*The Annals of Applied Statistics*
2013, Vol. 7, No. 1, 81–101
DOI: 10.1214/12-AOAS588


# GEOSTATISTICAL MODELING IN THE PRESENCE OF INTERACTION BETWEEN THE MEASURING INSTRUMENTS, WITH AN APPLICATION TO THE ESTIMATION OF SPATIAL MARKET POTENTIALS

BY FRANCESCO FINAZZI

*University of Bergamo*

This paper addresses the problem of recovering the spatial market potential of a retail product from spatially distributed sales data. In order to tackle the problem in a general way, the concept of spatial potential is introduced. The potential is concurrently measured at different spatial locations and the measurements are analyzed in order to recover the spatial potential. The measuring instruments used to collect the data interact with each other, that is, the measurement at a given spatial location is affected by the concurrent measurements at other locations. An approach based on a novel geostatistical model is developed. In particular, the model is able to handle both the measuring instrument interaction and the missing data. A model estimation procedure based on the expectation–maximization algorithm is provided as well as standard inferential tools. The model is applied to the estimation of the spatial market potential of a newspaper for the city of Bergamo, Italy. The estimated spatial market potential is eventually analyzed in order to identify the areas with the highest potential, to identify the areas where it is profitable to open additional newsstands and to evaluate the newspaper total market volume of the city.

**1. Introduction.** The market potential of a given retail product is the expected sales volume when the product is marketed. The spatial market potential is the spatial distribution of the market potential over a trading area. Sales are expected to be high if a store is opened at a spatial location characterized by a high spatial market potential, while they are expected to be low if the spatial location has a low spatial market potential.

With the goal to increase and to maximize the sales volume, a key issue is how to evaluate the spatial market potential. In this paper, it is assumed that the product is already marketed and that the sales data of spatially distributed stores are available. Thus, the aim is to estimate the spatial market potential by considering the sales data, the spatial characteristics of the trading area and the spatial interaction between the stores. The stores interact in the sense that the sales volume of each store is affected by the presence of all the others. As a consequence, the spatial market potential cannot be estimated ignoring the interaction.

Received March 2012; revised May 2012.

*Key words and phrases.* Geostatistics, spatial potential, spatial interaction, EM algorithm, Hessian estimation, geomarketing.

For all purposes and intents, the spatial market potential can be regarded as a spatial surface, as it is well defined for all the spatial locations of the trading area. Taking a statistical perspective, the spatial market potential is considered as a spatially continuous random field and the estimation of the spatial market potential is obtained through the estimation of the realization of the random field. Although well understood, however, no attempt has ever been made to address the problem following a geostatistical approach.

The estimation of a spatial market potential is an instance of the more general problem of recovering the realization of a spatially continuous random field in the case of interacting measuring instruments. The instruments interact in the sense that the measurement at a given spatial location is affected by the concurrent measurements at nearby locations.

A novel model able to handle both the interaction between the measuring instruments and the missing data is proposed. A case study is presented, in which the sales data of spatially distributed newsstands are used to estimate the spatial market potential of an economic daily newspaper for the city of Bergamo, Italy. The aim of the study is threefold: to identify the areas with the highest market potential, to identify the areas where it is profitable to open additional newsstands and to estimate the total market volume of the city with respect to the newspaper considered.

The rest of the paper is organized as follows: Section 2 provides the background and motivation for this work. Section 3 introduces a novel geostatistical model for the analysis of spatial point data in the case of interaction between the measuring instruments. Model estimation and inference are discussed in Section 4. Section 5 presents the case study while Section 6 provides conclusions. The technical aspects related to the model estimation are reported in the Appendices.

**2. Background.** In this section the spatial market potential estimation problem is discussed in terms of both the current state of the art and the available statistical methods. It turns out that the estimation of a spatial market potential from sales data received little attention in the past and that the classic geostatistical approach cannot be adopted to solve the problem.

2.1. *Spatial market potential estimation.* The problem of estimating the market potential of a retail product is not new in the literature. The state of the art is represented by the so-called spatial interaction models which describe the market potential in terms of flows between a set of origins (the customers) and a set of destinations (the stores). The interested reader is referred to the seminal papers of Reilly and Huff [Reilly (1931), Huff (1964)] and to the more recent literature [see, e.g., Davis (2006), Cliquet (2006), de Grange, Ibeas and Gonzalez (2011) and Fischer and Wang (2011)]. The spatial interaction models focus on the utility that consumers obtain from buying a retail product at a specific store. The utility is often a function of the attributes of the product, the attributes of the store and some attributes of separation such as the geographic distance, the transport cost and the transport time.

The main drawback of the spatial interaction models is that the spatial market potential is not explicitly modeled as a regionalized variable and it is defined only at the spatial location of the stores. This is in contrast with the concept of spatial market potential adopted in this paper, which is supposed to exist beyond the existence of the stores. Indeed, the market potential of a product at a given location in space can also be considered as the willingness of the consumers to reach that specific location in order to buy the product. The attributes of the store (including the price at which the store sells the product) may affect the way the spatial market potential is observed, but the spatial market potential is not driven by the stores.

The spatial interaction models literature also lacks of methods for estimating the model output uncertainty. This represents a critical issue, as, in practical applications, the available data are usually limited in number and the reliability of the model output must be provided.

Modeling the market potential as a regionalized variable, and, in particular, as a spatially continuous random field, allows to answer new and interesting questions. Denoting $q(\mathbf{s})$ the spatial market potential at the generic spatial location $\mathbf{s}$, the company that owns the stores may be interested in estimating $q(\mathbf{s})$ for each point of the trading area $\mathcal{D}$. For instance, the company may want to locate the maxima of $q(\mathbf{s})$ to be sure that it has a store near that location. If, instead, the company wants to open a new store, then it may want to evaluate the market potential conditioned on the presence of the actual stores. Moreover, the company may want to pursue both of the goals even if the sales data of some stores are missing. In this sense, when a latent spatial market potential has to be assessed and the uncertainty information must be provided, the geostatistical approach seems to be more appropriate than any approach based on the spatial interaction models.

2.2. *Geostatistical modeling*. The problem of estimating a spatially continuous random field from measurements collected at a finite number of locations in space is usually solved by considering geostatistical models and kriging techniques [see Cressie and Wikle (2011)].

The simplest geostatistical model described in Diggle and Ribeiro (2007), for instance, assumes an underlying stationary Gaussian random field $w(\mathbf{s})$ and observation $y(\mathbf{s}_i)$ which are realizations of conditionally mutually independent random variables $Y(\mathbf{s}_i)$ conditionally normally distributed with mean $E(Y(\mathbf{s}_i) \mid w(\cdot)) = w(\mathbf{s}_i)$ and variance $\sigma_\varepsilon^2$. In the spatial market potential case, however, due to the interaction between the stores, the conditional mutual independence of the random variables $Y(\mathbf{s}_i)$ is not met and, in general, $E(Y(\mathbf{s}_i) \mid q(\cdot)) \neq q(\mathbf{s}_i)$. A geostatistical approach for spatial interaction data can be found in Banerjee, Gelfand and Polasek (2000), though the interaction is defined in terms of flows between destinations and origins (cf. the previous paragraph) and the approach is not suitable for the problem addressed in this paper, where a "diffuse" origin is considered.

## 3. The geostatistical potential model.

3.1. *Introduction.* Before developing a suitable geostatistical model, the problem at hand is restated and generalized in the following way.

Let $q(\mathbf{s})$ be a spatial random field defined over the region of space $\mathcal{D} \subset \mathbb{R}^2$. The random field $q(\mathbf{s})$ is called here *potential*, though the term does not refer to any particular property of the field. The random field is concurrently measured at the set of spatial locations $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ and the observations $\mathbf{y}(\mathcal{S})$ are collected (possibly with missing data). The concurrency of the measurements is a key aspect in the sense that, in general, the observations $\tilde{\mathbf{y}}(\mathcal{S})$ collected in the case of nonconcurrent measurements differ from $\mathbf{y}(\mathcal{S})$.

The observations $\mathbf{y}(\mathcal{S})$ are supposed to be realizations of random variables $Y(\mathbf{s}_i)$ conditionally normally distributed with conditional mean

$$E\big(Y(\mathbf{s}_i) \mid q(\cdot), \mathcal{S}\big) = h\big(q(\mathbf{s}_i); \mathcal{S}\big); \qquad \mathbf{s}_i \in \mathcal{S}, i = 1, \dots, N,$$

and conditional variance $\sigma_\varepsilon^2$, where $h$ is a function modeling the interaction between the measuring instruments.

The interaction between the measuring instruments is said to be an *absorption interaction* if, for each $\mathbf{s} \in \mathcal{D}$, $h(q(\mathbf{s}_i); \mathcal{S}) < q(\mathbf{s})$. The interaction is such that, for each $\mathbf{s}_1 \in \mathcal{D}$, $h(q(\mathbf{s}_1); \mathcal{S}) \equiv q(\mathbf{s}_1)$ iff $\mathcal{S} = \{\mathbf{s}_1\}$, that is, the conditional mean of $Y(\mathbf{s}_1)$ is equal to $q(\mathbf{s}_1)$ if $\mathbf{s}_1$ is the only spatial location of $\mathcal{D}$ where $q$ is measured. Ultimately, it can be stated that the act of measuring the potential $q$ at a given $\mathbf{s}$ alters the (concurrent) measurements at other spatial locations.

At this point, the following distinction can be made: the potential $q(\mathbf{s})$ is the expected observed value when $q$ is measured only at the spatial location $\mathbf{s} \in \mathcal{D}$. On the other hand, the *conditional potential* $q(\mathbf{s}; \mathcal{S})$ is the expected observed value when $q$ is measured at the spatial location $\mathbf{s} \in \mathcal{D}$ given that it is concurrently measured at the set of locations $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$, $\mathbf{s}_i \in \mathcal{D}$, $N \geq 1$.

In the next paragraph, the way the potential and the conditional potential are modeled and estimated is discussed.

3.2. *Model definition.* The geostatistical potential model (GPM) is introduced here as the main statistical tool for the analysis of spatial data arising from concurrent measurements in the presence of interaction between the measuring instruments. In its general form, the GPM is described by the following hierarchy of equations:

$$\begin{aligned} y(\mathbf{s}; \mathcal{S}) &= h_{\boldsymbol{\vartheta}}\big(u(\mathbf{s}), \mathbf{s}, \mathcal{S}\big), \\ u(\mathbf{s}) &= q(\mathbf{s}) + \varepsilon(\mathbf{s}), \\ q(\mathbf{s}) &= \mu + \mathbf{x}(\mathbf{s})\boldsymbol{\beta} + \gamma w(\mathbf{s}). \end{aligned} \tag{3.1}$$

At the first stage of (3.1), $y(\mathbf{s}; \mathcal{S})$ is the measured conditional potential at the spatial location $\mathbf{s}$ while $h_{\boldsymbol{\vartheta}} : \mathbb{R} \times \mathcal{D} \times \mathbb{S} \longrightarrow \mathbb{R}$ is the *interaction function* which is parametrized by the parameter vector $\boldsymbol{\vartheta}$. The set $\mathbb{S}$ is the set of all finite spatial point patterns over $\mathcal{D}$ including the nonsimple patterns (i.e., patterns with overlapping points). At the second stage, $\varepsilon(\mathbf{s})$ represents an error component which is assumed to be i.i.d. $N(0, \sigma_\varepsilon^2)$ and is supposed to capture both the measuring error and the model error. Finally, at the third stage, the potential $q(\mathbf{s})$ is modeled by three summands, where $\mu$ is the mean, $\mathbf{x}(\mathbf{s})$ is a vector of covariates, $\boldsymbol{\beta}$ is the vector of coefficient, $w(\mathbf{s})$ is a zero-mean latent Gaussian process and $\gamma$ is a scale parameter. The covariance function of $w(\mathbf{s})$ is $\operatorname{cov}(w(\mathbf{s}), w(\mathbf{s}')) = \rho_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}')$, with $\rho_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}')$ a valid correlation function parametrized by the vector $\boldsymbol{\theta}$. The model parameter vector is $\Psi = (\mu, \boldsymbol{\beta}', \sigma_\varepsilon^2, \gamma, \boldsymbol{\theta}', \boldsymbol{\vartheta}')$ and it completely characterizes the GPM.

Note that, for the reasons discussed later on in the paper, it is assumed that, conditionally to the observed covariates $\mathbf{x}(\mathbf{s})$, the observed $y(\mathbf{s}; \mathcal{S})$ is not preferentially sampled [see Diggle, Menezes and Su (2010)] with respect to the latent variable $w$. As a consequence, the set of spatial locations $\mathcal{S}$ is treated as a constant rather than as the realization of a spatial point process, the local density of which is driven by $w$.

In order to have a better insight into the role of the interaction function $h_{\boldsymbol{\vartheta}}$, the following family of interaction functions is adopted:

$$h_{\boldsymbol{\vartheta}}(u(\mathbf{s}), \mathbf{s}, \mathcal{S}) = u(\mathbf{s}) \cdot \left(1 + \sum_{\mathbf{s}' \in \mathcal{S}} f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}')\right)^{-1} = u(\mathbf{s}) \cdot g_{\boldsymbol{\vartheta}}(\mathbf{s}; \mathcal{S}), \tag{3.2}$$

where $f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}') : \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}^+$ is a generic nonnegative binary function.

The function $f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}')$ can be any continuous function but, for practical applications, it should be monotonically decreasing with respect to distance. For instance,

$$f_{\boldsymbol{\vartheta}}(\mathbf{s}, \mathbf{s}') = f_{\boldsymbol{\vartheta}}(\|\mathbf{s} - \mathbf{s}'\|) = \exp\left(-\frac{\|\mathbf{s} - \mathbf{s}'\|}{\phi}\right)^\alpha, \tag{3.3}$$

where $\|\cdot\|$ is the Euclidean distance and $\boldsymbol{\vartheta} = (\phi, \alpha)'$ is the function parameter vector. In equation (3.3), $\phi$ defines the strength of the interaction while $\alpha > 0$ is a shape parameter.

Note that

$$\begin{aligned} y(\mathbf{s}; \mathcal{S}) &= u(\mathbf{s}) \cdot g_{\boldsymbol{\vartheta}}(\mathbf{s}; \mathcal{S}) \\ &= q(\mathbf{s}) \cdot g_{\boldsymbol{\vartheta}}(\mathbf{s}; \mathcal{S}) + \varepsilon(\mathbf{s}) \cdot g_{\boldsymbol{\vartheta}}(\mathbf{s}; \mathcal{S}) \\ &= q(\mathbf{s}; \mathcal{S}) + \varepsilon(\mathbf{s}; \mathcal{S}), \end{aligned} \tag{3.4}$$

namely, the observed potential is equal to the conditional potential $q(\mathbf{s}; \mathcal{S})$ plus a transformation of the error $\varepsilon(\mathbf{s})$. In particular, the second line of equation (3.4) follows directly from the second stage of model (3.1), while the third line is the second line rewritten in a more compact notation.

The term $g_{\boldsymbol{\vartheta}}(\mathbf{s}; \mathcal{S})$ is the key element of the interaction function and it deserves more explanation. If, as an example, the function (3.3) is considered and $\mathcal{S} \equiv \varnothing$, namely, if there are no measuring instruments, then $g_{\vartheta}(\mathbf{s};\mathcal{S})=1$ since the summand in equation (3.2) cannot be evaluated and it is equal to zero by definition. When a measuring instrument is added, $\mathcal{S}=\{\mathbf{s}_1\}$, the potential at $\mathbf{s}$ is a function of the distance between $\mathbf{s}$ and $\mathbf{s}_1$. In particular, if $\mathbf{s}=\mathbf{s}_1$, then $g_{\vartheta}(\mathbf{s};\mathcal{S})=0.5$. On the contrary, when $\|\mathbf{s}-\mathbf{s}_1\|\to\infty$, then $g_{\vartheta}(\mathbf{s};\mathcal{S})\to 1$. This reflects the fact that the action of absorbing the potential at site $\mathbf{s}_1$ influences the measure at the site $\mathbf{s}$. It is worth noting that $\mathbf{s}$ and $\mathbf{s}_1$ are exchangeable in the sense that absorbing and measuring the potential are equivalent actions and that the potential cannot be measured without being absorbed.

In this work, the measuring instruments are supposed to be *equally-effective*, that is, $g(\mathbf{s}_i;\{\mathbf{s}_j\})=g(\mathbf{s}_j;\{\mathbf{s}_i\})$ for all $\|\mathbf{s}_i-\mathbf{s}_j\|$. The property of equally-effectiveness is satisfied if the binary function $f_{\vartheta}(\mathbf{s},\mathbf{s}')$ is commutative,[1] which is the case of the function (3.3). In practical applications, the property may not be satisfied in the sense that a measuring instrument might be more effective in absorbing the potential than a second instrument close to it. Suppose equally-effective measuring instruments, however, simplify the model and any discrepancy from it is accounted for by the error term $\varepsilon$. Note that the measure of effectiveness is strictly related to the measure of attractiveness of the *spatial behavior of consumers models* typical of the geomarketing literature [see Cliquet (2006)].

To better understand the notions of potential and conditional potential, the following example is considered. Suppose that four measuring instruments are located at $\mathbf{s}_1=(0.2,0.2)$, $\mathbf{s}_2=(0.2,0.8)$, $\mathbf{s}_3=(0.8,0.2)$ and $\mathbf{s}_4=(0.8,0.8)$, $\mathbf{s}_i\in\mathcal{D}\equiv[0,1]\times[0,1]$, $i=1,\dots,4$. The GPM considered is

$$\begin{aligned} y(\mathbf{s};\mathcal{S}) &= u(\mathbf{s})\cdot g_{\vartheta}(\mathbf{s};\mathcal{S})=q(\mathbf{s};\mathcal{S}),\\ u(\mathbf{s}) &= q(\mathbf{s}),\\ q(\mathbf{s}) &= w(\mathbf{s}), \end{aligned} \tag{3.5}$$

namely, it is supposed that the conditional potential $q(\mathbf{s};\mathcal{S})$ is observed without error. Furthermore, suppose that

$$\rho_{\boldsymbol{\theta}}(\mathbf{s},\mathbf{s}')=\rho_{\theta}(\|\mathbf{s}-\mathbf{s}'\|)=\exp\left(-\frac{\|\mathbf{s}-\mathbf{s}'\|}{0.8}\right), \tag{3.6}$$

$$f_{\vartheta}(\mathbf{s},\mathbf{s}')=f_{\vartheta}(\|\mathbf{s}-\mathbf{s}'\|)=\exp\left(-\frac{\|\mathbf{s}-\mathbf{s}'\|}{0.3}\right) \tag{3.7}$$

and that $y(\mathbf{s}_i;\mathcal{S}\setminus\mathbf{s}_i)=10$.[2] The estimated potential and conditional potential are reported in the left and in the right parts of Figure 1, respectively. Regarding the potential, its value at the measuring instrument locations is equal to $13.2>10$. Each measuring instrument measures a potential equal to 10 since a fraction of it is ab-

[1] The binary function $f$ is commutative if $f(x,y)=f(y,x)$.

[2] Note that, in general, $y(\mathbf{s};\mathcal{S})$ should be simulated following equation (4.12). In this case, in order to better appreciate the role of the interaction function, $y(\mathbf{s}_i;\mathcal{S}\setminus\mathbf{s}_i)$ is supposed to be equal for all the locations.

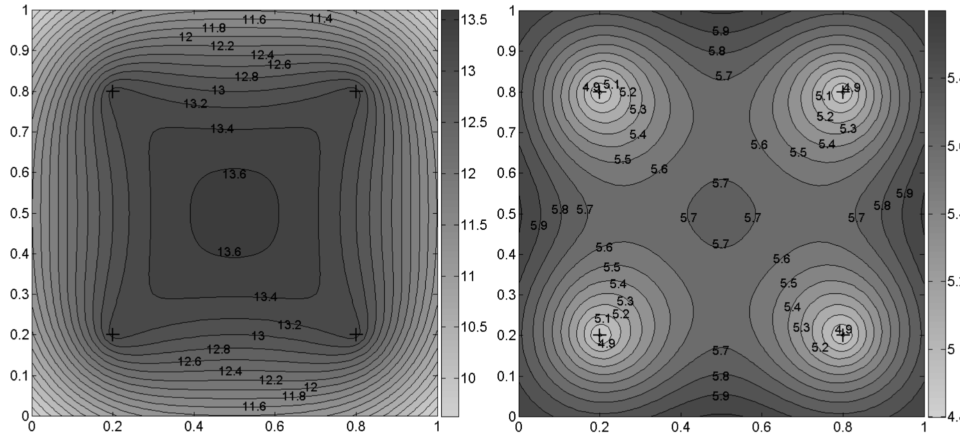

FIG. 1. (*Left*) *potential* $q(\mathbf{s})$; (*right*) *conditional potential* $q(\mathbf{s};\mathcal{S})$.

sorbed by the remaining measuring instruments. Indeed, the potential $q(\mathbf{s}_i) = 13.2$ would be measured by the single measuring instrument if the other instruments were not present. The conditional potential, as expected, has its lowest value at the measuring instrument locations and represents the potential that would be observed by a fifth measuring instrument if placed at the generic $\mathbf{s}$.

**4. Parameter estimation and inference.** Let $\mathbf{y} \equiv \mathbf{y}(\mathcal{S})$ be the $N \times 1$ vector of data collected at the sampling sites $\mathcal{S}$. The measurement equation for the vector $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{w} + \boldsymbol{\varepsilon}), \tag{4.1}$$

where $\mathbf{1}$ is the $N \times 1$ vector of ones, $\mathbf{X} \equiv \mathbf{X}(\mathcal{S})$ is the $N \times b$ matrix of covariates, $\mathbf{w} \equiv \mathbf{w}(\mathcal{S})$ is the latent Gaussian process at $\mathcal{S}$ with variance–covariance matrix $\boldsymbol{\Sigma}_{\mathbf{w}} \equiv \boldsymbol{\Sigma}_{\mathbf{w}}(\mathcal{S},\boldsymbol{\theta})$ and $\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}(\mathcal{S})$ is the measurement error at $\mathcal{S}$ with diagonal variance–covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma_\varepsilon^2 I_N$. Finally, $\mathbf{G} \equiv \mathbf{G}_{\boldsymbol{\vartheta}}(\mathcal{S})$ is the $N \times N$ diagonal matrix whose diagonal vector is

$$\mathbf{g} = \bigl(g_{\boldsymbol{\vartheta}}(\mathbf{s}_1;\mathcal{S}\setminus\mathbf{s}_1),\dots,g_{\boldsymbol{\vartheta}}(\mathbf{s}_N;\mathcal{S}\setminus\mathbf{s}_N)\bigr).$$

Furthermore, suppose that $\mathcal{S}$ is partitioned as $\{\mathcal{S}^{(1)},\mathcal{S}^{(2)}\}$, where $\mathcal{S}^{(1)}$ is the set of sites where the data are available and $\mathcal{S}^{(2)}$ is the set of sites where the data are missing. According to this, the vector $\mathbf{y}$ is partitioned as $\mathbf{y}^* = (\mathbf{y}^{(1)},\mathbf{y}^{(2)})'$, where $\mathbf{y}^{(1)} = \mathbf{L}\mathbf{y}$ is the subvector of the nonmissing data and $\mathbf{L}$ is the appropriate elimination matrix. The vector $\mathbf{y}^*$ is a permutation of $\mathbf{y}$ and $\mathbf{y} = \mathbf{D}\mathbf{y}^*$, with $\mathbf{D}$ the proper commutation matrix. The partitioned measurement equation becomes

$$\mathbf{y}^{(i)} = \mathbf{G}^{(i)}\bigl(\mathbf{1}^{(i)}\mu + \mathbf{X}^{(i)}\boldsymbol{\beta} + \gamma\mathbf{w}^{(i)} + \boldsymbol{\varepsilon}^{(i)}\bigr); \qquad i = 1,2,$$

and the variance–covariance matrix of the permuted errors is conformably partitioned as

$$\mathrm{Var}\big[\big(\boldsymbol{\varepsilon}^{(1)}, \boldsymbol{\varepsilon}^{(2)}\big)'\big] = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}.$$

In the sequel, given $\mathbf{b}$ a generic vector and $\mathbf{B}$ a generic matrix, $\mathbf{b}^{(1)}$ and $\mathbf{B}^{(1)}$ will stand for $\mathbf{Lb}$ and $\mathbf{LBL}'$, respectively, bearing in mind that, in general, $\mathbf{LB}^{-1}\mathbf{L}' \neq (\mathbf{LBL}')^{-1}$.

Given the data vector $\mathbf{y}$ and considering the GPM, the following inferential problems are of interest:

(1) to provide an estimate of the model parameter vector $\Psi$;
(2) to provide confidence intervals for the elements of $\hat{\Psi}$;
(3) to estimate the potential $q(\mathbf{s})$ over the region $\mathcal{D}$ and its uncertainty;
(4) to estimate the conditional potential $q(\mathbf{s}; \mathcal{S})$ over the region $\mathcal{D}$ and its uncertainty;
(5) to evaluate the expected total potential measured by a maximum of measuring instruments.

4.1. *Parameter estimation.* Problem 1 is tackled here following the maximum likelihood (ML) approach. With $w(\mathbf{s})$ being a latent process and due to possible missing data, the expectation–maximization (EM) algorithm is adopted to find the ML estimate $\hat{\Psi}$ of $\Psi$.

The EM algorithm is based on the complete-data likelihood function $L_\Psi(\mathbf{y}, \mathbf{w})$ and it provides an iterative procedure to update the model parameter estimate from $\hat{\Psi}^{(k)}$ to $\hat{\Psi}^{(k+1)}$ until convergence [see McLachlan and Krishnan (2008)]. In particular, for each iteration of the algorithm, the E-step computes the conditional expectation

$$Q\big(\Psi, \hat{\Psi}^{(k)}\big) = E_{\hat{\Psi}^{(k)}}\big[L_\Psi(\mathbf{y}, \mathbf{w}) \mid \mathbf{y}^{(1)}\big],$$

while, at the M-step, the following maximization is carried out:

$$\hat{\Psi}^{(k+1)} = \arg\max_{\Psi} Q\big(\Psi, \hat{\Psi}^{(k)}\big),$$

which is equivalent to solve the equation

$$\frac{\partial Q(\Psi, \hat{\Psi}^{(k)})}{\partial \Psi} = \mathbf{0}. \tag{4.2}$$

Considering the approach described in Fassò and Finazzi (2011), the following closed form updating formulas have been derived:

$$\hat{\mu}^{(k+1)} = \frac{\mathrm{tr}[(\hat{\mathbf{e}}^{(1)} + \mu^{(k)}\mathbf{1}^{(1)})(\mathbf{1}^{(1)})']}{N - N_m}, \tag{4.3}$$

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \big[(\mathbf{X}^{(1)})'\mathbf{X}^{(1)}\big]^{-1}(\mathbf{X}^{(1)})' \cdot \big(\hat{\mathbf{e}}^{(1)} + \mathbf{X}^{(1)}\boldsymbol{\beta}^{(k)}\big), \tag{4.4}$$

$$(\hat{\sigma}_{\varepsilon}^{2})^{(k+1)} = \frac{1}{N} \operatorname{tr}\begin{pmatrix} \hat{\mathbf{e}}^{(1)} \cdot (\hat{\mathbf{e}}^{(1)})' + (\gamma^{(k)})^{2}\hat{\mathbf{A}}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{pmatrix}, \tag{4.5}$$

$$\hat{\gamma}^{(k+1)} = \frac{\operatorname{tr}[(\hat{\mathbf{e}}^{(1)} + \gamma^{(k)}\hat{\mathbf{w}}^{(1)})(\hat{\mathbf{w}}^{(1)})']}{\operatorname{tr}[\hat{\mathbf{w}}^{(1)}(\hat{\mathbf{w}}^{(1)})' + \hat{\mathbf{A}}^{(1)}]}, \tag{4.6}$$

where $\hat{\mathbf{e}}^{(1)} = (\mathbf{G}^{(1)})^{-1}\mathbf{y}^{(1)} - \mu^{(k)}\mathbf{1}^{(1)} - \mathbf{X}^{(1)}\boldsymbol{\beta}^{(k)} - \gamma^{(k)}\hat{\mathbf{w}}^{(1)}$, $N_m$ is the number of missing data in $\mathbf{y}$ and

$$\hat{\mathbf{w}} = E_{\Psi^{(k)}}(\mathbf{w} \mid \mathbf{y}^{(1)}), \tag{4.7}$$

$$\hat{\mathbf{A}} = \operatorname{Var}_{\Psi^{(k)}}(\mathbf{w} \mid \mathbf{y}^{(1)}) \tag{4.8}$$

are the estimated latent variable and the estimation variance, respectively. The evaluation of (4.7) and (4.8) is reported in Appendix A.

The remaining model parameters can be updated by numerical optimization solving $(\hat{\boldsymbol{\theta}}^{(k+1)}, \hat{\boldsymbol{\vartheta}}^{(k+1)}) = \arg\max_{\boldsymbol{\theta},\boldsymbol{\vartheta}} Q(\Psi, \hat{\Psi}^{(k)})$. If both the correlation function $\rho_{\boldsymbol{\theta}}$ and the interaction function $h_{\boldsymbol{\vartheta}}$ have analytical form of the first and second derivative with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$, respectively, both the parameters can be updated by adapting the algorithm given in Xu and Wikle (2007).

Before concluding the paragraph, a point that is worth mentioning is how the preferential sampling problem can affect the model parameter estimation in the case of the GPM. The spatial data $\mathbf{y}(\mathcal{S})$ are preferentially sampled with respect to the potential $q(\mathbf{s})$ if the spatial pattern of $\mathcal{S}$ is not independent of $q(\mathbf{s})$. In practice, the spatial density of $\mathcal{S}$ can be higher at the spatial locations where $q(\mathbf{s})$ is known or expected to be high.

As discussed in Diggle, Menezes and Su (2010), if the data are preferentially sampled and the issue is not addressed, then the estimation of the parameter $\boldsymbol{\theta}$ related to the latent variable $w(\mathbf{s})$ is generally biased. With respect to the GPM, however, the model considered in Diggle, Menezes and Su (2010) does not include covariates. If the data are preferentially sampled with respect to the potential $q(\mathbf{s})$ but the covariates explain a good part of the variability of the potential, then $w(\mathbf{s})$ models only the "residual" random field $\tilde{e}(\mathbf{s}) = q(\mathbf{s}) - \mathbf{x}(\mathbf{s})\hat{\boldsymbol{\beta}}$ and the data $\mathbf{y}(\mathcal{S}) - \mathbf{X}(\mathcal{S})\hat{\boldsymbol{\beta}}$ can be assumed to be not preferentially sampled with respect to $\tilde{e}(\mathbf{s})$. In other words, even when the data are preferentially sampled with respect to $q(\mathbf{s})$, the adoption of good (spatial) covariates largely mitigates the problem. If no covariates are available and the data are suspected to be preferentially sampled, then the approach in Diggle, Menezes and Su (2010) should be considered.

4.2. *Parameter confidence intervals.* As known, the classic EM algorithm does not provide information about the uncertainty of the estimated parameter vector $\hat{\Psi}$. In order to avoid the more cumbersome supplemented EM algorithm [see Meng and Rubin (1991)], two methods are proposed to solve problem 2 of the above list, namely, to provide confidence intervals for the elements of $\hat{\Psi}$.

The first method is based on the fact that the maximum likelihood estimator has asymptotically normal distribution $N(\Psi_0, \mathbf{I}^{-1})$, with $\Psi_0$ the "true" value of $\Psi$ and $\mathbf{I}$ the Fisher information matrix. An approximation of the information matrix for multivariate normal variables can be evaluated as

$$
\begin{aligned}
\tilde{\mathbf{I}}_{ij} &= \partial_i \boldsymbol{\epsilon}' \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_j \boldsymbol{\epsilon} + \tfrac{1}{2}\operatorname{tr}\bigl(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_j \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\bigr) \\
&\quad + \tfrac{1}{4}\operatorname{tr}\bigl(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\bigr)\operatorname{tr}\bigl(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1} \partial_j \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\bigr)
\end{aligned} \tag{4.9}
$$

[see Shumway and Stoffer (2006)], where $\partial_i \boldsymbol{\epsilon}$ and $\partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ are short notation for $\partial \boldsymbol{\epsilon}(\Psi)/\partial \Psi_i$ and $\partial \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\Psi)/\partial \Psi_i$, respectively, and $1 \le i, j \le |\Psi|$.

In the case of the GPM, the vector

$$
\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta}) \tag{4.10}
$$

is normally distributed with variance–covariance matrix

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} &= \operatorname{Var}\bigl(\mathbf{y} - \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta})\bigr) \\
&= \operatorname{Var}(\gamma \mathbf{G}\mathbf{w} + \mathbf{G}\boldsymbol{\varepsilon}) \\
&= \mathbf{G}(\gamma^2 \boldsymbol{\Sigma}_{\mathbf{w}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\mathbf{G}' \\
&= \mathbf{g}\mathbf{g}' \odot \bigl(\gamma^2 \boldsymbol{\Sigma}_{\mathbf{w}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\bigr),
\end{aligned} \tag{4.11}
$$

where $\odot$ is the Hadamard product operator. The solution for the derivatives $\partial_i \boldsymbol{\epsilon}$ and $\partial_i \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is reported in Appendix B. In the presence of missing data, equation (4.9) is still valid, but $\boldsymbol{\epsilon}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ have to be replaced with $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{(1)}$, respectively.

With $\tilde{\mathbf{I}}$ available, approximated confidence intervals for the elements of $\hat{\Psi}$ are immediately provided by considering $N(\hat{\Psi}, \tilde{\mathbf{I}}^{-1})$. Note, however, that $N(\hat{\Psi}, \tilde{\mathbf{I}}^{-1})$ is a good approximation of the distribution $[\Psi \mid \mathbf{y}(\mathcal{S})]$ only when $N$ is large, which may not be the case in practical applications.

To solve this problem, following Fassò and Cameletti (2010), a second method based on the bootstrap technique is considered. Let $\hat{\Psi}$ be the estimated parameter vector. For each bootstrap run $m$, the vector $\mathbf{y}_{(m)} = \mathbf{D}[\mathbf{y}_{(m)}^{(1)} \;\; \mathbf{y}^{(2)}]'$ is considered, where

$$
\mathbf{y}_{(m)}^{(1)} = \mathbf{L} \cdot \mathbf{G} \cdot (\mathbf{1}\hat{\mu} + \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\gamma}\tilde{\mathbf{w}}_{(m)} + \tilde{\boldsymbol{\varepsilon}}_{(m)}) \tag{4.12}
$$

and where $\tilde{\mathbf{w}}_{(m)}$ and $\tilde{\boldsymbol{\varepsilon}}_{(m)}$ are realizations from the multivariate normal distributions $N(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\hat{\sigma}_{\varepsilon}^2))$ and $N(0, \boldsymbol{\Sigma}_{\mathbf{w}}(\hat{\boldsymbol{\theta}}))$, respectively. Note that $\mathbf{y}_{(m)}$ preserves the missing data pattern of the observed $\mathbf{y}$. The simulated $\mathbf{y}_{(m)}$ is used to estimate a new parameter vector $\hat{\Psi}_{(m)}$ through the EM algorithm and the set

$$
\hat{\Psi}_s = \{\hat{\Psi}_{(1)}, \ldots, \hat{\Psi}_{(M)}\} \tag{4.13}
$$

is considered as a sample from the distribution $[\Psi \mid \mathbf{y}(\mathcal{S})]$. If $M$ is large enough, then $\hat{\Psi}_s$ can be used to derive approximated confidence intervals for the elements of $\hat{\Psi}$ without normality assumptions.

4.3. *Potential and conditional potential estimation.* Following the plug-in approach, the estimated potential is obtained as

$$q_{\hat{\Psi}}(\mathbf{s}) = \hat{\mu} + \mathbf{x}(\mathbf{s})\hat{\boldsymbol{\beta}} + \hat{\gamma}\hat{\mathbf{w}}(\mathbf{s}); \qquad \mathbf{s} \in \mathcal{D}, \tag{4.14}$$

where $\hat{\mathbf{w}}(\mathbf{s}) = E_{\hat{\Psi}}(\mathbf{w}(\mathbf{s}) \mid \mathbf{y})$ is the kriging estimate of $\mathbf{w}(\mathbf{s})$, which is evaluated analogously to $\hat{\mathbf{w}}$ in equation (4.7).

The uncertainty of $q_{\hat{\Psi}}(\mathbf{s})$ is directly related to the uncertainty of $\hat{\Psi}$ which is expressed by $[\Psi \mid \mathbf{y}(\mathcal{S})]$. Again, approximated confidence intervals on $q_{\hat{\Psi}}(\mathbf{s})$ can be provided by repeatedly estimating $q_{\Psi}(\mathbf{s})$ with $\Psi$ extracted either from $N(\hat{\Psi}, \tilde{\mathbf{I}}^{-1})$ or from the set $\hat{\Psi}_s$ defined in equation (4.13). The estimated conditional potential is simply given by

$$q_{\hat{\Psi}}(\mathbf{s}; \mathcal{S}) = q_{\hat{\Psi}}(\mathbf{s}) \cdot g_{\hat{\boldsymbol{\vartheta}}}(\mathbf{s}; \mathcal{S}); \qquad \mathbf{s} \in \mathcal{D}.$$

Approximated confidence intervals on $q_{\hat{\Psi}}(\mathbf{s}; \mathcal{S})$ are provided following the same approach for $q_{\hat{\Psi}}(\mathbf{s})$.

4.4. *Total potential estimation.* The conditional potential $q(\mathbf{s}; \mathcal{S})$ provides information about the expected observation when a measuring instrument is placed at the generic location $\mathbf{s}$ given the existence of the other instruments. In practice, the following quantity is also of interest:

$$v = \max_{\mathcal{S} \in \mathbb{S}} \sum_{\mathbf{s}' \in \mathcal{S}} q(\mathbf{s}'; \mathcal{S} \setminus \mathbf{s}'); \qquad \mathcal{S} \neq \varnothing. \tag{4.15}$$

If, for example, $q(\mathbf{s})$ is the spatial market potential, then $v$ represents the maximum market volume for the trading area $\mathcal{D}$. Note that $v$ cannot be obtained by simply integrating $q(\mathbf{s})$ or $q(\mathbf{s}; \mathcal{S})$ over $\mathcal{D}$.

A simple way to estimate $v$ is to consider the estimated conditional potential $q_{\hat{\Psi}}(\mathbf{s}; \mathcal{S})$ with $\mathcal{S} = \varnothing$ and to sequentially populate $\mathcal{S}$ by choosing the spatial location of $\mathcal{D}$ where $q_{\hat{\Psi}}(\mathbf{s}; \mathcal{S})$ is maximum for the current $\mathcal{S}$. Following this approach, an estimation of $v$ is obtained for $|\mathcal{S}| \to \infty$. In practice, the value of $v$ stops to increase significantly after a finite number of iterations. Note that a by-product of (4.15) is the optimum $\mathcal{S}$ with respect to the maximization of $v$. When the main aim is the optimization of a retail network, however, the above approach should be adapted in order to impose a threshold on the minimum (geographic) distance between two elements of $\mathcal{S}$.

**5. Case study.** The GPM is considered here in order to estimate the market potential of an economic daily newspaper over the area of the city of Bergamo, northern Italy. The aim is to identify the areas with the highest market potential, to identify the areas where it would be profitable to open additional newsstands and to evaluate the maximum total market volume for the city.

The Italian daily newspaper market is characterized by 64 main newspaper heads with an average market volume of around 5.5 million daily copies. As far as the city of Bergamo concerns, only 16 out of 64 newspaper heads are commonly commercialized, as most of them are local heads referring to other Italian cities. The economic newspaper considered in this study represents 8% of the total sales volume for the Bergamo area in terms of daily copies. Moreover, it should be noted that the economic newspaper is of a clientele which differs from that of the most popular newspapers. This implies that the market potential of the economic newspaper is not necessarily reflected in the spatial distribution of the newsstands. In other words, conditionally to the observed covariates, the sales data are not preferentially sampled with respect to the market potential of the newspaper. This is also justified by the fact that the sale of daily newspapers represent only 20% of the total revenue of a newsstand.

The data available for the study consist of the yearly average daily number of copies sold on working days by $N = 75$ newsstands located over the Bergamo area. The sales data of 5 newsstands are unavailable though their location is known. The total daily average sales volume for the available newsstands is around 491 copies and it is believed that the maximum total volume attainable for the city of Bergamo is higher. The newsstand spatial locations are shown in Figure 2, along with the circle-plot of the average daily number of copies sold.

By considering the interpretation introduced in Section 3.1, it can be stated that the measuring system is represented by the newsstands and that the interaction between the measuring instruments is of the absorption type. In fact, once the customer has bought a copy of the newspaper, it is absorbed in the sense that the same

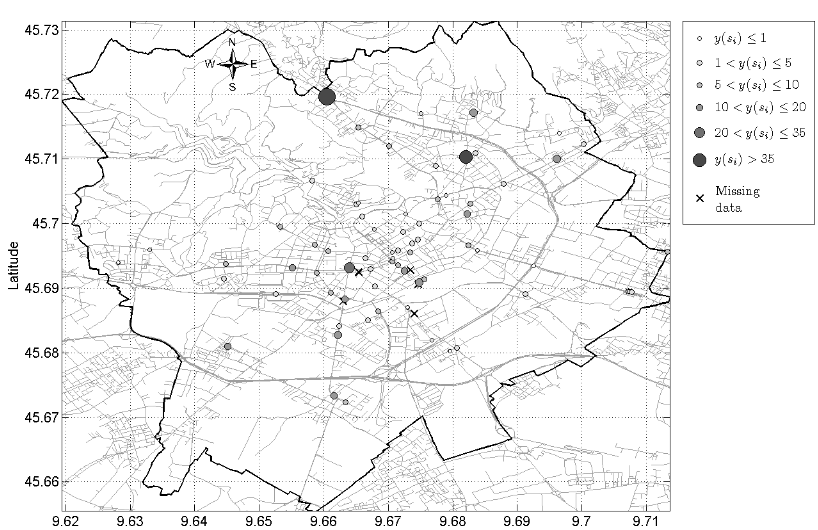


FIG. 2. *Newsstand locations and circle plot of the working day average daily number of copies sold.*

customer will not buy (during the same day) the same copy of the newspaper, neither at the same nor at a different newsstand. Since the newspaper price is fixed, the newsstands are considered equally effective and it is supposed that the customer chooses the nearest newsstand. Customer loyalty is admitted, but it is supposed that a newsstand is not more attractive than another.

The GPM model considered is the same defined in (3.1) but with $\mu \equiv 0$. This implies that the market potential goes to zero when moving far from the newsstand network, as $w(\mathbf{s})$ converges to its marginal mean which is zero. It follows that the market potential is zero (or very close to zero) over the areas where it would be unfeasible to have a newsstand. The spatial correlation function of the latent component $w$ is chosen to be

$$\rho_\theta(\mathbf{s}, \mathbf{s}') = \exp\left(-\frac{\|\mathbf{s}-\mathbf{s}'\|}{\theta}\right), \tag{5.1}$$

while the function (3.2) is considered as the interaction function, with

$$f_\phi(\mathbf{s}, \mathbf{s}') = \exp\left(\frac{-\|\mathbf{s}-\mathbf{s}'\|}{\phi}\right). \tag{5.2}$$

Two covariates are considered. The first covariate represents the spatial density of the joint-stock companies with registered offices in Bergamo. The companies are expected to induce a higher sales volume at the near newsstands. The second covariate is a function of the minimum Euclidean distance to the busiest street sections in terms of people and car traffic. In particular, the covariate value at the generic location $\mathbf{s}$ is given by $1/(d_{\min}(\mathbf{s}) + 0.1)$, where $d_{\min}(\mathbf{s})$ is the Euclidean distance (expressed in kilometers) from the location $\mathbf{s}$ to the nearest street section. Both the covariates are depicted in Figure 3. In order to make the $\boldsymbol{\beta}$ coefficients directly comparable, the covariates at the newsstand locations have been rescaled to the range [0, 1].

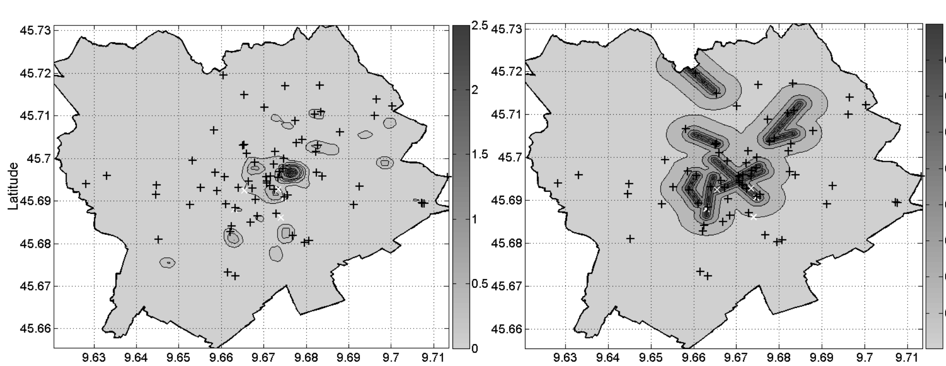


FIG. 3. *Model covariates*: (*left*) *spatial density of joint-stock companies*; (*right*) *function of the geographic distance to the nearest busy street section*.

TABLE 1
*Estimated model parameter and 95% bootstrap confidence interval*

| | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\sigma}_\varepsilon^2$ | $\hat{\gamma}$ | $\hat{\theta}$ | $\hat{\phi}$ |
|---|---|---|---|---|---|---|
| Estimated | 18.29 | 27.65 | 11.77 | 14.63 | 81.77 | 231.69 |
| LCL | −0.08 | 19.19 | 3.22 | 11.17 | 14.92 | 195.73 |
| UCL | 66.91 | 66.51 | 103.33 | 31.60 | 253.06 | 474.64 |

The model parameter vector $\Psi$ is estimated by means of the EM algorithm as discussed in Section 4.1 and by using the software provided in Finazzi (2013). The estimation result is reported in Table 1 with confidence intervals evaluated by following the bootstrap approach discussed in Section 4.2 and $M = 922$. Namely, 95% confidence intervals are obtained by evaluating empirical distributions on $\hat{\Psi}_s = \{\hat{\Psi}_{(1)}, \dots, \hat{\Psi}_{(M)}\}$. Note that the original number of bootstrap runs was $M = 1000$, but 78 runs have been ignored after testing the estimated parameters against anomalous values. In fact, the EM algorithm is not guaranteed to converge to the global maximum of the likelihood function. In this particular application, the condition $\phi > 1500m$ has been considered to identify bad estimation results, as values of $\phi$ higher than $1500m$ implies a very strong and unrealistic competition between the newsstands.

The empirical variance–covariance matrix of $\hat{\Psi}$ is reported in Table 2 and it can be compared with the approximated Hessian matrix evaluated by considering equation (4.9) and reported in Table 3. In particular, it can be noted that the approximated Hessian matrix tends to underestimate the variances related to the elements of $\hat{\Psi}$.

As expected, the $\boldsymbol{\beta}$ coefficients related to the covariates are both positive in sign. The coefficient $\beta_1$ related to the spatial density of the joint-stock companies is characterized by a larger confidence interval with lower control limit $-0.08$. Since the confidence interval is an approximation based on bootstrap runs, the covariate is retained.

TABLE 2
*Empirical variance–covariance matrix of* $\hat{\Psi}$ *based on* 922 *bootstrap runs*

| | $\beta_1$ | $\beta_2$ | $\sigma_\varepsilon^2$ | $\gamma$ | $\theta$ | $\phi$ |
|---|---|---|---|---|---|---|
| $\beta_1$ | 303.00 | 31.87 | 207.95 | 50.04 | 108.58 | 700.60 |
| $\beta_2$ | | 217.28 | 450.01 | 90.42 | 899.77 | 1091.76 |
| $\sigma_\varepsilon^2$ | | | 6180.13 | 206.79 | 4343.86 | 2725.23 |
| $\gamma$ | | | | 51.04 | 30.94 | 578.37 |
| $\theta$ | | | | | 8941.41 | 693.87 |
| $\phi$ | | | | | | 7268.92 |

TABLE 3
*Approximated Hessian matrix for $\hat{\Psi}$*

| | $\beta_1$ | $\beta_2$ | $\sigma_\varepsilon^2$ | $\gamma$ | $\theta$ | $\phi$ |
|---|---|---|---|---|---|---|
| $\beta_1$ | 158.83 | −2.97 | −46.78 | 14.78 | −23.55 | 231.79 |
| $\beta_2$ | | 66.68 | −46.53 | 14.70 | −23.42 | 230.56 |
| $\sigma_\varepsilon^2$ | | | 5274.86 | −207.60 | 2915.79 | −371.20 |
| $\gamma$ | | | | 14.10 | −99.79 | 117.26 |
| $\theta$ | | | | | 4412.62 | −186.89 |
| $\phi$ | | | | | | 1839.45 |

The estimated $\hat{\theta} \simeq 82m$ suggests that the potential $q$, net of the covariate, is not highly spatially correlated. Moreover, as supported by $\hat{\phi} \simeq 231m$, the competition between nearby newsstands is quite strong and two newsstands $200m$ apart measure/absorb (on average) only 70% of the actual market potential at their locations.

Given $\hat{\Psi}$, considering equation (4.14), the market potential $q_{\hat{\Psi}}(\mathbf{s})$ is estimated over the area of the city of Bergamo as depicted in Figure 4. For each $\mathbf{s} \in \mathcal{D}$, $q_{\hat{\Psi}}(\mathbf{s})$ provides the daily average newspaper number of copies that would be sold by a newsstand if placed at $\mathbf{s}$ without any other newsstand in $\mathcal{D}$. The maxima of $q_{\hat{\Psi}}(\mathbf{s})$ correspond to commercially strategic areas that should be served by at least one newsstand. The global maximum is equal to 79.86 yearly average newspaper copies and it is located at 45.6930° latitude and 9.6640° longitude. The estimated

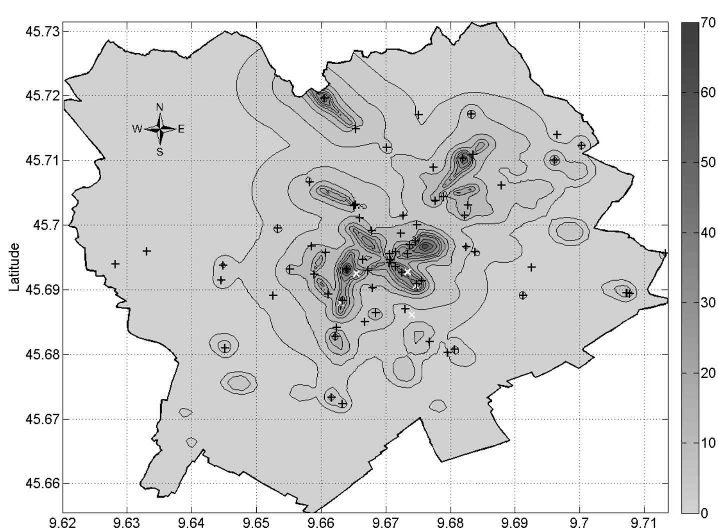


FIG. 4. *Estimated potential $q_{\hat{\Psi}}(\mathbf{s})$ (average daily number of copies) over the area of the city of Bergamo.*

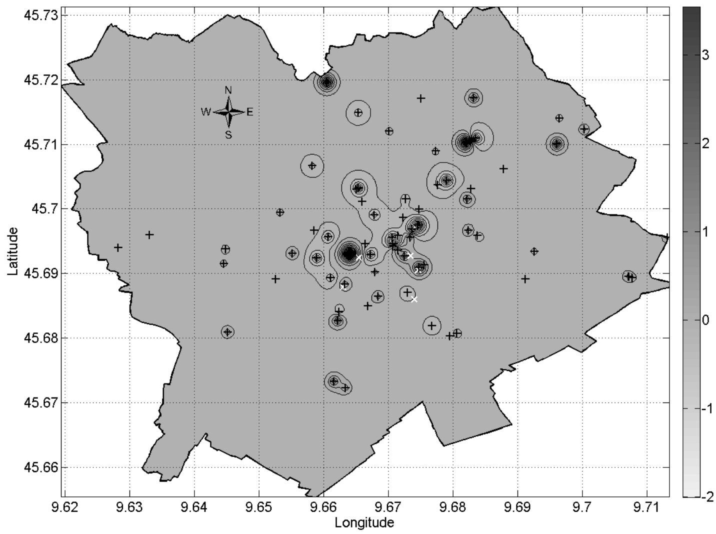


FIG. 5. *Estimated latent variable $\hat{w}(\mathbf{s})$ over the area of the city of Bergamo.*

latent variable $\hat{w}(\mathbf{s})$ is displayed in Figure 5. It can be noted that its role is more pronounced in the city center where, apparently, the covariates are less capable of explaining the observed market potential. The estimated conditional market potential $q_{\hat{\Psi}}(\mathbf{s};\mathcal{S})$, depicted in Figure 6, provides the daily average newspaper number of copies that would be sold by a newsstand if placed at $\mathbf{s}$ given the current news-

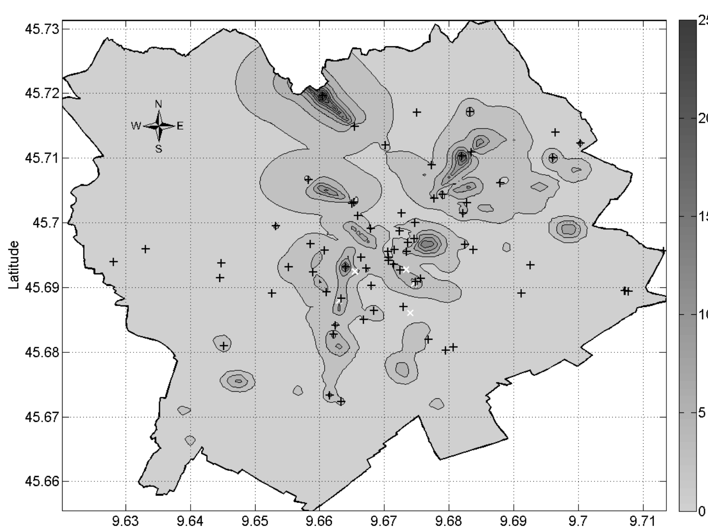


FIG. 6. *Estimated conditional potential $q_{\hat{\Psi}}(\mathbf{s};\mathcal{S})$ (average daily number of copies) over the area of the city of Bergamo.*

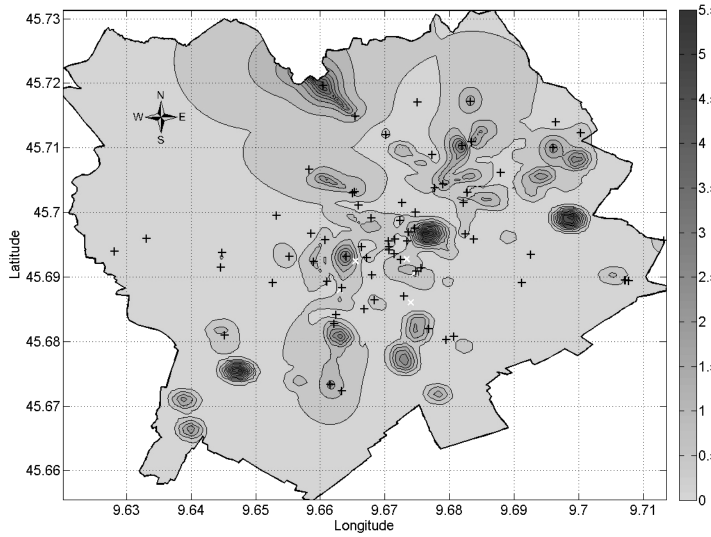


FIG. 7. *Bootstrap standard deviation map of the estimated conditional potential* $q_{\hat{\Psi}}(\mathbf{s}; \mathcal{S})$.

stands located at $\mathcal{S}$. Thus, the maxima of $q_{\hat{\Psi}}(\mathbf{s}; \mathcal{S})$ represent the spatial locations where it would be profitable to open additional newsstands. Finally, the bootstrap standard deviation of the estimated conditional market potential, representing its uncertainty, is shown in Figure 7.

The total market volume related to the economic newspaper and the Bergamo area has been evaluated following the procedure discussed in Section 4.4. Figure 8 shows the total market volume and its 95% confidence interval with respect to the number of newsstands. Note that the market volume stabilizes at around 688 average daily copies after 200 newsstands. This is a consequence of the fact that 200 newsstands absorb most of the market potential and adding more newsstands does not increase the total market volume significantly. Also, note that the optimized retail network of 75 newsstands absorbs a market volume equal to 646 copies, which corresponds to 94% of the maximum market volume and is 30.8% higher than the market volume absorbed by the actual retail network.

In light of this result, the publisher of the economic daily newspaper can consider improving the retail network in order to increase the daily market volume. Additional newsstands can be opened only with the consent of the municipal authority and, since the economic newspaper represents a small part of the daily revenue of a newsstand, it cannot be guaranteed that the additional newsstands would be opened where the market potential of the newspaper is high. Nevertheless, the Italian market of daily newspapers is undergoing a liberalization phase and the publisher should start thinking of new forms of retailing.

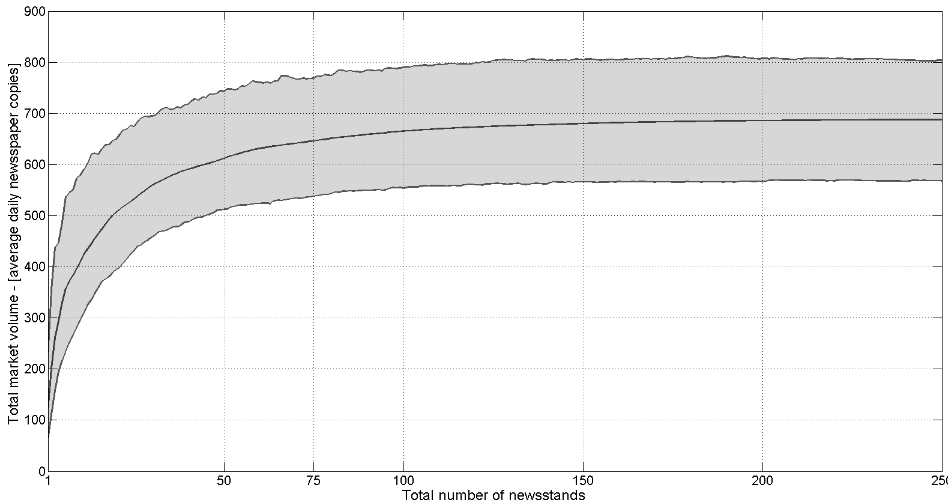


FIG. 8. *Total market volume with respect to the number of newsstands and* 95% *confidence interval based on* 922 *bootstrap runs*.

**6. Conclusions.** The geostatistical potential model has been proven to be an essential statistical tool for the estimation of the spatial market potential of a retail product from its sales data. The model output is immediately and easily interpretable (uncertainty included), as it is provided in the form of spatially continuous surfaces and with the same unit of measure of the original data.

The geostatistical potential model has been successfully applied to the estimation of the spatial market potential and to the total market volume of an economic daily newspaper for the city of Bergamo, Italy. The analysis of the results allows us to conclude that the daily sales volume can be significantly increased by focusing on the areas of the city which are characterized by a high market potential but they are not properly covered by the retail network.

Future extensions of the model include the introduction of the time variable, in order to describe and study the temporal fluctuations of the spatial market potential, and the relaxation of the equally-effectiveness property, in order to address the case of stores characterized by a different degree of attractiveness.

As a final remark, it is worth noting that the geostatistical potential model can be applied outside the geomarketing field. For instance, the sales data of the chemists of a city with respect to a drug can be analyzed in order to assess the diffusion of a disease in terms of a spatially continuous surface. More generally, the model can be applied in the case where the data related to a set of statistical units are available in aggregated form (e.g., due to privacy reasons) but they are georeferenced with respect to precise points in space.

## APPENDIX A: LATENT VARIABLE ESTIMATION

The Gaussian latent variable $\mathbf{w}$ is estimated by applying the usual formulas of the multivariate normal distribution. In particular,

$$\begin{aligned}\hat{\mathbf{w}} &= E_{\Psi^{(k)}}(\mathbf{w} \mid \mathbf{y}) \\ &= \boldsymbol{\Sigma}_{\mathbf{wy}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}[\mathbf{y} - E(\mathbf{y})] \\ &= \boldsymbol{\Sigma}_{\mathbf{wy}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}[\mathbf{y} - \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta})],\end{aligned} \tag{A.1}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{y}} &= \mathrm{Var}[\mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{w} + \boldsymbol{\varepsilon})] \\ &= \mathbf{G}\,\mathrm{Var}[\gamma\mathbf{w} + \boldsymbol{\varepsilon}]\mathbf{G}' \\ &= \mathbf{G}(\gamma^2\boldsymbol{\Sigma}_{\mathbf{w}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\mathbf{G}'\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{wy}} &= E[(\mathbf{w} - \mathbf{0}) \cdot [\mathbf{y} - E(\mathbf{y})]'] \\ &= E[\mathbf{w} \cdot (\gamma\mathbf{G}\mathbf{w})'] \\ &= \gamma\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{G}'.\end{aligned}$$

The variance of the estimated $\hat{\mathbf{w}}$ is given by

$$\begin{aligned}\hat{\mathbf{A}} &= \mathrm{Var}_{\Psi^{(k)}}(\mathbf{w} \mid \mathbf{y}) \\ &= \boldsymbol{\Sigma}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{wy}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}(\boldsymbol{\Sigma}_{\mathbf{wy}})'.\end{aligned} \tag{A.2}$$

When $\mathbf{y}$ is characterized by missing data, equations (A.1) and (A.2) become

$$\begin{aligned}\hat{\mathbf{w}} &= (\boldsymbol{\Sigma}_{\mathbf{wy}}\mathbf{L}')(\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}')^{-1}[\mathbf{L}(\mathbf{y} - \mathbf{G}(\mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta}))], \\ \hat{\mathbf{A}} &= \boldsymbol{\Sigma}_{\mathbf{w}} - (\boldsymbol{\Sigma}_{\mathbf{wy}}\mathbf{L}')(\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}')^{-1}(\mathbf{L}\Sigma_{\mathbf{wy}}).\end{aligned}$$

## APPENDIX B: VECTOR AND MATRIX DERIVATIVES

The evaluation of the approximate Fisher information matrix defined in equation (4.9) requires the computation of the vector derivatives $\partial\boldsymbol{\epsilon}(\Psi)/\partial\Psi_i$ and the matrix derivatives $\partial\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\Psi)/\partial\Psi_i$, $1 \le i \le |\Psi|$, with $\boldsymbol{\epsilon}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ defined in equations (4.10) and (4.11), respectively.

In the case of the spatial correlation function defined in equation (5.1) and the interaction function defined in equation (5.2), the following derivatives hold:

$$\frac{\partial\boldsymbol{\epsilon}(\Psi)}{\partial\Psi_i} = \begin{cases} -\mathbf{g}, & \text{if } \Psi_i = \mu, \\ -\mathbf{g} \odot \mathbf{x}_l, & \text{if } \Psi_i = \beta_l;\ 1 \le l \le b, \\ -\partial\mathbf{g}_\phi \odot (\mathbf{1}\mu + \mathbf{X}\beta), & \text{if } \Psi_i = \phi, \\ \mathbf{0}, & \text{otherwise,} \end{cases} \tag{B.1}$$

$$
\text{(B.2)} \qquad \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\Psi)}{\partial \Psi_i} = \begin{cases} \mathbf{g}\mathbf{g}' \odot I_N, & \text{if } \Psi_i = \sigma_\varepsilon^2, \\ 2\gamma \mathbf{g}\mathbf{g}' \odot \boldsymbol{\Sigma}_{\mathbf{w}}, & \text{if } \Psi_i = \gamma, \\ \gamma^2 \mathbf{g}\mathbf{g}' \odot \dfrac{\mathbf{H}}{\theta^2} \odot \boldsymbol{\Sigma}_{\mathbf{w}}, & \text{if } \Psi_i = \theta, \\ \tilde{\mathbf{G}} \odot (\gamma^2 \boldsymbol{\Sigma}_{\mathbf{w}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}), & \text{if } \Psi_i = \phi, \\ \mathbf{0}, & \text{otherwise,} \end{cases}
$$

where $\mathbf{x}_l$ is the $l$th column of the matrix $\mathbf{X}$ and $\mathbf{H}$ is the distance matrix based on $\mathcal{S}$.

Finally, the $(pq)$th element of the matrix $\tilde{\mathbf{G}}$ is given by $\partial g_p \cdot g_q + g_p \partial g_q$, where $g_p$ is the $p$th element of the vector $\mathbf{g}$ and while the $p$th element of the vector $\partial \mathbf{g}_\phi$ is given by

$$
\text{(B.3)} \qquad \partial g_p \equiv \frac{\partial g_p}{\partial \phi} = - \frac{\sum_{q \neq p}^N (h_{pq}/\phi^2) \exp(-\|\mathbf{s}_p - \mathbf{s}_q\|/\phi)}{[\sum_{q \neq p}^N \exp(-\|\mathbf{s}_p - \mathbf{s}_q\|/\phi)]^2}
$$

with $h_{pq}$ the $(pq)$th element of the matrix $\mathbf{H}$.

**Acknowledgments.** Special thanks to Alberto Saccardi of Nunatac Srl for providing the data, to Ilaria Cremonesi for the preliminary data analysis and to Alessandro Fassò whose advice is always precious.

## SUPPLEMENTARY MATERIAL

**Data set and Matlab® code** (DOI: 10.1214/12-AOAS588SUPP; .zip). Georeferentiated newsstand sales data and Matlab® code for the data analysis.

## REFERENCES

BANERJEE, S., GELFAND, A. E. and POLASEK, W. (2000). Geostatistical modelling for spatial interaction data with application to postal service performance. *J. Statist. Plann. Inference* **90** 87–105. MR1791583

CLIQUET, G. (2006). *Geomarketing: Methods and Strategies in Spatial Martketing*. ISTE, London.

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ. MR2848400

DAVIS, P. (2006). Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics* **37** 964–982.

DE GRANGE, L., IBEAS, A. and GONZALEZ, F. (2011). A hierarchical gravity model with spatial correlation: Mathematical formulation and parameter estimation. *Netw. Spat. Econ.* **11** 439–463.

DIGGLE, P. J., MENEZES, R. and SU, T.-L. (2010). Geostatistical inference under preferential sampling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **59** 191–232. MR2744471

DIGGLE, P. J. and RIBEIRO, P. J. JR. (2007). *Model-Based Geostatistics*. Springer, New York. MR2293378

FASSÒ, A. and CAMELETTI, M. (2010). A unified statistical approach for simulation, modeling, analysis and mapping of environmental data. *Simulation* **86** 139–154.

FASSÒ, A. and FINAZZI, F. (2011). Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics* **22** 735–748. MR2843140

FINAZZI, F. (2013). Supplement to "Geostatistical modeling in the presence of interaction between the measuring instruments, with an application to the estimation of spatial market potentials." DOI:10.1214/12-AOAS588SUPP.

FISCHER, M. M. and WANG, J. (2011). Spatial interaction models and spatial dependence. In *Spatial Data Analysis* 61–70. Springer, Berlin.

HUFF, D. L. (1964). Defining and estimating a trading area. *The Journal of Marketing* **28** 34–38.

MCLACHLAN, G. J. and KRISHNAN, T. (2008). *The EM Algorithm and Extensions*, 2nd ed. Wiley, Hoboken, NJ. MR2392878

MENG, X. L. and RUBIN, D. B. (1991). Using EM to obtain asymptotic variance–covariance matrices: The SEM algorithm. *J. Amer. Statist. Assoc.* **86** 899–909.

REILLY, W. J. (1931). *The Law of Retail Gravitation*. Reilly, New York.

SHUMWAY, R. H. and STOFFER, D. S. (2006). *Time Series Analysis and Its Applications: With R Examples*, 2nd ed. Springer, New York. MR2228626

XU, K. and WIKLE, C. K. (2007). Estimation of parameterized spatio-temporal dynamic models. *J. Statist. Plann. Inference* **137** 567–588. MR2298958

DEPARTMENT OF INFORMATION TECHNOLOGY
AND MATHEMATICAL METHODS
UNIVERSITY OF BERGAMO
VIALE MARCONI
5-24044 DALMINE (BG)
ITALY
E-MAIL: francesco.finazzi@unibg.it