Using Spatial Covariance Function for Antitrust Market Delineation*

Claudio R. Lucinda**
Arthur Barrionuevo Filho***

Abstract

In this paper, an estimation of a Spatial Covariance Function was proposed for determining the relevant geographic market for a merger. This methodology was applied to a proposed merger between two competing Brazilian supermarket chains. During this application, the shortcomings of the analysis carried out by the Brazilian Antitrust System were initially pointed out, including the geographic dimension of the relevant market, which was found to be separated in each municipality located on the shore of São Paulo state. The results, based on the estimated spatial covariance function using price data on 22 products in 43 supermarkets, indicate a single geographic market for all municipalities.

Keywords: Geographic Markets, Spatial Covariance Function, Merger Control.

JEL Codes: K21, C31.

*Submitted in November 2007. Revised in December 2009.

**Assistant Professor, Faculty of Economics, Business Administration and Accounting at Ribeirão Preto, University of São Paulo (USP). E-mail: claudiolucinda@fearp.usp.br

***Professor, São Paulo School of Economics and São Paulo School of Business Administration, Getulio Vargas Foundation (EESP and EAESP/FGV). E-mail: abarrio@fgvsp.br

Brazilian Review of Econometrics
v. 29, n^o 1, pp. 111–135 May 2009

1. Introduction

The analysis of mergers relies, to a great extent, on determining on which markets the resulting entity can exercise potential market power. Furthermore, given the emphasis economists place on the study of markets, it is to be expected that economic theory would have a reasonable reply to the question about determining market boundaries. However, in applied settings, many different techniques have been used, relying on various types of assumptions on the competitive structure of the markets involved, on the statistical properties of the data at hand and on the judgment of the economist. This paper aims to contribute to the debate on how to set the boundaries of geographic markets by presenting a new technique that can be used in the determination of the relevant market boundaries affected by a proposed merger, based on computation of the spatial correlation function, developed in Cressie (1993), Conley and Topa (2002) and Chen and Conley (2001). These papers measure the correlation between realizations of random variables as a function of a distance, measured according to a properly defined metric.

Even though the technique has been used on various subjects, to our knowledge, this paper is the first application of this methodology for the determination of the geographic dimension of the relevant market, focusing on a proposed merger between two competing Brazilian supermarket chains, with price data on 22 products in 43 supermarkets for 14 months. The results were compared to the conclusions of the analysis carried out by the Brazilian Antitrust System. Using the spatial covariance function, the results point out to a single market that encompasses all locations, in contrast to the decision of the Brazilian Antitrust System, which established that the municipalities composing the area should constitute separate markets, based on the area from which 75% of the customers of a given supermarket come.

In order to better define the contribution of this paper to the literature about antitrust market delineation, the starting point of the present paper consists in reviewing the procedures currently used for the determination of boundaries of relevant geographic markets. The most important concept used by antitrust authorities during merger analysis is the market power concept: the capacity a company has to unilaterally increase its prices after a merger operation. In order to make it operational, the first question to be addressed is: what is the relevant market for this merger operation? In other words, if this merger could possibly pose a threat of unilateral exercise of market power, on which market would this power be exercised?[1]

To answer that question, two dimensions of this relevant market are usually considered: the product and the geographic dimensions. In the applied antitrust

[1]This question must be addressed even if the approach used for the evaluation of competitive effects of the merger is based on simulations, as the parameters to be used in this simulation come from a set of assumptions about the relevant markets (i.e., the set of products whose prices are expected to be affected by the merger).

analysis they are currently defined using the so-called SSNIP (*Small, but Significant and Non-transitory Increase in Prices*) test, which implies that the product and geographic dimensions of a market are both the minimum set of substitute products (for the product dimension) and smallest geographic area (for the geographic dimension) which could act as a competitive check on a unilateral increase in prices of a given percentage. By competitive check, it is meant that the company which tries to increase its prices by that percentage finds it unprofitable to do so.[2]

There are a number of techniques available for this task. One of the first attempts to deal with this problem evaluates the *product flows* between different areas; if one does find that the product flows coming from outside a given geographic area are responsible for a small part of the consumption there, one does have evidence against the hypothesis that this area is part of a larger geographic market. The second part of the test – considered by Elzinga and Hogarty (1973, 1978) to be enough for the determination of market boundaries – consists in evaluating whether shipments to outside this geographic area account for a small part of the production carried out there. If so, this region can be considered a separate geographic market. This approach was criticized by Stigler and Sherwin (1985, p. 555), who point out that the existence or not of significant product flows between regions does not imply that they belong to the same market; hence, producers in one region act as a competitive check eliminating persistent price differences. They mentioned the fact that, even if two regions which are in different geographic markets could present large trade flows and, because of price discrimination – say, due to different demand elasticities – one could observe different prices. Furthermore, even if one does not find any product flow between these two regions, this does not imply they are different geographic markets; it is possible that producers in one region pose such a competitive threat to the other market that the same price is obtained in both regions, even though there are no actual shipments between them.

The second approach for the determination of market boundaries tries to estimate the *elasticities*[3] for a given product.[4] In this approach, some authors (such

[2]A survey on this topic is found in Motta (2004, chapter 2).

[3]Elasticity is a measure – adimensional – that gives the percentage change in the demanded quantity of a product in response to a given percentage change in another economic variable. The most common elasticities are the own-price elasticity (measuring the sensitivity of the demanded quantity to changes in its own price), the cross-price elasticity (measuring the sensitivity to changes in the prices of other products) and income elasticity (the sensitivity to changes in income).

[4]Usually, in order to have consistent estimates of elasticities, the econometric best practice recommends setting up a structural model – either explicitly or implicitly, by the choice of instruments. A related literature, not explored in this paper, also relying on structural econometric models, tries to shed some light upon the effects of a merger by simulating the likely effects on relevant variables, such as prices or markups. Examples of this approach are Nevo (2000), relying on estimation of price elasticities, and Werden and Froeb (2002), using calibration of relevant

as Werden, 1998) favor the use of the *own-price* elasticity of demand. Their reasoning, which is derived from the traditional economic theory, is based on the fact that there is a boundary in the own-price elasticity of demand that makes it unprofitable for a producer in a given geographic area to unilaterally increase its prices. That boundary – called critical elasticity of demand – depends on both the margins earned by these products and the hypothetical price increase defined on the SSNIP test. Another point of view on this approach involves the computation of the *elasticity of the residual demand* (Werden and Froeb, 1993, Scheffman and Spiller, 1987). The elasticity of the residual demand differs from the traditional price elasticity of demand in that it already considers all competitive responses from rivals to a given increase in prices by a producer. The empirical implementation of this technique requires both the setup of a structural model for the competition and the determination of a set of identification conditions for the estimation of the aforementioned elasticity.

The final way by which elasticities are used in the determination of relevant market boundaries involves the estimation of *cross-price* elasticities of demand. These estimates, which are expected to measure how much – as a percentage – the demand for a product changes in response to a given increase (also in percentage terms) in another product price, could provide a map of which regions could provide a competitive check on price increases by some producers. For instance, if one does find a positive elasticity between the price of a given product sold in region A and the quantity of the same product sold in region B, it might mean that the producers in region B can act as competitive constraint on the behavior of those in region A.

The "might" in the previous paragraph is due to a potential pitfall presented by all these approaches – the so-called *cellophane fallacy.* These previously discussed estimates are based on prices collected at the moment of the investigation. However, the SSNIP test mentions a hypothetical price increase from the price levels obtained under competitive conditions. As reported by Motta (2004, p. 105), the US Supreme Court determined that the relevant market for a proposed merger involving U.S. DuPont should include cellophane and other flexible wrapping materials, given the high cross-price elasticity of demand between cellophane and these wrapping materials. This decision was criticized on the grounds that such high cross-price elasticity was in itself a result of market power, by which DuPont increased its prices until consumers started to consider other wrapping materials as substitutes.

The third approach to determining the boundaries of relevant markets is based on the concept summarized by Stigler and Sherwin (1985, p. 555):

parameters. This approach tries to sidestep the problem with market definition posed here – in these papers, the relevant market is already defined – by directly focusing on the merger effects. However, this approach also has costs, as the results are affected by specification errors in the identification conditions and of the relevant market.

> "Consider the basic definition of a market: 'A market for a good is the area within the price of a good that tends to uniformity, an allowance made for transportation costs.' If there is a single price (allowing for transportation costs) in a given area, that must mean that either buyers or sellers (or both) can and do consider transactions at any point within the area to be an excellent (in the limit, a perfect) substitute for transactions at other points within the area. Hence, the market area embraces the buyers who are willing to deal with any seller, or the sellers who are willing to deal with any buyer, or both."[5]

This quotation sums up the reason for the usage of *price correlations* as a way to determine the geographic dimensions of a market, as in Stigler and Sherwin (1985) and Horowitz (1981). Considering a set of locations denoted by $\mathbf{S} = \{s_1, s_2, ...s_n\}$ corresponding to the locations of n sellers, and the prices charged by all these sellers collected in a vector $\mathbf{P} = \{P_{s1}, P_{s2}, ..., P_{sn}\}$ and one finds $\mathrm{Cov}(P_{si}, P_{sk}) > 0$, one might expect producers i and k to be in the same market. On the other hand, if we find that $\mathrm{Cov}(P_{si}, P_{sk}) = 0$, this might be interpreted as producers i and k not belonging in the same market.

Slade (1986) points out that correlations may be spuriously high if one does not control for the impact of other factors that affect the price behavior in different locations, or induce stochastic trends in the behavior of prices. The author uses Granger causality tests to address some of these problems. A very thoughtful critique of the usage of price correlation tests based on market analysis in antitrust is given by Werden and Froeb (1993). Their critique is based on the fact that price correlation tests are carried out without regard to an explicit modeling of the consumers' choices, and an informed application of the economic theory that underlies the SSNIP test could present better results; they also recommend the usage of residual demand elasticities as a tool for determining such boundaries.

Sherwin (1993), in a comment on Werden and Froeb's paper in the same journal issue, replies by pointing out that the implementation of the SSNIP test Werden and Froeb (1993) propose is also fraught with difficulties, and the authors do not provide conclusive evidence concerning the superiority of their proposal over price correlation tests. Finally, Sherwin (1993) also says Werden and Froeb (1993) proposal requires much economic analysis in order to identify the competitive structure of the market under scrutiny, so as to build an econometric structural model that obtains estimates of the relevant residual demand elasticities. As Sherwin (1993, p. 356) points out:

[5]The quotation used by Stigler and Sherwin in the excerpt above is from Cournot, and was also used by Marshall.

> "Such an approach ignores the fundamental purpose of the Guidelines in the first place, that is, to give guidance to those contemplating mergers and acquisitions. Indeed, why have the Guidelines at all? Instead, a fact-intensive economic analysis (whatever that means) could simply be used to evaluate directly the ultimate question of whether prices are likely to rise as a result of a merger. Why bother with market delineation and market share calculations, which are, after all, only intermediate inquiries?"

This quotation sums up the point that, even though the setup of an econometric structural model is required for recovering the so-called "deep parameters" of individual behavior, from the point of view of the Antitrust practice, it might not be effective to do so, given the time and data constraints usually found in such settings. Thus, it might be better to use techniques that are less intensive on assumptions derived from an intensive – and time-consuming – analysis of the market conditions to serve as broad guidelines for further analysis. These issues also point to the choice of a technique in this paper – spatial covariance function – for establishing the boundaries of the relevant markets.[6] The theoretical aspects of the estimation of such functions are presented in the next section.

2. Spatial Covariance Function

The responses given by Sherwin (1993) to the criticisms posed by Werden and Froeb (1993) point to a criterion for the establishment of a relevant geographic market for a proposed merger. These responses indicate the geographic market as the area for which one finds significant correlations or covariances between the prices for the products sold in this area, and non-significant correlations for the product prices sold outside this area.

However, there is a problem with the application of this concept: data on every location needed for determining the market boundary might not be available. For instance, if there is a positive covariance between the prices in two locations, s_i and s_j and one does not find a positive covariance between the prices recorded at s_i and s_k, the market boundary is expected to lie between s_j and s_k, but it is not possible to determine where exactly on the interval. To overcome these difficulties, and still be consistent with the principle set forth above, the use of the spatial covariance function as a guideline for setting the limits of the relevant market is proposed here. This function can be denoted as $C(d_{jk})$ and expresses

[6]This discussion has also some relevance for the discussion of a related point, important for sellers of differentiated products: the effects on product variety of a merger. The standard analysis focuses on the likely product variety effects of a merger after an analysis of the price effects, as one of the "efficiencies" considered as a counterweight to the price increase. For an integrated analysis, see Gandhi et al. (2008), whose empirical counterpart is still not developed and requires the setup of a structural model, which leads to the difficulties pointed above.

the covariance between the variables as a function of the distance between them – expressed as d_{jk}.

This spatial covariance function can be estimated non-parametrically by the local averaging method of Conley and Topa (2002) or by a shape preserving cardinal B-spline[7] sieve, as used in Chen and Conley (2001). Both of them assume the locations of the agents as exogenous to the data generating process of the variable studied, and the variable to be stationary and isotropic.[8] The shape preserving cardinal B-spline wavelet sieve, which will be used in this paper, is a special case of the method of sieves (Grenander, 1981). They consist in using a sequence of parametric families – in our case, the B-spline sieve – to approximate unknown functions. The unknown function to be approximated is the spatial covariance function, which can be written as a weighted sum of basis functions:

$$C(d_{ij}) = \sum_k b_k H_k(d_{ij}) \tag{1}$$

where b_k are coefficients to be estimated, required to be increasing in k,[9] and $H_k(d_{ij})$ is an approximation of the spectral measure for covariance stationary time series (Chen and Conley, 2001):

$$H_k(d_{ij}) = \int h(yd_{ij}) B'_{m,k}(d_{ij}) dy \tag{2}$$

where $B'_{m,k}$ is the first derivative of the $k-$th B-Spline of order m. This B-Spline can be represented as:

---

[7] A cardinal B-Spline can be defined from a nondecreasing sequence $\mathbf{t} = (\cdots, -2, -1, 0, 1, 2, \cdots)$, with generic element t_i. A cardinal B-Spline of order 1 for this knot sequence are the characteristic functions of this sequence, that is, the functions:

$$B_{i1}(t) = \begin{cases} 1 & if\, t_i \leq t \leq t_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

B-Splines of higher order can be obtained by recurrence:

$$B_{ik} = \omega_{ik} B_{i,k-1} + (1 - \omega_{i+1,k}) B_{i+1,k-1}$$

In which:

$$\omega_{ik}(t) = \begin{cases} \frac{t - t_i}{t_{i+k-1} - t_i} & if\, t_i \neq t_{i+k-1} \\ 0 & \text{otherwise} \end{cases}$$

Thus, a cardinal B-spline of order k consists of the sum of $k-1$ polynomials of order $k-1$. For example, a cardinal B-Spline of order two consists of two linear parts which join to form a piecewise linear function which vanishes outside the interval $[t_i, t_{i+2})$. A B-Spline of order three consists on a quadratic function.

[8] Isotropic means that the correlation between the realizations of the variable in two locations depends only on the distance between them and not on direction.

[9] Thus, $b_{k+1} > b_k$.

$$B_{m,k}(x) = \frac{1}{(1-m)!}\sum_{l=0}^{m}(-1)^l \begin{pmatrix} m \\ l \end{pmatrix} [\max(0, x-l)]^{m-1} \tag{3}$$

The $h(\cdot)$ function, on the other hand, could be written as:

$$h(yd_{ij}) = 2^{(n-2)/2}\Gamma\left(\frac{n}{2}\right)\frac{J_{(n-2)/2}(yd_{ij})}{(yd_{ij})^{(n-2)/2}}$$

where n represents the number of localities where prices could be collected. The $J(\cdot)$ is the Bessel Function and $\Gamma(\cdot)$ is the Gamma Function:

$$\begin{aligned} \Gamma(x) &= \int_0^\infty t^{x-1}e^{-t}dt \\ J_n(x) &\sim \frac{1}{2^n n!}x^n \end{aligned}$$

The relevant parameters are estimated using two-step sieve least squares, as described in Chen and Conley (2001), and result in the estimation of b_k, according to the restrictions above and to the following structure for the spatial covariance matrix:

$$\Sigma(d(\cdot)) = \begin{bmatrix} \sigma_1^2 + C(0) & C(d_{12}) & \cdots & C(d_{1N}) \\ C(d_{21}) & \sigma_2^2 + C(0) & \cdots & C(d_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ C(d_{N1}) & C(d_{N2}) & \cdots & \sigma_N^2 + C(0) \end{bmatrix} \tag{4}$$

where $C(\cdot)$ is equation (1), as defined above. Denoting $\mathbf{P}_t = (P_{1t}, \cdots, P_{Nt})$ as the price vector for time period t, the sieve estimation for the $\Sigma(D)$ matrix is based on the solution to the following minimization problem:

$$\begin{aligned} (\hat{\sigma_i}, \hat{C}(\cdot)) &= \arg \min_{(\sigma^2, C)\in(0,\infty)^N\times\mathcal{C}_T} \sum_{t=1}^{T}\left\{\sum_i (P_{it}^2 - [\sigma_i^2 + \sum_k b_k H_k(0)])^2 + \right. \\ &+ \left. \sum_i \sum_{i\neq j}(P_{it}P_{ij} - \sum_k b_k H_k(d_{ij}))^2 \right\} \end{aligned} \tag{5}$$

where $\mathcal{C}_t$ denotes the sieve for C, described in (Chen and Conley, 2001, p. 69). Given those estimates, the next step is to determine confidence intervals for those estimated covariances. The confidence intervals, as shown in the paper by Chen and Conley (2001), are derived for any desired level of confidence by bootstrapping.

This non-parametric estimation presents some advantages over the most common technique used in spatial econometric analyses – the Spatial Autoregressive (SAR) Model or the Conditionally Autoregressive (CAR) Model – for this sort of situation. Both approaches model the spatial dependence as a function of two elements: a neighborhood matrix, usually denoted by $\mathbf{W}$, and a spatial correlation parameter, commonly denoted by ρ. In order to apply the SAR or CAR methodologies in this case, one must start by establishing an exogenous criterion for the maximum distance between two locations for them to be considered as neighbors. This problem becomes especially acute considering that Wall (2004) asserts that small errors in the definition of the neighborhood matrix imply large changes in the estimated spatial covariances. This problem does not exist in the methodology set forth above, which provides small errors in the definition of the distances between agents (Conley, 1999) with robust results. Another advantage lies in the fact that this methodology was developed to deal with panel data, in contrast to the kernel regression approach in Conley and Topa (2002), which is applied to two time periods only. Since the dataset we used in the following application comprises repeated observations of the same supermakets, Chen and Conley's approach (2001) seemed more appropriate. The next section presents an application of this methodology to an antitrust case involving supermarkets.

3. Application: Brazilian Supermarkets

In February 03, 1999 two supermarket companies, CBD[10] and PERALTA filed a memorandum to the Brazilian Antitrust System in which CBD expressed its intent to purchase 38 supermarkets and one warehouse belonging to the PERALTA chain. The supermarkets involved in this operation were located in the following municipalities:[11]

Table 1
Supermarkets (PERALTA) and cities

City/town	Supermarkets	City/town	Supermarket
São Paulo	9	Praia Grande	4
Cubatão	3	São Bernardo do Campo	1
Santos	7	Itapecerica da Serra	1
Mongaguá	1	Caraguatatuba	1
Guarujá	3	Guarulhos	1
Peruíbe	1	São Sebastião	1
Itanhaém	1		

The Brazilian Antitrust System is composed of two Committees, one of them

[10] Acronym for its Brazilian name, *Companhia Brasileira de Distribuição*.

[11] These municipalities are also considered as part of a Metropolitan Statistical Area, called *Região Metropolitana da Baixada Santista*.

affiliated with the Ministry of Finance (SEAE),[12] and another one associated with the Ministry of Justice (SDE).[13] They are in charge of investigating the proposed merger and issuing an opinion on which parts are to be blocked and which remedies should be proposed. The opinion is forwarded to the deciding body (CADE),[14] whose judges have the final word. The memo sent by CBD and PERALTA was received by the committee affiliated with the Ministry of Justice (SDE), which carried out an analysis of the proposed merger, to be submitted to CADE.

In its analysis of the geographic dimension of the relevant market for this operation, the SDE chose to use different criteria, taking into account the fact that the supermarkets were located in cities of different sizes, including São Paulo – one of the largest cities in the world – and cities with less than 100,000 inhabitants. For cities with more than one million inhabitants, the Antitrust Authority relied on a study contracted by the Brazilian Supermarkets' Association, which tried to describe the area capable of attracting about 70% of the customers of a given store. For supermarkets with more than 20 checkouts, the study indicated that this area is of about 2.5 kilometers, and for supermarkets with more than 40 checkouts, this area was set at 5 kilometers. These values were used to establish the relevant geographic markets. For cities with less than one million inhabitants, the Authority defined the relevant geographic market as the boundaries of the municipality.

Based on these concepts, the SDE issued an opinion stating that the entity resulting from the merger has market power in five cities located on the southern shore of the State of São Paulo: Cubatão, São Vicente, Guarujá, Itanhaém and Praia Grande. Considering these markets, the SDE issued an opinion to the deciding body (CADE) requiring CBD to divest one of its supermarkets in Cubatão, disregarding the fact that the consumer who defines the relevant geographic market is not the average, but the marginal one, regardless of whether the consumer buys from one store or another. That consumer is relevant for the pricing decision of the "hypothetical monopolist" used in the Merger Guidelines discussed in the previous section and, with minor changes, applied by almost every one of the Antitrust Authorities in the world. Even if 70% of the consumers of a given supermarket are located in a perimeter of 5 kilometers around a given supermarket, one can easily deduce that, for a given demand configuration, a hypothetical monopolist refrains from increasing its prices even if the the same share of customers keep going to the same supermarket. In order to make this statement clearer, consider an example, based on a Hotelling model of spatial competition. Let us suppose a *continuum* of consumers distributed on a unit interval. These consumers are served by two sellers, located at points L1 and L2 in the picture below, buying from the producer with the lowest delivered price, composed of a factory price p0

[12] In Portuguese, *Secretaria Especial de Acompanhamento Econômico.*
[13] Acronym for its Brazilian name, *Secretaria de Direito Econômico.*
[14] In Portuguese, *Conselho Administrativo de Defesa Econômica.*

and a transportation cost, which gives the slope of the diagonal lines – marked t in the same figure.

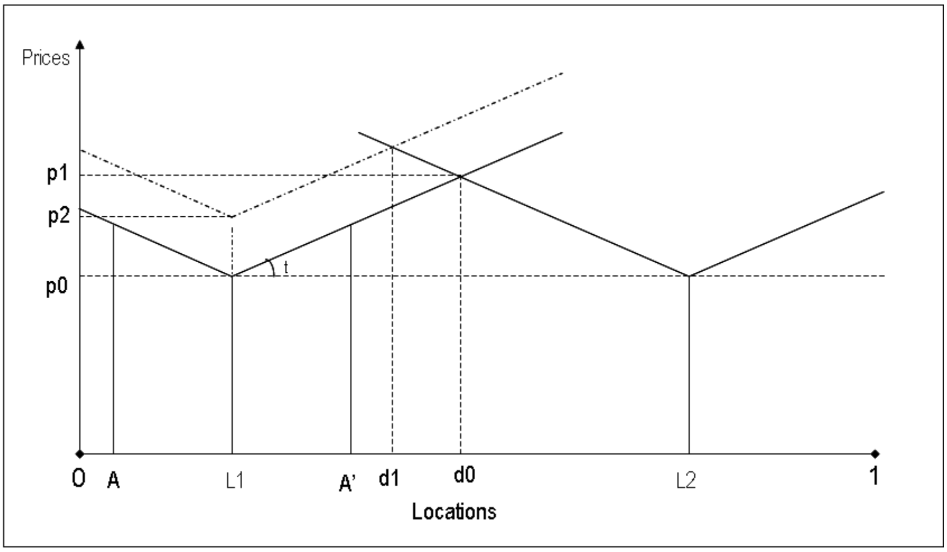

Figure 1
Spatial competition – example

This configuration gives the limits of the areas served by each producer. The seller located at L1 serves the area between 0 and d0, and the other one serves the remaining customers. The customer located at d0 is indifferent between the producers and pays a price equal to p1.

The approach pursued by the SDE implied that the relevant geographic market for producer 1 located at L1 comprises the consumers located on the line segment AA', which is inconsistent with the intent of the SSNIP test. The contradiction can be spotted by performing the following thought experiment. Suppose producer 1 is considering an increase in the factory price from p0 to p2. This course of action will be followed only if the decrease in profits due to lost sales for consumers located on the interval between d0 and d1 – the marginal consumers – is smaller than the increase in profits from the sales to customers between 0 and d1 – the inframarginal consumers. Only by chance either d0 or d1 equals A', which indicates the inadequacies of the approach taken by the SDE.

Considering these shortcomings of the geographic market definition espoused by the Authority, the next step is to determine it empirically, proposing a new methodology. This will be the subject of the following sections.

3.1 Data analysis

In order to investigate the hypothesis above, price data on 44 supermarkets were made available, whose geographical locations are presented in Appendix 1 and depicted in the following picture:

Figure 2
Geographic locations of the supermarkets

The latitudes and longitudes were used to compute the Euclidean distances – in kilometers – between supermarkets, which are depicted in Appendix 2.[15] The minimum distance between two supermarkets is about 200 meters and the longest distance is just over 93 kilometers. For each supermarket, the data on the prices of 500 products with most sales in BRL (Brazilian Reais) for a period of 14 months between January 2003 and February 2004 were made available. However, these 500 products were not the same in each location; thus, some of them had to be selected for the analysis.[16] The selected products, chosen because they appeared in the greatest number of months in the greatest number of supermarkets, are as follows:

Table 2
Selected products

Product	Brand	Packaging	Code
Hot chocolate mix	Toddy	400g	TODDY
Refined sugar	Uniao	1Kg	UNIAO
Rice	Tio Joao	5Kg	TIOJOAO
Crackers	Club Social	279g	CSOCIAL
Coffee	Pilao	500g	PILAO
Beer	Bavaria	350ML	BAVARIA
Beer	Brahma	350ML	BRAHMA
Beer	Kaiser	350ML	KAISER
Chocolate	BIS	150g	BIS
Meat		1Kg	MEAT
Detergent	OMO	1Kg	OMO
Poultry		1Kg	POULTRY
Milk	Parmalat	1L	PARMALAT
Condensed milk	Moca	395g	MOCA
Mayonnaise	Hellman's	500g	HELLMANS
Margarine	Qualy Crem	500g	QUALICREM
Watermelon		1Kg	WMELON
Tomato sauce	Pomarola	340g	POMAROLA
Mozzarela		1Kg	MOZZARELA
Cheese		1Kg	CHEESE
Soft drink	Guarana	2L	GUARANA
Soft drink	Coca-Cola	350ML	COLA

The descriptive statistics for the product prices – the only product-level characteristic available – are presented in the following table. These statistics are computed using the pooled data of each product, all supermarkets and time periods. Thus, we had a panel dataset of prices to which the information on the geographic locations of the supermarkets was matched:

[15] This metric was chosen due to the fact that information on other plausible metrics, such as travel times, was not available.

[16] As quantity data were not available, the scope for the setup of a structural econometric model analysis was severely limited.

Table 3
Descriptive statistics – prices (R$ per unit)

Code	Mean	Std. deviation	Minimum	Maximum	Number of obs.
TODDY	3.195	0.434	2.04	4.88	560
UNIAO	1.147	0.17	0.66	1.48	602
TIOJOAO	10.037	1.05	7.46	11.886	602
CSOCIAL	2.153	0.227	1.76	2.95	588
PILAO	3.687	0.214	2.409	4.226	462
BAVARIA	0.757	0.06	0.618	0.91	602
BRAHMA	0.94	0.055	0.78	1.076	602
KAISER	0.776	0.046	0.66	0.884	602
BIS	2.451	0.236	1.88	2.971	490
MEAT	8.8	1.11	6.488	12.05	602
OMO	5.566	0.413	4.7	6.892	546
POULTRY	2.324	0.21	1.801	2.906	602
PARMALAT	1.493	0.104	1.17	1.715	602
MOCA	1.749	0.131	1.331	2.118	602
HELLMANS	3.624	0.188	2.8	4.131	574
QUALICREM	2.806	0.188	2.205	3.64	574
WMELON	0.58	0.107	0.343	0.968	476
POMAROLA	1.375	0.15	0.967	1.808	420
MOZZARELA	10.452	1.161	7.514	13.43	532
CHEESE	11.05	1.178	8.451	14.04	532
GUARANA	1.79	0.106	1.53	1.995	588
COLA	0.884	0.052	0.69	0.99	574

The product prices for the sample were quite low, ranging from just below 0.60 BRL – in the case of one kilogram of watermelon – to just about 10.50 BRL in the case of one kilogram of mozzarela. These products are, for the most part, essential items, and these characteristics are consistent with these products being some of those with the highest sales. The differences in the number of observations are due to the lack of data for these products for all supermarkets and time periods. As for price variability, the product with the highest coefficient of variation (0.184) was watermelon, being consistent with the low price and supply characterized by highly competitive conditions.

As a preliminary analysis of the relationship of correlation between prices and distance, a linear regression model was tried, such as the following:

$$CORR_{ij} = \gamma_0 + \gamma_1 DIST_{ij} + \gamma_2 D_{1ij} + \gamma_3 D_{2ij} + \epsilon_{ij}$$

where the variable $CORR_{ij}$ represented the correlation between prices of supermarkets i and j, $DIST_{ij}$ is the distance in kilometers between the same supermarkets, D_{1ij} is a dummy variable with value one if the pair of supermarkets are located in the same city, and D_{2ij} is a dummy with value one if both supermarkets belong to the same chain. The regressions were carried out for each of the previously mentioned products, and the results were far from conclusive. Except for two products, in all cases, the hypothesis $\gamma_1 = 0$ could not be rejected. For these products in which $\gamma_1 \neq 0$, one was positive – indicating an increase in correlation as the distance increases, and another was negative. The next step was, for each of the products, to compute the spatial covariance as a function of Euclidean distance, as in the previous section. The covariances obtained were normalized by dividing them by the global variance.[17] A graphical example of the estimates and of the confidence intervals for the null hypothesis that covariance at that distance equals zero – for product 2, Uniao Refined Sugar – is presented in Appendix 3, and the values of the spatial convariance function for some distances are presented in the following table. The results of this paper were computed using eight third-order B-Splines evenly spaced over the support of the distance function.[18]

Table 4
Spatial covariance estimates

Code	Distance – in Km					
	0.22	10.01	20.11	30.20	50.09	92.00
TODDY	0.59947	0.60742	0.74686	0.85524	0.5822	0.80014
UNIAO	0.94081	0.93437	0.93543	0.96521	1.0028	0.9571
TIOJOAO	0.88844	0.91314	0.92393	0.90762	0.92458	0.88782
CSOCIAL	0.57532	0.5852	0.66484	0.64201	0.28372	0.57744
PILAO	0.42362	0.5572	0.45214	0.23734	0.4155	0.39352
BAVARIA	0.53426	0.57355	0.61191	0.63996	0.4736	0.57643
BRAHMA	0.45879	0.45955	0.41168	0.4696	0.47414	0.43027
KAISER	0.44943	0.49344	0.45199	0.48934	0.37327	0.36969
BIS	0.68389	0.71559	0.66271	0.71299	0.56572	0.51792
MEAT	0.72434	0.71683	0.69866	0.62802	0.56077	0.70425
OMO	0.51007	0.48723	0.47272	0.58807	0.54808	0.54028
POULTRY	0.76443	0.7742	0.79727	0.79434	0.71347	0.7695
PARMALAT	0.76177	0.78738	0.78833	0.86001	0.92414	0.80056
MOCA	0.53998	0.49633	0.49878	0.5628	0.51204	0.5267
HELLMANS	0.38283	0.41068	0.47144	0.44311	0.41565	0.39502
QUALICREM	0.60658	0.5802	0.61612	0.63633	0.53561	0.73912
WMELON	0.77716	0.77615	0.72515	0.7447	0.85412	0.78834
POMAROLA	0.49783	0.5991	0.65854	0.82389	0.60629	0.49433
MOZZARELA	0.44718	0.4698	0.50278	0.56978	0.41909	0.41138
CHEESE	0.56647	0.50989	0.55559	0.59291	0.47336	0.57242
GUARANA	0.62413	0.64224	0.56374	0.54167	0.65542	0.59559
COLA	0.47898	0.51001	0.5685	0.59343	0.48683	0.48299

[17]That means the variance of all observations of the selected product for all supermarkets and months. The rationale for this normalization is as follows: under the current model, the deviations from the means, denoted as $\mathbf{u}$, could be decomposed into two components, one collecting the spatially determined aspects of the variable (denoted as θ), and another one, the random components (denoted as ϵ). Thus, we can write this relation as:

$$\mathbf{u} = \theta + \epsilon$$

The normalization described in the text means the spatial covariance converges to the spatial correlation coefficient if θ corresponds to a larger fraction of the total variance of $\mathbf{u}$ then ϵ. All estimates were calculated using the MATLAB software, version 7.0.1. The code used in all estimates is available upon request.

[18]The analysis carried out in this paper was replicated by using seven and nine B-Splines and the results were essentially the same.

All of the estimates above are outside of the bootstrapped 95% confidence interval, indicating one can safely reject the hypothesis of non-correlation between prices at locations separated by these distances.[19] The fact that some correlations do not seem to decrease monotonically with distance might be the result of the lack of enough data in these distances – as noted before, not all products had price data for all supermakets in all time periods – together with the fact that the Euclidean distance is only an approximation of the true distance between supermarkets. Considering what was expounded before, this indicates that the decision the SDE made regarding the geographic market definition is incorrect.

However, some other points must be addressed before a final conclusion can be drawn. Some extensions of this result are presented in the following section.

3.2 Extensions

Despite indicating that the relevant geographic market is composed of all the cities together, these results must be tested further, as they could be subject of criticism. The first possible criticism concerns the fact that the products in the sample – being one of the 500 products with the highest sales – are the result of a distribution process and of marketing campaigns defined at the level of the supermarket chain. Even the SDE, in its final report for the deciding body (CADE), mentioned a study by Kwoka and White (2003), mentioning that only 1.5% of the products sold by the Toys 'R' Us retailer were priced according to local conditions.

Another important criticism is that local stores could select their portfolio of goods to be sold, and this portfolio could be a strategic variable for the competition between supermarkets; thus, the covariance could be subject to omitted variable bias. And finally, the companies' pricing decisions might be affected by the socioeconomic characteristics of the cities – or neighborhoods – in which the supermarkets are located.

[19] In addition, there seems to be some sort of positive correlation between the coefficient of variation and the spatial correlation estimates, which might indicate the role competition between manufacturers might have on price setting. However, given the aims of the paper, and the lack of a structural model, a definite causal explanation for these differences was left for further research.

In order to address these criticisms, the analysis of the previous section was extended, and some assumptions were made. The first one is that the geographic location decisions are exogenous to pricing decisions.[20] The second one is that marketing and distribution policies are decided on the chain level, and these decisions are constant throughout the period of analysis, being modeled as chain fixed effects. Finally, the sociodemographic characteristics of the cities were considered to be constant during the study period (14 months).[21] Thus, the differences in sociodemographic characteristics could be modeled as city dummies. Given that this area comprises some tourist attractions, changes in the sociodemographic profile of the cities were captured, in part, by a dummy for the holiday season. These assumptions lead to the following model for the price of each product i:

$$P_{it} = \alpha_0 + \beta_1 CHAIN_{it} + \beta_2 VACATION_{it} + \sum_{k=1}^{6} \gamma_k D_{kit} + \varepsilon_{it} \qquad (6)$$

where P_{it} refers to the price of each product selected at supermarket i in time period t, $CHAIN_{it}$ is a dummy with the value of one for the supermarkets belonging to the CBD chain and zero, otherwise. D_{kit} denotes a set of six dummies for the cities of São Vicente, Guarujá, Santos, Praia Grande, Cubatão and Peruíbe. $VACATION_{it}$ is another dummy variable, intended to capture the effects on the prices due to the holiday season (from December to February). The cities involved are coastal cities not far from the largest city in Brazil, São Paulo, and are important tourist destinations. Thus, the analysis herein tries to investigate the pattern of spatial covariance function after controlling for the differences mentioned above. The regression results[22] for each of the products are shown in the Table 5:

[20]This assumption is not unduly restrictive, since location changes were not observed during the period of analysis.

[21]This is not an unduly restrictive assumption, considering the small change in sociodemographic characteristics – such as age profile of the population and income distribution – of these cities during this period.

[22]With confidence levels indicated by asterisks calculated from Newey-West standard errors with one lag in the autoregression part of the estimates.

Table 5
Regression results

Code	Constant		CHAIN		Guarujá		Praia Grande		Peruíbe		São Vicente		Santos		VACATION		R^2
TODDY	3.222	**	0.003		0.160	**	0.202	**	0.216	*	0.101		0.041		-0.352	**	0.180
UNIAO	1.165	**	0.012		0.028		0.051	*	0.033		0.028		0.022		-0.137	**	0.158
TIOJOAO	9.776	**	0.041		0.190		0.202		0.149		0.219		0.205		0.205	**	0.013
CSOCIAL	2.085	**	0.012		0.122	**	0.192	**	0.140	**	0.172	**	0.133	**	-0.183	**	0.205
PILAO	3.551	**	0.037		0.129	**	0.174	**	0.185	**	0.148	**	0.070	**	0.071	**	0.110
BAVARIA	0.724	**	0.017	**	0.035	**	0.060	**	0.031	**	0.042	**	0.030	**	-0.019	**	0.111
BRAHMA	0.905	**	0.010	*	0.031	**	0.044	**	0.029	**	0.034	**	0.026	**	0.010	**	0.064
KAISER	0.760	**	0.009	**	0.028	**	0.038	**	0.026	**	0.025	**	0.025	**	-0.031	**	0.163
BIS	2.439	**	0.042	*	0.084	**	0.152	**	0.243	**	0.114	**	0.047		-0.224	**	0.269
MEAT	9.212	**	-0.464	**	-0.630	**	-0.681	**	-0.610	**	-0.858	**	-0.553	**	0.920	**	0.236
OMO	5.385	**	0.107	**	0.245	**	0.374	**	0.048		0.266	**	0.193	**	-0.177	**	0.115
POULTRY	2.242	**	0.018		0.088	**	0.097	**	0.030		0.063		0.045		0.048	**	0.035
PARMALAT	1.472	**	0.022	*	0.044	**	0.057	**	0.033		0.041	**	0.027		-0.060	**	0.110
MOCA	1.697	**	0.032	**	0.095	**	0.126	**	0.107	**	0.111	**	0.069	**	-0.115	**	0.271
HELLMANS	3.475	**	0.020		0.154	**	0.189	**	0.125	**	0.160	**	0.117	**	0.047	**	0.105
QUALICREM	2.834	**	-0.024		0.002		0.016		-0.038		-0.055		-0.040		-0.002		0.022
WMELON	0.567	**	0.000		0.002		0.031		0.027		0.011		0.014		0.003		0.010
POMAROLA	1.342	**	-0.011		-0.003		-0.015		0.053		0.012		0.006		0.092	**	0.098
MOZZARELA	10.552	**	0.043		0.123		0.336	*	-0.478	*	-0.382	*	-0.142		-0.215	*	0.048
CHEESE	11.224	**	-0.089		0.114		0.384	**	0.020		-0.445	*	0.012		-0.468	**	0.071
GUARANA	1.734	**	0.028	**	0.064	**	0.119	**	0.063	**	0.077	**	0.059	**	-0.052	**	0.162
COLA	0.862	**	0.015	**	0.032	**	0.058	**	0.024		0.036	**	0.027	**	-0.039	**	0.235

OBS: ** $p<0.05$ and * $p<0.1$.
The variables Guarujá, Praia Grande, Peruíbe, São Vicente and Santos are city dummies.

The spatial covariance of the residuals, also scaled by the global variance, is also presented on Table 6:

Table 6
Spatial covariance estimates

Product Code	Distance in Km 0.22	10.01	20.11	30.20	50.09	92.00
TODDY	0.518	0.540	0.651	0.782	0.491	0.705
UNIAO	0.936	0.924	0.920	0.956	0.994	0.958
TIOJOAO	0.889	0.912	0.922	0.908	0.923	0.889
CSOCIAL	0.470	0.489	0.563	0.537	0.205	0.470
PILAO	0.380	0.534	0.452	0.266	0.399	0.365
BAVARIA	0.533	0.554	0.577	0.617	0.482	0.577
BRAHMA	0.448	0.456	0.411	0.464	0.464	0.414
KAISER	0.375	0.400	0.359	0.401	0.305	0.347
BIS	0.557	0.579	0.561	0.597	0.437	0.538
MEAT	0.645	0.633	0.592	0.525	0.531	0.612
OMO	0.469	0.463	0.451	0.554	0.537	0.470
POULTRY	0.757	0.773	0.786	0.777	0.719	0.761
PARMALAT	0.761	0.765	0.752	0.826	0.900	0.821
MOCA	0.354	0.350	0.352	0.366	0.311	0.338
HELLMANS	0.385	0.395	0.443	0.435	0.407	0.398
QUALICREM	0.606	0.580	0.617	0.638	0.535	0.738
WMELON	0.777	0.776	0.725	0.744	0.854	0.789
POMAROLA	0.473	0.544	0.609	0.769	0.548	0.510
MOZZARELA	0.452	0.461	0.489	0.553	0.401	0.433
CHEESE	0.536	0.475	0.519	0.545	0.419	0.597
GUARANA	0.583	0.586	0.493	0.495	0.630	0.562
COLA	0.359	0.352	0.347	0.262	0.336	0.359

In order to address the problem of common trends in prices, two versions of equation (6) were also tried: one using first differences of prices, and another one using a time trend. In both versions, the substantive conclusions did not change; consequently, they were not reported.[23] The visual inspection of the results points out important decreases in the absolute values of the estimated covariances in only four products – CSOCIAL, MEAT, MOCA and COLA – among those listed in Table 4. For all other products, the estimated covariances are quite similar to those presented in the previous section.

As for the magnitudes of the covariances estimated, the higher values seem to point out to products such as rice, sugar, meat and milk, which tend to present a strong degree of competition on the producing side, which tends to make product prices more uniform in all locations than products which present a high degree of differentiation. This is also a possible explanation for the higher coefficient of variation for the same products.

[23]Results can be obtained from the authors upon request.

On the other hand, the lowest values are found for products such as beer and soft drinks, for which product differentiation plays a large role in defining the competitive landscape; they also tend to be the ones for which the covariances experienced the greatest reduction after controlling for city fixed effects and supermarket chain dummies, reflecting the roles of the bargaining power between the supermarket chain and manufacturer in prices.

Just as in the case presented above, each of the covariances is outside of the confidence interval for the null hypothesis of zero spatial covariance at that distance. Thus, one can conclude that the relevant market for the considered merger is composed of all the cities involved.

4. Conclusions

In this paper, a technique was proposed to identify the relevant geographic market for a proposed merger between two competing supermarket chains, CBD and PERALTA. In order to do so, the legal and theoretical aspects regarding the determination of the relevant market were initially reviewed, indicating the shortcomings of the analysis carried out by the Brazilian Antitrust Authority in its evaluation of the merger. This analysis established that the geographic dimension of the relevant market be separated in each of the municipalities on the shore of the Sao Paulo state.

After that, the methodology for determining the geographic dimension of the market by the estimation of a spatial covariance function was proposed. The estimation was carried out using a dataset including price data on 22 products for 43 supermarkets, from January 2003 to February 2004.

Initially, the calculated spatial covariances for different distances indicated significant covariances at all distances, which indicated a single geographic market embracing the eight cities (São Vicente, Santos, Guarujá, Praia Grande, Cubatão, Mongaguá, Itanhaém, Peruíbe) which were held as separate markets.

Given the shortcomings of the study, the analysis was extended in order to control for differences in the socioeconomic variables of the cities, as well as differences in the policies chosen by each supermarket chain, and seasonal patterns of consumption. The results confirmed those found in the previous section, indicating a single geographic relevant market for this merger, in opposition to the results of an analysis carried out by the Brazilian Antitrust Authority.

References

Bronnenberg, B., Dhar, S. K., & Dube, J.-P. (2005). Endogenous sunk costs and the geographic distribution of brand shares in consumer package goods industries. Working Paper University of Chicago GSB.

Chen, X. & Conley, T. (2001). A new semiparametric spatial model for panel time series. *Journal of Econometrics*, 105:59–83.

Conley, T. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics*, 92:1–45.

Conley, T. & Topa, G. (2002). Socio-economic distance and patterns of unemployment. *Journal of Applied Econometrics*, 17(4):1–45.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley Interscience, New York.

Elzinga, K. G. & Hogarty, T. F. (1973). The problem of geographic market definition in antimerger suits. *Antitrust Bulletin*, 18(1):45–81.

Elzinga, K. G. & Hogarty, T. F. (1978). The problem of geographic market definition revisited. *Antitrust Bulletin*, 23(1):1–18.

Gandhi, A., Froeb, L. M., Tschantz, S. T., & Werden, G. J. (2008). Post-merger product repositioning. *The Journal of Industrial Economics*, 56(1):49–67.

Grenander, U. (1981). *Abstract Inference*. Wiley, New York.

Horowitz, I. (1981). Market definition in antitrust analysis: A regression-based approach. *Southern Economic Journal*, 48(1):1–16.

Kwoka, J. E. & White, L. J. (2003). *The Antitrust Revolution: Economics, Competition and Policy*. Cambridge University Press, Cambridge.

Motta, M. (2004). *Competition Policy: Theory and Practice*. Cambridge University Press, Cambridge.

Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *Rand Journal of Economics*, 31(3):395–421.

Scheffman, D. T. & Spiller, P. (1987). Geographic market definition under the U.S. Department of Justice Merger Guidelines. *Journal of Law and Economics*, 30(1):123–147.

Sherwin, R. A. (1993). Comments on Werden and Froeb – Correlation, causality and all that jazz. *Review of Industrial Organization*, 8:355–358.

Slade, M. (1986). Exogeneity tests of market boundaries applied to petroleum produts. *The Journal of Industrial Economics*, 34(3):291–303.

Stigler, G. J. & Sherwin, R. A. (1985). The extent of the market. *Journal of Law and Economics*, 28(3):555–585.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121:311–324.

Werden, G. J. (1998). Demand elasticities in antitrust analysis. *Antitrust Journal*, 43:363–414.

Werden, G. J. & Froeb, L. M. (1993). Correlation, causality and all that jazz: The inherent shortcomings of price tests for antitrust market delineation. *Review of Industrial Organization*, 8:329–353.

Werden, G. J. & Froeb, L. M. (2002). The antitrust logit model for predicting unilateral competitive effects. *Antritrust Law Journal*, 70(1).

Appendix 1 – Stores Used

Table A.1
Stores used and coordinates

City	Code	Longitude	Latitude
Santos	CBD	-46.3231	-23.9590
Santos	PERALTA	-46.3091	-23.9713
Guaruja	CBD	-46.2586	-23.9919
Santos	CBD	-46.3083	-23.9843
Praia Grande	CBD	-46.4125	-24.0079
Guaruja	PERALTA	-46.2894	-23.9566
Sao Vicente	CBD	-46.3702	-23.9714
Santos	PERALTA	-46.3357	-23.9470
São Vicente	PERALTA	-46.3054	-23.9407
Praia Grande	PERALTA	-46.4608	-24.0188
Itanhaem	CBD	-46.7877	-24.1839
Guaruja	CBD	-46.2573	-23.9966
Peruibe	CBD	-46.9988	-24.3206
Cubatao	PERALTA	-46.4287	-23.8729
Cubatao	PERALTA	-46.4084	-23.9271
Santos	PERALTA	-46.3377	-23.9583
Santos	CBD	-46.3083	-23.9814
Cubatao	PERALTA	-46.4207	-23.8825
Santos	PERALTA	-46.3616	-23.9349
Sao Vicente	PERALTA	-46.4072	-23.9549
Santos	PERALTA	-46.3005	-23.9769
Mongagua	CBD	-46.6188	-24.0927
Guaruja	PERALTA	-46.2635	-23.9883
Santos	PERALTA	-46.3129	-23.9640
Peruibe	CBD	-46.9936	-24.3152
Itanhaem	CBD	-46.7829	-24.1805
Praia Grande	CBD	-46.4205	-24.0121
Sao Vicente	PERALTA	-46.4925	-23.9855
Guaruja	PERALTA	-46.2779	-23.9913
Guaruja	PERALTA	-46.2715	-23.9787
Sao Vicente	PERALTA	-46.3767	-23.9682
Santos	CBD	-46.3187	-23.9755
Santos	PERALTA	-46.3216	-23.9431
Praia Grande	PERALTA	-46.4798	-24.0270
Praia Grande	CBD	-46.4181	-24.0033
Sao Vicente	PERALTA	-46.3753	-23.9554
Guaruja	PERALTA	-46.2034	-23.9857
Guaruja	PERALTA	-46.2827	-23.9396
Guaruja	PERALTA	-46.2370	-23.9834
Santos	CBD	-46.3442	-23.9678
Guaruja	CBD	-46.2468	-23.9880
Santos	CBD	-46.3319	-23.9577
Praia Grande	CBD	-46.3986	-23.9878

OBS: Latitude and Longitude refer to geographic coordinates expressed as decimal fractions. For instance, a place located at 38° 53' 23" N, 77° 00' 32" W would appear as 38.889722°, -77.008889°.

Appendix 2 – Histogram of Distances

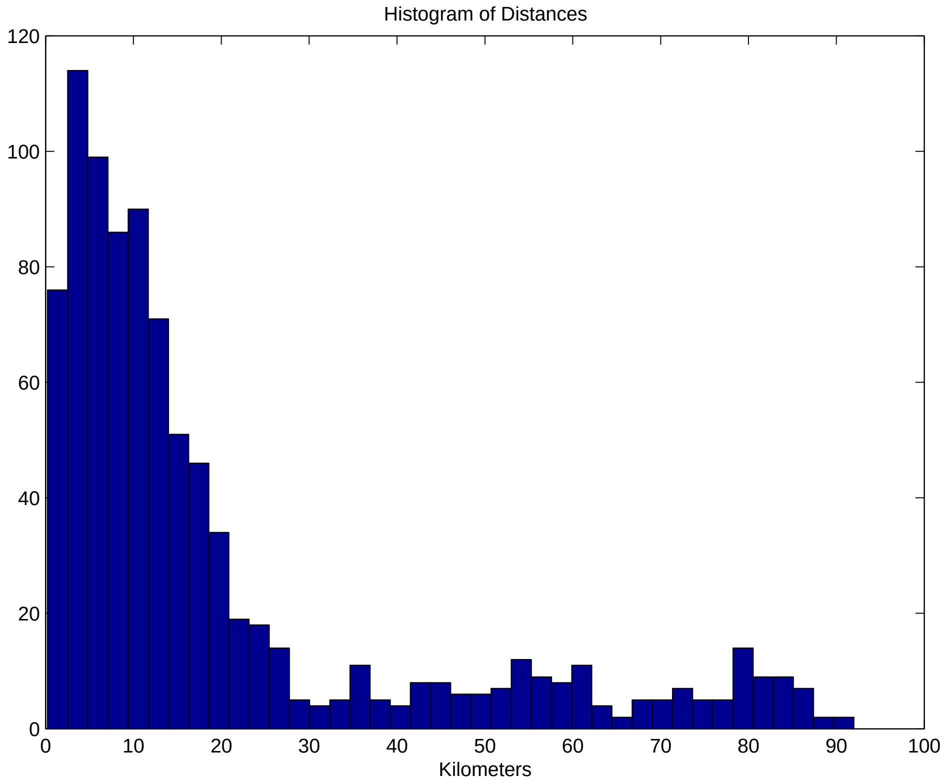

Figure A.1
Histogram of distances

Appendix 3 – Spatial Covariance Function

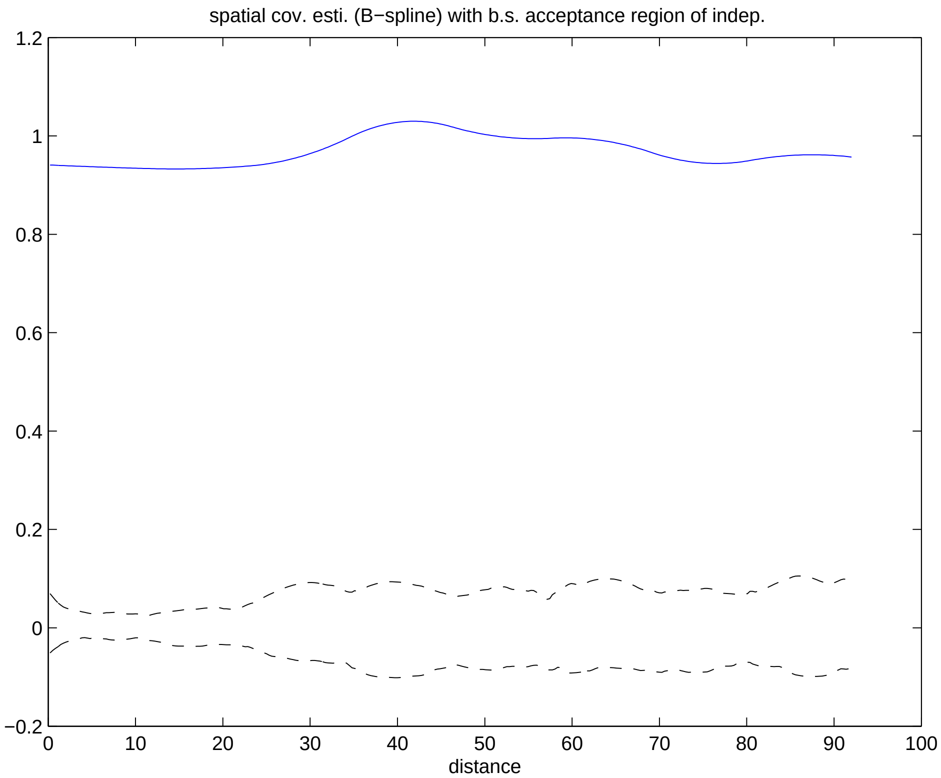

Figure A.2
Spatial covariance function