Peer effect analysis with latent processes

Vincent Starck[*]
LMU München

February 4, 2025

Abstract

I study peer effects that arise from irreversible decisions in the absence of a standard social equilibrium. I model a latent sequence of decisions in continuous time and obtain a closed-form solution for the likelihood. The method avoids regression on conditional expectations or linear-in-means regression – and thus reflection-type problems (Manski, 1993) or simultaneity issues – by modeling the (unobserved) realized direction of causality, whose probability is identified. Under a parsimonious parametric specification, I introduce a peer effect parameter meant to capture the causal influence of first-movers on their peers. Various forms of peer effect heterogeneity can be accommodated. Parameters are shown to be consistently estimated by maximum likelihood methods and lend themselves to standard inference.

Keywords: Peer effects, Continuous time, Heterogeneity, Causal Inference, Networks.

1 Introduction

Analyzing peer effects is notoriously difficult. In a seminal paper, Manski (1993) discusses the reflection problem that arises when one runs a regression on the conditional expectation, which induces restrictions on the regression coefficients that lead to tautological models or identification issues.

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

In applications that feature small group or friendship, the channel of influence plausibly operates through actual outcomes, which leads to Spatial Autoregressive (SAR) or linear-in-means (LIM) models. They alleviate reflection problems insofar as they allow for identification of peer effects (Bramoullé, Djebbari and Fortin, 2009; De Giorgi, Pellizzari and Redaelli, 2010; Blume et al., 2015), *e.g.*, through the use of non-overlaping peer groups that instruments away the simultaneity bias.

However, inference about the effect of peers using SAR models can be challenging. Manski (1993) points out difficulties in identifying and estimating peer effects that transcend reflection regression issues. Identification can be tenuous in SAR models as it relies on restrictions induced by functional form that are particularly sensitive to , *e.g.*, measurement error or mechanical relationships between individual outcomes and group means (Gibbons and Overman, 2012; Angrist, 2014). Although identification is typically the rule when the network is known (Blume et al., 2015), estimation issues may still preclude reliable inference about peer effects (Hayes and Levin, 2024). In the case of irreversible decisions, the standard notion of social equilibrium is inadequate. People opt in over time according to their characteristics and shocks, then cannot subsequently adjust their outcomes.

I propose a new framework by modeling the latent sequence of decisions in continuous time. It offers a more natural framework to discuss causality, as it breaks the simultaneity by separating first-movers that potentially generate a causal reaction from the subjects of influence. As we distinguish between sources and recipients of peer influence, exploring peer-effect heterogeneity can be more fruitful, as it provides parameters is more natural to interpret. Heterogeneity in peer effects is an important, but under-explored topic (recent work on the subject includes Mogstad, Torgovitsky and Volpe (2024)).

I formalize the counterfactual framework using potential outcomes, and define causal peer effect parameters that lend themselves to straightforward inference via maximum likelihood. Decisions are explicitly made asynchronously but are not assumed to be observed. Although the order in which decisions were taken is not identified absent additional information, it is possible to identify the probability of a sequence of adoption given individuals' characteristics. The order may be of interest on its own, but it also contributes to our understanding of peer effects and has potential policy-relevance, *e.g.*, for targeting or diffusion (He and Song, 2018).

It also enables discussion of the diffusion process and counterfactual analysis about, say, the expected time before some fraction of the population adopts a technology or the impact of policy interventions.

Frameworks for analyzing peer effects beyond linear models are scarce but often desirable. As summarized in Sacerdote (2014), "researchers have shown that linear-in-means model of peer effects is often not a good description of the world, although we do not yet have an agreed-upon model to replace it." Boucher et al. (2024) is a recent breakthrough that extends the response class to contain mean, maximum, and minimum outcome in situations of equilibrium. The method may also be useful to analyze staggered treatment adoption (Shaikh and Toulis, 2021; Athey and Imbens, 2022) by relaxing the common assumption that treatment adoptions arise independently.

2 Causality, stochastic process, and likelihood

2.1 Setup and causal peer effect parameters

I focus on irreversible[1] decisions: there is an initial default state, labeled 0, and the decision to opt-in leads to an absorbing state, labeled 1. For instance, the outcome y might represent vaccination status, technology adoption, retirement, migration decision, etc. The goal is to model and estimate peer effects, *i.e.* how decisions of peers alter an individual's probability of opting in. The peers are summarized by an adjacency matrix, W.

The adoption time of an individual i, denoted by T_i, is a random variable that depends on individuals' characteristics and their expectations. Observing a peer opting in modifies the likelihood that an individual does so as well. This can be due to conformity, information transmission, or other types of social influence. This implies that there is a change in the distribution of the adoption time of individual i, T_i, upon observing the adoption of a peer.

Adapting the potential outcome notation (Neyman, 1923; Rubin, 1974) to the current setup[2], the adoption time is represented by $T_i(\tau)$ for $\tau \in \mathbb{R}^{n-1}$: the adop-

[1] This can be because the action cannot be undone (*e.g.*, vaccination), is too costly to reverse, or because the focus is on first-time events.

[2] Potential outcomes have been used to formalize causality in peer effects in different contexts. The

tion time of individual i depends on the adoption times of other individuals. This dependence carries over to the outcomes $\mathbb{1}_{T_i \leq t}$, which are observed for $t = S > 0$: $y_i \stackrel{\text{def}}{=} \mathbb{1}_{T_i \leq S}$.

In general, the resulting change in the distribution of outcomes following a change in τ can be interpreted as a peer effect. The causal impact can be captured by parameters that summarize the effect of a change in τ on the distribution.

A parameter of interest is the analog of the average treatment effect: the mean change in the outcome following a change in τ:

$$\delta(\tilde{\tau}, \tau) \stackrel{\text{def}}{=} \mathbb{E}[Y_i(\tilde{\tau}) - Y_i(\tau)] = \mathbb{P}[Y_i(\tilde{\tau}) = 1] - \mathbb{P}[Y_i(\tau) = 1] \tag{1}$$

For instance, one could be interested in $\tilde{\tau} = \vec{0} \in \mathbb{R}^{n-1}$ and $\tau = \infty e_i$[3] or $\tau = \vec{\infty}$ so that the parameter describes the change in adoption probability induced by the initial adoption of one or all peers compared to them never adopting.

We could also be interested in counterfactual effects such as the expected adoption time, $\mathbb{E}[T_i]$, and various conditional versions, or the expected time before a fraction of the population opts in.

2.2 Stochastic process and unconfoundedness

Individual adoption is modeled with the following continuous-time stochastic process:

Definition 2.1 (Stochastic process)**.** *Let T_i^1 be a set of random variables over $\mathbb{R}^+$. When a first individual, say k, opts in (i.e., $k = \arg\min_i(T_i^1)$), she withdraws and the remaining random variables are updated to T_i^2, whose distribution may depend on the time elapsed and the event that k opted in. The process then goes on with the updated distributions, and so on. The time to adoption is therefore $T_i = \sum_{k=1} T_i^k$, where the*

literature on interference ("exogeneous peer effects" or sometimes spillovers), in which one's outcome depends on neighbors' treatment, addresses (versions of) the issue and identifies direct and indirect effects under various relaxations of SUTVA (Toulis and Kao, 2013; Sofrygin and van der Laan, 2016; Aronow and Samii, 2017; Arpino, Benedictis and Mattei, 2017; Liu et al., 2019; Forastiere, Airoldi and Mealli, 2020; Jackson, Lin and Yu, 2020; Sánchez-Becerra, 2021; Huber and Steinmayr, 2021). Potential outcome that depends on peer's outcome (*e.g.*, in Egami and Tchetgen Tchetgen (2024)) have been the object of less discussion. To my knowledge, the discussion of peer effects through the time dimension is new.

[3] e_i is a vector whose only nonzero entry is a 1 in place i. I adopt the convention $0 \cdot \infty = 0$.

sum ranges from 1 *to the round where* i *comes first. The outcomes are observed at time* S*:* $y_i = \mathbb{1}_{T_i \leq S}$.

The process decomposes the arrival time T_i of each individual i into a collection $T_i^* = (T_i^1, \ldots, T_i^n)$ of *latent partial times.* This process is quite general in terms of the dynamics it allows. It basically only assumes the arrow of time, ruling out feedback from the future: the information set corresponds to the filtration $\mathcal{F}$ generated by the outcomes.

Peer effects occur when the updated distributions do not coincide with the previous distribution. In the absence of peer effects, the potential outcomes do not vary with τ. Indeed, they are related to the latent partial times T_i^k *via*

$$T_i(\tau) = \sum_k \big(\mathbb{1}(T_i^k < \tau_{(k)} - \tau_{(k-1)})T_i^k + \mathbb{1}(T_i^k \geq \tau_{(k)} - \tau_{(k-1)})(\tau_{(k)} - \tau_{(k-1)})\big) \tag{2}$$

where $\tau_{(0)} \stackrel{\text{def}}{=} 0$ and the sum ranges from 1 to $\arg\min_k \{T_i^k \geq \tau_{(k)} - \tau_{(k-1)}\}$.

Latent partial times are usually correlated because characteristics influence both the choice of peers and the distribution of times, inducing, *e.g.*, homophily bias (Shalizi and Thomas, 2011). Controlling for these characteristics can restore independence:

Assumption 2.2 (Independence of latent partial times)**.** *The latent partial times are conditionally independent:* $T_{-i}^* \perp\!\!\!\perp T_{i'}^* | X, W, \mathcal{F}$

The independence of latent partial times is related to a notion of unconfoundedness. When adoption is described by the stochastic process, the time to adoption of individual i under exogenous assignment of other partial times to τ is $T_i(\tau)$. Then, unconfoundedness can be stated as follows.

Assumption 2.3 (Unconfoundedness of timings)**.** *Potential outcomes are independent of latent partial times:* $\forall i, \{T_{-i}^k, k = 1, \ldots, n\} \perp\!\!\!\perp \{T_i(\tau), \tau \in \mathbb{R}^{n-1}\} | X, W, \mathcal{F}$

where T_{-i} is the set of times of all individuals but i.

Then, from the formulation of potential outcomes, we obtain immediately the following result:

Proposition 2.4. *Unconfoundedness of timings is equivalent to independence of latent partial times.*

2.3 Likelihood

When observables are limited to the outcomes, the latent distributions are not identified. To make the model tractable and operational, I specify the distribution to be exponential and assume that only the unordered decision of peers affects the distribution of latent partial times:

Assumption 2.5. $T_i^k | X, W, \mathcal{F} \sim Exp(\lambda_i^{+C})$

where $C \subseteq \{1, \ldots, n\} \backslash \{i\}$ is the set of pre-step k adopters who are connected to i.

I focus on Exponential distributions for two reasons. First, exponential waiting times arise automatically under the assumption of a constant probability per unit of time, a natural point of departure. Second, exponential distributions are particularly attractive from an analytical standpoint and ensure tractability. Due to the memorylessness property, conditioning on elapsed time is irrelevant.

Importantly, exponential rates can be left unrestricted as functions of covariates. Given a distribution of the outcomes, they are nonparametrically identified under some conditions. As such, exponentials provide a convenient framework in which peer effects are easy to interpret while allowing for considerable heterogeneity. When time analysis is of special interest (*e.g.*, diffusion analysis, generation of counterfactuals, or identification based on times of adoption), the dynamics induced by exponential distributions have further implications as they have an influence on some estimates.

The exponential rates capture the heterogeneity across individuals and their changes over time reflect the influence of peers. Their changes can be directly interpreted as peer effects insofar as they capture the change in the shape of the distribution and the measure average reduction in adoption times. They also translate into practical formulae to compute parameters of interest.

Example 2.6. *Consider the change in probability of adoption:*

$$\mathbb{E}[Y_i(\vec{0}) - Y_i(\vec{\infty})|X] = e^{-\lambda_i S}(1 - e^{-(\lambda_i^{+} - \lambda_i)S)}) \tag{3}$$

where $+$, without a subscript, corresponds to the highest-level of peer effects – all peers have previously opted in.

To the first-order, this simplifies to

$$e^{-\lambda_i S}(\lambda_i^{+} - \lambda_i)S \tag{4}$$

The first term, $e^{-\lambda_i S}$, is the baseline probability of not opting in before time S, while $(\lambda_i^+ - \lambda_i)$ reflects the change in exponential rates induced by the prior adoption of friends.

The following examples consider simple homogeneous cases to illustrate the framework and build intuition for identification conditions, which are formalized in the next subsection.

Example 2.7. *Consider two (connected) individuals ($i = 1, 2$) and homogeneous rates λ, λ^+. The probabilities for the four possible outcomes are given by*

$$\begin{aligned}
p_{00} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = y_2 = 0 | W_{12} = 1] = e^{-2\lambda S} \\
p_{10} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = 1, y_2 = 0 | W_{12} = 1] = \lambda e^{-\lambda^+ S} g(2\lambda - \lambda^+) \\
p_{01} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = 0, y_2 = 1 | W_{12} = 1] = \lambda e^{-\lambda^+ S} g(2\lambda - \lambda^+) \\
p_{11} &\stackrel{\text{def}}{=} \mathbb{P}[y_1 = y_2 = 1 | W_{12} = 1] = 1 - p_{00} - p_{10} - p_{01}
\end{aligned}$$

where $g(\lambda) = \frac{1-e^{-\lambda S}}{\lambda}$ if $\lambda \neq 0$ and $g(0) = S$.

The probabilities are identified is one observes a sequence of independent draws of such pairs. In this case, the rates are identified since they can be recovered from the probabilities: $\lambda = -\frac{\ln(p_{00})}{2S}$ while λ^+ solves $\frac{\lambda S(p_{00} - e^{-\lambda^+ S})}{\ln(p_{00}) + \lambda^+ S} = p_{10}$, where the left-hand side is strictly decreasing.

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 strength and (ii) determine the probability of an individual moving first, which provides a probability for each direction of causality.

Example 2.8. *Consider a complete network of size n, $W = \iota\iota' - I$. People have baseline rates $\lambda = e^\beta$ for $\beta \in \mathbb{R}$ that are updated to $e^{\beta + \frac{k}{n-1}\delta}$ when k people have opted in.*

The outcomes $y_i \in \{0, 1\}$ only inform the fraction of people who opted in, which is insufficient to identify (β, δ) even as n grows. In this regime, positive peer effects cannot be distinguished from stronger baseline rates.

The complete network case thus requires more information, but can still be identified. For instance, if the times of adoption of early adopters are known, then λ is *identified from the behavior of the density around the origin. Given the knowledge of λ, δ is identified from the share of adopters.*

Because the source of non-identification is the absence of observed separate streams of information about baseline and peer effects, sparser networks offer other avenues for identification. If the identity of first-movers (people whose friends had not opted-in before they did) is known, then λ can be identified from a growing set of first-movers.

2.4 Identification

The sample comprises outcomes $(y_i, i = 1, \ldots, n)$, covariates $(x_i, i = 1, \ldots, n)$, and an adjacency matrix W ($W_{ij} = 1$ if and only if i and j are "neighbors") that determines the peer group of each individual.

Lemma 2.9. *Suppose one of the following holds.*
(i) The network has a block structure $b = 1, \cdots, B$. Blocks $b = 1, \ldots, B$ have sizes $N_b \leq \overline{N}$ for some $\overline{N} \in \mathbb{R}$ and the number of blocks grows large. $\mathbb{P}[Y_b = y_b | W_b, x_b] > 0$.
(ii) The order of adoption is known and the network is sparse.
Then, λ_i^{+C} is identified for all i and any set of people C who are connected to i with positive probability.

Remark 2.10 (Identification of order probabilities)**.** *Identification of the parameters and generally of the family of rates implies the probability of any sequence of adoption is also identified. Hence, although it is not possible to recover the identity of the first-movers or the realized order of adoptions, it is possible to assign a probability to such events.*

2.5 Likelihood

The likelihood induced by the stochastic process is available in closed form. This result is the object of the next theorem.

Theorem 2.11. *If the latent partial times are independent, then the log-likelihood under the exponential specification takes the form*

$$l \stackrel{\text{def}}{=} \ln \left(\sum_{p \in \mathcal{P}} \left(\prod_{i=1}^{G} \lambda_{p_i}^{+p_1, \cdots, p_{i-1} \cap W} \right) \sum_{g=1}^{G+1} \frac{e^{-c_{p_g} S}}{\prod_{h \neq g} c_{p_h} - c_{p_g}} \right) \tag{5}$$

where $\mathcal{P}$ contains permutations of the people who opted in and $c_g \stackrel{\text{def}}{=} \sum_{i=g}^{N} \lambda_i^{+\{1,\cdots,g-1\}\cap W}$.

The following example shows that the process is a generalization of a standard iid exponential draws. The heterogeneity in rates induce different distributions, while the peer effects reflected in $\lambda^+ \neq \lambda$ create spatial dependence.

Example 2.12. *Suppose that there is no heterogeneity nor peer effects: $\lambda_i^{+C} = \lambda$ for all i and any C. Then $c_g = \lambda(N-(g-1))$ and the likelihood of permutation p becomes*

$$\begin{aligned}
\prod_{i=1}^{G} \lambda_i^{+_1,\cdots,+_{i-1}} \sum_{g=1}^{G+1} \frac{e^{-c_g S}}{\prod_{h\neq g} c_h - c_g} &= \lambda^G \sum_{g=1}^{G+1} \frac{e^{-\lambda(N-(g-1))S}}{\prod_{h\neq g} \lambda(N-(h-1)) - \lambda(N-(g-1))} \\
&= \sum_{g=1}^{G+1} \frac{e^{\lambda(g-1)S}}{\prod_{h\neq g} h-g} \\
&= \sum_{g=1}^{G+1} \frac{e^{\lambda(g-1)S}}{(g-1)!\,(G+1-g)!}(-1)^{G+1-g} \\
&= \frac{e^{-\lambda NS}(e^{\lambda S}-1)^G}{G!} \\
&= \frac{e^{-\lambda(N-G)S}(1-e^{-\lambda S})^G}{G!}
\end{aligned}$$

Summing over all permutations yields $e^{-\lambda(N-G)S}(1-e^{-\lambda S})^G$, which is the likelihood of G i.i.d. exponentially-distributed variables falling below the cutoff, S.

Therefore, in the absence of peer effects, probabilities reduce to a standard exponential race with independent draws. Changes in exponential rates are a measure of dependence and social interaction; they reflect peer effects.

Remark 2.13 (Computation)**.** *Although summing over all permutations can lead to an impractical computational burden, the cost is significantly lower in practice. First, the likelihood factorizes based on the components of W, whose size can be much smaller (classrooms, villages, connected components of friends, etc.). Second, permutations are based on the number of adopters, rather than the size of the network. Hence, the complexity of permutations can be substantially reduced, especially if partial information about the ordering is available.*

Finally, approximations can further reduce the number of permutations. In par-

ticular, the average over all permutations can be estimated by a random sample[4] *of permutations by the law of large numbers.*

3 Asymptotic Theory

Although rates are nonparametrically identified, it is helpful to introduce parametric restrictions. This avoids major curse-of-dimensionality issues for estimation and simplifies interpretation.

A natural way to obtain a parsimonious model is to specify the baseline rates as $\lambda_i = g(W, x_i, \theta)$ for a finite-dimensional vector of parameters θ, and let the updated rates be obtained by a scaling factor. For instance, a simple specification is

$$\ln(\lambda_i^{+C}) = x_i^{'}\beta + d_i^{-1}|C|\delta \tag{6}$$

where $d_i = \sum_j W(i,j)$ is the degree of i. This reduces the dimensionality of the problem to $(\dim(x_i)+1)$, ensures positive rates, and lets peer effects be described by a parameter δ, which reflects how rates are scaled when the peer group opts in.

A natural way to explore heterogeneity in peer effects is then to add interaction terms or nonlinear terms. For instance, suppose a researcher is analyzing peer effects in vaccine uptake. They might conjecture that people respond differently depending on the level of education of their peers. Then, they can add the mean level of education among vaccinated peers to assess its impact on peer effect strength.

Example 3.1. *To the first-order in δ, we also have*

$$\mathbb{E}[Y_i(\vec{0}) - Y_i(\vec{\infty})|X] \approx \delta S \lambda_i e^{-\lambda_i S} \tag{7}$$

so that δ acts a scaling factor in the change in the probability of adoption induced by the peer group adopting at the onset.

---

[4] Instead of sampling uniformly over the set of permutations, one can also generate sequences of arrival according to the stochastic process. This generates draws according to the probability of the corresponding order of decisions, increasing speed of convergence. For a sample $\mathcal{P}_s$ of such draws, one can recover the likelihood from $\frac{1}{|\mathcal{P}_s|}\sum_{p\in\mathcal{P}_s} 1/\mathbb{P}[p] \to^p \frac{G!}{\sum_{p\in\mathcal{P}}\mathbb{P}[p]}$ using that $\frac{\mathbb{P}[p]}{\sum_{p\in\mathcal{P}}\mathbb{P}[p]}$ is the probability of sampling permutation p under this sampling scheme. The score can be obtained similarly.

I consider a sequence of networks with blocks or components of size N_b, $b = 1, \cdots, B$. With independent blocks, the likelihood factorizes as $l \stackrel{\text{def}}{=} \ln(\mathbb{P}[Y = y]) = \sum_{b=1}^{B} l_b$, where l_b is the log-likelihood of block b. The log-likelihood of each block is given by the formula established in the previous section. The score and Hessian are easily computed and are available in closed form. The details are provided in the Appendix.

The estimator inherits the usual properties of maximum likelihood estimators: it is consistent and asymptotically normal under regularity conditions. Formally,

Theorem 3.2 (Consistency). *Suppose the data-generating process is given by the stochastic process described in Section 2.2 and that the family of rates is identified and belongs to the interior of a compact set. Then, if blocks are drawn independently with $W_b \sim \mathcal{D}_W$ such that $N_b < \overline{N}$, the maximum likelihood estimator is consistent with $\hat{\lambda}_i^{+C} \to^p \lambda_i^*$, for any $C \subseteq \{1, \ldots, i-1, i+1, \ldots, n\}$, as $B \to \infty$.*

Moreover,

Theorem 3.3 (Asymptotic Normality). *Under the assumptions of Theorem 3.1, the maximum likelihood estimator is asymptotically normal as $B \to \infty$ with*

$$\sqrt{B}(\hat{\lambda} - \lambda) \to^d \mathcal{N}(0, \mathbb{E}[\partial l (\partial l)']) \tag{8}$$

These theorems allow for standard inference about the rates.

Remark: Spatial dependence For a given network structure, the strength of spatial dependence is controlled by the value of δ. In particular, observations are independent in the absence of peer effects ($\delta = 0$).

Remark: Discrete outcomes One can generalize the process to deal with discrete outcomes. For instance, one could have $\lambda_i = \sum_k \lambda_{ik}$ and if $T_i = t$ then choice k occurs with probability λ_{ik}/λ_i. 0 is then a baseline opt-out choice, the default.

Remark: Counterfactual analysis It is often of interest to assess the impact of policies, for instance targeting influential individuals in a network. The modeling of time dynamics allows for a direct evaluation of the 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 $k \neq i$.

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 $\lambda_i^{+k_1 \cdots + k_m} = e^{x_i'\beta + \frac{\sum_{j=1}^{m} W(i,k_j)}{d_i}\delta}$ 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 adoptions 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 adopted 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[5] 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

[5]The weighting matrix is row-normalized since the process suggests peer effects depend on the average number of adopted 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.

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.

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.

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.

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).

References

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

Aronow, Peter M, and Cyrus Samii. 2017. “Estimating average causal effects under general interference, with application to a social network experiment.” *The Annals of Applied Statistics*, 11(4): 1912–1947.

Arpino, Bruno, Luca De Benedictis, and Alessandra Mattei. 2017. “Implementing propensity score matching with network data: the effect of the General Agreement on Tariffs and Trade on bilateral trade.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(3): 537–554.

Athey, Susan, and Guido W Imbens. 2022. “Design-based analysis in difference-in-differences settings with staggered adoption.” *Journal of Econometrics*, 226(1): 62–79.

Blume, Lawrence E, William A Brock, Steven N Durlauf, and Rajshri Jayaraman. 2015. “Linear social interactions models.” *Journal of Political Economy*, 123(2): 444–496.

Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou. 2024. “Toward a general theory of peer effects.” *Econometrica*, 92(2): 543–565.

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.

Egami, Naoki, and Eric J Tchetgen Tchetgen. 2024. "Identification and estimation of causal peer effects using double negative controls for unmeasured network confounding." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(2): 487–511.

Forastiere, Laura, Edoardo M Airoldi, and Fabrizia Mealli. 2020. "Identification and estimation of treatment and interference effects in observational studies on networks." *Journal of the American Statistical Association*, 1–18.

Gibbons, Stephen, and Henry G Overman. 2012. "Mostly pointless spatial econometrics?" *Journal of regional Science*, 52(2): 172–191.

Hayes, Alex, and Keith Levin. 2024. "Peer effects in the linear-in-means model may be inestimable even when identified." *arXiv preprint arXiv:2410.10772*.

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

Huber, Martin, and Andreas Steinmayr. 2021. "A framework for separating individual-level treatment effects from spillover effects." *Journal of Business & Economic Statistics*, 39(2): 422–436.

Jackson, Matthew O, Zhongjian Lin, and Ning Neil Yu. 2020. "Adjusting for peer-influence in propensity scoring when estimating treatment effects." *Available at SSRN 3522256*.

Liu, Lan, Michael G Hudgens, Bradley Saul, John D Clemens, Mohammad Ali, and Michael E Emch. 2019. "Doubly robust estimation in observational studies with partial interference." *Stat*, 8(1): e214.

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

Mogstad, Magne, Alex Torgovitsky, and Oscar Volpe. 2024. "Peer Effects in Linear-in-Means Models with Heterogeneous Interaction Effects." *Working paper.*

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

Neyman, Jerzy. 1923. "On the application of probability theory to agricultural experiments. Essay on principles." *Ann. Agricultural Sciences*, 1–51.

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology*, 66(5): 688.

Sacerdote, Bruce. 2014. "Experimental and quasi-experimental analysis of peer effects: two steps forward?" *Annu. Rev. Econ.*, 6(1): 253–272.

Sánchez-Becerra, Alejandro. 2021. "Spillovers, Homophily, and Selection into Treatment: The Network Propensity Score." *Job market paper.*

Shaikh, Azeem M, and Panos Toulis. 2021. "Randomization tests in observational studies with staggered adoption of treatment." *Journal of the American Statistical Association*, 116(536): 1835–1848.

Shalizi, Cosma Rohilla, and Andrew C Thomas. 2011. "Homophily and contagion are generically confounded in observational social network studies." *Sociological methods & research*, 40(2): 211–239.

Sofrygin, Oleg, and Mark J van der Laan. 2016. "Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population." *Journal of causal inference*, 5(1).

Toulis, Panos, and Edward Kao. 2013. "Estimation of causal peer influence effects." 1489–1497, PMLR.

Appendix A: Proofs and Further results

2 individuals

Proof. The result is shown for possibly different rates λ_1, λ_2. Densities are denoted by p.

Consider the probability $\mathbb{P}[y_1 = 1, y_2 = 0]$. This corresponds to

$$
\begin{aligned}
\mathbb{P}[T_1 \leq S, T_2 > S] &= \int_0^S p[T_1 = t, T_2 > S] \; dt \\
&= \int_0^S \lambda_1 e^{-\lambda_1 t} \; p[T_2 > S | T_1 = t] \; dt \\
&= \int_0^S \lambda_1 e^{-\lambda_1 t} \; p[T_2^1 > t | T_1 = t] \; p[T_2^2 > S - t | T_1 = t] \; dt \\
&= \int_0^S \lambda_1 e^{-\lambda_1 t} \; p[T_2^1 > t] \; p[T_2^2 > S - t | T_1 = t] \; dt \\
&= \int_0^S \lambda_1 e^{-\lambda_1 t} e^{-\lambda_2 t} e^{-\lambda_2^+ (S-t)} \; dt
\end{aligned} \tag{9}
$$

If $\lambda_1 + \lambda_2 - \lambda_2^+ = 0$, this is simply $\lambda_1 S e^{-\lambda_2^+ S}$. If $\lambda_1 + \lambda_2 - \lambda_2^+ = 0$, then integrating yields $\lambda_1 e^{-\lambda_2^+ S}(1 - e^{-(\lambda_1 + \lambda_2 - \lambda_2^+)S})/(\lambda_1 + \lambda_2 - \lambda_2^+)$.

The remaining probabilities follow automatically. $\mathbb{P}[y_1 = 0, y_2 = 1]$ follows by symmetry, $\mathbb{P}[y_1 = 0, y_2 = 0] = e^{-(\lambda_1 + \lambda_2)S}$ trivially, and $\mathbb{P}[y_1 = 1, y_2 = 1]$ is deduced by the fact that the four probabilities sum up to 1.

□

Likelihood

Proof. The log-likelihood is given by

$$\ln(\mathbb{P}[Y_1 = y_1, \ldots, Y_N = y_N | X, W]) \tag{10}$$

Assume without loss that the individuals who opted in correspond to the first G observations. Then, summing over different orderings,

$$
\begin{aligned}
&\mathbb{P}[y_1 = \ldots = y_G = 1; y_{G+1} = \ldots = y_N = 0] \\
&= \sum_{p \in \mathcal{P}} \mathbb{P}[y_1 = \ldots = y_G = 1; y_{G+1} = \ldots = y_N = 0; y_{G+1}, \cdots, y_N > S; T_{p_1} < \ldots < T_{p_G}]
\end{aligned}
$$

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

Consider the representative term in the sum when individual i is the i-th to opt in. Using the independent of partial times and the exponential specification and following steps similar to the 2-individual examples, this term is given by

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

which, after some algebra, reduces to

$$e^{-\sum_{i=G+1}^N \lambda_i^{+1,\cdots,+G} S} \left(\prod_{i=1}^G \lambda_i^{+1,\cdots,+i-1} \right) I_{\{c_i, i=1,\cdots,G\}} \tag{11}$$

with $I_{\{h_i,i=1,\cdots,H\}} \overset{\text{def}}{=} \frac{1}{\prod_{g=1}^G \dot{c}_g} + (-1)^G \sum_{g=1}^G \frac{1}{\prod_{h\neq g}(\dot{c}_g - \dot{c}_h)} \frac{e^{-\dot{c}_g S}}{\dot{c}_g}$ and $\dot{c}_g \overset{\text{def}}{=} \sum_{i=g}^N \lambda_i^{+1,\cdots,+g-1} - \sum_{i=G+1}^N \lambda_i^{+1,\cdots,+G}$.

Introducing $\dot{c}_{G+1} = 0$ and then $c_g \overset{\text{def}}{=} \dot{c}_g + \sum_{i=G+1}^N \lambda_i^{+1,\cdots,+G} = \sum_{i=g}^N \lambda_i^{+1,\cdots,+g-1} \geq 0$ (with equality only if $g = G+1$ and $N = G$), the probability simplifies to[6]

$$\prod_{i=1}^G \lambda_i^{+1,\cdots,+i-1} \sum_{g=1}^{G+1} \frac{e^{-c_g S}}{\prod_{h\neq g} c_h - c_g} \tag{12}$$

□

Identification

Proof. In what follows, $\vec{0}$ denotes a 0-vector of comformable size. Consider the identified probability $p_{\vec{0}}(x_1, \ldots, x_n) = e^{-\sum_{i=1}^n (\lambda_i(x_i))S}$. Setting $x_i = x \ \forall i$, $\lambda(x) = \frac{-\ln(p_{00}(x,\ldots,x))}{nS}$ and thus the baseline rates are identified. Then, $p_{10}(x_1, x_2) = \lambda_1 e^{-\lambda_2^+ S} g(\lambda_1 + \lambda_2 - \lambda_2^+)$ allows us to identify λ_2^+ nonparametrically as $\lambda^+(x_2, x_1)$ because $e^{-xS} g(\lambda_1 + \lambda_2 - x) = e^{-(\lambda_1+\lambda_2)S} g(-(\lambda_1 + \lambda_2 - x))$ is an invertible function of x for any λ_1, λ_2.

If only person 1 activates,

[6] As in the example with 2 individuals, it can be checked from the integral computation that the limit for a term in a denominator converging to 0 gives the correct probability at the points where the function is undefined.

$$\lambda_1 \left(\frac{e^{-\sum_{i=2}^N \lambda_i^+ S}}{\sum_{i=1}^N \lambda_i - \sum_{i=2}^N \lambda_i^+} - \frac{e^{-\sum_{i=1}^N \lambda_i S}}{\sum_{i=1}^N \lambda_i - \sum_{i=2}^N \lambda_i^+} \right) \tag{13}$$

Identification under sparsity Define the first-movers F through $j \in F^c : \exists k, W_{kj} = 1, T_k < (T_j \wedge S)$.

$T|F,\lambda \sim T|W,T_j,\lambda \sim T_j^1|W,\lambda$ under unconfoundedness.

$$\mathbb{P}[y_j = 1 | j \in F] = \frac{\mathbb{P}[y_j = 1, j \in F]}{\mathbb{P}[j \in F]} \tag{14}$$

is identified if $\mathbb{P}[j \in F] \to c \in]0,1]$, i.e., the probability of being a first-mover does not vanish.

$\square$

Derivatives of the likelihood

The log-likelihood reads

$$l \stackrel{\mathsf{def}}{=} \ln \left(\frac{1}{G!} \sum_{p \in \mathcal{P}} \sum_{g=1}^{G+1} \left(\prod_{i=1}^{G} \lambda_{p_i}^{+p_1,\cdots,+p_{i-1}} \right) \frac{e^{-c_{p_g} S}}{\prod_{h \neq g} c_{p_h} - \ddot{c}_{p_g}} \right) \tag{15}$$

Let

$$a_{pg} \stackrel{\mathsf{def}}{=} \left(\prod_{i=1}^{G} \lambda_{p_i}^{+p_1,\cdots,+p_{i-1}} \right) \frac{e^{-c_{p_g} S}}{\prod_{h \neq g} c_{p_h} - c_{p_g}} \tag{16}$$

Then, the score is given by

$$\frac{\partial l}{\partial \theta} = \sum_{p \in \mathcal{P}} \sum_{g=1}^{G+1} \frac{a_{pg}}{\sum_p \sum_g a_{pg}} \left(\sum_{i=1}^{G} \frac{\partial_\theta \lambda_{p_i}^{+p_1,\cdots,+p_{i-1}}}{\lambda_{p_i}^{+p_1,\cdots,+p_{i-1}}} - \partial_\theta c_{p_g} S - \sum_{h \neq g} \frac{\partial_\theta (c_{p_h} - c_{p_g})}{c_{p_h} - c_{p_g}} \right) \tag{17}$$

and the Hessian is given by

$$\begin{aligned} H \stackrel{\mathsf{def}}{=} \frac{\partial l}{\partial \theta \partial \theta'} &= \sum_{p \in \mathcal{P}} \sum_{g=1}^{G+1} \partial_{\theta'} \left(\frac{a_{pg}}{\sum_p \sum_g a_{pg}} \right) \left(\sum_{i=1}^{G} \frac{\partial_\theta \lambda_{p_i}^{+p_1,\cdots,+p_{i-1}}}{\lambda_{p_i}^{+p_1,\cdots,+p_{i-1}}} - \partial_\theta c_{p_g} S - \sum_{h \neq g} \frac{\partial_\theta (c_{p_h} - c_{p_g})}{c_{p_h} - c_{p_g}} \right) \\ &+ \sum_{p \in \mathcal{P}} \sum_{g=1}^{G+1} \frac{a_{pg}}{\sum_p \sum_g a_{pg}} \left(\sum_{i=1}^{G} \partial_{\theta'} \left(\frac{\partial_\theta \lambda_{p_i}^{+p_1,\cdots,+p_{i-1}}}{\lambda_{p_i}^{+p_1,\cdots,+p_{i-1}}} \right) - \partial_\theta c_{p_g} S - \sum_{h \neq g} \partial_{\theta'} \left(\frac{\partial_{\theta\theta'} (c_{p_h} - c_{p_g})}{c_{p_h} - c_{p_g}} \right) \right) \end{aligned} \tag{18}$$

When $\lambda_i^{+k_1\cdots+k_m} = e^{x_i'\beta+d_i{}^{-1}\sum_{j=1}^m W(i,k_j)\delta}$ with $\theta \overset{\text{def}}{=} (\beta',\delta)'$, the derivatives take the form

$$\partial_\theta \lambda_i^{+k_1\cdots+k_m} = \lambda_i^{+k_1\cdots+k_m} \begin{pmatrix} x_i \\ d_i{}^{-1}\sum_{j=1}^m W(i,k_j) \end{pmatrix} \tag{19}$$

and

$$\partial_{\theta\theta'} \lambda_i^{+k_1\cdots+k_m} = \lambda_i^{+k_1\cdots+k_m} \begin{pmatrix} x_i x_i' & x_i d_i{}^{-1}\sum_{j=1}^m W(i,k_j) \\ x_i d_i{}^{-1}\sum_{j=1}^m W(i,k_j) & (d_i{}^{-1}\sum_{j=1}^m W(i,k_j))^2 \end{pmatrix} \tag{20}$$

Consistency and asymptotic normality

Proof. The proofs verify Newey and McFadden (1994)'s sufficient conditions for consistency and asymptotic normality of extremum estimators.

The parameter space is compact by assumption. The log-likelihood of block b is $l_b = \sum_{y\in 0,1^{N_b}} \mathbb{1}[Y_b = y]\ln(\mathbb{P}[Y_b = y])$, where the elements of the sum are bounded from below using that $N_b \leq \overline{N}$ and compactness of the rates and from above by 0. Since in addition $\mathbb{P}[Y_b = y]$ is continuous, uniform laws of large numbers implies (i) $\sup_{\{\lambda\}}|B^{-1}\sum_b l_b - \mathbb{E}[l_b]| \to 0$ and (ii) $\mathbb{E}[l_b]$ is continuous.

By continuity of the Hessian matrix and compactness of the parameter space, it follows that

$$\sup_{\{\lambda\}}|H_n - H| \to 0 \tag{21}$$

where H is continuous. Asymptotic normality of the score follows from laws of large numbers for triangular arrays, noting that $\mathbb{V}[l_b] \to \sigma^2$

By assumption, the parameters are identified and live in a compact set. The objective function converges uniformly to $\mathbb{E}[l_b]$, which is continuous, by uniforms law of large numbers (*e.g.*, Lemma 2.4 in Newey and McFadden (1994)). Indeed, l_b is continuous and

$$\min_{y,p,\lambda} \mathbb{P}[p(\lambda)] \leq e^{l_b} \leq 1, \tag{22}$$

where the minimization is over a finite set for y, p since $N_b \leq \overline{N}$ and the minimization over λ is over a compact set for a continuous function, so that $\sup|B^{-1}\sum_{b=1}^B l_b -$

$\mathbb{E}[l_b]| \to 0$.

$$\frac{1}{\sqrt{B}} \sum_{b=1}^{B} s_b \to^d \mathcal{N}(0; \mathbb{E}[s_b s_b']) \tag{23}$$

noting that the score has mean zero and

$$\begin{aligned}
\mathbb{V}[s_b] &= \mathbb{E}[s_b s_b'] \\
&\leq C\mathbb{E}[\partial L_b \partial L_b'] \\
&= C \sum_{W_b} \mathbb{P}[W_b]\mathbb{E}[\partial L_b \partial L_b' | W_b] \\
&= C \sum_{W_b} \mathbb{P}[W_b] \sum_{\tilde{y}} \mathbb{P}[Y = \tilde{y}] \partial L_b(\tilde{y}) \partial L_b'(\tilde{y}) \\
&\leq C \max_{\tilde{y}} \partial L_b(\tilde{y}) \partial L_b'(\tilde{y})
\end{aligned} \tag{24}$$

□