Semi-parametric Markov models for multi-type point patterns

Ib Thorsgaard Jensen[1], Jean-François Coeurjolly[2], and Rasmus Waagepetersen[1]

[1] *Department of Mathematical Sciences, Aalborg University, Thomas Mannsvej, 9220 Aalborg Øst, Denmark*

[2] *Laboratory Jean Kuntzmann, Université Grenobles Alpes, 150 Place du Torrent, 38400 Saint-Martin d'Hères, France*

Abstract

Multi-type Markov point processes offer a flexible framework for modelling complex multi-type point patterns where it is pertinent to capture both interactions between points as well as large scale trends depending on observed covariates. However, estimation of interaction and covariate effects may be seriously biased in the presence of unobserved spatial confounders. In this paper we introduce a new class of semi-parametric Markov point processes that adjusts for spatial confounding through a non-parametric factor that accommodates effects of latent spatial variables common to all types of points. We introduce a conditional pseudo likelihood for parameter estimation and show that the resulting estimator has desirable asymptotic properties. Our methodology not least has great potential in studies of industry agglomeration and we apply it to study spatial patterns of locations of two types of banks in France.

1 Introduction

Spatial multi-type (or multivariate) point pattern data where points represent objects or events of different types are abundant. Some examples of data sets are locations of rain forest trees of different species, crime scenes of different types of crimes, and locations of different types of industries or banks. The main interests for such data is to study possible dependence on spatial covariates and within- or between-type interactions between points. The statistical toolbox for this can roughly be divided into non-parametric cross summary statistics [e.g. Møller and Waagepetersen, 2003, Baddeley et al., 2014, Cronie and van Lieshout, 2016] and parametric methods.

Parametric modelling may comprise only the first and second order properties of a multi-type point process, more specifically the intensity function and the cross pair correlation functions [Møller and Waagepetersen, 2003]. It could also involve specification of the full multi-type point process distribution in terms of multi-type cluster processes [Jalilian et al., 2015], multi-type log Gaussian Cox processes [Brix and Møller, 2001, Waagepetersen et al., 2015], hierarchical Markov point processes [Högmander and Särkkä, 1999, Grabarnik and Särkkä, 2009], or multi-type Markov point processes [Picard et al., 2009, Goulard et al., 1996, Coeurjolly et al., 2012, Rajala et al., 2018]. Multi-type Markov point processes are typically interpreted in terms of the so-called Papangelou conditional intensity which essentially models the probability of occurrence of an event or object of a specific type and at a specific location conditional on an existing multi-type point configuration. The conditional intensity simultaneously comprises the effects of large scale trends (inhomogeneity) and interactions.

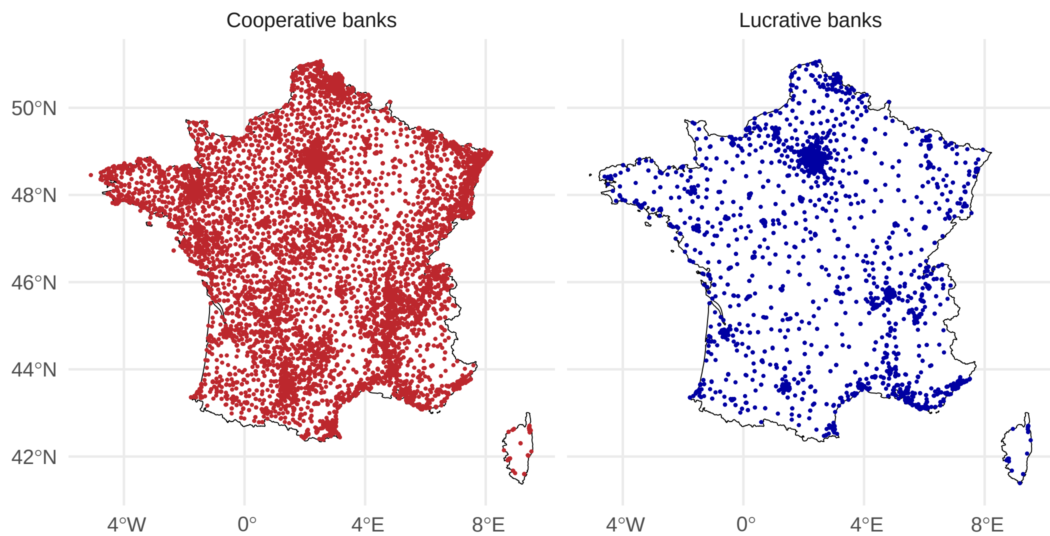
In this paper we introduce semi-parametric inference for multi-type Markov point processes. The basic structure of our proposed model is similar to the one considered in Rajala et al. [2018] but crucially, we enrich the conditional intensity with a non-parametric factor. This factor is common to all types of points and allows the modelling of large scale variation due to unobserved possibly confounding spatial variables. In the context of criminology or spatial econometrics, confounding variation could e.g. be caused by a varying population density which may be hard to quantify at a fine spatial resolution or by a complex urban environment. The problem of spatial confounding has received considerable interest in the context of random field data [e.g. Dupont et al., 2022] but much less so for spatial point pattern data which further supports the importance of our contribution. Due to the presence of the non-parametric factor, standard pseudo likelihood estimation [e.g. Goulard et al., 1996] is not applicable. We therefore propose a conditional pseudo likelihood function that does not depend on the non-parametric factor.

Our proposed model and estimation method is related to Hessellund et al. [2022a] who introduced semi-parametric models for the intensity functions of a multi-type point process and a composite multinomial logistic regression object function for parameter estimation. However, compared with Hessellund et al. [2022a] a distinct advantage of the new method is that first and second order properties can be inferred simultaneously. In contrast, the method in Hessellund et al. [2022a] for estimating the intensity function (i.e. first order properties) needs to be supplemented by a second step for inferring second-order properties [Hessellund et al., 2022b]. The method in Hessellund et al. [2022b] further does not work for multi-type point processes with repulsion between points of the same type. This is e.g. a serious restriction in relation to our data example.

We derive asymptotic properties of the conditional pseudo likelihood parameter estimator in the setting of increasing spatial domain. Since a parametric model for a multi-type Markov point process is a special case of the proposed semi-parametric model (with the non-parametric factor equal to one), we fill a gap in the existing literature on multi-type inhomogeneous Markov point processes, as the asymptotic results in Coeurjolly et al. [2012] and Ba and Coeurjolly [2023] are limited to respectively the stationary or the univariate case. Finite sample properties of parameter estimates and confidence intervals are investigated in a simulation study.

Our approach is strongly motivated by studies of industry clustering in spatial economet-

Figure 1: Locations of two types of French banks.



rics. Without explicitly mentioning point processes, the highly influential paper Duranton and Overman [2005] acknowledged the need to take into account common large scale trends when studying clustering of different types of industries in terms of non-parametric methods. This is in line with Hoover [1937] who defined industry co-localization as the agglomeration of an industry after controlling for general manufacturing. Later, Sweeney and Gómez-Antonio [2016] and Gómez-Antonio and Sweeney [2021] applied univariate Markov point process models to study within-type industry interactions as well as dependence on various economically relevant covariates. However, these papers used purely parametric methods and did not consider multi-type aspects of industry clustering. Our paper therefore also fills a gap in the spatial econometrics literature. We demonstrate the usefulness of our approach by a study of within and between interaction for a bivariate point pattern of so-called lucrative and cooperative banks in France, see Figure 1 and Section 6.

2 Background on point processes and multi-type Markov point processes

A multi-type point process Y on $\mathbb{R}^d$ is a random locally finite set of marked points $(u, m) \in \mathbb{R}^d \times \mathcal{M}$ where the mark space $\mathcal{M}$ is finite. Specifically we let $\mathcal{M} = \{1, \ldots, p\}$ for some $p \geq 1$. Letting $Y_B = Y \cap (B \times \mathcal{M})$ denote the marked points (u, i) in Y whose location part u falls in $B \subseteq \mathbb{R}^d$, local finiteness of Y means that Y_B is finite almost surely for any bounded $B \subset \mathbb{R}^d$. We define in the same way y_B for $y \subseteq \mathbb{R}^d \times \mathcal{M}$ and denote by $\mathcal{C}$ the set of all locally finite configurations and by $\mathcal{C}_B$ the restriction of $\mathcal{C}$ to configurations on $B \subseteq \mathbb{R}^d$, i.e. to $y \in \mathcal{C}$ with $y \subset B \times \mathcal{M}$. For $B \subseteq \mathbb{R}^d$, $|B|$ denotes the volume of B.

A multi-type point process Y can be identified with a multivariate point process $(X_1, \ldots, X_p)^\top$ where X_i consists of those points u for which $(u, i) \in Y$. Moreover, $X = \cup_{i=1}^p X_i$ is the point process of all points in Y with their marks removed. Similarly we use the generic notation $x_i = \{u \in \mathbb{R}^d | (u, i) \in y\}$ for the set of type i points in a multi-type configuration y. For any $B \subseteq \mathbb{R}$ we let $N_{X_i}(B)$ and $N_X(B)$ denote the number of points in X_i respectively X that fall in B (i.e. the numbers of points of type i respectively of any type in Y_B).

In the following sections we consider finite or infinite multi-type point processes specified in terms of a density function or the so-called Papangelou conditional intensity. To avoid excessive use of braces, we abuse notation and write $y \cup (u, i)$ for $y \cup \{(u, i)\}$ and $y \setminus (u, i)$ for $y \setminus \{(u, i)\}$.

2.1 Finite multi-type point processes

For a bounded $B \subset \mathbb{R}^d$, we consider finite multi-type point processes on $S = B \times \mathcal{M}$ specified by a density $f : \mathcal{C}_B \to [0, \infty[$ with respect to the unit rate multi-type Poisson point process on S. The density is of the form

$$f(y) = \frac{1}{c} g(y), \quad y \in \mathcal{C}_B, \tag{1}$$

where g is an unnormalized density defined as a function $g : \mathcal{C} \to [0, \infty]$ with the property that $g(y)$ is finite when y is finite. In Section 3.1 we specify g in more detail as a product of certain non-negative interaction functions. The normalizing constant c is usually intractable. The reason for using the domain $\mathcal{C}$ for g will become clear in Section 2.2 where we discuss infinite multi-type point processes on $\mathbb{R}^d$ and want to avoid reference to a specific B.

We exclusively consider the situation where g (or equivalently f) is hereditary, that is, $g(y) > 0 \Rightarrow g(y') > 0$ for finite $y \in \mathcal{C}$ and $y' \subset y$. Hence any subset y' of an allowed finite configuration y is allowed too. Then for any finite $y \in \mathcal{C}$ and $(u, i) \in S$ with $g\{y \cup (u, i)\} > 0$, the Papangelou conditional intensity is well-defined as

$$\lambda\{(u, i), y\} = \frac{g\{y \cup (u, i)\}}{g\{y \setminus (u, i)\}} \tag{2}$$

and heuristically, the quantity $\lambda\{(u, i), y\}\mathrm{d}u$ may be interpreted as the probability of observing a marked point with mark i in an infinitesimal neighbourhood of u given that Y outside this neighbourhood coincides with y. The definition of λ implies that $\lambda\{(u, i), y\} = \lambda\{(u, i), y \setminus (u, i)\}$ when $(u, i) \in y$. For the set of y and (u, i) with $g\{y \cup (u, i)\} = 0$ we arbitrarily define $\lambda\{(u, i), y\} = 0$.

The Papangelou conditional intensity is said to be locally stable (LS) if there exists $\bar{\lambda} < \infty$ such that for any $(u, i) \in S$ and finite $y \in \mathcal{C}$,

$$\lambda\{(u, i), y\} \leq \bar{\lambda}. \tag{LS}$$

Local stability (LS) ensures that g has finite integral c over $\mathcal{C}_B$ so that the density f is well-defined [e.g. Møller and Waagepetersen, 2003].

2.2 Multi-type point processes on $\mathbb{R}^d$

When $S = \mathbb{R}^d \times \mathcal{M}$, the density (1) does not make sense in general since $g(y)$ could be infinite for $y \in C \subset \mathcal{C}$ where C has positive probability under the unit rate multi-type Poisson process. Instead we extend the definition of the Papangelou conditional intensity (2) to $y \in \mathcal{C}$.

To ensure the validity of the definition (2) for infinite configurations y, we introduce a finite range property (also known as a Markov property), $\lim_{r\to\infty} g\{y_{B(u,r)} \cup (u,i)\}/g\{y_{B(u,r)}\} = g\{y_{B(u,R)} \cup (u,i)\}/g\{y_{B(u,R)}\}$ for some $R < \infty$ where $B(u,R)$ denotes the Euclidean ball with center u and radius R. The Papangelou conditional intensity is then well-defined in the infinite case as

$$\lambda\{(u,i), y\} = \lambda\left\{(u,i), y \cap B(u,R) \times \mathcal{M}\right\} \tag{FR}$$

where the right hand side is obtained from (2). In other words, the probability of observing a marked point (u,i) given a surrounding configuration y only depends on the R-neighbours of u in y. Multi-type point processes satisfying this condition are examples of multi-type Markov point processes [Møller and Waagepetersen, 2003].

We then say that Y is a multi-type point process on $\mathbb{R}^d$ with conditional intensity of the form (FR) for $y \in \mathcal{C}$ provided

$$\mathbb{E} \sum_{(u,i)\in Y} h\{(u,i), Y \setminus (u,i)\} = \mathbb{E} \sum_{i=1}^{p} \int h\{(u,i), Y\}\lambda\{(u,i), Y\}\mathrm{d}u \tag{3}$$

for any non-negative function $h : (\mathbb{R}^d \times \mathcal{M}) \times \mathcal{C} \to [0,\infty[$. The equation (3) is the so-called GNZ formula [Nguyen and Zessin, 1979, Georgii, 1979, 2011] for multi-type point processes. It is of course also valid for a finite multi-type point process as considered in the previous section.

Given the definitions of finite range and local stability we can state the following existence result.

Theorem 1. *[Jansen [2018, Theorem 5.6]] Suppose the Papangelou conditional intensity* λ :

$(\mathbb{R}^d \times \mathcal{M}) \times \mathcal{C} \to [0, \infty[$ *specified by* (2) *satisfies* (FR) *and* (LS). *Then there exists at least one infinite volume multi-type point process* Y *with Papangelou conditional intensity* λ *satisfying the GNZ equation* (3).

There may exist distinct infinite volume point processes sharing the same Papangelou conditional intensity but their conditional distributions on bounded B agree. For an infinite volume point process Y the conditional distribution of Y_B given $Y_{B^c} = z$, $z \in \mathcal{C}_{B^c}$, is specified by a density of the form

$$f_B(y|z) \propto g(y \cup z), \quad y \in \mathcal{C}_B \tag{4}$$

which is closely related to the finite case density (1).

In addition to the GNZ formula we also use the iterated GNZ formula

$$\begin{aligned} &\mathbb{E} \sum_{(u,i),(v,j)\in Y}^{\neq} h[(u,i),(v,j), Y \setminus \{(u,i),(v,j)\}] \\ =&\mathbb{E} \sum_{i,j=1}^{p} \int h\{(u,i),(v,j), Y\}\lambda\{(u,i),(v,j), Y\} \mathrm{d}u\mathrm{d}v \end{aligned} \tag{5}$$

for any non-negative function $h : \{(\mathbb{R}^d \times \mathcal{M}) \times \mathcal{C}\}^2 \to [0, \infty[$ and where $\lambda\{(u,i),(v,j),Y\} = \lambda\{(u,i),Y\}\lambda\{(v,j),Y \cup (u,i)\}$.

3 Semi-parametric modelling and methodology

3.1 Semi-parametric Markov point processes

We introduce a specific class of multi-type Markov point processes with points in a bounded domain $W \subset \mathbb{R}^d$ by specifying a particular model for the unnormalised density g in (1). We consider densities f of the form

$$f(y) \propto g(y) = \prod_{(u,i)\in y} \phi_i(u) \prod_{i,j=1}^{p} \phi_{ij}(x_i, x_j), \tag{6}$$

where $y \in \mathcal{C}_W$ is any configuration of marked points in $W \times \mathcal{M}$. The non-negative functions ϕ_i are used to model first order effects, i.e. spatial trends in the occurrence of type i points, and are assumed to be of the form

$$\phi_i(u) = \phi_0(u) \exp\{\gamma_{i0}^\top z_i(u)\} \tag{7}$$

where $z_i(u)$ and γ_{i0} respectively denote a vector of covariates at location u and a regression parameter vector. The function ϕ_0 is common to all X_i and is unrestricted except for being non-negative. This function constitutes a key component of our model since it can capture large scale variation common to all types of points yet not explained by the observed covariates.

The non-negative functions ϕ_{ij} model higher-order interaction terms. We assume that these interactions take the following log-linear form

$$\phi_{ij}(x_i, x_j) = \exp\{\gamma_{ij}^\top v_{ij}(x_i, x_j)\} \tag{8}$$

where the v_{ij} are (possibly) vector-valued functions depending on pairs of point patterns and the γ_{ij} are interaction parameter vectors. Our examples of v_{ij} below depend on an interaction range R_{ij} which may be considered known or a further target for estimation.

- Multi-type Strauss hard-core interaction: Define for interaction ranges $0 \leq R_{ij}$,

$$s(u, x_j, R_{ij}) = \left(\frac{1}{2}\right)^{1(i=j)} \sum_{v \in x_j \setminus u} 1(\|u - v\| \leq R_{ij})$$

to be the number of R_{ij} close neighbours of u in $x_j \setminus u$. Introducing further hard-core distances $0 \le r_{ij} \le R_{ij}$,

$$v_{ij}(x_i, x_j) = \begin{cases} \sum_{u \in x_i} s(u, x_j, R_{ij}) & \text{if } \|u - v\| \ge r_{ij} \text{ for all } u \in x_i, v \in x_j \setminus u \\ -\infty & \text{otherwise.} \end{cases}$$

With $r_{ij} = 0$ for all i, j, a multi-type Strauss interaction is obtained.

- Multi-type Geyer saturation interaction:

$$v_{ij}(x_i, x_j) = \sum_{u \in x_i} \min\{s(u, x_j, R_{ij}), c_{ij}\}$$

where $s(u, x_j, R_{ij})$ is defined as in the previous example and $0 < c_{ij} < \infty$ is the saturation parameter.

In the homogeneous case, i.e. when the ϕ_i are constant, it is known [see e.g. Baddeley et al., 2015, Ba and Coeurjolly, 2023], that the multi-type Strauss hard-core point process (with $r_{ij} > 0$ or $\gamma_{ij} < 0$ in case $r_{ij} = 0$) and the multi-type Geyer satisfy (LS) and (FR) with respectively $R = R^\vee, 2R^\vee$, where $R^\vee = \max_{i,j}\{R_{ij}\}$. If one adds simple conditions like $\sup_i \|z_i\|_\infty < \infty$ and $\|\phi_0\|_\infty < \infty$, the inhomogeneous versions of these models still satisfy (FR)-(LS) and so exist in the infinite volume according to Theorem 1. In addition to multi-type Strauss and multi-type Geyer, other options also exist such as the multi-type area interaction model [Picard et al., 2009, Nightingale et al., 2019].

Equations (7)-(8) allow us to rewrite the density as

$$f(y) \propto \left\{ \prod_{(u,i) \in y} \phi_0(u) \right\} \exp\left\{\gamma^\top v(y)\right\}$$

where γ and $v(y)$ are the stacked vectors given by $\gamma = (\gamma_1^\top, \ldots, \gamma_p^\top)^\top$ and

$v(y) = \{v_1(y)^\top, \ldots, v_p(y)^\top\}^\top$ with $\gamma_i = (\gamma_{i0}^\top, \gamma_{i1}^\top, \ldots, \gamma_{ip}^\top)^\top$ and

$$v_i(y) = \left\{ \sum_{u \in x_i} z_i(u)^\top, v_{i1}(x_i, x_1)^\top, \ldots, v_{ip}(x_i, x_p)^\top \right\}^\top.$$

Then the Papangelou conditional intensity (2) of Y becomes

$$\lambda\{(u,i), y\} = \phi_0(u) \ \exp\left[\gamma^\top v\{(u,i), y\}\right] \tag{9}$$

for $(u,i) \in \mathbb{R}^d \times \mathcal{M}$ and $y \in \mathcal{C}$, where we define

$$v\{(u,i), y\} = v\{y \cup (u,i)\} - v\{y \setminus (u,i)\} \tag{10}$$

which measures the local contribution of a marked point (u,i) to a configuration $y \setminus (u,i)$ or y depending on whether $(u,i) \in y$ or not. We define for $(u,i) \in \mathbb{R}^d \times \mathcal{M}$ and $y \in \mathcal{C}$ the following probability that does not depend on ϕ_0,

$$p\{(u,i), y\} = \frac{\lambda\{(u,i), y\}}{\Lambda(u,y)} = \frac{\exp\left[\gamma^\top v\{(u,i), y\}\right]}{\sum_{j=1}^p \exp\left[\gamma^\top v\{(u,j), y\}\right]}, \tag{11}$$

where $\Lambda(u,y) = \sum_{j=1}^p \lambda\{(u,j), y\}$ and where we take the convention $0/0 = 0$ in (11).

By the infinitesimal interpretation of $\lambda\{(u,i), y)\}$, $\Lambda(u,y)\mathrm{d}u$ can be interpreted as the conditional probability of observing a marked point of *any* type in the vicinity of u given the rest of the point configuration. It follows that $p\{(u,i), y\}$ can be interpreted as the probability that a point located at u is of type i given that a point from Y occurs at u and given that Y coincides with y in the surrounding of u.

3.2 Conditional pseudo likelihood

Assume we have a multi-type Markov point process model satisfying (LS)-(FR) observed on W. Since the density (6) is only specified up to an unknown normalizing constant, maximum likelihood estimation is not straightforward. The so-called pseudo likelihood [e.g. Goulard et al., 1996] formulated in terms of the Papangelou conditional intensity is therefore often employed instead. The conditional intensity does not depend on the normalizing constant but in our case the presence of the unspecified non-parametric component rules out the usual pseudo likelihood approach. Instead we propose a conditional pseudo likelihood approach as detailed in the following.

Let $D = W \ominus R$, which is the domain W eroded by R. It follows from (FR) that for any $(u,i) \in D \times \mathcal{M}$ and $y \in \mathcal{C}$, $\lambda\{(u,i), y\}$ and therefore $p\{(u,i), y\}$ can always be evaluated using y_W. To estimate the parameter vector γ, we propose to use the conditional pseudo likelihood $L(\gamma) = \prod_{(u,i)\in Y_D} p\{(u,i), Y \setminus (u,i)\}$. The conditional pseudo log-likelihood $\ell(\gamma) = \log L(\gamma)$ is

$$\ell(\gamma) = \sum_{(u,i)\in Y_D} \Big[\log \lambda\{(u,i), Y \setminus (u,i)\} - \log \Lambda\{u, Y \setminus (u,i)\} \Big]. \tag{12}$$

With the logistic form (11) for the probabilities p, $\ell(\gamma)$ is equivalent to the log likelihood of a multinomial logistic regression. However, in a standard multinomial logistic regression, the same set of covariates is usually shared among the p types, which requires $z_i = z_j$ and $v_{il}(\cdot,\cdot) = v_{jl}(\cdot,\cdot)$ for $i \neq j$ and each $l = 1, \ldots, p$. In this case parameter estimates can be obtained easily using standard software for multinomial logistic regression, e.g. the `VGAM` `R` package. However, if different covariates are used for different types or $v_{il}(\cdot,\cdot) \neq v_{jl}(\cdot,\cdot)$ for an l and some $i \neq j$, tailor made code is necessary. Also certain identifiability issues may necessitate reparametrizations, see Sections 3.3 and 3.5.

3.3 Identifiability

For the interaction components of $v\{(u,i),y\}$ given by (10) we use the shorthand notation $v_{ii}(u) = v_{ii}(x_i \cup u, x_i \cup u) - v_{ii}(x_i \setminus u, x_i \setminus u)$, $v^1_{ij}(u) = v_{ij}(x_i \cup u, x_j) - v_{ij}(x_i \setminus u, x_j)$, and $v^2_{ji}(u) = v_{ji}(x_j, x_i \cup u) - v_{ji}(x_j, x_i \setminus u)$, $i \neq j$. The Papangelou conditional intensity can then be written as

$$\lambda\{(u,i),y\} = \exp\left[\gamma_{i0}^\top z_i(u) + \gamma_{ii}^\top v_{ii}(u) + \sum_{\substack{l=1\\ l\neq i}}^{p}\left\{\gamma_{il}^\top v^1_{il}(u) + \gamma_{li}^\top v^2_{li}(u)\right\}\right].$$

In practice, we often have $z_i(u) = z_j(u)$ for all i, j. It is then clear that adding the same constant to all γ_{i0} will not change the log likelihood (12). This identifiability issue is typically resolved by considering contrasts $\beta_{i0} = \gamma_{i0} - \gamma_{p0}$ or, equivalently, fixing $\gamma_{p0} = 0$.

Suppose further that we choose the same type of interaction (e.g. Strauss hard-core with same interaction ranges and hard-core distances) between x_i and x_l for $i = 1, \dots, p$. Then $v^1_{il}(u) = v^1_{jl}(u)$ for all i, j, and we cannot identify the parameters γ_{il}, $i = 1, \dots, p$ (and with similar considerations for the γ_{li}). Again this could be resolved by fixing one parameter, e.g. $\gamma_{pl} = 0$. Since we in practice estimate the interaction ranges from the given data, it is unlikely to encounter the situation $v_{il}(u) = v_{jl}(u)$ for all $i \neq j$ which means that the identifiability issue is less of a concern for the interaction parameters γ_{il}.

In case of symmetry, $v_{ij}(x_i, x_j) = v_{ji}(x_j, x_i)$, the parameters γ_{ij} and γ_{ji} are obviously not identifiable. In general symmetry is not satisfied in which case one could consider $\gamma_{ij} \neq \gamma_{ji}$. One example is a multi-type Strauss process with $R_{ij} < R_{ji}$. This becomes a kind of multiscale model where $\gamma_{ij} + \gamma_{ji}$ is a parameter related to pairs of type i and j points with interpoint distance less than R_{ij} and γ_{ji} is a parameter related to pairs with distance in $]R_{ij}, R_{ji}]$. Another example with $v_{ij}(x_i, x_j) \neq v_{ji}(x_j, x_i)$ is a multi-type Geyer saturation process with $c_{ij} = c_{ji} = 1$ and $R_{ij} = R_{ji}$ (consider the case $x_i = \{u\}$ and $x_j = \{v, w\}$ where u, v, w are all within distance

$R_{ij} = R_{ji}$ of each other). In the interest of model parameter parsimony one may nevertheless prefer to restrict to the symmetric parametrization $\gamma_{ij} = \gamma_{ji}$.

3.4 Unbiasedness and approximate covariance matrix

The basic requirement for consistency of the conditional pseudo likelihood estimator obtained by maximizing (12) is unbiasedness of the score function. Let λ' and Λ' denote the derivative (i.e. gradient vector) of λ and Λ with respect to γ. The conditional pseudo score of (12) is

$$e(\gamma) = \frac{\mathrm{d}\ell(\gamma)}{\mathrm{d}\gamma} = \sum_{(u,i)\in Y_D} h\{(u,i), Y \setminus (u,i)\} \tag{13}$$

with

$$h\{(u,i),y\} = \frac{\lambda'\{(u,i),y\}}{\lambda\{(u,i),y\}} - \frac{\Lambda'(u,y)}{\Lambda(u,y)} = v\{(u,i),y\} - E\{u, y \setminus (u,i)\}$$

and $E(u,y) = \sum_{l=1}^{p} v\{(u,l),y\}p\{(u,l),y\}$ for $(u,i) \in \mathbb{R}^d \times \mathcal{M}$ and $y \in \mathcal{C}$ with $(u,i) \notin y$. Hence $E(u,y)$ can be viewed as a conditional expectation of $v\{(u,I),y\}$ for a random index I with $\mathbb{P}(I = i) = p\{(u,i),y\}$, and $h\{(u,i),y\}$ can be viewed as a score residual vector. This suggests that $e(\gamma)$ has mean zero which can be formally verified using the GNZ formula (3),

$$\begin{aligned}
\mathbb{E}\, e(\gamma) &= \mathbb{E} \sum_{(u,i)\in Y_D} h\{(u,i), Y \setminus (u,i)\} = \mathbb{E} \sum_{i=1}^{p} \int_D h\{(u,i),Y\}\lambda\{(u,i),Y\}\,\mathrm{d}u \\
&= \mathbb{E} \int_D \left[\sum_{i=1}^{p} \lambda'\{(u,i),Y\} - \frac{\Lambda'(u,Y)}{\Lambda(u,Y)} \sum_{i=1}^{p} \lambda\{(u,i),Y\} \right] \mathrm{d}u \\
&= \mathbb{E} \int_D \Big\{ \Lambda'(u,Y) - \Lambda'(u,Y) \Big\}\, \mathrm{d}u = 0.
\end{aligned}$$

Following the asymptotic results in Section 4, an approximate covariance matrix of the estimator $\hat{\gamma}$ is given in terms of the sensitivity matrix $S(\gamma) = -\mathbb{E}\mathrm{d}e(\gamma)/\mathrm{d}\gamma^\top$ and the variance-covariance matrix $\Sigma(\gamma) = \mathbb{V}\mathrm{ar}\{e(\gamma)\}$ of the score function. We verify in Section A in the supplementary material that the sensitivity matrix is

$$S(\gamma) = \mathbb{E}\int_D \Lambda(u,Y)V(u,Y)\mathrm{d}u \tag{14}$$

where for $y \cap \{(u,1),\dots,(u,p)\} = \emptyset$, $V(u,y) = E^2(u,y) - E(u,y)E(u,y)^\top$ can be viewed as a conditional variance with $E^2(u,y) = \sum_{i=1}^p v\{(u,i),y\}v\{(u,i),y\}^\top p\{(u,i),y\}$. We also obtain in Section A in the supplementary material, under the finite range assumption (FR), that the score covariance matrix is

$$\begin{aligned}&\Sigma(\gamma) = S(\gamma)+\\ &\mathbb{E}\sum_{i,j=1}^{p}\int_D\int_{D\cap B(u,R)} h\{(u,i),Y\cup(v,j)\}h\{(v,j),Y\cup(u,i)\}^\top\lambda[\{(u,i),(v,j)\},Y]\mathrm{d}u\mathrm{d}v\end{aligned} \tag{15}$$

where we remind that $\lambda[\{(u,i),(v,j)\},Y] = \lambda\{(u,i),Y\}\lambda\{(v,j),Y\cup(u,i)\}$.

The expressions (14)-(15) cannot be evaluated due to the dependence on the unknown non-parametric factor ϕ_0. Following Coeurjolly and Rubak [2013], one may, however, apply GNZ (3) and iterated GNZ (5) backward to obtain the following unbiased estimators

$$\hat{S}(\gamma) = \sum_{(u,i)\in Y_D} V\{u, Y\setminus(u,i)\} \tag{16}$$

$$\hat{\Sigma}(\gamma) = \sum_{\substack{(u,i),(v,j)\in Y_D:\\ \|u-v\|\le R}}^{\neq} h\{(u,i),Y\setminus(u,i)\}h\{(v,j),Y\setminus(v,j)\}^\top. \tag{17}$$

3.5 Reparametrization

As discussed in Section 3.3, we may need to represent the full parameter vector $\gamma \in \mathbb{R}^k$ in terms of a lower dimensional parameter vector $\beta \in \mathbb{R}^{k'}$, $k' < k$, i.e. $\gamma = T(\beta)$ where $T : \mathbb{R}^{k'} \to \mathbb{R}^k$ is a full rank linear mapping. Then an estimator $\hat{\beta}$ is obtained by maximizing the log likelihood

$\ell\{T(\beta)\}$. With an abuse of notation the resulting score function $e(\beta) = T^\top e\{T(\beta)\}$ is still unbiased where we also let T denote the $k \times k'$ Jacobian matrix of the mapping T. Continuing this abuse of notation, the resulting sensitivity matrix and score variance matrix become

$$S(\beta) = T^\top S\{T(\beta)\}T \quad \text{and} \quad \Sigma(\beta) = T^\top \Sigma\{T(\beta)\}T, \tag{18}$$

and the approximate covariance matrix for $\hat{\beta}$ becomes $S(\beta)^{-1}\Sigma(\beta)S(\beta)^{-1}$. For the asymptotic results in Section 4 we consider the reparametrized model with β varying in an open convex subset of $\mathbb{R}^{k'}$.

3.6 Kernel estimation of ϕ_0

Let k be a probability density on $\mathbb{R}^d$ with support $B(0, \omega)$ for $\omega > 0$. By GNZ,

$$\mathbb{E} \sum_{(u,i)\in Y} \frac{k(u-v)}{\exp\left[\gamma^\top v\{(u,i), y \setminus (u,i)\}\right]} = p\mathbb{E} \int_{\mathbb{R}^d} \phi_0(u)k(u-v)\mathrm{d}u \approx p\phi_0(v),$$

assuming $\phi_0(\cdot)$ is roughly constant on $B(v, \omega)$. Hence we have the kernel estimator

$$\hat{\phi}_0(v) = \frac{1}{p} \sum_{(u,i)\in Y} \frac{k(u-v)}{\exp\left[\gamma^\top v\{(u,i), y \setminus (u,i)\}\right]}.$$

It is an open question how to choose an optimal ω and the theoretical properties of the kernel estimator are unknown. We use the kernel estimator for purely exploratory purposes in Section 6.

4 Asymptotic results

We consider a multi-type Markov model as defined in Section 3.1. We assume it satisfies (LS) and (FR) and thus there exists at least one multi-type Markov point process on $\mathbb{R}^d \times \mathcal{M}$ which encourages us to consider an increasing domain framework. Following Section 3.5 we consider the reparametrization $\gamma = T(\beta)$ and consider the properties of the estimator $\hat{\beta} \in \mathbb{R}^{k'}$. We consider a sequence of increasing observation windows $(W_n)_{n\geq 1}$ and define as previously $D_n = W_n \ominus R$. We also add the index n to emphasize the dependence of $\hat{\beta}_n$, ℓ_n, e_n, S_n, and Σ_n on the observation window W_n.

For our asymptotic results we need the following quite mild conditions (see discussion of conditions in the final paragraph of this section).

C.1 The sequence of eroded observation windows D_n is increasing with $\lim_{n\to\infty} |D_n| = \infty$.

C.2 The Papangelou conditional intensity function has the log-linear specification given by (9)-(10). A reparametrization $\gamma = T(\beta)$ is used with $\beta \in B \subseteq \mathbb{R}^{k'}$ where T is a full rank linear mapping and B is an open convex subset of $\mathbb{R}^{k'}$. The property (LS) is satisfied for the true value of $\gamma = T(\beta)$. The finite range property (FR) is satisfied for the conditional intensity (and hence for $v\{(u,i),y\})$ for some $0 < R < \infty$.

C.3 Recall the notation in Section 3.3 for the interaction components of $v\{(u,i),y\}$. We assume that there exist $c_1 > 0$ and $c_2 \geq 0$ such that for any $(u,i) \in \mathbb{R}^d \times \mathcal{M}$ and y with $\lambda\{(u,i),y\} > 0$, $\sup_{u\in\mathbb{R}^d} \|z_i(u)\| < c_1$, $i = 1,\dots,p$, and

$$\begin{aligned}
&\|v_{ii}(u)\| \leq c_1 N_i\{B(u,R)\}^{c_2}, \quad \|v^1_{ij}(u)\| \leq c_1 N_j\{B(u,R)\}^{c_2} \\
&\|v^2_{ji}(u)\| \leq c_1 N_j\{B(u,R)\}^{c_2}.
\end{aligned} \tag{19}$$

C.4 The scaled limiting sensitivity and score variance matrices satisfy

$$\varphi_S^* = \liminf_{n\to\infty} \varphi_{\min}\big\{|D_n|^{-1}S_n(\beta)\big\} > 0 \text{ and } \varphi_\Sigma^* = \liminf_{n\to\infty} \varphi_{\min}\big\{|D_n|^{-1}\Sigma_n(\beta)\big\} > 0,$$

where $\varphi_{\min}(M)$ denotes the smallest eigenvalue of a matrix M.

The following result states the asymptotic behavior of the conditional pseudo likelihood estimator.

Theorem 2. *Under the assumptions C.1-C.4, the following holds as $n \to \infty$.*
(i) There exists a sequence of local maxima $\hat\beta_n$ of $\ell_n(\beta)$ such that $\|\hat\beta_n - \beta\| = O_{\mathbb{P}}(|D_n|^{-1/2})$.
(ii) The sequence $\{\hat\beta_n\}_{n\geq 1}$ is asymptotically normal,

$$\Sigma_n(\beta)^{-1/2}S_n(\beta)(\hat\beta_n - \beta) \xrightarrow{d} N(0, I_{k'}), \tag{20}$$

where $\Sigma_n(\beta)$ and $S_n(\beta)$ are given by (18).

The proofs of these results are given in Section B in the supplementary material. By (20), the covariance matrix of $\hat\beta_n$ can be approximated by $V_n(\beta) = S_n^{-1}(\beta)\Sigma_n(\beta)S_n(\beta)^{-1}$. In practice we replace S_n and V_n by the estimates (16) and (17) and plug in $\hat\beta_n$ for β. This allows us to derive asymptotic confidence intervals for the components of β.

Consider the assumptions C.2-C.3 for the models mentioned in Section 3.1 for which properties (LS)-(FR) are valid as already discussed when covariates and the non-parametric factor are bounded. The condition (19) is also satisfied for the aforementioned models with $c_2 = 0$ for the multi-type Strauss hard-core interaction (with $r_{ij} > 0$), the multi-type Geyer saturation interaction, and the multi-type area-interaction, and with $c_2 = 1$ for the multi-type Strauss interaction. Assumption C.4 depends on the behaviour of the non-parametric factor and the covariates on the whole of $\mathbb{R}^d$ and thus cannot be verified. However, for C.4 to hold, it is required that the intersection of D_n with the support of ϕ_0 grows at the same rate as D_n.

5 Simulation Study

We demonstrate the usefulness of the conditional pseudo likelihood approach by a simulation study considering a semi-parametric multi-type Markov model with $\mathcal{M} = \{1, 2, 3\}$, and first order effects depending on an intercept and one covariate $z = \{z(u)\}_{u \in W}$ common to all three types of points. The baseline function φ_0 and the covariate z are obtained as realizations of two independent Gaussian random fields (GRF). Both are simulated with an exponential covariance function, $C(u, v) = \sigma^2 \exp(-\|u - v\| / \phi)$. For φ_0 we use a GRF (truncated at zero if needed) with mean 350, $\sigma^2 = 900$, and $\phi = 0.1$, and for z we use a zero mean GRF with $\sigma^2 = 0.2$ and $\phi = 0.1$.

For the interaction part, we consider an inhomogeneous multi-type Poisson process, a multi-Strauss process, and a multi-Geyer saturation process (for brevity, Poisson, Strauss, and Geyer in the following). For all three processes, 1800 replications are simulated on $W_1 = [0, 1]^2$ and $W_2 = [0, 2]^2$ using the R package `spatstat` [Baddeley et al., 2015]. From each simulation we obtain parameter estimates as well as estimated standard errors and approximate 95% confidence intervals based on the asymptotic results in Section 4. The number 1800 yields a Monte-Carlo standard error of 0.005 for the estimated coverage rates provided the true coverage rate is 95%. Single realizations of φ_0 and z are simulated on W_2 and used across all replications. For all three processes we use parameters $\gamma_{10,2} = 1/2$, $\gamma_{20,2} = -1/2$, $\gamma_{30,2} = 0$ for the covariate z. For the intercepts we use $\gamma_{10,1} = \gamma_{20,1} = \gamma_{30,1} = 0$ for Poisson, $\gamma_{10,1} = \gamma_{20,1} = \gamma_{30,1} = \log 1.6$ for Strauss, and $\gamma_{10,1} = \log(1.3/1.4)$, $\gamma_{20,1} = \log(1.3/1.6)$, $\gamma_{30,1} = \log 1.3$ for Geyer.

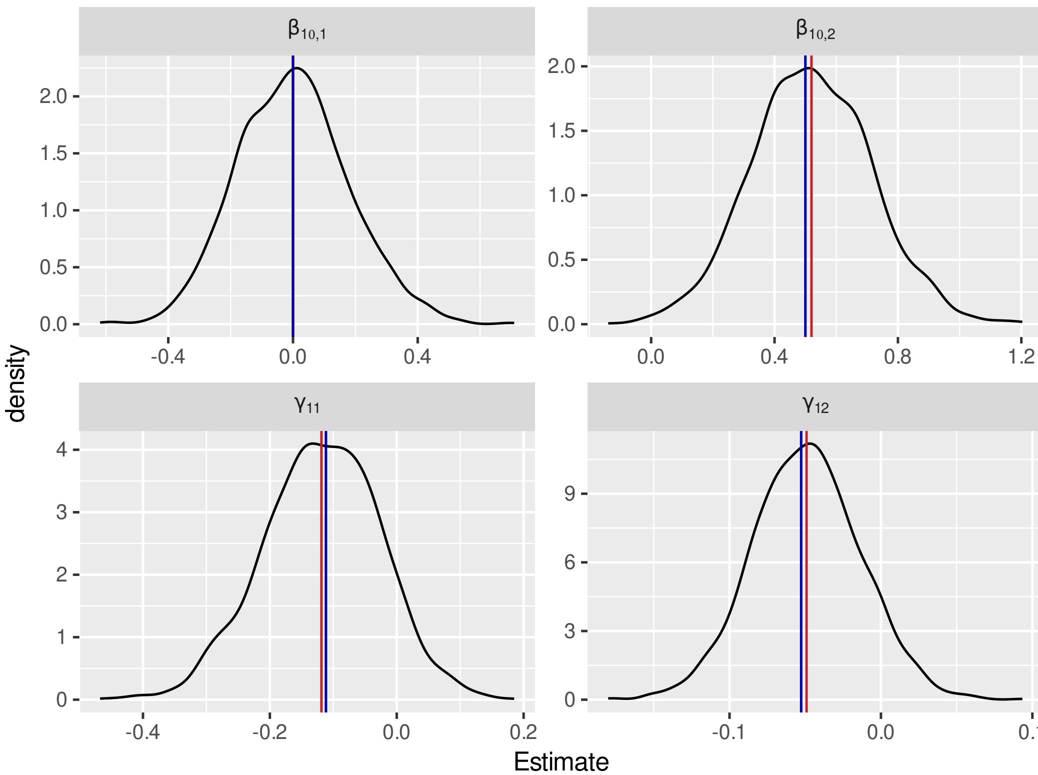
For all three processes, we use a within-type interaction range of 0.02 and a between-type interaction range of 0.04 for both the simulation (where relevant) and model fitting. For Strauss, we use interaction parameters $\gamma_{ii} = \log 0.8$ and $\gamma_{ij} = \log 0.9$, $i \neq j$. No function for simulating Geyer multi-type processes with between-type interaction exists in `spatstat`. Instead we simulate three independent single-type saturation processes, and thus no between-type interaction is present. We choose within-type interaction parameters $\gamma_{11} = \log 1.1$, $\gamma_{22} = \log 1.2$, $\gamma_{33} = \log 0.8$, and a common saturation parameter 10 for all three types of points. The expected numbers of each type of points across the three models fall in the range 360-385 in case of W_1 and in the range 1417-1519 in case of W_2.

The Poisson process is nested within the Strauss process (with the interaction parameters equal to zero) and we fit the Strauss process for both the Poisson and the Strauss simulations. We impose symmetry on the interaction parameters $\gamma_{ij} = \gamma_{ji}$ for any $i \neq j$ and we take the third type as the reference type for the intercept and covariate parameters, i.e. we estimate $\beta_{i0} = \gamma_{i0} - \gamma_{30} \in \mathbb{R}^2$, $i = 1, 2$. Figure 2 shows a few representative kernel density plots of parameter estimates obtained from simulations of the Strauss process in case of the window W_1. The plots show that the estimates are close to being normally distributed with negligible bias. Further plots (not shown) confirm this for the remaining parameter estimates under Poisson, Strauss, and Geyer simulations.

Figure 3 shows standard errors of the parameter estimates as well as means of estimated standard errors for all three models. For Poisson and Strauss, the means of estimated standard errors are in good agreement with the actual standard errors. For Geyer, the standard errors for the parameters associated with the spatial covariates and intercepts are somewhat underestimated on W_1, while all standard errors are well-estimated on W_2. In accordance with the asymptotic results, the ratios between standard errors for W_2 and W_1 are close to $\sqrt{|W_2|/|W_1|} = 2$.

Figure 4 shows estimates of coverage rates for approximate 95% confidence intervals based on asymptotic normality as well as 95% Monte Carlo probability intervals for the coverage rates assuming that the true coverage rate is 95%. On W_1, coverage rates are very close to the nominal 95% for Poisson and Strauss and a bit too low for Geyer. On W_2, coverage rates are close to the nominal level for all models.

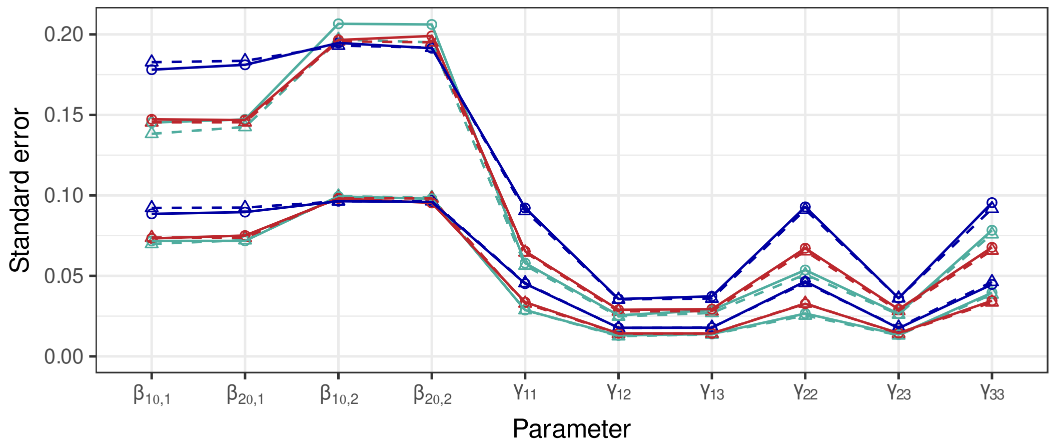
Figure 2: Kernel density plots of estimates of the parameters $\beta_{10,1}, \beta_{10,2}, \gamma_{11}$, and γ_{12} for simulations of the multi-type Strauss process on $W_1 = [0,1]^2$. The true parameter value and the mean parameter estimate are shown with blue and red vertical bars.



6 Clustering of French banks

Most French banks can be divided into two types: cooperative and lucrative. Cooperative banks originated within local communities with the purpose of providing loans to local retail customers and moderately sized enterprises. Lucrative banks are limited liability companies typically founded within larger cities. We consider a dataset of locations of 11,244 cooperative banks and 3,448 lucrative banks recently collected by Artis et al. [2025], see Figure 1. The bank locations are aggregated from three different cooperative banks and two different lucrative banks. As of 01/01/2023, these banks cover 92% of French clients.

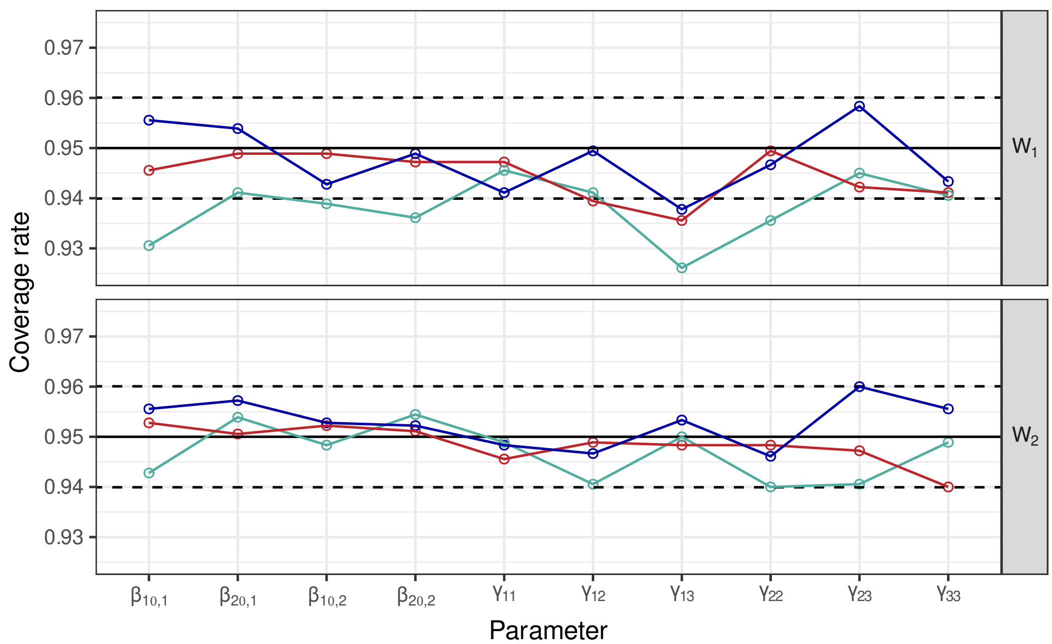
Figure 3: Standard errors (points and solid lines) and means of estimated standard errors (triangles and dashed lines) for Poisson (red), Strauss (blue), and Geyer (teal) and windows W_1 (upper curves) and W_2 (lower curves). Curves are added for easier visual interpretation.



In addition to the bank locations, socio-economic variables and estimated population density are available for 1681 administrative units covering France [Artis et al., 2025]. Detailed definitions of these covariates are given in Section C in the supplementary material (see also Figure 5). We treat these variables as piece-wise constant covariates with constant value within administrative units.

We model the data using a bivariate Geyer saturation model which allows for both positive and negative interactions within and between the two types of banks. In the context of banks, the Geyer saturation parameter models an effect of diminishing returns after reaching a ceiling - e.g. a possible positive effect on the conditional intensity from adding a neighbouring bank vanishes when a neighbourhood is already saturated with banks. Our first aim is to study possible differences regarding the association between the bank locations and the socio-economic variables. While the structures of the spatial patterns of the two types of banks may differ, it is clear that the intensities of the banks are strongly dependent on urbanization. If this trend is of a similar form for cooperative and lucrative banks, it can be accounted for by

Figure 4: Coverage rates for nominal 95% confidence intervals in case of W_1 (upper plot) and W_2 (lower plot). Red, blue, and teal for simulations of Poisson, Strauss, and Geyer processes. 95% Monte Carlo probability intervals for the coverage rates are shown with black dashed lines. Curves are added for easier visual interpretation.



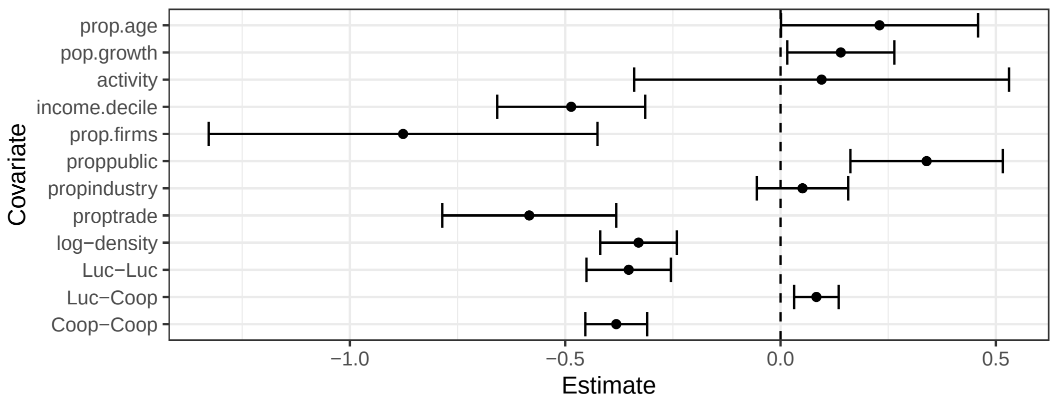
the non-parametric factor in our model. Urbanization is likely closely related to population density so that including both the non-parametric factor as well as population density might seem superfluous. However, when population density is included as a covariate it only explains discrepancies between the conditional intensities of one type of bank relative to the other. Thus the non-parametric factor may capture absolute effects of population density among other spatial factors common to both types of banks. Secondly, we assess possible unexplained interactions within and between the types of banks by considering estimates of the interaction parameters.

As mentioned in Section 3.3, it is not possible to identify the absolute effects of the spatial covariates on the intensities of each type of banks. We use the lucrative banks as a reference and estimate the effects of the spatial covariates on the cooperative banks relative to the lucrative banks. It is on the other hand possible to identify all interaction effects provided not all interaction ranges are identical. We use a common value $R_w = R_{11} = R_{22}$ for the within-type interaction ranges and another value $R_b = R_{12} = R_{21}$ for the between-type interaction ranges. We moreover use a common saturation parameter c for all combinations of types of banks. We estimate R_w, R_b, and c by fitting the model over a grid of combinations of interaction ranges and saturation parameter and choosing the combination with the largest conditional pseudo likelihood. This results in values $R_w = 0.006$, $R_b = 0.004$, and $c = 4$. The bank locations are given in terms of longitude and latitude, and the interaction ranges correspond to approximately 300 and 200 metres respectively. To avoid edge effects, we erode the observation window by a distance of twice the largest interaction range included in the grid search, in this case 0.02.

Figure 5 illustrates the parameter estimates. For `activity` and `propindustry`, the confidence intervals include zero. The confidence intervals for `prop.age`, `pop.growth`, and `proppublic` are confined to positive parameter values, and the confidence intervals for `income.decile`, `prop.firms`, `proptrade`, and `log-density` to negative parameter values. This indicates that administrative units with a higher proportion of younger people relative to older people, greater recent population growth, and a high proportion of public-sector jobs tend to have a higher proportion of cooperative banks. Conversely, administrative units with higher income inequality, a greater proportion of small businesses, a higher proportion of jobs in the trade sector, and higher population density tend to have a higher proportion of lucrative banks. A plot showing the fitted probability that a bank is cooperative conditional on the rest of the point pattern can be found in Figure S1 in the supplementary material.

Both estimated within-type interactions are negative, while the estimated between-type interaction is positive. It is well-known in the economics literature that banks tend to cluster [Lord and Wright, 1981, Chang et al., 1997]. One might hypothesize many potential explanations for this phenomenon, such as proximity to a qualified labour-pool or beneficial infrastructure. The

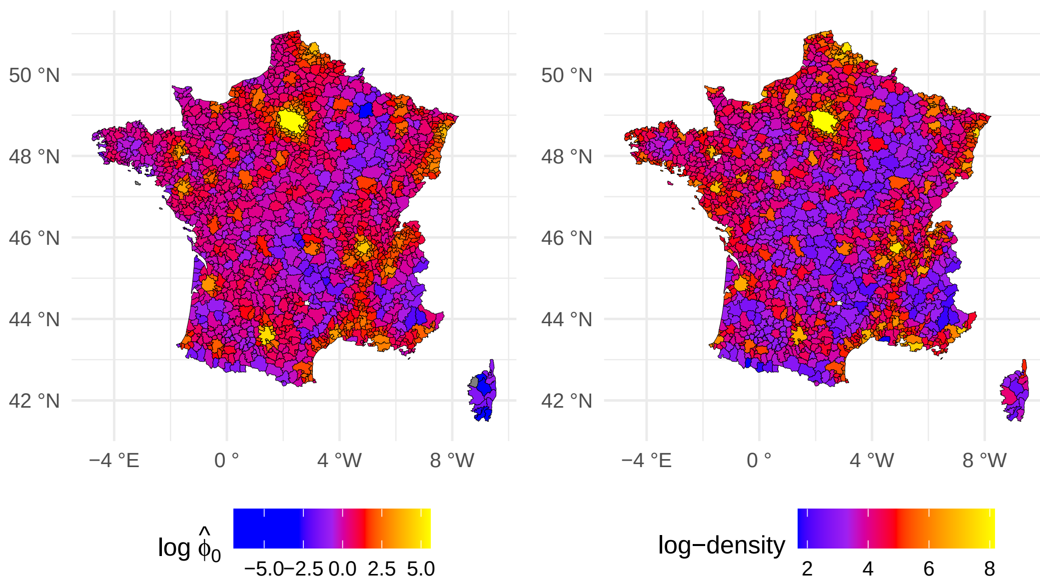
Figure 5: Parameter estimates (points) and confidence intervals (lines) from the semi-parametric Markov model. Parameters associated with the spatial covariates represent the effect on the cooperative banks relative to the effect of the lucrative banks. The 95% confidence intervals (CI) are calculated using the standard errors estimated based on the asymptotic results.



findings of Lord and Wright [1981] indicate that a contributing factor might be so-called *rational herding*. This is a phenomenon that occurs when individuals or firms believe that others possess more information than themselves. As a result, firms may mimic one another, even if the resulting outcome is not optimal. On the other hand, banks also compete with each other, which can explain negative interactions. Between these effects, the semi-parametric Markov model suggests that after taking spatial covariates and the common trends embedded in the non-parametric factor into account, competition dominates for banks of the same type. The positive cross-type interaction between lucrative and cooperative banks might indicate that they avoid competition by targeting different customers, while retaining some of the effects that lead to clustering.

Finally, we estimate ϕ_0 as described in Section 3.6. Figure 6 shows a plot of $\log \hat{\phi}_0$ across France with band width $\omega = 0.25$ chosen large enough to avoid zero estimates of ϕ_0. The estimates of $\phi_0(u)$ for locations u in France were averaged within administrative regions to ease visual comparison with the log population density, also shown in Figure 6. Visual inspection and a correlation of 0.73 between $\log \hat{\varphi}_0$ and log population density suggest that the non-parametric factor ϕ_0 is effective in picking up effects of population density common to both types of banks. This is reassuring for further applications where important spatial variables might be unobserved. Moreover, although ϕ_0 seems to be strongly associated with population density, it may capture effects of further unobserved spatial variables as well in the current bank example.

Figure 6: Estimate of $\log \phi_0$ obtained with band width $\omega = 0.25$ and an Epanechnikov kernel averaged over administrative units (left) along with log-population density (right).



7 Discussion

In this paper we significantly expand the toolbox of spatial statistics by introducing a semi-parametric approach for inhomogeneous multi-type Markov point processes. We do this by introducing a non-parametric factor adjusting for unobserved spatial variables that might otherwise confound parameter estimation. We moreover introduce a computationally efficient conditional pseudo likelihood method for parameter estimation. Asymptotic results and simulation studies demonstrate the good properties of the resulting parameter estimates. The application to French banks shows how the model can be used to decompose spatial patterns according to impacts of spatial covariates and within- and between-type interactions where the latter can be linked to economic theories of industry co-agglomeration.

Model assessment is an open problem for the semi-parametric Markov model. Commonly used point process residuals [Baddeley et al., 2005] and second order point process summary statistics like the K-function require a consistent estimator of the intensity function. Plugging in the kernel estimate of the non-parametric factor is not sufficient since this kernel estimate is not consistent. In Section 3.4 we discussed zero-mean conditional pseudo score residuals whose properties might be explored further following ideas in Coeurjolly and Lavancier [2013].

8 Acknowledgments

Ib T. Jensen and R. Waagepetersen were supported by grants VIL57389, Villum Fonden, and NNF23OC0084252, Novo Nordisk Foundation.

References

A. Artis, J.-F. Coeurjolly, F. Letué, and A. Choiruddin. Marked point processes intensity estimation using sparse group lasso method applied to locations of lucrative and cooperative banks in mainland France. *submitted*, 2025.

I. Ba and J.-F. Coeurjolly. Inference for low-and high-dimensional inhomogeneous Gibbs point processes. *Scandinavian Journal of Statistics*, 50(3):993–1021, 2023.

A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(5):617–666, 2005.

A. Baddeley, A. Jammalamadaka, and G. Nair. Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C*, 63(5): 673–694, 2014.

A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. CRC Press, 2015.

A. Brix and J. Møller. Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scandinavian Journal of Statistics*, 28(3):471–488, 2001.

A. E. Chang, S. Chaudhuri, and J. Jayaratne. Rational herding and the spatial clustering of bank branches: an empirical analysis. Research Paper 9724, Federal Reserve Bank of New York, 1997.

J.-F. Coeurjolly and F. Lavancier. Residuals and goodness-of-fit tests for stationary marked Gibbs point processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(2):247–276, 2013.

J.-F. Coeurjolly and E. Rubak. Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scandinavian Journal of Statistics*, 40(4):669–684, 2013.

J.-F. Coeurjolly, D. Dereudre, R. Drouilhet, and F. Lavancier. Takacs–Fiksel method for stationary marked Gibbs point processes. *Scandinavian Journal of Statistics*, 39(3):416–443, 2012.

O. Cronie and M. van Lieshout. Summary statistics for inhomogeneous marked point processes. *Annals of the Institute of Statistical Mathematics*, 68:905–928, 2016.

E. Dupont, S. N. Wood, and N. H. Augustin. Spatial+: a novel approach to spatial confounding. *Biometrics*, 78(4):1279–1290, 2022.

G. Duranton and H. G. Overman. Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106, 2005.

H.-O. Georgii. Canonical Gibbs measures, volume 760 of Lecture Notes in Mathematics, 1979.

H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.

M. Gómez-Antonio and S. Sweeney. Testing the role of intra-metropolitan local factors on knowledge-intensive industries' location choices. *The Annals of Regional Science*, 66:699–728, 2021.

M. Goulard, A. Särkkä, and P. Grabarnik. Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics*, pages 365–379, 1996.

P. Grabarnik and A. Särkkä. Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. *Ecological Modelling*, 220(9):1232–1240, 2009.

K. B. Hessellund, G. Xu, Y. Guan, and R. Waagepetersen. Semi-parametric multinomial regression for multivariate point pattern data. *Journal of American Statistical Association*, 117:1500–1515, 2022a.

K. B. Hessellund, G. Xu, Y. Guan, and R. Waagepetersen. Second-order semi-parametric inference for multivariate log Gaussian Cox processes. *Journal of the Royal Statistical Society, Series C*, 71:244–268, 2022b.

H. Högmander and A. Särkkä. Multitype spatial point patterns with hierarchical interactions. *Biometrics*, 55(4):1051–1058, 1999.

J. Hoover, Edgar M. Spatial price discrimination. *The Review of Economic Studies*, 4(3): 182–191, 06 1937. ISSN 0034-6527. doi: 10.2307/2967453.

A. Jalilian, Y. Guan, J. Mateu, and R. Waagepetersen. Multivariate product-shot-noise Cox models. *Biometrics*, 71:1022–1033, 2015.

S. Jansen. Gibbsian point processes. Lecture Notes for a course at Ludwig Maximilian University of Munich, 2018.

J. D. Lord and D. B. Wright. Competition and location strategy in branch banking: Spatial avoidance or clustering. *Urban Geography*, 2(3):189–200, 1981. doi: 10.2747/0272-3638.2.3.189.

J. Møller and R. Waagepetersen. *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton, 2003.

X. X. Nguyen and H. Zessin. Integral and differential characterizations of the Gibbs process. *Mathematische Nachrichten*, 88(1):105–115, 1979.

G. F. Nightingale, J. B. Illian, R. King, and P. Nightingale. Area interaction point processes for bivariate point patterns in a Bayesian context. *Journal of Environmental Statistics*, 9(2), 2019.

N. Picard, A. Bar-Hen, F. Mortier, and J. Chadoeuf. The multi-scale marked area-interaction point process: a model for the spatial pattern of trees. *Scandinavian journal of statistics*, 36 (1):23–41, 2009.

T. Rajala, D. J. Murrell, and S. C. Olhede. Detecting multivariate interactions in spatial point patterns with Gibbs models and variable selection. *Journal of the Royal Statistical Society: Series C*, 67:1237–1273, 2018.

S. Sweeney and M. Gómez-Antonio. Localization and industry clustering econometrics: An assessment of Gibbs models for spatial point processes. *Journal of Regional Science*, 56(2): 257–287, 2016.

R. P. Waagepetersen, Y. Guan, A. Jalilian, and J. Mateu. Analysis of multispecies point patterns by using multivariate log-Gaussian Cox processes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 65:77–96, 2015.