Computational Statistics
https://doi.org/10.1007/s00180-021-01090-7

ORIGINAL PAPER

# Reduced-bias estimation of spatial autoregressive models with incompletely geocoded data

**Flavio Santi**[1] · **Maria Michela Dickson**[2] · **Diego Giuliani**[2] · **Giuseppe Arbia**[3] · **Giuseppe Espa**[2]

Received: 27 March 2020 / Accepted: 19 February 2021


## Abstract

The application of spatial Cliff–Ord models requires information about spatial coordinates of statistical units to be reliable, which is usually the case in the context of areal data. With micro-geographic point-level data, however, such information is inevitably affected by locational errors, that can be generated intentionally by the data producer for privacy protection or can be due to inaccuracy of the geocoding procedures. This unfortunate circumstance can potentially limit the use of the spatial autoregressive modelling framework for the analysis of micro data, as the presence of locational errors may have a non-negligible impact on the estimates of model parameters. This contribution aims at developing a strategy to reduce the bias and produce more reliable inference for spatial models with location errors. The proposed estimation strategy models both the spatial stochastic process and the coarsening mechanism by means of a marked point process. The model is fitted through the maximisation of a doubly-marginalised likelihood function of the marked point process, which cleans out the effects of coarsening. The validity of the proposed approach is assessed by means of a Monte Carlo simulation study under different real-case scenarios, whereas it is applied to real data on house prices.

**Keywords** Spatial statistics · Cliff–Ord modelling · Model fitting · Geocoding · Coarsening · Marginal likelihood

✉ Flavio Santi
flavio.santi@univr.it

1 Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, VR, Italy

2 Department of Economics and Management, University of Trento, Trento, TN, Italy

3 Department of Statistical Sciences, Catholic University of the Sacred Heart, Rome, Italy

Published online: 13 March 2021

## 1 Introduction

Cliff–Ord (Cliff and Ord 1969) spatial models are based on the implicit assumption that the information about the spatial location of statistical units is accurate. Whilst this circumstance is the norm in the context of areal data (such as municipalities, counties or regions), it is rarely met when the observations are points in space (such as firms, houses or facilities), whose locations may be either missing or affected by locational errors (see Zimmerman 2008; Zimmerman and Li 2010; Arbia et al. 2019b).

Although geolocation may fail for some units because of technical reasons, incomplete positioning arises more frequently in geocoding processes, especially in those circumstances where coordinates of units are obtained by matching postal addresses with georeferenced street maps (see e.g. Kravets and Hadden 2007). Clearly, the quality of the resulting geolocation depends both on the correctness and completeness of postal addresses, as well as on the effectiveness of matching algorithms and software, nonetheless, if position of some units is uncertain, this fact should be properly considered in the estimation process.

When an incomplete address is geocoded, unit position is conventionally imputed to the centroid of the area where unit is located, as it can be known from address information. Such areas may be counties, municipalities, or, more frequently, ZIP code areas (Zimmerman 2008). From a statistical point of view, the presence of locational errors due to coarsened locations may have a significant impact on parameter estimates of spatial models based on the Cliff–Ord approach (Cliff and Ord 1969), as positional errors lead to downward biased estimates for the spatial autoregressive parameters and inconsistent estimates for covariates coefficients (Arbia et al. 2016).

This paper tackles the problem of estimating spatial models where part of units is affected by coarsening. In particular, we focus on the spatial autoregressive model (see e.g. Cressie 2015, ch. 6). The proposed estimation strategy models both the spatial stochastic process and the coarsening mechanism by means of a marked point process whose intensity function is estimated according to the coarsened-data estimator proposed by Zimmerman (2008). The model is then fitted through the maximisation of a doubly-marginalised likelihood function of the marked point process, which cleans out the effects of coarsening.

The first marginalisation of the likelihood function allows the dimensionality of the model to be consistently reduced to non-coarsened points, and it is derived analytically. The second marginalisation is performed via Monte Carlo simulations over the locations of coarsened points.

The modelling approach and Monte Carlo experiments presented in the paper show the validity of the proposed estimation method compared to other estimation approaches. In particular, the comparison concerns the parameter estimates and the direct and indirect effects of model covariates on the dependent variable (Arbia et al. 2019a).

The paper is organised as follows. Section 2 describes the modelling approach and the notation adopted. Section 3 illustrates and discusses the proposed estimation approach. Section 4 illustrates the results of Monte Carlo simulations where the finite properties of parameter estimators and estimators of direct and indirect impacts of

regressors are studied. In Sect. 5 the method is applied to spatial hedonic models for house prices in Beijing. Section 6 concludes the paper.

## 2 Modelling approach and notation

Consider $n$ units $i = 1, \ldots, n$ for which a quantitative characteristic of interest $y_i \in \mathbb{R}$ and $k$ regressors $x_i \in \mathbb{R}^k$ are known. Assume that postal addresses are available for all $n$ units, however only $p < n$ of them are complete, whereas $n - p$ are incomplete. Assume also, that the $n - p$ units with incomplete addresses can be assigned to, say, the ZIP areas they actually belong to.

Under these conditions, if a spatial Cliff–Ord model is used for modelling $y$ (a thorough illustration of the reasons why a spatial modelling approach may be necessary is available e.g. in LeSage and Pace 2009, ch. 2), the coarsening of the $n - p$ unit locations only affects the specification of the spatial weight matrix, as $y_i$ and $x_i$ are known for all units $i = 1, \ldots, n$.

Consider, for example, the following isotropic spatial autoregressive model (SAR):

$$\begin{cases} y = \rho W y + X\beta + \varepsilon \\ \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n) \end{cases} \tag{1}$$

where $X \in \mathbb{R}^{n \times k}$ is the design matrix which includes $k$ regressors, and $W \in \mathbb{R}^{n \times n}$ is the usual zero-diagonal spatial weight matrix whose elements $w_{ij}$ take positive values according to some proximity criterion, whereas equal zero if units $i$ and $j$ are not considered as neighbours.

It can be verified that, if $p/n$ is the proportion of non-coarsened units, the share of elements of $W$ not affected by coarsening is only about $(p/n)^2$, whereas all elements change if $W$ is stochastic (that is, if $W$ is row-standardised).

Although the magnitude of the effects of coarsening on the spatial weight matrix is the cause of bias of estimators of the autoregressive parameter $\rho$ (Arbia et al. 2016), it is worth stressing that the bias of the estimators of $\rho$ is not just originated from perturbations in the values of weights within neighbourhoods, but it is mainly ascribable to modifications in the neighbourhood relations amongst units, as shown in Santi et al. (2020).

On the other hand, the biasedness of the estimator of the autoregressive parameter $\rho$ in model (1) gives rise to biasedness of estimators of the other parameters too. It can be proved (see Appendix A.1) that

$$\mathbb{E}(\hat{\beta} - \beta | \hat{\rho}) = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} \left( \sum_{j=1}^{\infty} \rho^j W^j \right) X\beta \, \frac{\hat{\rho} - \rho}{\rho}, \tag{2a}$$

so that the bias on $\beta_j$ (for any $j = 1, \ldots, k$) gets larger as the relative bias of $\hat{\rho}$ grows, provided that $\beta_j \neq 0$; whereas the conditional bias of $\hat{\sigma}^2$ has the form (see

Appendix A.2):

$$\mathbb{E}(\hat{\sigma}^2 - \sigma^2|\hat{\rho}) = -\frac{k}{n}\sigma^2 + \left[\frac{(Q_\rho X\beta)^{\mathrm{T}}(Q_\rho X\beta)}{n} + \sigma^2\frac{\mathrm{tr}(Q_\rho^{\mathrm{T}}Q_\rho)}{n}\right](\hat{\rho}-\rho)^2 - 2\sigma^2\frac{\mathrm{tr}(Q_\rho)}{n}(\hat{\rho}-\rho), \tag{2b}$$

where $Q_\rho \equiv (I_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})W(I_n - \rho W)^{-1}$. In this case, unlike estimator $\hat{\beta}$, the maximum likelihood estimator $\hat{\sigma}^2$ is biased even if $\hat{\rho}$ is unbiased, because of the second term on the right-hand side of Eq. (2b).

Equations (2) show that biasedness of estimator $\hat{\rho}$ reverberates on estimates of other parameters, and thus on derived quantities such as covariate impacts (see Sect. 3.2), confidence intervals, and prediction intervals. This motivates the adoption of estimation methods aiming at reducing the bias of $\hat{\rho}$ originated from coarsening of unit locations.

The estimation method proposed in this paper basically reduces the dimensionality of the model by concentrating the likelihood on the $p$ non-coarsened units, thus limiting the effects of the coarsened locations on model estimates and, at the same time, exploiting the available information about covariates and zone-based location of the coarsened units.

The problem is modelled as a marked point process where both the stochastic spatial process and the coarsening process are specified conditionally on the underlying point process.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $Z \in \mathbb{R}^{n\times 2}$ be a realisation of $n$ points from a 2-dimensional point process $\{Z(s,\omega): s \in S\}$ defined over a bounded metric space $(S, \|\cdot\|)$, where $S \subset \mathbb{R}^2$. Let $\lambda: S \to \mathbb{R}^+$ be the intensity function of $\{Z(s,\omega): s \in S\}$ defined as:

$$\lambda(s) = \lim_{|\mathrm{d}s|\to 0}\frac{\mathbb{E}(N(s,\mathrm{d}s))}{\mathrm{d}s},$$

in which $N(s,\mathrm{d}s)$ is the count function for points in the neighbour $\mathrm{d}s \subset S$ centred in $s \in S$ (see e.g. Illian et al. 2008).

Conditionally on $Z$, the isotropic SAR (1) is defined for the spatial process $y$, where the spatial weight matrix $W$ is row-standardised, and its elements $w_{ij}$ are defined as follows:

$$w_{ij} = \begin{cases} \frac{\kappa(\|z_i - z_j\|)}{\sum_{v=1}^n \kappa(\|z_i - z_v\|)} & \text{if } i \neq j \text{ and } \sum_{v=1}^n \kappa(\|z_i - z_v\|) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

for any $i, j \in \{1,\ldots,n\}$, and some non-increasing function $\kappa: \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that $\lim_{x\to\infty}\kappa(x) = 0$. Common choices for $\kappa$ are $\kappa(d) = \alpha/d$, $\kappa(d) = \alpha/d^2$, $\kappa(d) = \mathrm{e}^{-\alpha d}$, $\kappa(d) = \mathrm{e}^{-\alpha d^2}$, $\alpha$ being some positive constant. Often, a cut-off distance $\bar{d}$ is also specified, so that $\kappa(d) = \mathbb{1}_{\{d\le\bar{d}\}}\alpha/d^2$. — See Anselin (1988) and Anselin (2002) for a discussion on the alternative specifications of spatial weight matrices (and

thus of the decay function $\kappa$); whereas LeSage and Pace (2014) and Santi et al. (2020) analyse the effects of spatial weight matrix misspecification.

The coarsening process can be either dependent on the intensity function $\lambda$ and the realisation of the point process $\{ Z(s, \omega) : s \in S \}$ or independent from them. Here we just assume that the coarsening is modelled by means of a random vector $\Phi$, which is a realisation of $n$ Bernoulli random variables independent from the spatial process $y$ conditionally on the point process $Z$. Thus, the components $\Phi_i$ of the random vector $\Phi$ are defined as follows:

$$\Phi_i \sim \mathcal{B}(p_i) \tag{4}$$

for $i = 1, \ldots, n$, and take value $\Phi_i = 0$ if point $i$ is coarsened, whereas $\Phi_i = 1$ if point $i$ has been correctly geocoded.

Finally, let $\mathcal{S} = \{ S_1, S_2, \ldots, S_R \}$ be a partition of the space $S$ into $R$ regions such that, for any unit $i$ with coordinate $z_i \in S$, it exists one region $S_r$ such that $z_i \in S_r$.[1] It is assumed that, for each coarsened unit $i$, the region $S_r$ where $i$ is located is known.

To sum up, for all units $i = 1, \ldots, n$ the values of the dependent variable $y_i$ and the covariates $x_i$ are known. For non-coarsened units $i = 1, \ldots, p$ the coordinates $z_i \in S$ are known, whereas it is known the coarsening area $S_r$ of each coarsened unit $i = p+1, \ldots, n$ such that $z_i \in S_r$. Other missing or unknown information such as the values of parameters and the coordinates of coarsened units about model (1) should be either learnt (through estimation) or made it non-relevant (through marginalisation).

Before illustrating our proposal for tackling the estimation problem, we introduce the notation that will be used throughout the rest of the paper.

We denote by subscript $P$ and subscript $C$ non-coarsened and coarsened points respectively (that is, points where $\Phi_i = 1$ and $\Phi_i = 0$ respectively). Conditionally on the random vector $\Phi$, SAR (1) can be restated as it follows:

$$\begin{bmatrix} y_P \\ y_C \end{bmatrix} = \rho \begin{bmatrix} W_{PP} & W_{PC} \\ W_{CP} & W_{CC} \end{bmatrix} \cdot \begin{bmatrix} y_P \\ y_C \end{bmatrix} + \begin{bmatrix} X_P \\ X_C \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_P \\ \varepsilon_C \end{bmatrix} \tag{5}$$

provided the original SAR is properly permuted by means of a suitable permutation matrix $P_\Phi \in \{0, 1\}^{n \times n}$, that is:

$$\begin{bmatrix} y_P \\ y_C \end{bmatrix} = P_\Phi y, \quad \begin{bmatrix} W_{PP} & W_{PC} \\ W_{CP} & W_{CC} \end{bmatrix} = P_\Phi W P_\Phi, \quad \begin{bmatrix} X_P \\ X_C \end{bmatrix} = P_\Phi X, \quad \begin{bmatrix} \varepsilon_P \\ \varepsilon_C \end{bmatrix} = P_\Phi \varepsilon. \tag{6}$$

Restatement (5) permits observations about coarsened ($C$) and non-coarsened ($P$) points to be organised in block matrices.

[1] In fact, this assumption is not crucial in our analysis, and can be easily generalised by assuming $\mathcal{S}$ to be a cover of $S$ such that $S \in \mathcal{S}$. This generalisation permits various degrees of incompleteness in postal addresses to be modelled, including the situation where some units are only known to be located in $S$. The estimation method proposed later can be applied with no modifications also to this framework, however, for the sake of notational simplicity, in the rest of the paper only the case where $\mathcal{S}$ is a partition of $S$ is discussed.

We also define matrix $A \equiv I_n - \rho P_\Phi W P_\Phi \in \mathbb{R}^{n \times n}$, so that:

$$A = \begin{bmatrix} A_{PP} & A_{PC} \\ A_{CP} & A_{CC} \end{bmatrix} = \begin{bmatrix} I_p - \rho W_{PP} & -\rho W_{PC} \\ -\rho W_{CP} & I_{n-p} - \rho W_{CC} \end{bmatrix}. \tag{7}$$

Finally, it can be proved (see e.g. Lu and Shou 2002) that the following relations hold for the inverse matrix $A^{-1}$:

$$A^{-1} = \begin{bmatrix} A_{PP}^{-1} + A_{PP}^{-1} A_{PC} \tilde{\Xi}^{-1} A_{CP} A_{PP}^{-1} & -A_{PP}^{-1} A_{PC} \tilde{\Xi}^{-1} \\ -\tilde{\Xi}^{-1} A_{CP} A_{PP}^{-1} & \tilde{\Xi}^{-1} \end{bmatrix} \tag{8}$$

where $\tilde{\Xi} \equiv A_{CC} - A_{CP} A_{PP}^{-1} A_{PC}$ is the Schur complement of $A_{PP}$ and

$$\begin{aligned} A^{-1} &\equiv \begin{bmatrix} (A^{-1})_{PP} & (A^{-1})_{PC} \\ (A^{-1})_{CP} & (A^{-1})_{CC} \end{bmatrix} \\ &= \begin{bmatrix} \Xi^{-1} & -\Xi^{-1} A_{PC} A_{CC}^{-1} \\ -A_{CC}^{-1} A_{CP} \Xi^{-1} & A_{CC}^{-1} + A_{CC}^{-1} A_{CP} \Xi^{-1} A_{PC} A_{CC}^{-1} \end{bmatrix} \end{aligned} \tag{9}$$

where $\Xi \equiv A_{PP} - A_{PC} A_{CC}^{-1} A_{CP}$ is the Schur complement of $A_{CC}$ (see e.g. Horn and Johnson 2013).

## 3 Estimation strategy

### 3.1 Model fitting

The reduced form of model (5) based on inversion (8) permits the following equation to be derived:[2]

$$y_P = \rho W_{PP} y_P + X_P \beta + \varepsilon_P + A_{PC} \tilde{\Xi}^{-1} \left[ A_{CP} A_{PP}^{-1} (X_P \beta + \varepsilon_P) - (X_C \beta + \varepsilon_C) \right]. \tag{10}$$

Left-hand side term of Eq. (10) together with the first three terms of the right-hand side perfectly describe a SAR amongst correctly geo-referenced points, sharing the same parameters of the complete model (1). Unfortunately, the last term on the right-hand side makes things more complicated.

The fourth term on the right-hand side of Eq. (10) proves that, in general, any subset of observations of a SAR does not follow a SAR. Indeed, it makes the estimation process of a SAR with coarsened points particularly tricky, since Eq. (10) includes blocks of matrix $A$ which depend on the (unknown) coordinates of the coarsened points.

As previously stated, the estimation strategy proposed in this paper relies on a double marginalisation of the likelihood function of the SAR (1). In particular, the

[2] See the Appendix A.3 for a proof.

former marginalisation should be made with respect to $y_P$, thus concentrating the information about coarsened points into a lower dimensional space. A similar approach to the marginalisation of the SAR has already proved to be successful in the context of variance estimation in 2-dimensional systematic sampling (see Espa et al. 2017). The latter marginalisation should instead be made with respect to the point process of non-coarsened points $Z_P$, so as to include direct and indirect effects of positional errors in the (marginal) probability distribution of $y_P$.

The first marginalisation can be derived in closed form from the reduced form of model (1) based on inversion (9), and equals

$$y_P = \Xi^{-1} X_P \beta + \Xi^{-1} \varepsilon_P - \Xi^{-1} A_{PC} A_{CC}^{-1} (X_C \beta + \varepsilon_C),$$

which implies that:

$$\mathbb{E}(y_P | Z, \Phi) = \Xi^{-1} X_P \beta + \rho\, \Xi^{-1} W_{PC} A_{CC}^{-1} X_C \beta, \tag{11a}$$

$$\operatorname{cov}(y_P | Z, \Phi) = \sigma^2\, \Xi^{-1} (I_p + \rho^2\, W_{PC} (A_{CC}^{\mathrm{T}} A_{CC})^{-1} W_{PC}^{\mathrm{T}}) (\Xi^{-1})^{\mathrm{T}}, \tag{11b}$$

so that the log-likelihood function $\ln \mathcal{L}(\rho, \beta, \sigma^2 | y, X, Z, \Phi)$ of the of the model (1) marginalised with respect to $y_P$ equals:

$$\begin{aligned} \ln \mathcal{L}(\rho, \beta, \sigma^2 | y_P, X, Z, \Phi) = & -\frac{p}{2} \ln(2\pi\sigma^2) + \ln |\Xi| \\ & -\frac{1}{2} \ln |I_p + \rho^2\, W_{PC} (A_{CC}^{\mathrm{T}} A_{CC})^{-1} W_{PC}^{\mathrm{T}}| \\ & -\frac{1}{2\sigma^2} (\Xi y_P - X\beta - \rho W_{PC} A_{CC}^{-1} X_C \beta)^{\mathrm{T}} \cdot \\ & \cdot (I_p + \rho^2\, W_{PC} (A_{CC}^{\mathrm{T}} A_{CC})^{-1} W_{PC}^{\mathrm{T}})^{-1} \cdot \\ & \cdot (\Xi y_P - X\beta - \rho W_{PC} A_{CC}^{-1} X_C \beta). \end{aligned} \tag{12}$$

The second marginalisation requires the intensity function $\lambda$ to be estimated, so as to characterise the first-order properties of the spatial point process $\{ Z(s, \omega) : s \in S \}$ and, in turn, the probabilistic law of the spatial weight matrix $W$ under coarsened geocoding.

The point process $\{ Z(s, \omega) : s \in S \}$ along with the coarsening process $\{ \Phi_i : i = 1, \ldots, n \}$ defines a bivariate point pattern (Illian et al. 2008).[3] According to Zimmerman (2008), for any $s \in S$, the intensity function $\lambda : S \to \mathbb{R}^+$ of the spatial point pattern affected by incomplete geocoding can be estimated as follows:

$$\hat{\lambda}(s) = \sum_{i=1}^{n} [\hat{\phi}(z_i)]^{-1} K_h(s - z_i), \tag{13}$$

[3] Such bivariate point process is either a Cox process or an inhomogeneous $\phi$-thinned process, according to whether $\phi$ is stochastic or not (see Illian et al. 2008, ch. 6). In this paper $\phi$ (and thus the probabilities $\mathbb{P}(\Phi_i = 1), i = 1, \ldots, n$) are treated as non-random, however all the results presented in the paper holds also if $\phi$ is a realisation of a random field (see Zimmerman 2008, for details).

where $K_h$ is some kernel function with bandwidth $h$, $z_i$ is the observed location of unit $i$, and $\hat{\phi}$ is an estimate of the geocoding propensity function $\phi\colon S \to (0, 1]$, whose reciprocal $(1/\hat{\phi})$ is used as the weighting criterion of the kernel estimator.

As for $K_h$, Zimmerman (2008) uses a Gaussian kernel whose bandwidth is automatically selected by minimising the mean-square error statistic defined in Diggle (1985) through cross-validation (Berman and Diggle 1989).

In operative terms, Zimmerman (2008) estimates the intensity function $\lambda$ through the R (R Core Team 2020) function `density.ppp` of package `spatstat` (Baddeley et al. 2015), whereas the bandwidth is computed by means of the function `bw.diggle` (of package `spatstat` as well). Monte Carlo simulations illustrated in Sect. 4 and the application to real data discussed in Sect. 5 use the same functions.

The geocoding propensity function $\phi$ can be estimated in various ways, according to the available information about the coarsening process. In this paper, the values of the coarsening probabilities in (4) are assumed to be such that $p_i = \phi(z_i)$, given the coordinate $z_i \in S$ of the unit $i$. It follows that:

$$\hat{\phi}(s) = \frac{\sum_{r=1}^{R} \sum_{i=1}^{n} \Phi_i \mathbb{1}_{\{z_i \in S_r\}} \mathbb{1}_{\{s \in S_r\}}}{\sum_{r=1}^{R} \sum_{i=1}^{n} \mathbb{1}_{\{z_i \in S_r\}} \mathbb{1}_{\{s \in S_r\}}}, \tag{14}$$

so that $\hat{\phi}$ is constant over each region $S_r \in \mathcal{S}$, and equals the proportion of non-coarsened points in $S_r$.

The solution we propose in this paper consists in five steps which are summarised in Algorithm 1.

**Algorithm 1** (Double-marginalisation estimation)

1. *the geocoding propensity function $\phi$ is estimated over $\mathcal{S}$ through estimator* (14)*;*
2. *the intensity function $\lambda$ of the coarsened point process $Z$ is estimated according to* Zimmerman (2008) *through estimator* (13)*;*
3. *the likelihood of SAR* (1) *marginalised with respect to $y_P$ is derived from* (11)*; we denote that likelihood function by $\mathcal{L}(\rho, \beta, \sigma^2 | y_P, X, Z, \Phi)$;*
4. *the likelihood $\mathcal{L}(\rho, \beta, \sigma^2 | y_P, X, Z, \Phi)$ is marginalised with respect to $Z_P$, that is:*

$$\mathcal{L}(\rho, \beta, \sigma^2 | y_P, X, Z_P, \Phi) = \int_{S^{n-p}} \mathcal{L}(\rho, \beta, \sigma^2 | y_P, X, Z_P, z_C, \Phi)\, \hat{\varrho}(z_C | Z_P)\, \mathrm{d}z_C \tag{15}$$

   *where $\hat{\varrho}\colon S^{n-p} \to \mathbb{R}^+$ is the conditional probability density function of $Z_C | Z_P$ implied by the estimated intensity function $\hat{\lambda}$;*
5. *marginal likelihood $\mathcal{L}(\rho, \beta, \sigma^2 | y, X, Z_P, \Phi)$ is maximised with respect to $\rho$, $\beta$ and $\sigma^2$.*

As anticipated, marginalisation (15) has to be performed numerically since it seems impossible to compute it analytically. Anyway, two issues may make the outlined method computationally unfeasible.

Firstly, the high-dimensional integration space in (15) may substantially deteriorate the performances of Monte Carlo integration methods.

Secondly, the need to evaluate integral (15) at every step of the optimisation procedure dramatically exacerbates the problem outlined in the previous point.

In order to overcome both problems (and the second in particular), we rely on the cross-entropy algorithm for the optimisation of noisy functions (Rubinstein and Kroese 2004). Unlike other numerical optimisation methods such as the Expectation-Maximisation algorithm (Dempster et al. 1977; Robert and Casella 2004), at each iteration the cross-entropy algorithm simultaneously performs the marginalisation and the optimisation of the likelihood function $\mathcal{L}(\rho, \beta, \sigma^2 | y, X, Z_P, Z_C, \Phi)$. This leads to a substantial reduction of the computational burden required by the optimisation routine.

Results of Monte Carlo simulations discussed in the next section have been performed adopting the same parameters and instrumental distributions of the cross-entropy algorithm as described in Bee et al. (2017), where the method has been applied to maximum likelihood estimation of generalised linear multilevel models (the only exception is in the number $N$ of draws, as it will be clarified later).

### 3.2 Impact estimators

According to LeSage and Pace (2009), the effects of covariates on the dependent variable of a SAR do not solely depend on regression coefficients $\beta$, as the spatially-lagged dependent variable induces an indirect effect resulting from the autoregressive parameter $\rho$ and the spatial weight matrix $W$. It follows that the overall impact of a regressor on the value of the dependent variable can be decomposed in a direct and an indirect impact, which, however, it is not constant amongst all units. For these reasons, averages of total ($T(\beta)$), direct ($D(\beta)$), and indirect ($M(\beta)$) impacts are usually computed (LeSage and Pace 2009):

$$T(\beta) = n^{-1}\, \iota_n^{\mathrm{T}} (I - \rho W)^{-1} \iota_n \beta, \tag{16a}$$

$$D(\beta) = n^{-1}\, \mathrm{tr}(I - \rho W)^{-1} \beta, \tag{16b}$$

$$M(\beta) = T(\beta) - D(\beta). \tag{16c}$$

According to the model we have described in Sect. 2, some elements of the spatial weight matrix $W$ are uncertain when geocoding is not complete. It follows that impacts should be estimated via Monte Carlo simulations, where the weight matrices are defined according to the realisations of a point process $Z$ with estimated intensity function $\hat{\lambda}$. Thus, the Monte Carlo estimators of the impact measures (16) can be defined as follows:

$$\widehat{(A^{-1})} = \frac{1}{N} \sum_{k=1}^{N} (I - \hat{\rho} W_k)^{-1}, \qquad \hat{T}(\hat{\beta}) = n^{-1}\, \iota_n^{\mathrm{T}} \widehat{(A^{-1})} \iota_n \hat{\beta},$$

$$\hat{D}(\hat{\beta}) = n^{-1}\, \mathrm{tr}\, \widehat{(A^{-1})} \hat{\beta}, \qquad \hat{M}(\hat{\beta}) = \hat{T}(\hat{\beta}) - \hat{D}(\hat{\beta}).$$

Since Monte Carlo estimation of matrix $\widehat{(A^{-1})}$ may be computationally demanding, because of the inversions of the weight matrices $W_k$, a truncated geometric series of $(I - \hat{\rho} W_k)^{-1}$ may reduce substantially the computational burden of the simulation:

$$\widehat{(A^{-1})} = \frac{1}{N} \sum_{k=1}^{N} \sum_{h=0}^{m} \hat{\rho}^h W_k^h.$$

where $m$ represents the truncation order.

### 3.3 Asymptotics and generalisations

As stated in the introduction, this paper aims at proposing an estimation method for spatial models *à la* Cliff–Ord (Cliff and Ord 1969), where a portion of data is affected by coarsening, thus the primarily interest is devoted to the parameters of the model, as well as other measures of covariates' effects (like, e.g. direct, indirect and total impacts, which are discussed in Sect. 3.2).

The double marginalisation performed in Algorithm 1 derives from the following marginalisation of the probability density function of the marked point process:

$$f_{Y_P|Z_P}(y_P|z_P) = \int \left[ \int f_{Y_P,Y_C|Z_P,Z_C}(y_P, y_C|z_P, z_C)\, \mathrm{d}y_C \right] \varrho(z_C|z_P)\, \mathrm{d}z_C, \quad (17)$$

where conditioning of all probability density functions with respect to the coarsening vector $\Phi$ has been omitted for notational simplicity, and $\varrho(z_C|z_P) = f_{Z_C|Z_P}(z_C|z_P)$ has been denoted consistently to the notation of Eq. (15). The inner integral of Eq. (17) corresponds to the first marginalisation described at point 3 of Algorithm 1, whereas the outer integral determines the marginalisation of Eq. (15).

The inner integral of Eq. (17) is a marginalisation of a model whose maximum likelihood estimators have been proved to be consistent by Lee (2004), provided that specific requirements on the asymptotic specification of the spatial weight matrix $W$ are satisfied. Yet, the asymptotic behaviour of the double-marginal estimator depends also on both the geocoding propensity function estimator (14) and the intensity function estimator (13); both estimators are consistent, and their asymptotic properties are discussed in Zimmerman (2008), however this is not enough to guarantee that the double-marginal estimator is consistent too. The reason for this is that the spatial weight matrix is built according to both the spatial point pattern and the coarsening process, and its asymptotic behaviour is fully determined by the properties of that two processes. To our knowledge, at the moment, there are no theoretical results which can be exploited in order to prove (or refuse) the consistency of double-marginal estimator.

As for the applicability of the double-marginal estimator, it is worth pointing out that it can be easily adapted or generalised to other coarsening mechanisms, point processes, or stochastic spatial processes, as it is only required that the model can be identified and marginalised. Thus, if a spatial model other than the SAR is considered, Algorithm 1 changes in step 3, where the likelihood function of the model is

marginalised with respect to non-coarsened units, whereas the rest of the algorithm does not change.

A special case is represented by the spatial Durbin model (SDM), which generalises the SAR model by including some (or all) spatially lagged covariates amongst the regressors. In this case, both Algorithm 1 and likelihood function (12) are valid without modifications, provided that the design matrix $X$ is properly redefined so as to include extra covariates.

The family of Cliff–Ord spatial models consists of other specifications which somehow include other forms of spatial dependence or allows for other specifications of the covariate effects. An extensive review of the existing Cliff–Ord spatial models can be found in Cressie (2015), Anselin (1988), and LeSage and Pace (2009). Here it is worth reminding the *general nesting spatial model* (GNSM) defined in Elhorst (2014):

$$\begin{cases} y = \rho W y + \alpha \iota_n + X\beta + WX\theta + u \\ u = \lambda W u + \varepsilon \\ \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n) \end{cases} \tag{18}$$

Although the GNSM (18) is not identifyable, it deserves consideration, as it includes the main Cliff–Ord spatial models as special cases, if one or more restrictions are apllied to its parameters.—For example, the SDM is obtained when $\lambda = 0$, whereas the SAR model (1) results if $\lambda = 0$ and $\delta = 0$ (the constant vector $\iota_n$ can be included into the design matrix $X$).

The full log-likelihood of the GNSM (18) can be proved to be:

$$\ln \mathcal{L}(\rho, \lambda, \alpha, \beta, \delta, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|A_\rho| + \ln|A_\lambda| - \frac{[A_\lambda(A_\rho y - \alpha\iota_n - X\beta - WX\theta)]^{\mathrm{T}}[A_\lambda(A_\rho y - \alpha\iota_n - X\beta - WX\theta)]}{2\sigma^2}, \tag{19}$$

where $A_\rho = I_n - \rho W$ and $A_\lambda = I_n - \lambda W$.

The likelihood of models nested in GNSM can be derived from (19), whereas the first analytical marginalisation can be derived through inversions (8) and (9). Once the first marginalisation has been derived, Algorithm 1 can be applied just as illustrated above.

## 4 Monte Carlo simulations

The performances of the proposed estimation approach in finite samples have been studied by means of Monte Carlo simulations. The complication of both the modelling setting and estimation method considerably widens the variety of scenarios which should be considered for studying the estimators' properties in finite samples.

In this section twelve different scenarios are considered:

(A) a point pattern with $n = 250$ points is generated over an irregular area $S$ according to an inhomogeneous Poisson process with the intensity function $\lambda$ represented

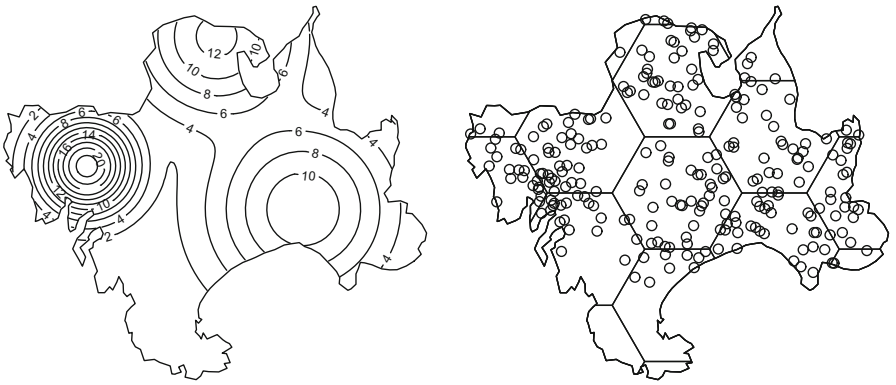

**Fig. 1** Intensity function $\lambda$ used for generating the point process (left) and the realisation of the process for $n = 250$ with hexagonal partition ($R = 17$) of the space (right)

in Fig. 1. The surface $S$ is partitioned into $R = 17$ hexagonal regions of equal size excepting for border zones (see Fig. 1). The SAR includes two regressors (generated as realisations of a standard normal distribution) and a constant term, so that $X \in \mathbb{R}^{n\times 3}$. The parameters of the SAR are $\rho = 0.5$, $\beta = [1, 1, -1]^{\mathrm{T}}$, $\sigma^2 = 1$, whereas the spatial weight matrix $W$ is computed according to (3), and $\kappa(x) = \mathbb{1}_{\{x\leq 0.5\}}$ (note that sides of hexagons measure 1.5). Each unit of the point pattern is independently coarsened with probability 0.4, hence the expected number of coarsened units is $\mathbb{E}(p) = \mathbb{E}(\Phi^{\mathrm{T}}\iota_n) = 0.4 \cdot n = 100$. Simulations are based on $N = 300$ replications, each of which share the same point pattern and design matrix $X$. Models are fitted through the cross-entropy algorithm for noisy functions (Rubinstein and Kroese 2004) as implemented in Bee et al. (2017), but for the number of draws (denoted by $N$ in Bee et al. 2017) which equals 200 for the first iteration and 100 for subsequent iterations instead of 1000 for all iteration as suggested in Bee et al. (2017). Simulations have been performed by means of the software R (R Core Team 2020), whereas cross-entropy optimisation has been carried out through the R package noisyCE2 (Santi 2020);

(B) the same simulation settings as in point (A), except that $\rho = 0.3$;
(C) the same simulation settings as in point (A), except that $\rho = 0.7$;
(D) the same simulation settings as in point (A), except that $\sigma^2 = 2$;
(E) the same simulation settings as in point (A), except that $n = 500$ and $\kappa(x) = \mathbb{1}_{\{x\leq\sqrt{1/8}\}}$. Function $\kappa$ has been redefined so that the average number of neighbours per unit is the same as in case (A);
(F) the same simulation settings as in point (A), except that $n = 1000$ and $\kappa(x) = \mathbb{1}_{\{x\leq 0.25\}}$. Function $\kappa$ has been redefined so that the average number of neighbours per unit is the same as in case (A);
(G) the same simulation settings as in point (A), except that $\phi(s) \propto 0.8\,\lambda(s)$. Function $\phi$ is set so that the coarsening probability ranges between 0.2 and 0.75, whereas its average equals 0.4, in line with all the other simulation scenarios;
(H) the same simulation settings as in point (A), except that $\phi(s) \propto -0.8\,\lambda(s)$. Function $\phi$ is set so that the coarsening probability ranges between 0.04 and 0.60, whereas its average equals 0.4, in line with all the other simulation scenarios;

(I) the same simulation settings as in point (A), except that each unit of the point pattern is independently coarsened with probability 0.1 (instead of 0.4), hence the expected number of coarsened units is $\mathbb{E}(p) = \mathbb{E}(\Phi^{\mathrm{T}}\iota_n) = 0.1 \cdot n = 25$;
(J) the same simulation settings as in point (A), except that each unit of the point pattern is independently coarsened with probability 0.2 (instead of 0.4), hence the expected number of coarsened units is $\mathbb{E}(p) = \mathbb{E}(\Phi^{\mathrm{T}}\iota_n) = 0.2 \cdot n = 50$;
(K) the same simulation settings as in point (A), except that each unit of the point pattern is independently coarsened with probability 0.8 (instead of 0.4), hence the expected number of coarsened units is $\mathbb{E}(p) = \mathbb{E}(\Phi^{\mathrm{T}}\iota_n) = 0.6 \cdot n = 150$;
(L) the same simulation settings as in point (A), except that the sides of hexagons measure 1, thus the number of regions is $R = 29$.

For each scenario five estimation methods are considered:

- the maximum likelihood estimator based on a dataset where locations of all units are known, and there is no coarsening. Hereinafter this estimator is referred to as NCM, which stands for *non-coarsened model*;
- the proposed estimator based on double marginalisation (hereinafter DME);
- the maximum likelihood estimator of the SAR based only on non-coarsened units (hereinafter PDM, which stands for *purged data model*). In this case the weight matrix is computed using the same $\kappa$ function as the data generating process, but no standardisation is performed;
- the maximum likelihood estimator of the SAR based only on non-coarsened units. Unlike the previous case, the spatial weight matrix is row-standardised (hereinafter SPDM, which stands for *standardised PDM*);
- the maximum likelihood estimator of the SAR based on all points. Location of coarsened points is imputed to the centroids of regions where points are located, and a row-standardised weight matrix is derived according to the same $\kappa$ function as the data generating process. Hereinafter, this method is referred to as CIP, which stands for *centroid imputed position*.

Clearly, the NCM estimates are obtained under the ideal condition of no-uncertainty about unit locations, they are thus expected to be the most efficient amongst the others considered in the simulations.

PDM and SPDM are only based on correctly-georeferenced units, however all the information about dependent and independent variables is lost for coarsened units, which are not involved in the estimation process. This results in smaller effective sample size, in an alteration of part of the elements of the spatial weight matrix $W$ due to row-standardisation, as well as in an alteration of the dependence structure amongst all units induced by the inversion of matrix $I_n - \rho W$.

On the other hand, the CIP estimates use all the information about dependent and independent variables, whereas the imputation of the unit locations typically alters the actual neighbourhood relations both amongst coarsened units, and between coarsened and non-coarsened units.

DME is compared to all the estimation methods in order to verify whether it produces estimates which are more efficient than those obtained through PDM, SPDM and CIP. NCM estimates are used as the high benchmark.

Results of simulations are summarized in terms of relative root mean squared error (RRMSE) and relative bias in Tables 1, 2 and 3. For reasons of space, impacts estimates about the first regressor only are reported, since estimates on other regressors' impacts are similar.

As expected, in all scenarios the NCM estimator is the best performer for all parameters both in terms of bias and RMSE, and it is not commented in the following if not for the purpose of making comparisons to other estimation approaches.

Estimates in Tables 1, 2 and 3 show two general results which basically hold under all scenarios.

Firstly, the estimates obtained from all estimation methods are rather stable under all simulation settings for most parameters and impacts. The only remarkable exception is represented by the estimates of the error variance, which are rather sensitive with respect to the value of parameter $\rho$ and $\sigma^2$.

Secondly, the rank of estimation methods in terms of both bias and RMSE is basically the same whatever the scenario we consider, although some differences emerge amongst parameters.

If covariate coefficients are considered (that is $\beta_0, \beta_1, \beta_2$), DME estimator is the best performer in terms of relative bias. On the other hand, the CIP estimator exhibits the smallest RRMSE, followed by the SPDM estimator, whereas larger RRMSEs result from DME and PDM estimator. Anyway, both in case of relative bias and RRMSE, differences amongst estimators are rather small, if we consider covariates coefficients $\beta_1$ and $\beta_2$, whereas larger variability emerges for $\beta_0$.

Things change if the autoregressive parameter $\rho$ is considered. In this case, the DME clearly outperforms all other estimators both in terms of relative bias and RRMSE in all considered scenarios, whereas the second-best estimator is SPDM estimator followed by CIP and PDM estimators. Unlike regressors coefficients, differences amongst estimation methods are large in terms of relative bias and RRMSE.

If the error dispersion parameter $\sigma$ is considered, the four estimation methods for coarsened data can be gathered into two groups. The former includes the best performers which are DME and SPDM, the latter consists in CIP and PDM estimators, which almost double the relative bias and the RRMSE of estimators in the other group. Interestingly, estimators of each group exhibit very similar relative bias and RRMSE.

The performances of estimators on assessing impacts of covariates clearly reflect the statistical performances on parameters $\rho$, $\beta_1$, and $\beta_2$. Thus CIP, PDM, and SPDM estimators perform well in estimating the direct impact, whereas the DME definitely outperforms the others when indirect impact is estimated. The efficiency of DME on indirect impact estimation is large enough to make DME the most efficient estimator also for the total impact. Analogous results hold also in terms of bias.

Although relative performances of estimators are pretty stable amongst scenarios considered in the simulations, it is worth analysing more in depth the results of the simulations.

The comparison between results of scenarios B, A and C enables to assess the effect of parameter $\rho$ on the estimators, which turns out to be marked for all parameters and most estimation methods. In particular, as $\rho$ gets larger (note that $\rho_{(B)} = 0.3$, $\rho_{(A)} = 0.5$, $\rho_{(C)} = 0.7$), both the relative bias (in absolute value) and the RRMSE of $\hat{\rho}$ decrease for all estimation methods except for PDM, whereas the absolute value of

**Table 1** Relative root mean squared error and relative bias (in parenthesis) of parameter and impact estimators for scenarios A, B, C, D of Monte Carlo simulations (see Sect. 4) for various estimation methods

| Method | $\rho$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\sigma$ | $D(\beta_1)$ | $M(\beta_1)$ | $T(\beta_1)$ |
|---|---|---|---|---|---|---|---|---|
| *Scenario A* | | | | | | | | |
| NCM | 4.78 | 10.40 | 2.29 | 3.85 | 4.20 | 2.43 | 9.57 | 5.22 |
| | (−0.40) | (−0.70) | (-0.11) | (-0.15) | (-0.59) | (-0.14) | (-0.46) | (-0.28) |
| DME | 23.74 | 24.58 | 3.99 | 6.16 | 25.41 | 4.59 | 37.50 | 18.81 |
| | (−22.25) | (−17.98) | (0.68) | (0.28) | (24.09) | (−2.57) | (−36.00) | (−17.89) |
| PDM | 83.95 | 39.36 | 4.13 | 5.88 | 40.72 | 4.11 | 49.03 | 23.78 |
| | (−83.89) | (−34.71) | (2.00) | (−0.02) | (39.54) | (−1.67) | (−43.44) | (−20.82) |
| SPDM | 29.67 | 30.19 | 3.80 | 5.80 | 25.70 | 3.97 | 48.34 | 23.10 |
| | (−28.50) | (−24.85) | (0.74) | (0.82) | (24.23) | (−1.33) | (−47.44) | −22.46 |
| CIP | 33.01 | 27.50 | 3.14 | 4.57 | 44.09 | 3.77 | 47.87 | 23.46 |
| | (−32.11) | (−23.75) | (1.48) | (1.36) | (43.36) | (−2.61) | (−47.06) | (−22.98) |
| *Scenario B* | | | | | | | | |
| NCM | 9.84 | 9.93 | 2.17 | 3.50 | 4.51 | 2.20 | 13.81 | 4.63 |
| | (−0.19) | (−0.63) | (−0.05) | (−0.23) | (−0.85) | (−0.03) | (0.22) | (0.04) |
| DME | 29.15 | 17.39 | 3.18 | 5.01 | 9.79 | 3.42 | 36.19 | 11.44 |
| | (−25.81) | (−9.36) | (−0.22) | (0.06) | (7.32) | (−1.27) | (−32.69) | (−10.08) |
| PDM | 84.15 | 20.88 | 3.06 | 4.87 | 12.89 | 3.35 | 49.32 | 15.05 |
| | (−84.05) | (−14.62) | (0.00) | (−0.23) | (10.99) | (−1.16) | (−45.28) | (−13.53) |
| SPDM | 36.51 | 19.10 | 3.00 | 4.84 | 9.71 | 3.19 | 47.20 | 14.26 |
| | (−34.22) | (−12.70) | (−0.18) | (0.38) | (7.33) | (−0.96) | (−45.38) | (−13.42) |
| CIP | 38.88 | 17.78 | 2.26 | 3.62 | 14.22 | 2.71 | 46.52 | 14.27 |
| | (−36.64) | (−13.23) | (−0.19) | (−0.18) | (13.07) | (−1.50) | (−44.43) | (−13.54) |

**Table 1** continued

| Method | $\rho$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\sigma$ | $D(\beta_1)$ | $M(\beta_1)$ | $T(\beta_1)$ |
|---|---|---|---|---|---|---|---|---|
| *Scenario C* | | | | | | | | |
| NCM | 2.42 | 10.99 | 2.21 | 3.76 | 4.92 | 2.36 | 7.79 | 5.54 |
| | (−0.34) | (−0.93) | (−0.20) | (−0.13) | (−0.86) | (−0.33) | (−0.91) | (−0.70) |
| DME | 17.58 | 42.81 | 4.86 | 8.64 | 58.44 | 7.08 | 43.65 | 30.03 |
| | (−16.66) | (−33.12) | (2.04) | (1.03) | (56.86) | (−5.79) | (−42.98) | (−29.48) |
| PDM | 84.69 | 86.62 | 8.58 | 8.46 | 109.38 | 5.54 | 48.59 | 32.33 |
| | (−84.66) | (−81.77) | (7.35) | (2.63) | (108.13) | (−2.27) | (−40.87) | (−26.86) |
| SPDM | 22.30 | 51.97 | 5.29 | 7.21 | 59.68 | 4.53 | 51.00 | 33.02 |
| | (−21.60) | (−44.62) | (3.24) | (2.08) | (57.89) | (−0.96) | (−50.39) | (−32.45) |
| CIP | 28.41 | 49.80 | 6.70 | 7.43 | 110.29 | 5.39 | 53.90 | 36.01 |
| | (−27.92) | (−45.83) | (5.81) | (5.18) | (109.49) | (−4.42) | (−53.50) | (−35.69) |
| *Scenario D* | | | | | | | | |
| NCM | 6.35 | 14.79 | 3.30 | 4.97 | 4.87 | 3.30 | 12.12 | 6.51 |
| | (−0.76) | (0.69) | (0.37) | (0.54) | (−0.80) | (0.30) | (−0.42) | (−0.03) |
| DME | 24.23 | 27.68 | 5.00 | 7.58 | 15.30 | 5.13 | 37.42 | 18.65 |
| | (−21.96) | (−16.14) | (1.15) | (1.23) | (13.30) | (−2.06) | (−35.01) | (−17.16) |
| PDM | 84.20 | 39.79 | 5.34 | 7.35 | 25.13 | 4.86 | 48.32 | 23.31 |
| | (−84.14) | (−31.45) | (2.56) | (0.82) | (23.89) | (−1.18) | (−42.58) | (−20.15) |
| SPDM | 30.77 | 31.76 | 4.79 | 7.48 | 16.03 | 4.60 | 48.73 | 23.08 |
| | (−29.17) | (−22.53) | (1.40) | (1.84) | (14.30) | (−0.77) | (−47.46) | (−22.17) |
| CIP | 34.08 | 29.41 | 4.20 | 6.01 | 27.21 | 4.12 | 48.42 | 23.50 |
| | (−32.67) | (−22.32) | (2.03) | (2.09) | (26.27) | (−2.09) | (−47.14) | (−22.74) |

Direct (D), indirect (M) and total (T) impact estimates refer to the second regressor (whose coefficient is $\beta_1$). All values are multiplied by 100

**Table 2** Relative root mean squared error and relative bias (in parenthesis) of parameter and impact estimators for scenarios E, F, G, H of Monte Carlo simulations (see Sect. 4) for various estimation methods

| Method | $\rho$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\sigma$ | $D(\beta_1)$ | $M(\beta_1)$ | $T(\beta_1)$ |
|---|---|---|---|---|---|---|---|---|
| *Scenario E* | | | | | | | | |
| NCM | 3.42 | 7.01 | 1.47 | 2.36 | 3.44 | 1.52 | 6.68 | 3.52 |
| | (−0.17) | (−0.20) | (0.03) | (−0.07) | (−0.70) | (0.02) | (−0.08) | (−0.02) |
| DME | 26.50 | 24.73 | 3.36 | 5.17 | 25.81 | 3.46 | 40.36 | 19.73 |
| | (−25.62) | (−20.79) | (1.87) | (2.65) | (24.89) | (−2.09) | (−39.42) | (−19.14) |
| PDM | 85.04 | 34.12 | 4.10 | 7.02 | 41.29 | 2.95 | 37.40 | 17.85 |
| | (−85.02) | (−31.62) | (3.18) | (5.72) | (40.67) | (−0.21) | (−26.80) | (−12.35) |
| SPDM | 29.77 | 26.93 | 3.17 | 4.97 | 25.99 | 2.68 | 47.67 | 22.28 |
| | (−29.01) | (−23.84) | (1.94) | (2.86) | (25.02) | (−0.65) | (−47.01) | (−21.83) |
| CIP | 33.62 | 24.32 | 2.96 | 5.52 | 40.96 | 2.91 | 48.02 | 23.23 |
| | (−33.00) | (−21.96) | (2.34) | (4.48) | (40.44) | (−2.26) | (−47.43) | (−22.90) |
| *Scenario F* | | | | | | | | |
| NCM | 2.66 | 5.59 | 1.10 | 1.65 | 2.14 | 1.15 | 5.18 | 2.74 |
| | (−0.17) | (0.18) | (−0.03) | (0.14) | (0.12) | (−0.04) | (−0.23) | (−0.13) |
| DME | 26.43 | 24.93 | 2.27 | 3.90 | 24.32 | 3.20 | 40.64 | 20.07 |
| | (−25.93) | (−23.09) | (1.17) | (2.37) | (23.91) | (−2.59) | (−40.15) | (−19.79) |
| PDM | 88.26 | 43.37 | 3.51 | 5.57 | 39.51 | 2.49 | 55.11 | 26.31 |
| | (−88.25) | (−42.31) | (3.10) | (4.85) | (39.22) | (−1.85) | (−54.37) | (−25.90) |
| SPDM | 31.69 | 30.18 | 2.24 | 3.73 | 23.90 | 2.07 | 50.01 | 23.64 |
| | (−31.37) | (−28.95) | (1.53) | (2.75) | (23.52) | (−1.28) | (−49.78) | (−23.49) |
| CIP | 33.96 | 26.72 | 2.81 | 5.00 | 36.06 | 2.38 | 48.42 | 23.29 |
| | (−33.67) | (−25.60) | (2.49) | (4.55) | (35.86) | (−2.00) | (−48.15) | (−23.14) |

**Table 2** continued

| Method | $\rho$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\sigma$ | $D(\beta_1)$ | $M(\beta_1)$ | $T(\beta_1)$ |
|---|---|---|---|---|---|---|---|---|
| *Scenario G* | | | | | | | | |
| NCM | 4.67 | 10.23 | 2.14 | 3.72 | 4.36 | 2.18 | 9.09 | 4.82 |
| | (− 0.10) | (0.27) | (0.01) | (− 0.02) | (− 0.86) | (0.03) | (0.22) | (0.12) |
| DME | 21.86 | 24.29 | 3.66 | 5.96 | 24.63 | 4.08 | 35.26 | 17.55 |
| | (− 20.49) | (− 17.05) | (0.84) | (0.33) | (23.43) | (− 2.12) | (− 33.81) | (− 16.64) |
| PDM | 80.09 | 33.36 | 3.67 | 5.86 | 37.55 | 3.72 | 46.57 | 22.36 |
| | (− 80.05) | (− 27.57) | (1.55) | (0.81) | (36.58) | (− 1.32) | (− 44.73) | (− 21.21) |
| SPDM | 27.07 | 29.30 | 3.56 | 5.75 | 24.73 | 3.50 | 45.52 | 21.30 |
| | (− 26.00) | (− 23.52) | (1.15) | (0.85) | (23.40) | (− 0.34) | (− 44.62) | (− 20.63) |
| CIP | 31.00 | 25.80 | 2.95 | 4.30 | 41.10 | 3.57 | 45.71 | 22.40 |
| | (− 30.14) | (− 21.94) | (1.43) | (1.28) | (40.46) | (− 2.40) | (− 44.85) | (− 21.86) |
| *Scenario H* | | | | | | | | |
| NCM | 5.04 | 10.24 | 2.11 | 3.85 | 4.58 | 2.19 | 9.70 | 5.14 |
| | (− 0.48) | (− 1.40) | (0.05) | (− 0.19) | (− 1.01) | (0.01) | (− 0.42) | (− 0.19) |
| DME | 24.43 | 27.55 | 3.63 | 6.04 | 24.82 | 4.26 | 38.00 | 19.01 |
| | (− 22.95) | (− 21.33) | (0.90) | (0.13) | (23.33) | (− 2.63) | (− 36.53) | (− 18.17) |
| PDM | 89.09 | 50.58 | 4.86 | 5.73 | 43.11 | 4.09 | 54.25 | 26.19 |
| | (− 89.06) | (− 46.77) | (3.27) | (− 1.09) | (42.11) | (− 1.71) | (− 48.39) | (− 23.10) |
| SPDM | 31.01 | 33.33 | 3.50 | 5.69 | 25.33 | 3.56 | 49.13 | 23.46 |
| | (− 29.85) | (− 28.52) | (1.26) | (0.54) | (23.89) | (− 1.41) | (− 48.23) | (− 22.86) |
| CIP | 32.87 | 27.86 | 3.04 | 4.67 | 44.74 | 3.34 | 47.38 | 23.09 |
| | (− 31.94) | (− 24.23) | (1.86) | (1.34) | (44.01) | (− 2.34) | (− 46.54) | (− 22.59) |

Direct (D), indirect (M) and total (T) impact estimates refer to the second regressor (whose coefficient is $\beta_1$). All values are multiplied by 100

**Table 3** Relative root mean squared error and relative bias (in parenthesis) of parameter and impact estimators for scenarios I, J, K, L of Monte Carlo simulations (see Sect. 4) for various estimation methods

| Method | $\rho$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\sigma$ | $D(\beta_1)$ | $M(\beta_1)$ | $T(\beta_1)$ |
|---|---|---|---|---|---|---|---|---|
| *Scenario I* | | | | | | | | |
| NCM | 4.78 | 10.43 | 2.06 | 3.75 | 4.53 | 2.20 | 9.50 | 5.10 |
| | (−0.16) | (−0.58) | (0.16) | (−0.04) | (−1.03) | (0.18) | (0.29) | (0.23) |
| DME | 7.28 | 13.66 | 2.54 | 4.47 | 8.26 | 2.60 | 13.68 | 7.03 |
| | (−3.83) | (−3.11) | (0.04) | (0.01) | (5.93) | (−0.67) | (−9.36) | (−4.65) |
| PDM | 86.67 | 26.44 | 3.04 | 4.27 | 33.63 | 2.80 | 38.25 | 18.46 |
| | (−86.65) | (−21.59) | (2.04) | (−1.04) | (32.90) | (−0.81) | (0.02) | (−0.43) |
| SPDM | 8.79 | 14.00 | 2.34 | 4.03 | 7.46 | 2.37 | 15.96 | 7.80 |
| | (−6.44) | (−5.42) | (0.41) | (0.29) | (5.13) | (−0.01) | (−12.24) | (−5.62) |
| CIP | 11.97 | 13.99 | 2.39 | 4.05 | 18.79 | 2.63 | 20.11 | 10.07 |
| | (−10.03) | (−6.56) | (0.31) | (0.20) | (17.30) | (−1.03) | (−17.12) | (−8.40) |
| *Scenario J* | | | | | | | | |
| NCM | 5.06 | 9.60 | 2.06 | 3.40 | 4.47 | 2.20 | 9.92 | 5.28 |
| | (−0.50) | (0.04) | (−0.02) | (0.16) | (−0.53) | (−0.07) | (−0.52) | (−0.27) |
| DME | 11.91 | 14.89 | 2.76 | 4.79 | 14.31 | 3.09 | 21.37 | 10.84 |
| | (−9.78) | (−6.59) | (0.24) | (0.35) | (12.56) | (−1.32) | (−18.71) | (−9.29) |
| PDM | 85.89 | 28.55 | 3.36 | 4.17 | 36.13 | 3.06 | 31.53 | 15.58 |
| | (−85.88) | (−24.18) | (2.12) | (−0.57) | (35.25) | (−1.08) | (−17.68) | (−8.68) |
| SPDM | 14.99 | 16.94 | 2.66 | 4.27 | 13.40 | 2.73 | 26.63 | 12.72 |
| | (−13.51) | (−10.34) | (0.62) | (0.67) | (11.74) | (−0.33) | (−24.77) | (−11.53) |
| CIP | 18.84 | 16.15 | 2.59 | 3.98 | 28.81 | 3.00 | 30.17 | 14.91 |
| | (−17.59) | (−10.87) | (0.65) | (0.96) | (27.80) | (−1.69) | (−28.63) | (−14.03) |

**Table 3** continued

| Method | $\rho$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\sigma$ | $D(\beta_1)$ | $M(\beta_1)$ | $T(\beta_1)$ |
|---|---|---|---|---|---|---|---|---|
| *Scenario K* | | | | | | | | |
| NCM | 4.89 | 9.31 | 2.02 | 3.55 | 4.42 | 2.09 | 9.43 | 4.97 |
| | (−0.48) | (0.20) | (−0.02) | (0.13) | (−1.20) | (−0.06) | (−0.52) | (−0.27) |
| DME | 38.47 | 38.15 | 5.21 | 8.66 | 36.57 | 5.80 | 53.69 | 26.64 |
| | (−36.89) | (−30.17) | (1.58) | (1.74) | (35.05) | (−3.31) | (−52.39) | (−25.80) |
| PDM | 81.42 | 50.07 | 5.18 | 8.30 | 45.17 | 5.04 | 65.30 | 31.27 |
| | (−81.29) | (−43.86) | (2.43) | (1.42) | (43.80) | (−1.86) | (−63.47) | (−30.09) |
| SPDM | 45.94 | 45.60 | 4.95 | 8.20 | 38.18 | 4.97 | 67.31 | 32.06 |
| | (−44.76) | (−39.29) | (1.69) | (1.98) | (36.70) | (−1.84) | (−66.77) | (−31.60) |
| CIP | 44.04 | 34.64 | 3.41 | 4.47 | 52.76 | 3.57 | 59.21 | 28.69 |
| | (−43.14) | (−31.81) | (2.49) | (2.29) | (52.13) | (−2.76) | (−58.55) | (−28.33) |
| *Scenario L* | | | | | | | | |
| NCM | 4.84 | 9.44 | 2.01 | 3.40 | 4.93 | 2.16 | 9.57 | 5.13 |
| | (−0.27) | (−0.47) | (0.01) | (−0.23) | (−0.60) | (0.00) | (−0.07) | (−0.03) |
| DME | 20.28 | 22.43 | 3.52 | 6.00 | 24.59 | 4.16 | 33.16 | 16.73 |
| | (−18.59) | (−15.55) | (0.47) | (0.14) | (22.91) | (−2.41) | (−31.33) | (−15.66) |
| PDM | 84.05 | 38.60 | 4.28 | 5.87 | 40.73 | 4.02 | 50.74 | 24.49 |
| | (−84.00) | (−34.50) | (2.54) | (−0.43) | (39.48) | (−1.16) | (−42.00) | (−19.87) |
| SPDM | 29.18 | 29.64 | 3.50 | 5.48 | 25.88 | 3.48 | 47.55 | 22.48 |
| | (−28.07) | (−24.61) | (1.18) | (0.57) | (24.12) | (−0.88) | (−46.65) | (−21.85) |
| CIP | 27.94 | 22.79 | 3.12 | 4.21 | 40.35 | 3.27 | 41.89 | 20.44 |
| | (−26.98) | (−19.15) | (1.76) | (1.55) | (39.58) | (−2.05) | (−40.91) | (−19.86) |

Direct (D), indirect (M) and total (T) impact estimates refer to the second regressor (whose coefficient is $\beta_1$). All values are multiplied by 100

the relative bias and the RRMSE grow for all the other parameters as $\rho$ gets larger. The magnitude of relative bias and the RMSE of impact estimators grow as well with $\rho$ for all estimation methods with the exception of the indirect impact of PDM which is basically independent of $\rho$.

The effect of $\sigma^2$ can be assessed by comparing scenario A (where $\sigma = 1$) and scenario D (where $\sigma = 2$). The increase in the variance error leads to greater RRMSE for both the regressor coefficients and the direct impacts, whereas the RRMSE of the autoregressive parameter $\rho$ and other impacts (indirect and total) remain basically unchanged. On the other hand, both the relative bias and the RRMSE of $\hat{\sigma}$ improve substantially.

It is worth noting that the relative bias and the RRMSE of the estimators of the autoregressive parameter $\rho$ and the error dispersion parameter $\sigma$ are associated. In particular, the larger $(\hat{\rho} - \rho)/\rho$, the smaller the relative bias and the RRMSE of $\hat{\sigma}$. Such a relation is fully consistent with with the relation in Eq. (2b), where the negative bias of $\hat{\rho}$ entails a positive bias of $\hat{\sigma}$.

An analogous relation emerges for the regression coefficients $\beta_0, \beta_1, \beta_2$, according to Eq. (2a), however in this case the magnitude of the effect is not particularly marked, probably due to the fact that Eq. (2a) is linear in $\hat{\rho}$, unlike Eq. (2a), which is quadratic.

The effect of the sample size $n$ emerges if the results of scenario A ($n = 250$), E ($n = 500$) and F ($n = 1000$) are compared. As expected, both the relative bias, and the RRMSE of all estimators of non-coarsened model (NCM) gets smaller (in absolute value) as $n$ increases. On the other hand, none of the other estimators show a similar pattern, as neither the relative bias nor the RRMSE exhibits any convergent trend.

If scenarios A ($\phi(s) = 0.4$), G ($\phi(s) \propto 0.8\,\lambda(s)$), and H ($\phi(s) \propto -0.8\,\lambda(s)$) are compared, no clear pattern emerges, although it seems that RRMSE tends to slightly increase as we move from scenario G to A, and from A to H, suggesting that better estimates can be obtained if coarsening is more frequent in areas where the intensity of the point process is higher (scenario G), whereas the opposite is true if the intensity of the point process and the coarsening probability are inversely related (scenario H).

The comparison between scenarios A, I, J and K allows one to assess the effect of the proportion of coarsened units on the estimators. The coarsening function $\phi$ of all scenarios is constant over the domain $S$, and it is defined so that the expected share of coarsened points equals 10%, 20%, 40% and 60% for scenarios I, J, A and K, respectively. The results of the simulations shows that the estimators of all methods (except NCM, obviously) deteriorates in terms of efficiency as the coarsening probability gets larger. If the autoregressive parameter $\rho$ is considered, a strong bias towards zero emerges as the coarsening probability increases, in line with the empirical and theoretical results provided in Arbia et al. (2016). Clearly, the relative biases of indirect and total impacts derive from the behaviour of the estimators for $\rho$.

It is worth noting that the DME estimators are more efficient and less biased than other estimation methods, however such advantage tends to diminish as the proportion of coarsened points exceeds 50%. Monte Carlo simulations based on 80% of coarsened units (not presented in this paper for the sake of brevity) have shown that DME and CIP estimators have has similar performances, whereas they are about 20%–40% more efficient than other estimation methods.

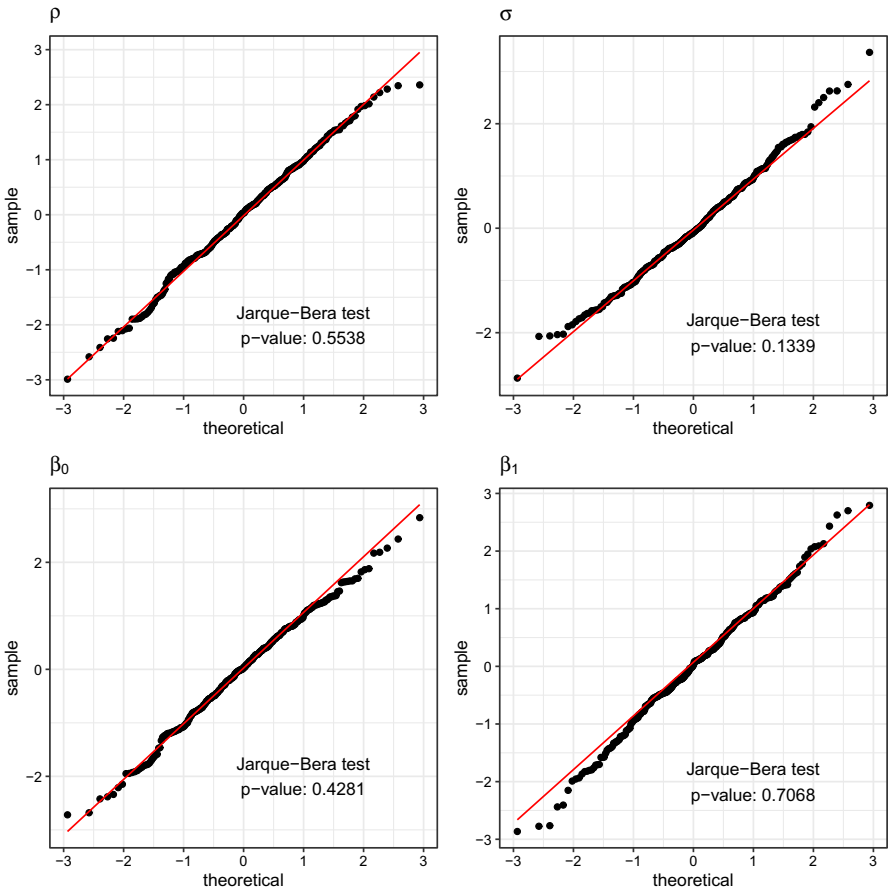


**Fig. 2** Normal quantile-quantile plot of parameter estimates from scenario A. In each plot the $p$-value of the Jarque-Bera normality test (Jarque and Bera 1980) has been annotated

Finally, if the size of the regions is reduced (scenario L vs. A), the effects of coarsening are more limited, and this turns in to an improved efficiency and relative bias for all the estimators. This is an expected result, since the loss of information due to coarsening is lower if the regions where points are located are smaller.

As for the distribution of double-marginal estimators, the goodness of fit to the Gaussian distribution is generally fairy good. See Fig. 2 for an example based on simulations from scenario A.

## 5 Application to hedonic models for house prices

In this section, the estimation method based on double marginalisation is applied to an hedonic model for house prices in Beijing.

The dataset used for fitting the model consists of a sample of 361 transactions made freely available by Qichen (2019) and collected through web-scraping from

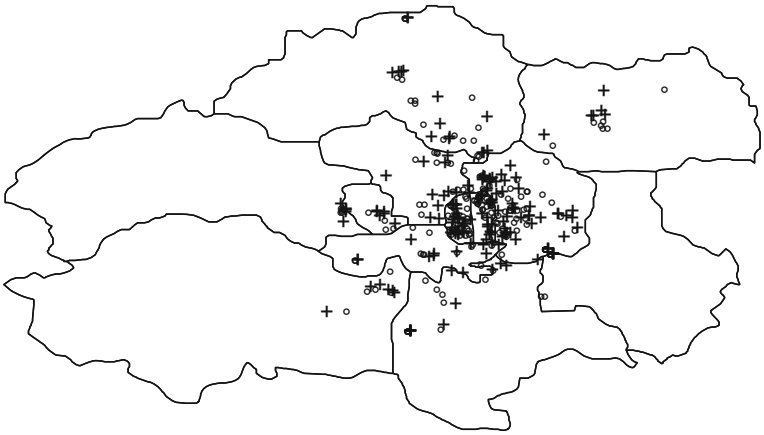

**Fig. 3** Map of the central districts of Beijing (Changping, Chaoyang, Daxing, Dongcheng, Fangshan, Fengtai, Haidian, Mentougou, Shijingshan, Shunyi, Tongzhou, Xicheng) and location of houses considered in the model. Position of correctly located houses is drawn with a cross, whereas (unreliable) position of coarsened locations is drawn by means of a circle

`bj.lianjia.com`. For each transaction, it is available information on sale price, the number of living and drawing rooms, the number of bathrooms, the type of building, the construction time, the type of building structure, whether the house is close to a subway station, and whether the elevator is available. Moreover, the Beijing's district and the geographical coordinates where houses are located are available.

Figure 3 shows the map of points where the houses are located. A group of 65 points (about 18%) in the map has been marked differently because in those cases longitude and latitude are not consistent with the district where the house should be located. For this reason, those points are considered as coarsened.

Table 4 shows the estimates of regression coefficients and total impacts for a SAR model where the logarithm of house price is regressed over the covariates listed before. The spatial weight matrices has been defined according to the $k$-nearest neighbour criterion with $k = 20$. The model has been fitted through the double-marginal estimator (DME), the standardised purged data model (SPDM) and the centroid imputed position model (CIP). The NCM model has not been considered, as correct locations of coarsened units is not available, whereas the purged data model (PDM) has not been fitted since in case of equally weighted matrices $W$ based on $k$-nearest neighbour criterion the SPDM and the PDM models are equivalent.

Results in Table 4 exhibit a good agreement in terms significance and sign of point estimates amongst the three estimation methods, nevertheless differences in magnitudes between several regression coefficients emerge. The estimates of the autoregressive parameter $\rho$ are rather different from one method to another; as simulations pointed out, as we move from the DME estimator to the CIP estimator and from the latter to the SPDM estimator, the point estimates of $\rho$ tends to get closer and closer to zero.

**Table 4** Autoregressive parameter $\rho$ and regression coefficients $\beta_j$ estimates of a SAR model on log-transformed prices of 361 houses in Beijing

| Regressor | DME | SPDM | CIP |
|---|---|---|---|
| Autoregressive parameter $\rho$ | 0.575*** | 0.398 | 0.500*** |
| Intercept | 1.325*** | 2.411*** | 1.729*** |
| Two living rooms | 0.299*** | 0.258*** | 0.361*** |
| More than two living rooms | 0.608 | 0.519*** | 0.602 |
| Two drawing rooms | 0.108 | 0.079 | 0.087 |
| More than two drawing rooms | 0.773*** | 0.742** | 0.577* |
| More than one bathroom | 0.319*** | 0.391*** | 0.379*** |
| Building type 2 | 0.111 | 0.252 | 0.169 |
| Building type 3 | 0.000 | −0.002 | −0.017 |
| Building type 4 | 0.020 | 0.014 | 0.048 |
| Building type "other" | 0.367 | 0.352 | 0.428* |
| Construction time in (1989, 1999] | 0.109 | 0.094 | 0.145*** |
| Construction time in (1999, 2019] | −0.006 | −0.013 | 0.038 |
| Renovation condition 2 | 0.441** | 0.434* | 0.400*** |
| Renovation condition 3 | 0.411*** | 0.377*** | 0.346*** |
| Renovation condition 4 | 0.506*** | 0.477*** | 0.434*** |
| Building structure type 6 | 0.084 | 0.080 | 0.105 |
| Other type of building structure | 0.183 | 0.169 | 0.135 |
| Elevator | 0.356*** | 0.321*** | 0.273** |
| Subway station | 0.185** | 0.224*** | 0.228*** |

The locations of 65 houses (18%) are coarsened, thus the model has been fitted through DME, SPDM and CIP

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

## 6 Conclusions

The estimation method proposed in this paper for tackling the problem of incompletely geocoded data is based on a modelling approach which integrates the point process, the coarsening process and the spatial process through a marked point process model whose likelihood function is then marginalised twice so as to clean out the effects of coarsening.

Monte Carlo simulations for the spatial autoregressive model have shown that the proposed method is basically equivalent to other methods in terms of bias and RMSE in the estimation of regressor coefficients, whereas it returns more efficient and less biased estimates for the spatial autoregressive parameter, the error variance, the indirect impacts, and the total impacts. Gains in efficiency and biasedness are substantial and they clearly emerge under the various simulation settings. The proposed methodology can be generalised in various directions to account for other forms of data incompleteness typically emerging when analysing large spatial datasets related to individual economic agents.

**Acknowledgements** The authors thank two anonymous reviewers for their valuable comments and suggestions, that improved an earlier version of the article.

**Funding** Open access funding provided by Università degli Studi di Verona within the CRUI-CARE Agreement.

**Availability of data and material** Not applicable.

## Declarations

**Conflicts of interest** The authors declare that they have no conflict of interest.



## A Proofs

### A.1 Proof of Equation (2a)

The reduced form of model (1) is:

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\varepsilon, \tag{20}$$

if we define $A \equiv I_n - \rho W$, it follows that:

$$y \sim \mathcal{N}\left(A^{-1}X\beta, \quad \sigma^2 (A^{\mathrm{T}}A)^{-1}\right),$$

thus, the log-likelihood function of the model is

$$\ln \mathcal{L}(\rho, \beta, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|A| - \frac{1}{2\sigma^2}(Ay - X\beta)^{\mathrm{T}}(Ay - X\beta).$$

The maximum likelihood estimation of the SAR model is often carried out by concentrating the likelihood function with respect to $\rho$ (LeSage and Pace 2009). Thus, the first order condition on $\beta$ depends on the point estimate of $\rho$ as it follows:

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\hat{A}y.$$

where $\hat{A} \equiv I_n - \hat{\rho}W$. It follows that:

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Ay + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\hat{A} - A)y$$

$$
\begin{aligned}
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Ay + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\hat{\rho}-\rho)Wy \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}A(A^{-1}X\beta + A^{-1}\varepsilon) + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\hat{\rho}-\rho)W(A^{-1}X\beta + A^{-1}\varepsilon) \\
&= \beta + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\hat{\rho}-\rho)W(A^{-1}X\beta + A^{-1}\varepsilon)
\end{aligned}
$$

where $y$ has been substituted by its reduced form.

The expected value of the error term $\mathbb{E}(\varepsilon) = 0$ and the geometric series expansion $A^{-1}(I_n - \rho W)^{-1} = \sum_{j=0}^{\infty} \rho^j W^j$ motivate the following calculations:

$$
\begin{aligned}
\mathbb{E}(\hat{\beta}-\beta|\hat{\rho}) &= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(\hat{\rho}-\rho)WA^{-1}X\beta \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}W\left(\sum_{j=0}^{\infty}\rho^j W^j\right)X\beta(\hat{\rho}-\rho) \\
&= (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\left(\sum_{j=1}^{\infty}\rho^j W^j\right)X\beta\,\frac{\hat{\rho}-\rho}{\rho}.
\end{aligned}
$$

This completes the proof. □

### A.2 Proof of Equation (2b)

First of all, note that, according to the notation introduced in Appendix A.1:

$$
\hat{A}A^{-1} = [A + (\hat{A} - A)]A^{-1} = I_n - (\hat{\rho}-\rho)WA^{-1}. \tag{21}
$$

Secondly, define $H \equiv I_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$, and note that the first order condition on $\sigma^2$ leads to the following estimator:

$$
\hat{\sigma}^2 = \frac{1}{n}(\hat{A}y - X\hat{\beta})^{\mathrm{T}}(\hat{A}y - X\hat{\beta}),
$$

which can be restated as it follows:

$$
\hat{\sigma}^2 = \frac{1}{n}(H\hat{A}y)^{\mathrm{T}}(H\hat{A}y),
$$

since $\hat{A}y - X\hat{\beta} = \hat{A}y - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\hat{A}y = H\hat{A}y$.

Thirdly, note that $HX = X - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X = 0$, thus the reduced form (20) and the Eq. (21) allow us to write:

$$
\begin{aligned}
H\hat{A}y &= H\hat{A}A^{-1}(X\beta + \varepsilon) \\
&= H(I_n - (\hat{\rho}-\rho)WA^{-1})(X\beta + \varepsilon) \\
&= -(\hat{\rho}-\rho)HWA^{-1}X\beta + H\varepsilon - (\hat{\rho}-\rho)HWA^{-1}\varepsilon \\
&= -(\hat{\rho}-\rho)Q_\rho X\beta + H\varepsilon - (\hat{\rho}-\rho)Q_\rho\varepsilon.
\end{aligned}
$$

Finally, note that $H$ is symmetric and idempotent, thus:

$$\begin{aligned}\mathbb{E}(n\hat{\sigma}^2|\hat{\rho}) &= n\,\mathbb{E}((H\hat{A}y)^{\mathrm{T}}(H\hat{A}y))\\ &= (\hat{\rho}-\rho)^2(Q_\rho X\beta)^{\mathrm{T}}(Q_\rho X\beta)+\mathbb{E}(\varepsilon^{\mathrm{T}}T^{\mathrm{T}}H\varepsilon)-(\hat{\rho}-\rho)\,\mathbb{E}(\varepsilon^{\mathrm{T}}H^{\mathrm{T}}Q_\rho\varepsilon)\\ &\quad-(\hat{\rho}-\rho)\,\mathbb{E}(\varepsilon^{\mathrm{T}}Q_\rho^{\mathrm{T}}H\varepsilon)+(\hat{\rho}-\rho)^2\,\mathbb{E}(\varepsilon^{\mathrm{T}}Q_\rho^{\mathrm{T}}Q_\rho\varepsilon)\end{aligned}$$

Note that $H$ is a projection matrix, thus it is symmetric ($H = H^{\mathrm{T}}$) and idempotent ($HH = H$), and this also implies that $HQ_\rho = Q_\rho$. It follows that:

$$\begin{aligned}\mathbb{E}(n\hat{\sigma}^2|\hat{\rho}) &= n\,\mathbb{E}((H\hat{A}y)^{\mathrm{T}}(H\hat{A}y))\\ &= (\hat{\rho}-\rho)^2(Q_\rho X\beta)^{\mathrm{T}}(Q_\rho X\beta)+\mathbb{E}(\varepsilon^{\mathrm{T}}H\varepsilon)-(\hat{\rho}-\rho)\,\mathbb{E}(\varepsilon^{\mathrm{T}}H^{\mathrm{T}}Q_\rho\varepsilon)\\ &\quad-(\hat{\rho}-\rho)\,\mathbb{E}(\varepsilon^{\mathrm{T}}Q_\rho^{\mathrm{T}}H\varepsilon)+(\hat{\rho}-\rho)^2\,\mathbb{E}(\varepsilon^{\mathrm{T}}Q_\rho^{\mathrm{T}}Q_\rho\varepsilon)\\ &= (\hat{\rho}-\rho)^2(Q_\rho X\beta)^{\mathrm{T}}(Q_\rho X\beta)+\sigma^2\,\mathrm{tr}(H)\\ &\quad-2\sigma^2(\hat{\rho}-\rho)\,\mathrm{tr}(Q_\rho)+\sigma^2(\hat{\rho}-\rho)^2\,\mathrm{tr}(Q_\rho^{\mathrm{T}}Q_\rho),\end{aligned}$$

from the fact that $\mathrm{tr}(H) = n - k$, Eq. (2b) follows. □

### A.3 Proof of Equation (10)

According to (6), Eq. (5) can be restated as follows:

$$P_\Phi y = \rho P_\Phi W P_\Phi y + P_\Phi X\beta + P_\Phi\varepsilon,$$

hence, the reduced form of $P_\Phi y$ is:

$$P_\Phi y = A^{-1}(P_\Phi X\beta + P_\Phi\varepsilon), \tag{22}$$

where $A$ is defined in Eq. (7).

If the inversion (8) is used for $A$, the block of non-coarsened observations of Eq. (22) becomes:

$$y_P = (A_{PP}^{-1}+A_{PP}^{-1}A_{PC}\tilde{\Xi}^{-1}A_{CP}A_{PP}^{-1})(X_P\beta+\varepsilon_P)+(-A_{PP}^{-1}A_{PC}\tilde{\Xi}^{-1})(X_C\beta+\varepsilon_C). \tag{23}$$

Now, both sides of (23) are premultiplied by $A_{PP}$, and terms rearranged as follows:

$$A_{PP}y_P = X_P\beta+\varepsilon_P+(A_{PC}\tilde{\Xi}^{-1}A_{CP}A_{PP}^{-1})(X_P\beta+\varepsilon_P)-(A_{PC}\tilde{\Xi}^{-1})(X_C\beta+\varepsilon_C). \tag{24}$$

Finally, $A_{PP}$ is replaced by its definition (7), whereas matrix $A_{PC}\tilde{\Xi}^{-1}$ is gathered from third and fourth term on the right hand side of (24):

$$y_P - \rho W_{PP}y_P = X_P\beta+\varepsilon_P+A_{PC}\tilde{\Xi}^{-1}\left[A_{CP}A_{PP}^{-1}(X_P\beta+\varepsilon_P)-(X_C\beta+\varepsilon_C)\right],$$

if term $\rho W_{PP}y_P$ is added to both sides of the previous equation, Eq. (10) is obtained. □

## References


Anselin L (1988) Spatial econometrics: methods and models. Springer, Berlin

Anselin L (2002) Under the hood: Issues in the specification and interpretation of spatial regression models. Agric Econ 27(3):247–267. https://doi.org/10.1111/j.1574-0862.2002.tb00120.x

Arbia G, Espa G, Giuliani D (2016) Dirty spatial econometrics. Ann Reg Sci 56(1):177–189. https://doi.org/10.1007/s00168-015-0726-5

Arbia G, Bera A, Osman D, Suleyman T (2019) Testing impact measures in spatial autoregressive models. Int Regional Sci Rev 43:40–75

Arbia G, Espa G, Giuliani D (2019) Spatial microeconometrics. Routledge, London

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC Press, London

Bee M, Espa G, Giuliani D, Santi F (2017) A cross-entropy approach to the estimation of generalised linear multilevel models. J Comput Graph Stat 26(3):695–708. https://doi.org/10.1080/10618600.2016.1278003

Berman M, Diggle PJ (1989) Estimating weighted integrals of the second-order intensity of a spatial point process. J R Stat Soc B 51:81–92

Cliff AD, Ord JK (1969) The problem of spatial autocorrelation. London papers in regional science 1, studies in regional science 25–55

Cressie NAC (2015) Statistics for spatial data, revised. Wiley, New York City

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B (Methodol) 39(1):1–38

Diggle PJ (1985) A kernel method for smoothing point process data. J R Stat Soc Ser C 34:138–147

Elhorst JP (2014) Spatial econometrics. Springer briefs in regional science. Springer, Berlin

Espa G, Giuliani D, Santi F, Taufer E (2017) Model-based variance estimation in two-dimensional systematic sampling. Metron 75(3):265–275. https://doi.org/10.1007/s40300-017-0125-z

Horn RA, Johnson CR (2013) Matrix analysis, 2nd edn. Cambridge University Press, New York

Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns. Wiley, Chichester

Jarque CM, Bera AK (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econ Lett 6(3):255–259

Kravets N, Hadden WC (2007) The accuracy of address coding and the effects of coding errors. Health Place 13(1):293–298. https://doi.org/10.1016/j.healthplace.2005.08.006

Lee LF (2004) Asymptotic of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6):1899–1925

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. Chapmann&Hall/CRC, Boca Raton

LeSage JP, Pace RK (2014) The biggest myth in spatial econometrics. Econometrics 2:217–249. https://doi.org/10.3390/econometrics2040217

Lu TT, Shou SH (2002) Inverses of $2 \times 2$ block matrices. Comput Math Appl 43:119–129

Qichen Q (2019) Housing price in Beijing. https://www.kaggle.com/ruiqurm/lianjia. Accessed 2020-11-03

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Robert CP, Casella G (2004) Monte Carlo statistical methods. Springer, Berlin

Rubinstein RY, Kroese DP (2004) The cross-entropy method. Springer, New York

Santi F (2020) noisyCE2: cross-entropy optimisation of noisy functions. https://CRAN.R-project.org/package=noisyCE2. R package version 1.1.0

Santi F, Dickson MM, Espa G, Taufer E, Mazzitelli A (2020) Handling spatial dependence under unknown unit locations. Spat Econ Anal 10(1080/17421772):1769171

Zimmerman DL (2008) Estimating the intensity of a spatial point process from locations coarsened by incomplete geocoding. Biometrics 64(1):262–270

Zimmerman DL, Li J (2010) The effects of local street network characteristics on the positional accuracy of automated geocoding for geographic health studies. Int J Health Geogr 9(1):1–11