	0.741
Pereira	0.453	0.740	0.897
Pasto	0.429	0.791	0.759
Valledupar	0.399	0.647	
Riohacha	0.395	0.682	
Quibdo	0.372	0.830	

The index calculated by Romero (2005) is only estimated for 11 cities.