Definition and Estimation of Peer Effects through Latent Processes

Vincent Starck*
LMU München

November 17, 2023

Abstract

I propose a framework to analyze peer effects by modeling a latent sequence of decisions in continuous time. The method avoids regression on conditional expectations – and thus reflection type problems (Manski, 1993) – and 'outcome on means' regression by positing a direction of realized causality, which may not be observed but is subject to probabilistic quantification. I define a peer effect parameter meant to capture the causal peer influence of the first-movers. The parameter – and possibly covariates' coefficient – is shown to be consistently estimated by maximum of likelihood methods and lends itself to standard inference.
Keywords: Peer effects, Continuous time, Networks, Spatial econometrics.

1 Introduction

Peer effects have traditionally been investigated via estimation of models of the form $\mathbb{E}[y|x,z] = f(\mathbb{E}[y|x], z)$, where x are characteristics used to define peer group and z are covariates directly affecting the outcome. The practice has been criticized both on theoretical and empirical grounds (Manski, 1993; Angrist, 2014). Indeed, this type of model – even in nonlinear forms – is

*I gratefully acknowledge financial support from the European Research Council (Starting Grant No. 852332)

tautological as it suffers from the reflection problem (Manski, 1993), which threatens its well-definedness and leads to identification issues.

Spatial Autoregressive Models (SAR; Lee (2004)) models bypass some of these concerns. Reflection issues disappear when the conditional expectation is replaced by a weighted average of outcomes – typically not meant to estimate the conditional expectation, but rather to characterize a system's equilibrium relationship. Relatedly, the use of non-overlapping peer groups may alleviate reflection concerns and allow peer effect identification and estimation (Bramoullé, Djebbari and Fortin, 2009; De Giorgi, Pellizzari and Redaelli, 2010).

Nevertheless, these models are not immune to all empirical concerns brought in Angrist (2014) and may generate spurious peer effects. In addition, such models most naturally represent simultaneous decisions or equilibrium results; they are less suited for sequential, irreversible decisions. They do not provide a framework to discuss causality questions, diffusion, or identification of first movers and (probability of the) sequence of arrivals. For many applications, it is likely that peer effects hinge on the sequential nature of decisions and that first movers exert social influence, rather than on the generic influence of a peer group that leads to simultaneity issues.

I propose a new framework to analyze peer effects and model the latent sequence of decisions in continuous time. I define a causal peer effect parameter that can be consistently estimated by maximum of likelihood methods. Decisions are explicitely made asynchronously, though their timing may not be observed. The paper applies to both spatial and network applications, and as such has also connections to diffusion problems in the network literature (*e.g.*, He and Song (2018)).

2 Model and Likelihood

2.1 Motivation and identification in a simple case

Consider two individuals ($i = 1, 2$) who can make an irreversible decision (*e.g.*, getting a vaccine) at any point during a time frame represented by the interval $[0; S]$, where 0 is the start of eligibility and S is the time to observation, and indicate 'activation' by $y_i = 1$. Let the distribution of the time before decision be exponential with rates λ_1, λ_2. The decision of an individual may affect the probability that the other opts in, modifying the rate of time to activation to $\lambda_1^{+_2}$ or $\lambda_2^{+_1}$ from that point on. Here, the subscript on the + indicates whose move triggered a change in rate. Since there are only two people in this introductory example, the subscript can and will be dropped in the remained of the subsection. The resulting change in the distribution of outcomes can be viewed as a causal peer effect.

The probabilities for the four possible outcomes follows from algebra:

$$\begin{aligned} p_{00} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = y_2 = 0] = e^{-(\lambda_1+\lambda_2)S} \\ p_{10} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = 1, y_2 = 0] = \frac{\lambda_1 e^{-\lambda_2^+ S}}{(\lambda_1 + \lambda_2 - \lambda_2^+)}(1 - e^{-(\lambda_1+\lambda_2-\lambda_2^+)S}) \\ p_{01} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = 0, y_2 = 1] = \frac{\lambda_2 e^{-\lambda_1^+ S}}{(\lambda_1 + \lambda_2 - \lambda_1^+)}(1 - e^{-(\lambda_1+\lambda_2-\lambda_1^+)S}) \\ p_{11} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = y_2 = 1] = 1 - p_{00} - p_{10} - p_{01} \end{aligned}$$

If $\lambda_1 = \lambda_2$ (or more generally if $\lambda_i = f(x_i; \theta)$ for a function f, observed covariates x_i, and finite dimensional identifiable parameter θ; $\lambda_1 = \lambda_2$ corresponding to the intercept case with $x_1 = x_2$), the family of rates can be obtained given the probabilities. In this case, $\lambda_1 = \lambda_2 = -\frac{\ln(p_{00})}{2S}$ and λ_2^+ can be recovered from $\frac{-\lambda_1 S(e^{-\lambda_2^+ S} - p_{00})}{\ln(p_{00}) + \lambda_2^+ S} = p_{10}$, where the left-hand side is strictly decreasing[1]. λ_1^+ can be obtained analogously.

Although it is not possible to identify the identity of the first mover – the individual who may have exerted a peer effect on the other individual – when $y_1 = y_2 = 1$, it is possible to (i) estimate the peer effect strengths and (ii) determine the probability of an individual moving first whenever the λ's are identified.

Note that in the absence of peer effects – $\lambda_i^+ = \lambda_i$ for $i = 1$ and $i = 2$ – the probabilities reduce to a standard exponential race with independent. The change in the exponential rates is thus a measure of dependence and social interaction, which will later be captured by a parameter $\delta \in \mathbb{R}$.

Moving beyond this simple example, one might attempt to extend the analysis to arbitrary group sizes and to obtain an estimator of the family

[1] Its derivative reads $\lambda_1 S \frac{-p_{00} + e^{-\lambda_2^+ S} + e^{-\lambda_2^+ S}\lambda_2^+ S + \ln(p_{00}) e^{-\lambda_2^+ S}}{(\ln(p_{00}) + \lambda_2^+ S)^2}$ and the numerator is positive.

of rates. In the following subsections, I formalize the underlying stochastic process and show that, given data on $y_i, i = 1, \cdots, n$ (and possibly covariates x_i) the probability of the sample $P(y_1, \ldots, y_n | x_1, \ldots, x_n, \{\lambda_i^{+k_1,\ldots,+k_m})\})$ has a closed form solution. This suggests a maximum of likelihood estimator whose asymptotic properties are investigated in Section 3. Consistency of the family of λ's follows upon necessary restrictions for identification.

2.2 Framework and peer effects

Consider a sample of (binary) variable y_i, covariates x_i, and an adjacency matrix W ($W_{ij} = 1$ if and only if i and j are 'neighbors' or 'friends') that determines individuals' peer group.

The goal is to model and estimate peer effects, *i.e.* how the decision of peers alters the decision probability of an individual. I consider a 'default' situation with $y_i = 0$ (typically, 'doing nothing', the inertia decision) and the decision to 'activate' (switch to 1) is irreversible (similarly, that only first time event matters). Observing someone activating modifies the likelihood that others do so. For instance, y might represent vaccination status, a decision to exert effort, technology adoption, a decision to attend college or an event, etc.

Individual activation is modelled with the following continuous time stochastic process:
Let T_i^1 be an Exponential distribution with rate λ_i for all i. When a first individual, say k, activates ($T_k^1 = \min(T_i^1)$), the rates for activation evolve to λ_i^{+k} for all of k's neighbors. The process then repeats with the altered rates, and so on. The time to activation is thus $T_i = \sum_{t=1}^{\tau_i} T_{k_t}^t$, where k_t is the index of the 'winner' at stage t, and we observe $y_i = \mathbb{1}_{T_i \leq S}$ at time S. S is a timeline during which each individual had the opportunity to activate. S can be the time up to a deadline, a school year, or the time elapsed from availability/eligibility to observation by the researcher.

I focus on Exponential distributions for two reasons. First, exponential waiting times arise automatically under the assumption of a constant probability per unit time, a natural point of departure. Second, this process offers computational niceties because of the properties of the exponential. When k activates, the clock should be restarted for all of his or her friends, but not for anyone else. However, due to the memorylessness property, the clock may be re-started for everyone, with the rate of non-friend left unchanged. This can make both theoretical analysis and simulations simpler.

The model is viewed as causal, with elements of a potential outcome framework, in that some individuals will be subject a to peer effect - "treatment" - and some will not. There are two main departures from a classical treatment effect framework, which stem from the nature of the problem: one's outcome induces another's treatment and this direction is often unobserved because information about the order of activations is typically limited, and there are multiple possible treatments that correspond to all possible dynamic selection of peers.

Peer effect are defined at the individual level through the changes in λ_i induced by neighboring activations. The model posits the existence (ex post) of a direction of causality, though the realized direction may not be observed by the researcher. This puts peer effects on stronger theoretical grounds as it defines a structural peer effect parameter and avoids ill-defined peer groups, which tend to be subject to the reflection problem (Manski, 1993).

2.3 Likelihood

Assume without loss that the activated observations are $1, \ldots, G$. Since the sequence of arrivals is unknown, the probability of the sample corresponds to

$$\begin{aligned}
&\mathbb{P}[y_1, \ldots, y_G \leq S; y_{G+1}, \ldots, y_N > S] \\
&= \sum_{p \in \mathcal{P}} \mathbb{P}[y_1, \ldots, y_G \leq S; y_{G+1}, \cdots, y_N > S; T_{p_1} < \ldots < T_{p_G}]
\end{aligned}$$

where the sum is over all permutations (with generic element $p = (p_1, \ldots, p_G)$) of the G first arrivals.

The representative term in the permutation, (displayed for convenience with individual 1 arriving first, followed by individual 2, etc.) is then computed as

$$\int_0^S \int_{t_1^1}^\infty \cdots \int_{t_1^1}^\infty \prod_{i=1}^N \lambda_i e^{-\lambda_i t_i^1} \quad \int_0^{S-t_1^1} \cdots \int_{t_2^2}^\infty \prod_{i=2}^N \lambda_1^{+1} e^{-\lambda_i^{+1} t_i^2} \cdots$$
$$\int_0^{S-\sum_{g=1}^{G-1} t_g^g} \cdots \int_{t_G^G}^\infty \prod_{i=G}^N \lambda_i^{+1\cdots+G-1} e^{-\lambda_i^{+1\cdots+G-1} t_i^G}$$
$$\int_{S-\sum_{g=1}^G t_g^g}^\infty \cdots \int_{S-\sum_{g=1}^G t_g^g}^\infty \prod_{i=G+1}^N \lambda_i^{+1\cdots+G} e^{-\lambda_i^{+1\cdots+G} t_i^{G+1}}$$
$$dt_1^1 \cdots dt_N^1 dt_2^2 \cdots dt_N^2 \cdots dt_G^G \cdots dt_N^G \cdots dt_1^{G+1} \cdots dt_N^{G+1}$$

which, after some algebra, reduces to

$$e^{-\sum_{i=G+1}^N \lambda_i^{+1,\cdots,+G} S} \left(\prod_{i=1}^G \lambda_i^{+1,\cdots,+i-1} \right) I_{\{c_i, i=1,\cdots,G\}} \tag{1}$$

where $I_{\{h_i,i=1,\cdots,H\}} \stackrel{\text{def}}{=} \frac{1}{\prod_{g=1}^G c_g} + (-1)^G \sum_{g=1}^G \frac{1}{\prod_{h\neq g}(c_g - c_h)} \frac{e^{-c_g S}}{c_g}$ and $c_k \stackrel{\text{def}}{=} \sum_{i=k}^N \lambda_i^{+1,\cdots,+k-1} - \sum_{i=G+1}^N \lambda_i^{+1,\cdots,+G}$.

Introducing $c_{G+1} = 0$, the term simplifies to $\sum_{g=1}^{G+1} \frac{e^{-c_g S}}{\prod_{h\neq g} c_h - c_g}$ and then the likelihood can be further reduced to

$$\prod_{i=1}^G \lambda_i^{+1,\cdots,+i-1} \sum_{g=1}^{G+1} \frac{e^{-\ddot{c}_g S}}{\prod_{h\neq g} \ddot{c}_h - \ddot{c}_g} \tag{2}$$

where $\ddot{c}_g \stackrel{\text{def}}{=} c_g + \sum_{i=G+1}^N \lambda_i^{+1,\cdots,+G} = \sum_{i=k}^N \lambda_i^{+1,\cdots,+k-1} \geq 0$ (with equality only if $g = G+1$ and $N = G$).

In this form, the model is general and handles considerable heterogeneity, albeit with too many parameters to be identified as such. One may restrict the heterogeneity in various ways, depending on empirical concerns and the goal of the analysis (for instance, whether heterogeneity in initial rates, in peer effects, in the identity of first-mover, etc. are more relevant).

A leading case of interest naturally arises from the desire to define a simple peer effect parameter while accounting for individual heterogeneity through observed covariates x_i. A way to make the model parsimonious and to incorporate covariates is to specify $\lambda_i^{+k_1\cdots+k_L} = e^{x_i'\beta + N_i(k_1,\cdots,k_L)\delta}$ (or any sensible link function in place of the exponential), where $N_i(k_1, \dots, k_L)$ is the number of neighbors for i among $k_1, \dots, k_L$. Various related models can be specified to deal with the specifics of peer effects in a given application. For instance, if there are different 'types' of links with varying peer effect strengths, one can replace by $N_i\delta$ by $\sum_T \delta_j N_i^T$. Such a case occurs when peer effects are expected to be, say, stronger with friends than family members and the modeler would expect N_i^{friend} to be associated to a higher coefficient, δ_{friend}, than that of N_i^{family}, δ_{family}.

Interaction terms of the form $(N_i x_{ik})$ may also be added to account for the effect of a covariate, x_{ik}, on the strength of peer effects and N_i may be filtered through a nonlinear function to reflect, *e.g.*, decreasing peer strength in the activated friend count. Finally, the functional form may also be modified, for instance by substituting the exponential by the positive part or using the average number of activated neighbors instead of the total.

Although summing over all permutations can lead to an impractical computational burden, the cost is significantly smaller in practice. First, the likelihood factorizes if W can be divided into non-connected subgraphs, which occurs frequently with separations into distinct classrooms, villages, etc. Second, additional information - such as partial knowledge of the order of activations - substantially reduces the complexity of permutations.

Third, approximations can further cut the number of permutations. In particular, maximizing the log-likelihood amounts to maximizing
$\ln\left(\frac{1}{G!}\sum_{p\in\mathcal{P}}\left(\prod_{i=1}^{G}\lambda_{p_i}^{+p_1,\cdots,+p_{i-1}}\right)\sum_{g=1}^{G+1}\frac{e^{-\ddot{c}_{p_g}S}}{\prod_{h\neq g}\ddot{c}_{p_h}-\ddot{c}_{p_g}}\right)$
where the average over all permutations can be estimated by a random sample of permutations by the law of large numbers.

Note finally that when individuals are broadly similar or when the network exhibits symmetry, it may be possible to group terms or simplify intermediate computations.

If sampling is skewed/tilted, we have for a sample $\mathcal{P}_s$ that
$\frac{1}{|\mathcal{P}_s|}\sum_{p\in\mathcal{P}_s}\partial_\theta\mathbb{P}[p]/\mathbb{P}[p] \to^p \sum_{p\in\mathcal{P}}\frac{\partial_\theta\mathbb{P}[p]}{\sum_{p\in\mathcal{P}}\mathbb{P}[p]}$
$\frac{1}{|\mathcal{P}_s|}\sum_{p\in\mathcal{P}_s}1/\mathbb{P}[p] \to^p \frac{G!}{\sum_{p\in\mathcal{P}}\mathbb{P}[p]}$ using that $\frac{\mathbb{P}[p]}{\sum_{p\in\mathcal{P}}\mathbb{P}[p]}$ is the probability of sampling permutation p under tilted sampling.

2.4 Particular case

Suppose there is no heterogeneity nor peer effects: $\lambda_i^{\{+\}} = \lambda$ for all i and collection of $+$.

Then $\ddot{c}_g = \lambda(N-(g-1))$ and the likelihood of ordering p becomes

$$\begin{aligned}\prod_{i=1}^{G} \lambda_i^{+_1,\cdots,+_{i-1}} \sum_{g=1}^{G+1} \frac{e^{-\ddot{c}_g S}}{\prod_{h\neq g} \ddot{c}_h - \ddot{c}_g} &= \lambda^G \sum_{g=1}^{G+1} \frac{e^{-\lambda(N-(g-1))S}}{\prod_{h\neq g} \lambda(N-(h-1)) - \lambda(N-(g-1))} \\ &= \sum_{g=1}^{G+1} \frac{e^{\lambda(g-1)S}}{\prod_{h\neq g} h-g} \\ &= \sum_{g=1}^{G+1} \frac{e^{\lambda(g-1)S}}{(g-1)!\,(G+1-g)!}(-1)^{G+1-g} \\ &= \frac{e^{-\lambda N S}(e^{\lambda S}-1)^G}{G!} \\ &= \frac{e^{-\lambda(N-G)S}(1-e^{-\lambda S})^G}{G!}\end{aligned} \tag{3}$$

Summing over all permutations yields $e^{-\lambda(N-G)S}(1-e^{-\lambda S})^G$, which is the likelihood from iid exponential random variables.

3 Asymptotic Theory

I focus on the simple model suggested in the previous section, *i.e.* the family of rates are determined by $\lambda_i^{+_{k_1}\cdots+_{k_m}} = e^{x_i^{'}\beta+\sum_{j=1}^{m} W(i,k_j)\delta}$.

More complex models can be handled similarly, the main threat to consistent estimation coming from (lack of) identification when the number of parameters is too large. Identification is primarily a matter of imposing structure on the family of lambda's as to limit the number of parameters. It is easily achieved in parsimonious models as the introductory example suggests.

The asymptotic behavior of the estimator can be analyzed conditionally on the network W and covariates x_i. Consider a distribution of 'blocks' or 'classes' of size n_b, which will be assumed bounded by $\overline{n_b}$. We observe the blocks $b = 1, \cdots, B$ with size n_b so that the sample size is $N \stackrel{\text{def}}{=} \sum_{b=1}^{B} n_b$. Any individual i in the sample belongs to one block, denoted by $b(i)$, and a block contains a set of n_b individuals, listed through $i(b)$. I will make use the notation $ij(b)$ to denote all the pairs in block b, *i.e.* $ij(b) \stackrel{\text{def}}{=} \{(i,j) : i \neq j, b(i) = b(j) = b\}$. To ease notation, conditioning on covariates and network structure is omitted.

The log-likelihood factorizes as
$l \stackrel{\text{def}}{=} \sum_{y \in \Delta_N} \mathbb{1}_{Y=y} \ln(\mathbb{P}[Y = y]) = \sum_{b=1}^{B} \sum_{y_{i(b)} \in \Delta_{n_b}} \ln(\mathbb{P}[Y_{i(b)} = y_{i(b)}])$, where $\Delta_N \stackrel{\text{def}}{=} \{0,1\}^N$, and asymptotics will rely on B (and thus N) going to infinity.

The log-likelihood can be derived for each block from the formula established in the previous section. For illustration, consider the first block and let $\mathcal{P}$ contain the permutations of $1, \ldots, G$, whose terms I represent by $p \stackrel{\text{def}}{=} \{p_1, \ldots, p_G\}$. Then, the probability in the first block, $\ln(\mathbb{P}[Y_{i(1)} = y_{i(1)}])$, is given by

$$\ln\left(\sum_{p \in \mathcal{P}} e^{\sum_{i=1}^{G_1} x'_{p_i}\beta + N_{p_i}(p_1, \cdots, p_{i-1})\delta} I_{\{c_k^p, k=1,\cdots,G_1\}}\right) - \sum_{i=G_1+1}^{n_1} e^{x'_i\beta + N_i(1,\cdots,G_1)} S \quad (4)$$

where $c_k^p \stackrel{\text{def}}{=} \sum_{i=k}^{n_1} e^{x'_{p_i}\beta + N_{p_i}(p_1,\ldots,p_{k-1})\delta} - \sum_{i=G_1+1}^{n_1} e^{x'_i\beta + N_i(1,\ldots,G_1)\delta}$.

Its derivatives, from which the score can be easily constructed, are given by

$$\begin{aligned}\frac{\partial \ln(\mathbb{P}[Y_{i(1)} = y_{i(1)}])}{\partial \beta} =& \frac{\sum_{p \in \mathcal{P}} e^{\sum_{i=1}^{G_1} x'_{p_i}\beta + N_{p_i}(p_1,\ldots,p_{i-1})\delta} \partial_\beta I_{\{c_k^p, k=1,\ldots,G_1\}}}{\sum_{p \in \mathcal{P}} e^{\sum_{i=1}^{G} x'_{p_i}\beta + N_{p_i}(p_1,\ldots,p_{i-1})\delta} I_{\{c_k^p, k=1,\ldots,G_1\}}} \\ &+ \sum_{i=1}^{G_1} x_{p_i} - \sum_{i=G_1+1}^{n_1} e^{x'_i\beta S + \sum_{j=1}^{G} W(i,j)\delta} S x_i \end{aligned} \quad (5)$$

and

$$\frac{\partial \ln(\mathbb{P}[Y_{i(1)}=y_{i(1)}])}{\partial \delta} = \frac{\sum_{p\in\mathcal{P}} e^{\sum_{i=1}^{G_1} x'_{p_i}\beta + N_{p_i}(p_1,\ldots,p_{i-1})\delta} I_{\{c_k^p, k=1,\ldots,G_1\}} \sum_{i=1}^{G} N_{p_i}(p_1,\ldots,p_{i-1})}{\sum_{p\in\mathcal{P}} e^{\sum_{i=1}^{G} x'_{p_i}\beta + N_{p_i}(p_1,\ldots,p_{i-1})\delta} I_{\{c_k^p, k=1,\ldots,G_1\}}} + \frac{\sum_{p\in\mathcal{P}} e^{\sum_{i=1}^{G_1} x'_{p_i}\beta + N_{p_i}(p_1,\ldots,p_{i-1})\delta} \partial_\delta I_{\{c_k^p, k=1,\ldots,G_1\}}}{\sum_{p\in\mathcal{P}} e^{\sum_{i=1}^{G_1} x'_{p_i}\beta + N_{p_i}(p_1,\ldots,p_{i-1})\delta} I_{\{c_k^p, k=1,\ldots,G_1\}}} - \sum_{i=G_1+1}^{n_1} e^{x'_i\beta S + \sum_{j=1}^{G_1} W(i,j)\delta} S N_i(1,\ldots,G_1) \tag{6}$$

In these expressions, the derivative of the I-term can be computed as $\partial I_{\{h_i,i=1,\ldots,H\}} = \frac{-\sum_{g=1}^{G}\frac{\partial c_g}{c_g}}{\prod_{g=1}^{G} c_g} - (-1)^G \sum_{g=1}^{G} \frac{1}{\prod_{h\neq g}(c_g-c_h)} \frac{e^{-c_g S}}{c_g}\left[\partial c_g S + \frac{\partial c_g}{c_g} + \sum_{h\neq g}\frac{\partial c_g - \partial c_h}{c_g - c_h}\right]$

The maximum likelihood estimator is consistent and asymptotically normal under regularity conditions. Specifically, one can verify Newey and McFadden (1994)'s sufficient conditions for extremum estimators, assuming $(\beta,\delta)\in\mathcal{B}\times\mathcal{S}$, a compact set.

First, (β,δ) is identified, *e.g.* β is identified from $\mathbb{P}[Y_b=0|W,x_i]$ (as function of the observed x_i) and then δ is identified from $\mathbb{P}[Y_b=e_1]$. The space $(\beta,\delta)\in\mathcal{B}\times\mathcal{S}$ is compact by assumption and the limit objective function is continuous.
Finally, $\frac{1}{B}\sum_{b=1}^{B}\sum_{y_b\in\{0,1\}^{n_b}}\mathbb{P}[Y_b=y_b]\ln(\mathbb{P}[Y_b=y_b]) \leq \frac{1}{B}\sum_{b=1}^{B}\ln(n_b) \leq \ln(\overline{n_b})$, bounding the entropy by that of the uniform distribution and then using the bound on group size. As a result, uniform laws of large numbers apply and the sample objective function converges uniformly in probability.

This implies the following theorem:

Theorem 3.1 (Consistency). *Suppose the data generating process is given by the stochastic process described in Section 2.2 with rates given by* $\lambda_i^{+k_1\cdots+k_m} = e^{x'_i\beta+\sum_{j=1}^{m}W(i,k_j)\delta}$ *and* $(\beta,\delta)\in\mathcal{B}\times\mathcal{S}$*, a compact set. Then, the proposed maximum likelihood estimator is consistent with* $(\hat\beta,\hat\delta)\to^p(\beta,\delta)$ *as* $B\to\infty$*.*

Moreover,

Theorem 3.2 (Asymptotic Normality). *Under the assumptions of Theorem 3.1, the maximum likelihood estimator is asymptotically normal as* $B\to\infty$

with

$$\sqrt{B}\left(\begin{pmatrix}\hat{\beta}\\ \hat{\delta}\end{pmatrix}-\begin{pmatrix}\beta\\ \delta\end{pmatrix}\right)\to^{d}\mathcal{N}\left(\begin{pmatrix}0\\ 0\end{pmatrix};\mathbb{E}\left[\begin{pmatrix}\frac{\partial l_b}{\partial\beta}\frac{\partial l_b}{\partial\beta'} & \frac{\partial l_b}{\partial\beta}\frac{\partial l_b}{\partial\delta}\\ \frac{\partial l_b}{\partial\delta}\frac{\partial l_b}{\partial\beta'} & \frac{\partial l_1}{\partial\delta}\frac{\partial l_b}{\partial\delta}\end{pmatrix}\right]\right) \tag{7}$$

As a result, inference can be carried out about both the effects of covariates and peer effects. Another consequence of the theorems is the consistent estimation of the exponential rates. It may be of interest to identify the 'leaders' and the 'followers' in any sub-block of connected activated people, *i.e.* individuals who were activated first and those who responded later. While the identity of the first activated individual is not identifiable, consistent estimation of the λ's allows one to recover the probability that i was the first to activate as well as (probabilistic) relative rankings of order of activation.

Remark: Spatial dependence Spatial dependence is determined by the value of δ in addition to the network structure. In particular, observations are independent in the absence of peer effects ($\delta = 0$). This suggests that blocks with weak ties (relative to δ for given structure and initial rates) could sometimes be treated as approximately independent. Although a detailed analysis of these types of argument is beyond the scope of the paper, it may be a useful relaxation of the block structure for some applications.

Remark: Continuous outcomes Continuous random variables can be handled by adding an optimization over unobserved binary activation status. For instance, let Y be the binary activation status and y be the final (continuous) outcome given by $y = Z + \mathbb{1}_{Y\leq S}\mu$, where for instance $Z|X \sim \mathcal{N}(X'\gamma, \sigma^2)$ For any candidate vector $Y \in \{0,1\}^N$, this leads to a sum over all possible activation patterns and the resulting likelihood can be optimized over to obtain a maximum of likelihood estimator.

Remark: Forecasts One can use the estimated β and δ and determine the most likely next activated person, average time to activation, etc. This can allow some tests on the model if there is a subsequent measurement after S.

Remark: Counterfactual analysis It is often of interest to assess the impact of policies, for instance targeting influential individuals in a network. The peer effect parameter allows a direct evaluation of the direct effect of imposing $y_i = 1$ at a given point. It is also easy to simulate the evolution of a network with and without imposing $y_i = 1$ for a group of selected individuals at time 0 and to assess the change in probabilities that $y_k = 1$ for $k \neq i$.

Remark: Exogenous peer effects One may add a "WX term" to the argument of the exponential to account for so-called exogenous peer effects, which affect one's outcome through the characteristics of peers.

4 Simulations

I now assess the performance of the estimator in simulations. I first consider 'correctly specified' models in which all relevant covariates are available. In the second subsection, I investigate the robustness of the estimator to omitted variables, measurement errors, and group heterogeneities.

4.1 Simulations with block and homophilic network formation

I simulate the stochastic process described in Section 2.2 with an underlying network structure of either 'classrooms' or homophilic matching type, both of which are common in empirical studies.

First, I construct a network with 1000 individuals and a 'block' structure $((W = I \otimes \iota\iota'))$ with groups of size 5, 10, and 20. In the previous section's notation, this means $N = 1000$, $n_b = 5\ \forall b$, and $B = 1000/n_b$ and individuals are connected to all individuals within the same block. I set the family of rates to obey $\lambda_i^{+k_1 \cdots + k_m} = e^{x_i'\beta + \frac{\sum_{j=1}^{m} W(i,k_j)}{\sum_{j=1}^{n} W(i,j)}\delta}$ with two covariates (uniformly on $[-1;1]$ and (standard) normally distributed, respectively), various levels of peer effect strength δ (-0.5, 0, and 0.5), and $\beta = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$.

There is no information about the order of activations so all permutations *a priori* matter. I make use of the random sampling over permutations mentioned in Section 2 to alleviate the computational burden whenever the number of activated people in a group exceeds 8.

Estimates are compared to SAR estimates from a simple (endogenous) regression on x_i and $W_i y$ and to the SAR maximum likelihood estimator[2]

[2]The weighting matrix is row-normalized since the process suggests peer effects depend

These are frequent estimates of peer effects which can be hoped to capture whether there is a peer influence, but cannot be expected to be consistent given incorrect specification; they would not capture the true nature of the peer effects, but could detect their sign and presence. Coefficients on covariates, which are also reported, have similarly no direct counterparts in a linear model and are not expected to be consistently estimated by regression or SAR MLE.

The results are reported in Table 1. The maximum of likelihood estimator described in the previous section performs well in all instances and exhibits very low bias. A standard regression is usually able to pick up the correct sign of peer effects in this specific setup, but cannot recover the structural coefficient. The SAR MLE broadly follows the same lines.

on the average number of activated friends. Notice, however, that the model using sums has an equivalent representation using averages when groups have the same size: it amounts to scaling δ by group size.

Table 1: Simulations with 'classrooms' network structure

			Bias			Standard deviation			RMSE		
n_b	δ		Reg	SAR	Exp	Reg	SAR	Exp	Reg	SAR	Exp
5	-0.5	$\hat{\delta}$	0.39	0.43	0.03	0.04	0.03	0.12	0.39	0.43	0.13
		$\hat{\beta}_1$	-0.70	-0.37	0.00	0.02	0.03	0.09	0.70	0.38	0.09
		$\hat{\beta}_2$	-0.36	-0.19	0.00	0.01	0.02	0.05	0.36	0.20	0.05
	0	$\hat{\delta}$	-0.02	-0.01	0.00	0.08	0.04	0.10	0.08	0.04	0.10
		$\hat{\beta}_1$	-0.70	-0.37	0.00	0.02	0.03	0.09	0.70	0.37	0.09
		$\hat{\beta}_2$	-0.36	-0.20	0.00	0.01	0.02	0.05	0.36	0.20	0.05
	0.5	$\hat{\delta}$	-0.35	-0.41	-0.01	0.07	0.04	0.09	0.36	0.42	0.09
		$\hat{\beta}_1$	-0.71	-0.37	0.00	0.02	0.03	0.09	0.71	0.37	0.09
		$\hat{\beta}_2$	-0.36	-0.21	0.00	0.01	0.02	0.05	0.36	0.21	0.05
10	-0.5	$\hat{\delta}$	0.32	0.41	0.00	0.14	0.07	0.13	0.35	0.41	0.13
		$\hat{\beta}_1$	-0.69	-0.37	0.00	0.02	0.04	0.09	0.69	0.38	0.09
		$\hat{\beta}_2$	-0.35	-0.19	0.00	0.01	0.02	0.05	0.35	0.20	0.05
	0	$\hat{\delta}$	-0.01	-0.01	0.00	0.12	0.07	0.11	0.12	0.07	0.11
		$\hat{\beta}_1$	-0.70	-0.37	0.00	0.02	0.04	0.09	0.70	0.37	0.09
		$\hat{\beta}_2$	-0.36	-0.20	0.00	0.01	0.02	0.05	0.36	0.20	0.05
	0.5	$\hat{\delta}$	-0.37	-0.42	0.00	0.10	0.06	0.10	0.38	0.43	0.10
		$\hat{\beta}_1$	-0.71	-0.37	0.01	0.02	0.04	0.08	0.71	0.37	0.08
		$\hat{\beta}_2$	-0.36	-0.21	0.01	0.01	0.02	0.06	0.36	0.21	0.06
20	-0.5	$\hat{\delta}$	0.30	0.40	0.00	0.21	0.10	0.13	0.36	0.41	0.13
		$\hat{\beta}_1$	-0.70	-0.37	0.01	0.02	0.06	0.08	0.70	0.37	0.09
		$\hat{\beta}_2$	-0.36	-0.20	0.01	0.01	0.02	0.05	0.36	0.20	0.05
	0	$\hat{\delta}$	-0.06	-0.03	-0.01	0.18	0.10	0.11	0.19	0.10	0.11
		$\hat{\beta}_1$	-0.70	-0.36	0.00	0.02	0.06	0.08	0.70	0.36	0.08
		$\hat{\beta}_2$	-0.36	-0.20	0.00	0.01	0.02	0.05	0.36	0.20	0.05
	0.5	$\hat{\delta}$	-0.38	-0.42	-0.02	0.15	0.09	0.10	0.40	0.43	0.11
		$\hat{\beta}_1$	-0.71	-0.36	0.07	0.02	0.06	0.18	0.71	0.37	0.19
		$\hat{\beta}_2$	-0.36	-0.21	0.03	0.01	0.02	0.08	0.36	0.21	0.09

I now consider another network structure, in which individuals within groups decide whether to make a connection based on their characteristics. Specifically, I consider a homophilic link formation process in which individual match according to their similarities: $W_{ij} = 1$ iff $\frac{\|X_{1i}-X_{1j}\|+\|X_{2i}-X_{2j}\|}{2} < \eta_{ij})$, where the collection of $\eta_{ij} = \eta_{ji}$ forms an array of independent uniform random variables. Group sizes are $5, 20$, or a larger group of 100 and the sample size is gain $N = 1000$.

The results are displayed in the next table.

Table 2: Homophilic

			Bias			Standard deviation			RMSE		
n_b	δ		Reg	SAR	Exp	Reg	SAR	Exp	Reg	SAR	Exp
5	-0.5	$\hat{\delta}$	0.39	0.42	0.03	0.04	0.03	0.13	0.39	0.43	0.13
		$\hat{\beta}_1$	−0.70	−0.38	−0.01	0.02	0.02	0.09	0.70	0.38	0.09
		$\hat{\beta}_2$	−0.36	−0.20	0.00	0.01	0.02	0.05	0.36	0.20	0.05
	0	$\hat{\delta}$	0.00	0.00	−0.02	0.04	0.03	0.11	0.04	0.03	0.11
		$\hat{\beta}_1$	−0.69	−0.38	0.02	0.02	0.02	0.09	0.69	0.38	0.09
		$\hat{\beta}_2$	−0.36	−0.19	0.01	0.01	0.02	0.05	0.36	0.20	0.05
	0.5	$\hat{\delta}$	−0.38	−0.41	−0.03	0.03	0.02	0.10	0.38	0.41	0.11
		$\hat{\beta}_1$	−0.71	−0.38	0.02	0.02	0.02	0.08	0.71	0.38	0.08
		$\hat{\beta}_2$	−0.36	−0.20	0.00	0.01	0.02	0.05	0.36	0.21	0.05
10	-0.5	$\hat{\delta}$	0.41	0.44	0.03	0.05	0.03	0.12	0.41	0.44	0.12
		$\hat{\beta}_1$	−0.69	−0.39	0.00	0.02	0.02	0.09	0.70	0.39	0.09
		$\hat{\beta}_2$	−0.35	−0.20	0.00	0.01	0.02	0.05	0.35	0.20	0.05
	0	$\hat{\delta}$	0.00	0.00	−0.02	0.04	0.03	0.10	0.04	0.03	0.10
		$\hat{\beta}_1$	−0.70	−0.38	0.00	0.02	0.02	0.08	0.70	0.38	0.08
		$\hat{\beta}_2$	−0.36	−0.20	0.00	0.01	0.02	0.05	0.36	0.20	0.05
	0.5	$\hat{\delta}$	−0.39	−0.42	−0.05	0.05	0.03	0.10	0.39	0.43	0.11
		$\hat{\beta}_1$	−0.71	−0.37	0.01	0.02	0.02	0.08	0.71	0.37	0.08
		$\hat{\beta}_2$	−0.36	−0.21	0.00	0.01	0.02	0.05	0.36	0.21	0.05

Although the performance of OLS or SAR-MLE in terms of bias and RMSE in the absence of peer effects ($\delta = 0$) suggests that these estimators may successfully detect the absence of social influence, notice that estimates are generally attenuated compared to the structural parameter and that decisions will eventually be based on tests or confidence intervals. As a result, the coverage performance of the confidence intervals may be a more relevant benchmark and will be analyzed in the next subsection.

Interestingly, OLS and SAR-MLE feature attenuation bias with respect to the structural parameter. As a result, they may seem to perform better in terms of RMSE in the absence of peer effect. In practice, however, what matters is the test for the presence of peer effect or, equivalently, the resulting confidence intervals. In the next subsection, I explore the coverage performance of the three estimators to assess their ability to (correctly) not reject a null hypothesis of no peer effects in both correctly specified and misspecified models.

4.2 Mispecifications type of results

Peer effect studies are often subject to criticism due to modeling (Manski, 1993; Angrist, 2014) and empirical (Angrist, 2014) concerns. While it is hoped that the framework developed in this paper alleviates modeling concerns - in particular, by avoiding reflection problems -, it is of interest to evaluate the behavior of the estimator under frequent empirical difficulties: missing or omitted covariates, group level heterogeneity, or measurement error.

I focus here on the peer effect parameter δ, which will typically the parameter of interest.

Because OLS and SAR cannot identify the structural coefficient but could still detect the presence of peer effects, it is of interest to look at the coverage performance. I look at the frequency at which a 95% confidence interval contains 0, indicating the absence of peer effects, under the generating process in which peer effects are indeed absent ($\delta = 0$).

Table 3 reports the coverage of a 95% confidence interval under the homophilic network structure when the researcher (i) observes both covariates, (ii) observes only the first covariate, (iii) observes a mismeasured (with ($\mathcal{N}(0; 0.25)$) error) first covariate, and (iv)/(v)/(vi) there is (uniform on $[-1; 0]$) group heterogeneity (added to the argument of the exponential) in the (i)/(ii)/(iii) scenario.

The coverage performance of the estimator developed in the paper is far better than that of OLS and SAR-MLE. Although the most serious issues (lack of covariate and measurement error combined with heterogeneity issues) can lead to severe size distortions, spurious peer effects are unlikely under more standard scenarios. The test for the presence of peer effect is adequately sized in the case of correct specification and is moderately distorted under measurement error or group heterogeneity.

Both OLS and SAR-MLE have a tendency to spuriously detect peer effects at a rate higher than the pre-specified level, even with homogeneous groups and adequate covariates. Any empirical difficulty such as measurement error, unobserved covariate, or heterogeneity leads by itself to a high risk of unwarranted rejection of the null of no peer effects, echoing critiques in Angrist (2014).

Table 3: Coverage analysis with potential misspecification

			Coverage		
n_b	δ		Reg	SAR	Exp
5	0	Size	0.90	0.79	0.95
		Size	0.72	0.62	0.90
		Size	0.69	0.55	0.90
		Size	0.83	0.70	0.89
		Size	0.62	0.47	0.71
		Size	0.56	0.42	0.67

Table 4: Coverage performance of a 95% confidence interval from OLS with clustered standard errors, SAR-MLE, and maximum of likelihood on latent exponential processes.

References

Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics*, 30: 98–108.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of econometrics*, 150(1): 41–55.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of social interactions through partially overlapping peer groups." *American Economic Journal: Applied Economics*, 2(2): 241–75.

He, Xiaoqi, and Kyungchul Song. 2018. "Measuring diffusion over a large network." *arXiv preprint arXiv:1812.04195.*

Lee, Lung-Fei. 2004. "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models." *Econometrica*, 72(6): 1899–1925.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The review of economic studies*, 60(3): 531–542.

Newey, Whitney K, and Daniel McFadden. 1994. "Large sample estimation and hypothesis testing." *Handbook of econometrics*, 4: 2111–2245.