Submitted to *Econometrica*

DISCRETE CHOICE WITH GENERALIZED SOCIAL INTERACTIONS

OSCAR VOLPE
Department of Economics, Harvard University

This paper studies social interactions in discrete choice models where individuals interact differently with different network members, conforming to some while distinguishing from others. Under this generalized framework, I show how to learn about endogenous interaction effects from data on individual choices. I propose a partial identification strategy that leverages within-network variation in individual characteristics to account for unobserved contextual effects. I also show how to derive internal instruments to correct for measurement error bias, a key source of endogeneity in models with incomplete information. Lastly, I apply my approach to data from two empirical settings: classroom peer effects in Tennessee primary schools and spillovers in deworming treatment uptake in Kenya. In both settings, I find that differences in social interaction effects, where individuals are more likely to conform to similar peers, play an important role in shaping economic outcomes.

KEYWORDS: social interaction effects, discrete choice, strategic complements and substitutes, contextual effects, measurement error bias, partial identification.

1. INTRODUCTION

The role of social interactions in shaping individual behavior is an active research topic in economics. Most empirical work has focused on cases with uniform, positive interactions, where all agents conform in the same way to the average action of others in their network. Meanwhile, less attention has been given to cases with nonuniform or mixed-sign interactions, where agents are affected differently by different network members—conforming to some while distinguishing from others.These types of interactions can be harder to evaluate empirically; yet, they are increasingly recognized as important in many economic settings, where individual identity shapes how people's actions are influenced by those around them.

Oscar Volpe: ovolpe@fas.harvard.edu

I am grateful to Steven Durlauf for his guidance and support. I also thank Isaiah Andrews, Stéphane Bonhomme, Ozan Candogan, Arun Chandrasekhar, Michael Dinerstein, Ali Hortaçsu, Robert Moffitt, Magne Mogstad, Francesca Molinari, Philip Reny, and Alex Torgovitsky for helpful comments and suggestions, as well as seminar participants in the North American UEA Meeting, the SITE Conference on Empirical Implementation of Theoretical Models, and the Applied Micro working group, I/O working group, and third-year research seminar at the University of Chicago. I am grateful to three anonymous referees for invaluable feedback. All errors are my own.

In this paper, I develop new empirical tools to study discrete choice models with *generalized social interactions*, where agents experience both positive and negative spillovers. I build an equilibrium framework that extends the classical model of Brock and Durlauf (2001) by allowing the sign and intensity of interaction effects to vary among agents in a network. Under this framework, I show how to recover key economic quantities from individual choice data, and I construct an estimator for endogenous interaction effects that can be readily applied to standard datasets. I then apply my methods to data from two large-scale education experiments to draw inference about the role of nonuniform interaction effects.

In Section 2, I begin by modeling a network of agents who make binary choices subject to social interaction effects and private payoffs. As in Brock and Durlauf (2001), I assume agents act with incomplete information, responding to expected rather than realized choices of others. Yet, unlike previous work, I allow interaction effects to vary among agents and to take negative values. To achieve this, I extend the model to allow for a matrix of interaction effects $\mathbf{J} = [J_{k\ell}]_{k,\ell}$, where $J_{k\ell}$ specifies how one type of agent ℓ influences another type k.

By imposing no restrictions on $\mathbf{J}$, I show that the model can be used to study a range of economic decision problems across different network structures, featuring both conformity and differentiation. I characterize the uniqueness and local stability of equilibria under this generalized framework, showing how these properties depend on the signs and magnitudes of the interaction effects.[1] I also show that, in the presence of multiple equilibria, negative interactions can introduce welfare tradeoffs, in which no equilibrium is broadly favorable.

Next, in Section 3, I turn to the question of identification. That is, given data on individual choices across multiple networks, what can be learned about the role of social interactions? The identification of discrete choice network models with incomplete information is studied extensively by Brock and Durlauf (2001, 2007), and more recently by Bhattacharya et al. (2023). Graham (2015), Paula (2017) and Kline and Tamer (2020) give recent surveys of the literature. My analysis builds on this work by tackling two key identification challenges.

First, I address the issue of unobserved network effects—contextual factors that obscure the role of social interactions on group-level outcomes. For example, given data on student achievement, it is hard to isolate the role of peer effects from unobserved class characteristics like teacher quality. Much of the applied literature rules out or places strong parametric restrictions on these unobservables.[2] However, this risks introducing omitted variable bias.

[1] While these properties are well-studied for uniform, positive interactions, they do not readily transfer to cases with negative interactions; see Jackson and Zenou (2015) for discussion. My analysis contributes by showing how key results associated with positive interactions extend to a broader set of models involving negative interactions.

[2] For example, Brock and Durlauf (2001, 2007) assume that the contextual effect is a constant linear function of observed variables, whereas Bhattacharya et al. (2023) assume that it has a specific linear factor structure. Others,

To handle this issue, I propose a partial identification strategy to recover key features of $\mathbf{J}$ without imposing any restrictions on unobserved network effects. This method exploits a panel structure inherent to the model, where agents in the same network interact in varying ways. Since these agents are exposed to the same contextual factors, I can control for network unobservables by comparing outcomes of different agents in the same network. This method allows me to draw insight about systematic variation in social interaction effects.

Second, I address a core issue of network models with incomplete information, which is that agents' expectations are never directly observed. Instead, researchers only see agents' realized choices in finite networks. Since the average action in the network converges to an expectation as the network size grows large, previous work has treated this as an estimation issue rather than as a threat to identification.[3] However, I argue this approach is inadvisable, because for any finite network, the average action is endogenous in the model due to measurement error. I show that by failing to account for this endogeneity, standard estimation methods would produce highly biased estimates, even as the number of networks grows large.

I approach the issue as a classic measurement error problem (Wooldridge, 2013, Section 15.4), noting that observed average choices are always noisy proxies for true expectations and are thus endogenous. To correct for endogeneity, I use internal instruments: randomly partitioning each network into two parts, and using the average action in one part as an instrument for the average action in the other. As long as networks are partitioned randomly, this method yields an IV estimator for social interactions that is consistent in any finite network setting. I demonstrate that this estimator performs well in Monte Carlo simulations.

Finally, in Section 4, I apply my framework to data for two empirical applications. First, I analyze Project STAR—a large-scale education experiment in Tennessee that randomized students into classrooms—to study how peer effects differ by gender. I find strong evidence that boys and girls are more likely to conform to their own gender than to the other gender. Next, I use data from the Primary School Deworming Project, previously studied by Miguel and Kremer (2004), to estimate how spillovers in treatment uptake differ by age and gender. Again, I find that students prefer to conform to similar peers. Moreover, I estimate that these peer effect differences are strongest among younger students. In both settings, I find that the generalized interactions model uncovers systematic variation in interaction effects that most standard frameworks would overlook. Accounting for this variation allows for a richer understanding of how social interactions distort the effects of targeted policies in networks.

---

such as Hoxby (2000), propose using a panel to difference-out these fixed effects between time periods. However, this approach requires both access to panel data and the assumption that contextual effects are invariant over time.

[3] Previously, estimation has relied on double asymptotic arguments, requiring both the number and size of the networks to tend to infinity; e.g., see Lee et al. (2014), Guerra and Mohnen (2022), and Bhattacharya et al. (2023).

2. MODEL WITH GENERALIZED INTERACTION EFFECTS

I study a binary choice model with interaction effects that vary based on individual identity. This model generalizes the framework of Brock and Durlauf (2001) by allowing agents to be influenced affected by different members of their network. For example, an agent might conform to some types of individuals while deviating from others, and the strength of these influences may also vary. Unlike previous work, my analysis is nonparametric with respect to the distribution of idiosyncratic preferences. Therefore, the equilibrium properties and subsequent identification results are robust to a wider range of functional form assumptions.

2.1. *Agent Preferences*

Consider a network of agents, indexed by i, where each agent belongs to one of K types. These types can represent social identities, such as gender, class, race, or partisanship, or they can represent organizational roles within the network, such as workers and supervisors, suppliers and manufacturers, patients and clinicians, or villagers and community leaders.

Each agent chooses a binary action $\omega_i \in \{0, 1\}$ at a common time. When making choices, agents are influenced by their beliefs about the actions of others in the network. Let $\bar{\omega}^k$ be the average action among agents of type k, and let $\bar{\omega}^k_{-i}$ be the average action among agents of type k excluding i. The utility of choosing an action ω_i for any agent i of type k is:

$$U_i(\omega_i|k) = v_k(\omega_i) + J_{kk}\omega_i \operatorname{E}_i(\bar{\omega}^k_{-i}) + \sum_{\ell \neq k} J_{k\ell}\omega_i \operatorname{E}_i(\bar{\omega}^\ell) + \epsilon_i(\omega_i). \tag{1}$$

In this utility function, $v_k(\omega_i)$ is a deterministic private payoff, which could depend on the agent's type, and $\epsilon_i(\omega_i)$ is an idiosyncratic payoff that varies among individuals. The terms $\operatorname{E}_i(\bar{\omega}^k_{-i})$ and $\operatorname{E}_i(\bar{\omega}^\ell)$ are agent i's subjective beliefs about $\bar{\omega}^k_{-i}$ and $\bar{\omega}^\ell$, respectively. Under this framework, utility exhibits proportional spillovers, such that there is a multiplicative interaction between the agent's choice and the expected average choice within each type.[4]

The parameter $J_{k\ell}$ represents the interaction effect, specifying how agents of type k are influenced by agents of type ℓ. These effects can be collected in a matrix $\mathbf{J} \in \mathbb{R}^{K \times K}$, where:

$$\mathbf{J} = \begin{bmatrix} J_{11} & J_{12} & \cdots & J_{1K} \\ J_{21} & J_{22} & \cdots & J_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ J_{K1} & J_{K2} & \cdots & J_{KK} \end{bmatrix}. \tag{2}$$

[4] Another common way to model spillovers is through a quadratic conformity effect, $-\frac{1}{2}J_{k\ell}[\omega_i - \operatorname{E}_i(\bar{\omega}^\ell)]^2$, as studied by Bernheim (1994). This parameterization is equivalent in that it leads agents to make the same choices.

When analyzing model behavior, I will refer to $\mathbf{J}$ as the *interaction matrix*. The entries of this matrix characterize the direction, sign, and intensity of social influence across types.[5]

With binary outcomes, the private payoffs can be written as linear functions of the choice: $v_k(\omega_i) = h_k\omega_i + \eta_k$ and $\epsilon_i(\omega_i) = \varepsilon_i\omega_i + \xi_i$, where h_k represents the deterministic private preference for choosing $\omega_i = 1$, and ε_i is the idiosyncratic taste for this action. These tastes are assumed independent across agents and their distributions can differ among agent types:

$$\mathrm{P}(\varepsilon_i \leq z|k) = F_{\varepsilon|k}(z), \quad \text{for } k = 1, \ldots, K, \tag{3}$$

where $F_{\varepsilon|k}$ is strictly increasing, real analytic, and symmetric about zero. No further parametric assumptions are imposed throughout the paper. Therefore, the framework can accommodate a wide range of empirical specifications, including logistic and Gaussian errors.

2.2. *Economic Examples of Generalized Interactions*

The classical model of social interactions for binary outcomes (Brock and Durlauf, 2001) imposes two key restrictions on the interaction effects. First, it assumes uniformity ($J_{k\ell} = J$ for all k and ℓ) so that all agents influence each other in the same way, regardless of type. Second, it assumes all interaction effects are positive ($J_{k\ell} \geq 0$ for all k and ℓ) so that each agent's payoff rises when others take the same action. This rules out strategic substitutability where agents differentiate from, rather than conform to, some members of their network.

I relax the first restriction by allowing interaction effects to differ among types and also by not imposing symmetry ($J_{k\ell} = J_{\ell k}$), so the influence of type k on type ℓ can differ from that of type ℓ on type k. This generalization makes the model well-suited to study a variety of economic decision problems across different network structures. Table 1, Panel A, gives examples of four commonly studied networks in economics, along with the conditions on $\{J_{k\ell}\}_{k,\ell}$ that define each one. The first is a *tree*, used to study spillovers in organizational hierarchies and supply chains, as well as the spread of information or disease. The second is a *circle*, relevant for analyzing local interactions among close contacts (Ellison, 1993) and spatial competition between nearby firms (Salop, 1979). The third is a *bipartite* network, used to study two-sided markets such as among workers and firms or buyers and sellers. The fourth is a *complete* network where all agents interact, as is assumed in classical models of peer effects. Each of these networks is nested under the generalized interactions framework.

I relax the second restriction by allowing the model to handle any combination of positive or negative interaction effects. This extension makes the model applicable to environments

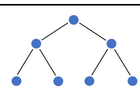
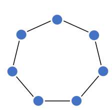
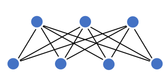
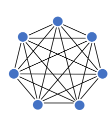
[5] $\mathbf{J}$ can be interpreted as the adjacency matrix of a directed graph, where nodes represent different agent types.

where agents seek to differentiate their behavior from others. Examples, shown in Table 1, Panel B, include *free-riding* (Bramoullé and Kranton, 2007), where agents contribute less to a public good when others contribute more; *social distinction* (Akerlof and Kranton, 2000), where agents differentiate their choices to reinforce identity; *segregation* (Schelling, 1971), where agents prefer actions that result in the exclusion of particular peers; and *polarization* (Boxell et al., 2022), where agents deviate from actions taken by those with opposing views. In each case, an agent's preference to distinguish her choice from others can be modeled as a negative interaction effect. Thus, by allowing for a mix of positive and negative interaction effects, the generalized framework can accommodate both conformity and differentiation.

Relative to the classical model, the generalized interactions framework admits a new set of target parameters that offer insight into how social interactions differ among individuals. For example, the difference $J_{k_1\ell} - J_{k_2\ell}$ indicates whether agents of type k_1 are more or less likely than agents of type k_2 to conform to type ℓ. These terms can be used to construct an economic measure of polarization, $\delta_{k_1k_2} = J_{k_1k_1} + J_{k_2k_2} - J_{k_1k_2} - J_{k_2k_1}$, which quantifies how much more agents of types k_1 and k_2 conform to their own type than to the other type.

TABLE I

COMMONLY STUDIED ECONOMIC NETWORKS AND INTERACTION EFFECTS

	Illustration ($K = 7$)	Structural Representation	Economic Examples
Panel A: Network Structures			
Tree		$\sum_{k\neq\ell} \mathbb{1}\{J_{k\ell} \neq 0\} = K - 1$ and $\mathbf{J}$ irreducible	hierarchy, vertical production, contagion, info diffusion
Circle		$J_{k\ell} \neq 0$ if $\ell = k \pm 1 \pmod{K}$	local interactions, spatial competition
Bipartite Network		$J_{k\ell} \neq 0$ if $k \in A$ and $\ell \in B$ for disjoint $A, B \in \{1, \dots, K\}$	labor market, product market, college admission/matching
Complete Network		$J_{k\ell} \neq 0$ for all k, ℓ	classroom peer effects, public goods settings, oligopoly
Panel B: Modes of Social Influence			
Strategic Complementarity	$\frac{\partial[U_i(1\|k) - U_i(0\|k)]}{\partial\, \mathrm{E}_i(\bar{\omega}^\ell)} \geq 0$	$J_{k\ell} \geq 0$	conformity/imitation, social learning, technology adoption
Strategic Substitutability	$\frac{\partial[U_i(1\|k) - U_i(0\|k)]}{\partial\, \mathrm{E}_i(\bar{\omega}^\ell)} \leq 0$	$J_{k\ell} \leq 0$	free-riding, social distinction, segregation, polarization

Notes. This table lists examples of networks and spillovers encompassed by the generalized interactions model.

2.3. *Equilibrium under Noncooperative Decisionmaking*

My analysis focuses on pure strategy Bayesian Nash equilibria where agents act noncooperatively—that is, without any coordination. Agents of type k choose $\omega_i = 1$ with probability:

$$\mathrm{P}(\omega_i = 1|k) = F_{\varepsilon|k}\Big(h_k + J_{kk}\,\mathrm{E}_i(\bar{\omega}^k_{-i}) + \sum_{\ell \neq k} J_{k\ell}\,\mathrm{E}_i(\bar{\omega}^\ell)\Big). \tag{4}$$

Since ω_i takes values in the set $\{0, 1\}$, the expected action $\mathrm{E}(\omega_i|k)$ also equals $\mathrm{P}(\omega_i = 1|k)$.

Agents act with incomplete information. Specifically, an agent i knows her own idiosyncratic taste ε_i and the deterministic preference parameters (h_k, $\{J_{k\ell}\}_\ell$, $F_{\varepsilon|k}$) for each type, but not the idiosyncratic tastes ε_j of other agents $j \neq i$. This information structure can be interpreted in two ways. First, it may be that agents do not see everyone else's choice, either because actions are taken simultaneously, such as during a test, or because the network is too large to track individual actions, such as in a school or village. Under this interpretation, ε_i reflects the uncertainty that agents have about others' choices. A second interpretation of the information structure is that agents respond to social norms—the behavior of a "typical" agent of a given type—while ignoring idiosyncratic deviations from those norms. This interpretation is consistent with economic theories of social influence mechanisms (Akerlof and Kranton, 2000), where norms are understood to reflect shared values of a social group.

An equilibrium is characterized by the expected average choices that are consistent with agents' optimal decisions. I assume agents have consistent beliefs, so that they all correctly infer each others' actions in expectation: $\mathrm{E}_i(\omega_j|k) = \mathrm{E}(\omega_j|k)$ for all i, j, and k. By symmetry of the choice probabilities (4), it follows that $\mathrm{E}(\omega_i|k) = \mathrm{E}(\omega_j|k)$ for all i, j, and k. Letting m^k denote $\mathrm{E}(\bar{\omega}^k)$, an equilibrium $m = (m^1, \ldots, m^K)$ is defined as the solution to:

$$m^k = F_{\varepsilon|k}\Big(h_k + \sum_{\ell=1}^{K} J_{k\ell} m^\ell\Big), \quad \text{for } k = 1, \ldots, K. \tag{5}$$

Since $F_{\varepsilon|k}$ is continuous, Brouwer's fixed point theorem ensures that an equilibrium exists.

2.4. *Equilibrium Uniqueness and Local Stability*

I now examine two key equilibrium properties of the model: uniqueness and local stability. Uniqueness asserts that, for a fixed set of parameters, there is one solution to the equilibrium system. Local stability concerns whether iteration on agents' best responses converges to an equilibrium. Formally, m is *locally stable* if it is a limiting solution to the dynamical system $m^k_t = F_{\varepsilon|k}(h_k + \sum_{\ell=1}^{K} J_{k\ell} m^\ell_{t-1})$ for $k = 1, \ldots, K$, where the starting value m_0 lies in some sufficiently small neighborhood of m. Meanwhile, m is *unstable* if there is a neighborhood of m where, for any starting value m_0, an eventual iterate m_t lies outside the neighborhood.

Both properties are fundamental to comparative statics. If a locally stable equilibrium is perturbed slightly, then behavioral outcomes would return to that equilibrium. Meanwhile, if an equilibrium is unstable, then nearby outcomes would diverge from it. In settings with multiple equilibria, a locally stable equilibrium is one that a researcher observes in practice, while an unstable equilibrium acts as a tipping point between stable equilibrium outcomes.

2.4.1. *Networks with a Unique Locally Stable Equilibrium*

I begin by establishing conditions for the equilibrium to be unique and locally stable. Under the generalized interactions framework, these conditions depend on both the structure of the network and the magnitude of the social interaction effects relative to private preferences. Specifically, I find that a unique and locally stable equilibrium will exist if network ties are sparse—limiting the potential for individual behavior to become self-reinforcing—or if the interaction effects are weak—so that private preferences tend to dominate an agent's utility.

To obtain these conditions, I first derive the following property, which demonstrates that the equilibrium is unique and locally stable if and only if there exist no unstable equilibria.

PROPERTY 1: *For almost all distributions $\{F_{\varepsilon|k}\}_k$, the number of equilibria is finite and odd. If there are d_s locally stable equilibria, there are at least $d_s - 1$ unstable equilibria.*

Building on this property, I now characterize what kinds of interaction effects rule out any unstable equilibria. To do so, I analyze the *Jacobian matrix* of the equilibrium system (5):

$$\mathbf{D}(m) = \boldsymbol{\beta}(m)\mathbf{J}, \tag{6}$$

where $\boldsymbol{\beta}(m)$ is a diagonal matrix whose kth entry is $f_{\varepsilon|k}\big(h_k + \sum_{\ell=1}^{K} J_{k\ell} m^{\ell}\big)$, representing the expected share of type k agents near the decision threshold at equilibrium m. Hence, the Jacobian equals a positively weighted interaction matrix $\mathbf{J}$, where each row k is scaled by the likelihood that agents of type k are close to being indifferent between the two actions.

Equilibrium stability is governed by the *spectral radius* of the Jacobian matrix, defined as the largest of its eigenvalues in absolute value: $\rho(\mathbf{D}(m)) = \max\{|\lambda| : \lambda \text{ eigenvalue of } \mathbf{D}(m)\}$. This parameter measures how expansive the Jacobian matrix is as a linear operator, thereby quantifying the total intensity of spillovers in the network. When $\rho(\mathbf{D}(m))$ is less than one, the equilibrium m is locally stable; otherwise, if it exceeds one, the equilibrium is unstable.[6]

LEMMA 1: *If* $\rho(\mathbf{D}(m)) < 1$*, then* m *is locally stable; if* $\rho(\mathbf{D}(m)) > 1$*, then* m *is unstable.*

Together, these results imply that a unique, locally stable equilibrium exists if network distortions are contained, such that $\rho(\mathbf{D}(m)) < 1$ at any equilibrium m. I list two special cases.

Example 1 (One-Way Interactions). Suppose social influence flows only in one direction, so that all feedback effects—given by products such as $J_{kj_1}J_{j_1j_2}\cdots J_{j_Mk}$—are zero. In this setting, there is no sequence of interactions through which an agent can be influenced, even indirectly, by her own type. As a result, individual behavior cannot become self-reinforcing. It follows that $\rho(\mathbf{D}(m)) = 0$, which implies that the equilibrium is unique and locally stable.

Example 2 (Weak Interactions). Suppose the social interaction effects are weak relative to private payoffs. Specifically, for each type k, let $\sum_{\ell=1}^{K} |J_{k\ell}| < S_k$, where S_k is a function of $\{h_k\}_k$ that reduces to $S_k = 1/f_{\varepsilon|k}(0)$ if $h_k = 0$ for all k. In this case, $\rho(\mathbf{D}(m)) < 1$ for any solution m to system (5). Therefore, the equilibrium must be unique and locally stable.

2.4.2. *Networks with Multiple Locally Stable Equilibria*

I now analyze when the model has multiple locally stable equilibria. This question is well-studied in the classical setting where all interaction effects are positive ($J_{k\ell} \geq 0$ for all k and ℓ). In this case, whenever the social interactions are strong enough to generate an unstable equilibrium, the model will have multiple equilibria, with at least two being locally stable.[7]

In general, this property does not extend to cases with negative interactions. In fact, some models with negative interaction effects have a unique but unstable equilibrium, meaning that agents would never converge on a single set of choices. To illustrate this phenomenon, suppose $J_{kk} < 0$ so that agents of type k seek to differentiate themselves from members of their own type. As $-J_{kk}$ grows large, all equilibria become unstable, because when most of type k selects the high action, they would rather switch to the low action, and vice versa.[8] A similar sort of instability can also result from specific patterns of between-type interactions.

[6]The setting $\rho(\mathbf{D}(m)) = 1$ represents a knife-edge case that occurs with probability zero and has no particular economic relevance in the model. In such cases, equilibrium stability depends on higher-order derivatives of (5).

[7]The microeconomic foundations for models of complementarity are largely developed by Donald Topkis (see Topkis (1998) for a summary), Vives (1990), and Milgrom and Roberts (1990). This work uses key results from lattice theory, e.g., Tarski's fixed point theorem, to study equilibrium behavior under uniform complementarities.

[8]In a simplified model where $h_k = 0$ and $J_{k\ell} = 0$ for $\ell \neq k$, this global instability occurs if $-J_{kk} > 1/f_{\varepsilon|k}(0)$. As another example, if $\mathbf{J}$ is symmetric and its eigenvalues all have non-positive real parts, then there is just one equilibrium, which is locally stable if $\rho(\mathbf{D}(m)) < 1$ and unstable if $\rho(\mathbf{D}(m)) > 1$. See the Appendix for details.

For example, if $J_{k\ell} > 0$ and $J_{\ell k} < 0$, then agents of type k seek to conform to type ℓ, while agents of type ℓ seek to differentiate from type k.[9] If these effects are strong enough, they can once again generate global instability, where each type's response undermines the other.

The contribution of my analysis is to pinpoint the source of global instability in models with negative interactions, and use this insight to classify which models can sustain locally stable equilibria, even with strong spillovers. The examples above suggest that global instability occurs when agents seek to deviate from their own type. In this case, agents never settle on a choice as they are always dissatisfied with the outcome. So, to rule out global instability, it is critical that social interactions do not lead to self-contradictory preferences. This means that in aggregate—that is, accounting for all feedback effects $J_{kj_1}J_{j_1j_2}\cdots J_{j_Mk}$—no agent should be repelled by her own type. The condition below formalizes this requirement.

ASSUMPTION A: There exists an invertible matrix $\mathbf{B}$ such that $\mathbf{BJB}^{-1} \geq 0$ entrywise.

Assumption A states that $\mathbf{J}$ can be transformed, under a change of basis, into a non-negative matrix. Through this transformation, I can reinterpret the underlying network as one where all interaction effects are positive, that is, where all agents conform. Crucially, this reinterpretation preserves the stability of equilibria, as determined by the spectral properties of $\mathbf{J}$. Leveraging this relationship, I can extend the stability properties of models with uniformly positive interactions to those that allow for negative interactions. This leads the next result.

PROPERTY 2: *Suppose Assumption A holds. Then, for almost all distributions $\{F_{\varepsilon|k}\}_k$, there is exactly one more locally stable equilibrium than there are unstable equilibria.*

This property central to my analysis. It guarantees that under Assumption A, there is always a locally stable equilibrium, and that multiple locally stable equilibria will exist if and only if the network distortions are strong enough to generate an unstable equilibrium. Note that Assumption A encompasses a wide range of settings that may be relevant to policymakers. I illustrate two special cases below, and I also provide additional examples in the Appendix.

Example 1 (No Negative Feedback). Suppose all feedback effects are non-negative, such that $J_{kj_1}J_{j_1j_2}\cdots J_{j_Mk} \geq 0$ for any types $k, j_1, j_2, \ldots, j_M$.[10] This restriction applies to many commonly studied economic networks. For a tree, it applies if the interaction effects among any two types k and ℓ (weakly) share the same sign: $J_{k\ell}, J_{\ell k} \geq 0$ or $J_{k\ell}, J_{\ell,k} \leq 0$. For a circle, it applies if the number of interactions with $J_{k\ell} < 0$ is even. For a bipartite network,

[9]This setting is strategically similar to a matching pennies game, which has no pure strategy Nash equilibrium.

[10]This condition represents a special case of Assumption A in which the change-of-basis matrix $\mathbf{B}$ is diagonal. Indeed, as I prove in the Appendix, it applies if and only if $\mathbf{BJB}^{-1}$ is non-negative for some diagonal matrix $\mathbf{B}$.

it applies if all interaction effects are either non-negative or non-positive. For a complete network, it applies if each indirect effect $J_{km}J_{m\ell}$ has the same sign as the direct effect $J_{k\ell}$. In this way, it requires that complementarity and substitutability are transitive across types.

Example 2 (Own-Type Conformity). Suppose $\mathbf{J}$ is *diagonally dominant*, meaning that $J_{kk} \geq \sum_{\ell \neq k} |J_{k\ell}|$ for all types k. In this case, an agent's preference to conform to her own type is stronger than the combined influences from other types. As this example indicates, Assumption A does not require ruling out all negative feedback; rather, it suffices that the accumulation of feedback effects does not cause agents to be repelled by their own types.

2.5. *Comparisons of Social Welfare across Types*

I conclude this section by characterizing agent welfare across different equilibria. Note that, in this model, there is no strict Pareto ranking over equilibria because extreme realizations of the idiosyncratic taste ε_i can dominate an agent's utility. Therefore, to measure welfare, I analyze the expected utility over ε_i for each agent type k evaluated at any equilibrium m.

$$\mathcal{W}_k(m) = \mathrm{E}\left(\max_{\omega_i}\left\{h_k\omega_i + \eta_k + \sum_{\ell=1}^{K} J_{k\ell}\omega_i m^\ell + \varepsilon_i\omega_i + \xi_i\right\}\right). \tag{7}$$

To facilitate welfare comparisons, I rescale choices so that $\omega_i \in \{-1, 1\}$. This change does not impact behavioral outcomes, but it implies that, without private incentives ($h_k = \varepsilon_i = 0$ for all i, k), individual utility is the same whether everyone selects the high action or everyone selects the low action. Therefore, this rescaling ensures there is no negative externality inherent to an equilibrium when all agents are fully ambivalent between the two choices.[11]

The next property characterizes how social welfare outcomes relate to specific patterns of individual behavior. It shows that, for each type, social welfare is highest at the equilibrium where agents are most unified in their choices. Also, if agents are privately biased toward a specific action, then social welfare tends to be highest when most agents choose that action.

PROPERTY 3: *Let $\mathcal{M}$ denote the set of equilibria, and define $m^* = \operatorname{argmax}_{m\in\mathcal{M}} \mathcal{W}_k(m)$. Then $m^* = \operatorname{argmax}_{m\in\mathcal{M}} |\mathrm{E}(\bar{\omega}^k)|$. Moreover, there always exists a threshold $T_k \in \mathbb{R}$ where:*

$$m^* = \begin{cases} \operatorname{argmax}_{m\in\mathcal{M}} \mathrm{E}(\bar{\omega}^k), & \text{if } h_k > T_k \\ \operatorname{argmin}_{m\in\mathcal{M}} \mathrm{E}(\bar{\omega}^k), & \text{if } h_k < T_k. \end{cases}$$

[11] Welfare analysis in network-based models is especially sensitive to how utility is specified, even if different specifications yield the same choices. The reason is that agents' payoffs depend on underlying drivers of spillover effects, such as conformity versus social learning, not just the effects themselves; see Bhattacharya et al. (2023).

In the presence of multiple equilibria, a key consideration for policy is how social welfare rankings differ for different types of agents. Namely, is there a single favorable equilibrium that maximizes welfare for all types, or are there tradeoffs such that improving welfare for one type would always reduce it for another? This question matters for whether policy can resolve coordination failures by shifting an economy toward a more favorable equilibrium.

In the classical model where all the interaction effects are positive, the set of equilibria is well-ordered, forming a complete lattice. From a welfare perspective, this property implies that there is always one equilibrium that maximizes social welfare for all types, provided all types are privately biased toward the same action.[12] As I demonstrate in the next property, this relationship does not generally apply to environments with negative interaction effects.

PROPERTY 4: *Suppose all feedback effects are non-negative. Then for any types k and ℓ:*

(i) If $J_{k\ell} > 0$, then $\mathrm{E}(\bar{\omega}^k)$ and $\mathrm{E}(\bar{\omega}^\ell)$ are maximized (minimized) at the same equilibrium.

(ii) If $J_{k\ell} < 0$, then $\mathrm{E}(\bar{\omega}^k)$ is maximized at the same equilibrium where $\mathrm{E}(\bar{\omega}^\ell)$ is minimized.

This property shows that negative interactions can introduce tradeoffs in improving welfare. For example, consider two types of agents, k and ℓ, each deciding whether to take a medical treatment, such as a vaccine. Suppose both types, in isolation, would benefit from treatment ($h_k, h_\ell > 0$), but each one prefers to differentiate from the other type ($J_{k\ell}, J_{\ell k} < 0$). If these interaction effects are strong, then multiple equilibria can exist: in one, a majority of type k is treated; in another, a majority of type ℓ is treated. Each type prefers the equilibrium where a majority of its members are treated, but both cannot achieve this outcome simultaneously.

In this setting, no equilibrium outcome is broadly favorable. Therefore, it is less obvious when a policy intervention can achieve net welfare gain. Any attempt to increase treatment uptake for one type of agent imposes a negative externality by reducing uptake for the other.

3. IDENTIFICATION AND ESTIMATION OF ENDOGENOUS INTERACTION EFFECTS

I now show how to recover features of the interaction matrix $\mathbf{J}$ using individual choice data. My analysis follows two key steps. First, I prove identification when the expected choices $\mathrm{E}(\omega_i|k)$ are observed. This allows me to focus on separating endogenous interactions from unobserved network effects. Second, I extend my analysis to the case where $\mathrm{E}(\omega_i|k)$ is unobserved, but can be approximated by the average choice $\bar{\omega}^k$ in a finite network. I show how to construct internal instruments to correct for the resulting measurement error bias. I then derive a linear IV estimator for the interaction effects, which performs well in simulations.

[12] Specifically, if $h_k > 0$ for all k, then $\mathrm{argmax}_{m\in\mathcal{M}}\, \mathcal{W}_k(m) = \mathrm{argmax}_{m\in\mathcal{M}}\, \mathcal{W}_\ell(m)$ for any types k and ℓ.

3.1. *Econometric Specification*

Consider a setting with many networks $n = 1, \ldots, N$, where agents interact only with others in their network. In practice, these networks could be schools, workplaces, neighborhoods, or villages. Suppose that the researcher has data on individual choices ω_i, either for the entire population or for a random sample of agents drawn from each network. In addition, the researcher knows the network n to which each agent belongs, as well as each agent's type k.

To ensure broad applicability, I allow each agent's private utility to vary across networks by replacing the deterministic private preference term h_k with $h_k + \alpha_n$. Therefore, for any agent i of type k in network n, private utility depends on: (1) individual-specific factors ε_i, (2) type-specific factors h_k, and (3) contextual network effects α_n. Crucially, each of these factors—ε_i, h_k, and α_n—is unobserved by the researcher. An agent i's action is given by:

$$\omega_i = \mathbb{1}\left\{ h_k + \alpha_n + J_{kk}\,\mathrm{E}_i(\bar{\omega}^k_{n,-i}) + \sum_{\ell \neq k} J_{k\ell}\,\mathrm{E}_i(\bar{\omega}^\ell_n) + \varepsilon_i \geq 0 \right\}, \tag{8}$$

where h_k, α_n, ε_i, and $\{J_{k\ell}\}_{k,\ell}$ are unknown coefficients. The model can also be extended to include observed covariates or prices at the agent- or network-level; however, as these do not meaningfully change my results, I omit them for now and give details in the Appendix.

I maintain two assumptions about the error structure. First, I assume the random payoffs ε_i and ε_j are independent for any agents i and j within and across networks. Hence, there is no covariation in the error terms. Second, I assume that, given an agent's type k, the error term ε_i is independent of α_n. This condition rules out self-selection into networks based on unobserved idiosyncratic payoffs.[13] Both these assumptions are standard in the literature on social interactions, and they can also be relaxed to allow for selection on observables.[14]

ASSUMPTION B.1: (i) $\{\varepsilon_i\}_i$ are pairwise independent; (ii) $\mathrm{P}(\varepsilon_i \leq z | k, \alpha_n) = F_{\varepsilon|k}(z)$.

As before, agents act with incomplete information and consistent beliefs. In equilibrium:

$$m^k_n = F_{\varepsilon|k}\left(h_k + \alpha_n + \sum_{\ell=1}^{K} J_{k\ell} m^\ell_n \right), \quad \text{for } k = 1, \ldots, K \text{ and } n = 1, \ldots, N. \tag{9}$$

[13] This restriction is stronger than needed for identification. As shown by Manski (1988) and also discussed by Horowitz (2009), full conditional independence can be replaced by a weaker quantile independence assumption.

[14] Let $\mathrm{P}(\varepsilon_i \leq z | W_n, \alpha_n, k) = F_{\varepsilon|W_n,k}(z)$ for some observable network-level variable W_n. Even if $F_{\varepsilon|W_n,k}$ does not equal $F_{\varepsilon|k}$, all identification arguments still follow by comparing networks that are observably similar.

As discussed in Section 2, multiple equilibria can exist if network spillovers are sufficiently strong. In these cases, I assume that agents know which equilibrium is realized, so there is no coordination in selecting an equilibrium.[15] Note that nonuniqueness will not interfere with identification since my analysis does not require directly solving the equilibrium system (9). Indeed, as shown in Property 1, the model has only finitely-many equilibria, and the preference parameters $(\{h_k\}_k, \{\alpha_n\}_n, \mathbf{J}, \{F_{\varepsilon|k}\}_k)$ are invariant across them all. Therefore, the model remains identified as long as there is only one set of preference parameters that rationalizes the data under the equilibrium conditions—even if that data comes from one of several possible equilibria. In the analysis that follows, I prove that the preference parameters are identified and may be consistently estimated under any realized equilibrium.[16]

In most cases, it is also reasonable to assume the equilibrium is locally stable (see Section 2.4 for conditions); otherwise, it is unlikely to be observed in practice. However, while this matters for interpretation, my identification analysis does not rely on equilibrium stability.

3.2. *Identification with Known Expected Average Choices*

Suppose researchers can observe the expected average choices m_n^k. Then, the main barrier to identification is the unobserved network effect α_n, which obscures the role of social interactions on network-level outcomes. This issue is noted by Brock and Durlauf (2007), Prop. 2, who show that network unobservables prevent point identification of interaction effects.

I propose a partial identification strategy to recover key features of $\mathbf{J}$ without placing any added assumptions on α_n. This method involves differencing-out α_n among different types of agents in the same network. The intuition behind this strategy is captured in the following lemma. I give the proof alongside the result so my approach can be more clearly interpreted.

LEMMA 2—Sufficiency Claim: $\mathrm{E}(\omega_i|k, \alpha_n, \{m_n^\ell\}_\ell) = \mathrm{E}(\omega_i|k, \{m_n^\ell\}_\ell)$ *for all* n *and* k.

Proof. Fix any $\tilde{k} \neq k$. Since $F_{\varepsilon|\tilde{k}}$ is strictly increasing, equation (9) can be inverted so that:

$$\alpha_n = F_{\varepsilon|\tilde{k}}^{-1}(m_n^{\tilde{k}}) - h_{\tilde{k}} - \sum_{\ell=1}^{K} J_{\tilde{k}\ell} m_n^\ell. \tag{10}$$

Substituting this expression into the definition of $\mathrm{E}(\omega_i|k, \alpha_n, \{m_n^\ell\}_\ell)$ yields the equation:

[15] This assumption can be rationalized by noting that, whenever the realized equilibrium is locally stable, agents can compute it independently through fixed-point iteration on (9), as long as they begin with a common prior m_0.

[16] This contrasts with two-step approaches used in the empirical industrial organizational literature, where identification relies on computing equilibrium conditions. See Bhattacharya et al. (2023), Section 5.5, for discussion.

$$\mathrm{E}(\omega_i|k, \alpha_n, \{m_n^\ell\}_\ell) = F_{\varepsilon|k}\left(h_k - h_{\tilde{k}} + \sum_{\ell=1}^{K}(J_{k\ell} - J_{\tilde{k}\ell})m_n^\ell + F_{\varepsilon|\tilde{k}}^{-1}(m_n^{\tilde{k}})\right). \tag{11}$$

The network effect cancels, so $\mathrm{E}(\omega_i|k, \alpha_n, \{m_n^\ell\}_\ell)$ becomes a constant function of $\{m_n^\ell\}_\ell$.
Q.E.D.

This lemma shows how observable differences among agents in the same network may be used to control for contextual effects. Consider two agents i and j of different types, k_1 and k_2, who both reside in network n. Since they share the same network, all contextual factors affecting i also apply to j. Any difference in i and j's choices must be due to differences in type-specific preferences, $(h_{k_1}, \{J_{k_1\ell}\}_\ell)$ and $(h_{k_2}, \{J_{k_2\ell}\}_\ell)$ or idiosyncratic tastes, ε_i and ε_j.

This framework offers a natural panel structure, allowing me to control for network-level factors by comparing the outcomes of different agents in the same network. This approach is conceptually similar to the method of differencing-out incidental parameters in panel data models; see, for example, Chamberlain (1980).[17] By removing α_n, I achieve identification of the relative type-specific parameters, $h_{k_1} - h_{k_2}$ and $\{J_{k_1\ell} - J_{k_2\ell}\}_\ell$, which provide insight into how different types of agents interact. At the same time, I also lose information about the absolute levels of h_k and $J_{k\ell}$, as well as the network effects themselves, which, as discussed by Chamberlain (1984), limits the range of probability statements that one can make.

I present two versions of my identification result. First, I give conditions for semiparametric identification, where the error distributions $F_{\varepsilon|k}$ are known to the researcher. Next, I give conditions for nonparametric identification, where $F_{\varepsilon|k}$ are unknown. While the nonparametric version allows for additional model flexibility, it also requires substantially more variation in the data. In both versions, identification is achieved without α_n being observed.

Conditions for Semiparametric Identification

Before stating the semiparametric identification result, I present the following assumption.

ASSUMPTION B.2: The contextual network effect α_n is continuously distributed on $\mathbb{R}$.

This assumption, based on conditions X.1 in Manski (1988) and A.4 in Brock and Durlauf (2007), ensures there is enough variation across n to reveal the nonlinear relationship between m_n^k and α_n. This nonlinearity is a core feature of discrete choice models, and it marks

[17] Bonhomme (2012) proposes a systematic approach to obtain moment restrictions that are free of fixed effects via functional differencing. Additionally, see Bonhomme and Dano (2023) and Dano (2025) for recent advances.

a key distinction from linear simultaneous equations models. In the linear case, identification is impeded by the *reflection problem* (Manski, 1993), which enters when m_n^k depends linearly on α_n. In this case, even if α_n were a constant function of observables, it would be impossible to isolate the role of social interactions without additional exclusion restrictions.

By contrast, the reflection problem does not enter in discrete choice settings. As shown by Brock and Durlauf (2001, 2007), discrete choice inherently implies a nonlinear relationship between m_n^k and α_n, which acts as a built-in exclusion restriction. This nonlinearity allows me to distinguish social interactions from network effects, provided there is enough variation in α_n. Note that, while the unbounded support assumption is convenient for delivering this variation, it is much stronger than necessary. It suffices that α_n has a support large enough to rule out collinearity in m_n^k; see Horowitz (2009), Section 4.2, for details.[18]

THEOREM 1: *Suppose Assumptions B.1 and B.2 hold, and assume m_n^k is observed for all networks n and agent types k. If the distribution functions $\{F_{\varepsilon|k}\}_k$ are known, then:*

(i) Without further assumptions, $\{h_{k_1} - h_{k_2}\}_{k_1,k_2}$ and $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$ are identified.

(ii) If $\alpha_n = W_n' d$ for some observed vector W_n, then d, $\{h_k\}_k$, and $\{J_{k\ell}\}_{k,\ell}$ are identified.

This theorem has two parts. Part *(i)* draws on Lemma 2 to show that the model is partially identified even if α_n is unknown. In this case, I can recover the differences in type-specific preferences and interaction effects, $h_{k_1} - h_{k_2}$ and $\{J_{k_1\ell} - J_{k_2\ell}\}_\ell$, for any two types k_1 and k_2.[19] This finding is new to the literature and is the theorem's main contribution. Part *(ii)* considers a special case —studied by Brock and Durlauf (2001, 2007) and others—where α_n is a constant function of observables. In this case, all the parameters are point identified.

Conditions for Nonparametric Identification

For nonparametric identification, I require an extra assumption: there must be an exogenous individual-level covariate that varies continuously on an unbounded support. This variation allows me to trace out each of the error distributions $F_{\varepsilon|k}$, which can then be used to recover the rest of the model parameters. This approach closely follows Brock and Durlauf (2007).

To set ideas, I first modify the choice equation to include exogenous covariates $X_i \in \mathbb{R}^r$. Let $\omega_i = \mathbb{1}\{X_i'c + h_k + \alpha_n + \sum_{\ell=1}^K J_{k\ell} m_n^\ell + \varepsilon_i \geq 0\}$, with $\mathrm{P}(\varepsilon_i \leq z | X_i, k, \alpha_n) = F_{\varepsilon|k}(z)$ and $\mathrm{P}(X_i \leq x | k, \alpha_n) = \mathrm{P}(X_i \leq x | k)$. Suppose that X_i satisfies the following assumption.

[18] Reflection can also be avoided in linear models with intransitive triads, where not all agents interact. As shown by Bramoullé et al. (2009), these missing links offer exclusion restrictions that can be exploited for identification.

[19] If $F_{\varepsilon|k}$ are known exactly (not up-to-scale), then $J_{k\ell}$ are identified up-to-scale by normalizing the own-type effects J_{kk} to one. Similarly, private payoffs h_k are identified relative to some reference type k^*, where $h_{k^*} = 0$.

ASSUMPTION B.3: For any k, $\text{supp}(X|k)$ is not contained in a proper linear subspace of $\mathbb{R}^r$; also, there is a component x_j of X—with a nonzero coefficient c_j—such that, for almost all $x_{-j|k}$, the distribution of $x_{j|k}$ given $x_{-j|k}$ has positive density everywhere on $\mathbb{R}$.

This assumption, based on Manski (1988), guarantees there is enough variation in at least one individual-level covariate to recover the error distributions $F_{\varepsilon|k}$ up-to-scale, even when these distributions are fully unknown. Using this assumption, I arrive at the following result.

THEOREM 2: *Suppose Assumptions B.1, B.2, and B.3 hold, and assume m_n^k is observed for all networks n and agent types k. Then $\{F_{\varepsilon|k}\}_k$ and c are identified up-to-scale. Also:*

(i) Assuming nothing else, $(\{h_{k_1} - h_{k_2}\}_{k_1,k_2}, \{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell})$ is identified up-to-scale.

(ii) If $\alpha_n = W_n' d$ for some observed W_n, then $(d, \{h_k\}_k, \{J_{k\ell}\}_{k,\ell})$ is identified up-to-scale.

3.3. *Identification with Unknown Expected Average Choices*

In practice, agents' expectations m_n^k are never directly observed. Instead, a researcher only sees the average choices $\bar{\omega}_n^k$ among agents in finite networks. These averages serve as noisy approximations of the true expectations, $\bar{\omega}_n^k = m_n^k + u_n^k$, where u_n^k is the measurement error.

If the researcher substitutes m_n^k with $\bar{\omega}_n^k$ without accounting for measurement error, then the previous identification arguments break down. To see why, I re-write the model (8) as:

$$\omega_i = \mathbb{1}\left\{h_k + \alpha_n + \sum_{\ell=1}^{K} J_{k\ell}\bar{\omega}_n^\ell + \tilde{\varepsilon}_i \geq 0\right\}, \quad \text{where } \tilde{\varepsilon}_i = \varepsilon_i - \sum_{\ell=1}^{K} J_{k\ell} u_n^\ell. \tag{12}$$

In this model, the observed average choices $\{\bar{\omega}_n^\ell\}_\ell$ are endogenous, since they are correlated with the error term $\tilde{\varepsilon}_i$. Indeed, even if agent i's choice ω_i were omitted when constructing $\bar{\omega}_n^k$, it is still the case that $\text{Cov}(\bar{\omega}_n^\ell, \tilde{\varepsilon}_i) = -J_{k\ell} \times \text{Var}(u_n^\ell) \neq 0$ for each type $\ell \in \{1, \dots, K\}$. This endogeneity is an inherent feature of network-based models with incomplete information. Indeed, for any finite networks, replacing m_n^k with $\bar{\omega}_n^k$ leads to measurement error bias.

I propose a strategy to correct for this measurement error bias using internal instruments. This procedure involves two steps. First, within each network n and type k, I randomly split agents into two subsamples: a and b. The partitioning method and the relative sizes of these subsamples do not matter for identification. Second, I compute the average action in each subsample, $\bar{\omega}_{n,a}^k$ and $\bar{\omega}_{n,b}^k$, using $\bar{\omega}_{n,a}^k$ as an endogenous regressor and $\bar{\omega}_{n,b}^k$ as an instrument.

By construction, $\bar{\omega}_{n,a}^k$ and $\bar{\omega}_{n,b}^k$ are noisy approximations of m_n^k, and their measurement errors are independent. This independence is a consequence of the incomplete information setting, where agents respond to rational expectations rather than to the realized choices of others. Leveraging this property, I show that the IV procedure is valid and yields consistent parameter estimates as the number of networks grows large—for any finite network sizes.

Before stating my identification result, I will first motivate it with an illustrative example.

Example (Brock and Durlauf, 2001). In the paper, Brock and Durlauf (2001) study the case:

$$m_n = \tanh\left(h + W_n' d + J m_n\right), \quad \text{for } n = 1, \ldots, N, \tag{13}$$

where W_n is an observed vector and h, d, and J are unknown coefficients.[20] In analyzing the model, it is common practice to assume m_n is observable. The basis for this assumption is that, as network sizes grow large, $\bar{\omega}_n$ provides a consistent estimate of m_n. Hence, a researcher's inability to see m_n is treated as an estimation issue, separate from identification.

In practice, however, m_n is unobserved for finite networks. So, a researcher would need to replace m_n with $\bar{\omega}_n$ in estimation. Without accounting for measurement error, standard estimation methods, such as maximum likelihood or OLS, would produce biased estimates. To illustrate this point, consider an OLS estimand constructed from the observed means $\bar{\omega}_n$. Defining $u_n = \bar{\omega}_n - m_n$ and $v_n = \tanh^{-1}(\bar{\omega}_n) - \tanh^{-1}(m_n)$, I can re-write the model as:

$$\tanh^{-1}(\bar{\omega}_n) = h + W_n' d + J\bar{\omega}_n + \tilde{\xi}_n, \quad \text{where } \tilde{\xi}_n = -J u_n + v_n. \tag{14}$$

To ease notation, let $\tilde{m}_n$, $\tilde{\bar{\omega}}_n$, and $\tilde{Y}$ denote the residuals from a least squares regression of m_n, $\bar{\omega}_n$, and $\tanh^{-1}(\bar{\omega}_n)$, respectively, on the vector $(1, W_n')$. The OLS estimand for J is:

$$J^{\text{OLS}} = \frac{\text{Cov}(\tilde{\bar{\omega}}_n, \tilde{Y})}{\text{Var}(\tilde{\bar{\omega}}_n)} = J \times \frac{\text{Var}(\tilde{m}_n)}{\text{Var}(\tilde{m}_n) + \text{Var}(u_n)} + \frac{\text{Cov}(\tilde{\bar{\omega}}_n, v_n)}{\text{Var}(\tilde{m}_n) + \text{Var}(u_n)}. \tag{15}$$

This estimand is "doubly-biased" since both the explanatory variable and outcome are measured with error. Hence, for any finite networks, OLS fails to recover the interaction effect.

As illustrated in Figure 1, the bias can be substantial even for very large networks. When setting $J = 0.5$, I find that OLS overstates the interaction effect by a factor of 4 in networks with 100 agents and by a factor of 2 in networks with 500 agents. The bias becomes small only when network sizes reach 5,000 agents, a scale rarely attainable in most applications.

To overcome this issue, I propose IV estimation: splitting each network into two parts, a and b, and using the average choice in part b as an instrument for the average choice in part

[20] The authors allow for individual-level covariates X_i, which I omit in this example for notational simplicity.

a. By construction, the instrument $\bar{\omega}_{n,b}$ is independent of measurement error in $\bar{\omega}_{n,a}$, and it satisfies instrument relevance, as $\mathrm{Cov}(\bar{\omega}_{n,a}, \bar{\omega}_{n,b}) = \mathrm{Var}(m_n) \neq 0$. By standard arguments, J can be recovered by following procedure: (1) regress $\bar{\omega}_{n,a}$ on $(1, W_n', \bar{\omega}_{n,b})$ to obtain fitted values $\mathbf{L}(\bar{\omega}_{n,a}|1, W_n', \bar{\omega}_{n,b})$ and (2) regress the outcome Y_n on $(1, W_n', \mathbf{L}(\bar{\omega}_{n,a}|1, \bar{\omega}_{n,b}))$. To ensure that this jointly identifies J, d', and h, I address a further complication: owing to the nonlinearity of the function $\tanh^{-1}$, the measurement error v_n in $\tanh^{-1}(\bar{\omega}_n)$ is not mean-zero. To correct for this, I define Y_n as a specific linear combination of $\tanh^{-1}(\bar{\omega}_{n,a})$ across subsamples, exploiting the known dependence of $\mathrm{E}[v_{n,a}]$ on the subsample size. As I demonstrate below, this procedure jointly recovers J, d', and h for any finite network sizes.

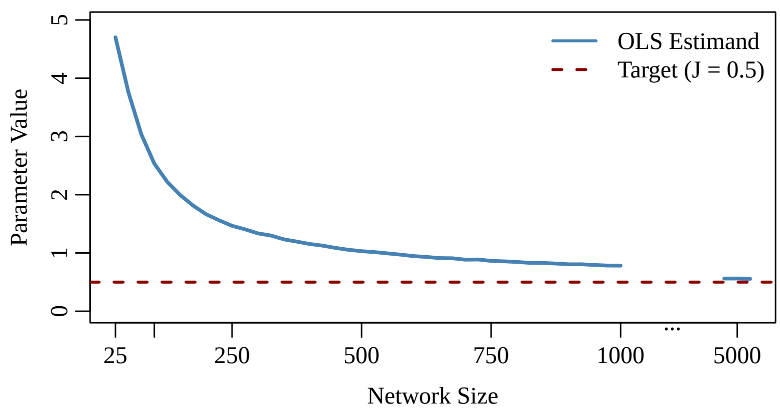


FIGURE 1.—Bias of OLS estimand for J at different network sizes, setting $J = 0.5$, $h = 1$, and $d = 1$.

An IV Estimand for Endogenous Interaction Effects

To formalize my identification strategy, I first define $\bar{\omega}_{n,a}^k$ and $\bar{\omega}_{n,b}^k$ to be the average choices in each randomly selected subset of network n and type k. Throughout my analysis, I use $\bar{\omega}_{n,a}^k$ as the endogenous regressor and $\bar{\omega}_{n,b}^k$ as the instrument. I also define the outcomes as:

$$Y_n^k = \Omega_{n,a}^k F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,a}^k) + \Omega_{n,a'}^k F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,a'}^k), \tag{16}$$

where a' is a strict subset of a, and $(\Omega_{n,a}^k, \Omega_{n,a'}^k)$ are known weights that depend on the sizes of a and a', denoted by $I_{n,a}^k$ and $I_{n,a'}^k$, respectively. To define these weights, I take two steps. First, I define $P \geq 1$ unique combinations of subgroup pairs (a, a') across networks. Each combination p yields a sequence of subsample sizes $\{I_{n,\kappa(p,n)}^k\}_n$, where $\kappa(p, n)$ indicates if a or a' is selected in network n. Next, I define $\Omega_{n,\tilde{a}}^k = \lim_{P\to\infty} \sum_p [\mathbf{E}_k^{-1}]_{p,1} \mathbb{1}\{I_{n,\kappa(p,n)}^k = I_{n,\tilde{a}}^k\}$ for $\tilde{a} \in \{a, a'\}$, where $\mathbf{E}_k \in \mathbb{R}^{P\times P}$ is a matrix whose entries are $[\mathbf{E}_k]_{i,j} = \mathrm{E}[(I_{n,\kappa(j,n)}^k)^{1-i}]$. Details on deriving and computing these weights are provided in the Appendix. Note that, while identification formally requires the number of partitioning rules P to be large, I find that in practice, even $P=1$ or $P=2$ is enough to yield estimates with effectively no bias.[21]

THEOREM 3: *Suppose Assumptions B.1 and B.2 hold and $\{F_{\varepsilon|k}\}_k$ are identified. Then:*

(i) $\beta_{k_1,k_2} = [h_{k_1} - h_{k_2}, J_{k_1 1} - J_{k_2 1}, ..., J_{k_1 K} - J_{k_2 K}]'$ *is identified for any* (k_1,k_2)*. Also:*

$$\beta_{k_1,k_2} = \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n[Y_n^{k_1} - Y_n^{k_2}]), \quad \textit{where} \begin{cases} X_n = (1, \bar{\omega}_{n,a}^1, \ldots, \bar{\omega}_{n,a}^K)' \\ Z_n = (1, \bar{\omega}_{n,b}^1, \ldots, \bar{\omega}_{n,b}^K)'. \end{cases}$$

(ii) If $\alpha_n = W_n' d$ *for observed* W_n*,* $\beta_k = [h_k, d', J_{k1}, ..., J_{kK}]'$ *is identified for any* k*. Also:*

$$\beta_k = \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n Y_n^k), \quad \textit{where} \begin{cases} X_n = (1, W_n', \bar{\omega}_{n,a}^1, \ldots, \bar{\omega}_{n,a}^K)' \\ Z_n = (1, W_n', \bar{\omega}_{n,b}^1, \ldots, \bar{\omega}_{n,b}^K)'. \end{cases}$$

Theorem 3 shows that social interaction effects are identified even when networks are finite. Moreover, these parameters can be recovered via linear IV estimation, where $\bar{\omega}_{n,b}^k$ is used as an instrument for $\bar{\omega}_{n,a}^k$, and where outcomes Y_n^k are defined in (16). By the sample analogue principle, I can construct estimators for β_k and β_{k_1,k_2} that will converge in probability as the number of networks grows large, for any fixed network sizes.[22] In the Appendix, I derive asymptotic properties of these estimators, demonstrating that they are capable of inference.

In Table 2, I assess the performance of the IV estimators using Monte Carlo simulations, varying both the network sizes and the number of networks in the data. These results illustrate two key properties of the estimation strategy. First, the estimators perform better with larger networks. This is because, as the network size grows large, the sample average $\bar{\omega}_{n,a}^k$ will be a better approximation of the true expectation m_n^k, thus reducing measurement error.

Second, the IV estimators perform better as the number of networks in the data increases. This property is implied by Theorem 3, which shows that, for any fixed network sizes, the model parameters can be consistently estimated through IV estimation. Crucially, this result does not apply to OLS. Indeed, while the OLS estimator remains biased even as the number of networks tends to infinity, the IV estimator will converge to the true model parameters.

In principle, the IV estimator can be applied to data with networks of any size. However, in practice, for datasets without many networks, it is important that network sizes are large enough to yield a strong first stage; that is, for any type k, the average choice in one subset

---

[21] If $P=1$, then outcomes reduce to $Y_n^k = F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,a}^k)$. Derivations for $P>1$ are provided in the Appendix.

[22] Consistency relies only on measurement errors having mean zero and being mean independent: $\mathrm{E}(u_{n,a}^k)=0$ and $\mathrm{E}(u_{n,a}^k u_{n,b}^\ell)=0$ for all k,ℓ. These moment conditions are not affected by the possibility of multiple equilibria.

of a network should strongly predict the average choice in the other subset. If there are very few agents in a network, the IV estimator can suffer from weak instruments. In general, I find that with < 500 networks, the IV estimator is best suited to settings with ≥ 10 agents per network and type. This applies if the network is a classroom, school, or village, but it limits applicability to very small networks, such as households. Note that instrument relevance is testable, and it depends not only on the network size but also on the variance of ε_i.

TABLE II

MEAN SQUARED ERROR OF OLS AND IV ESTIMATORS BY NETWORK SIZE AND NUMBER

		Specification A ($K = 1$)						Specification B ($K = 2$)					
Number of	Agents per	$\widehat{h}$		$\widehat{d}$		$\widehat{J}$		$\widehat{h_1 - h_2}$		$\widehat{J_{11} - J_{12}}$		$\widehat{J_{22} - J_{21}}$	
Networks	Network	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
N = 200	30	8.74	3.57	0.53	0.27	16.06	6.26	0.52	0.31	5.67	2.58	5.60	2.45
	100	2.07	0.26	0.10	0.02	4.00	0.47	0.14	0.04	1.49	0.23	1.48	0.23
	200	0.73	0.17	0.04	0.01	1.40	0.29	0.05	0.01	0.53	0.07	0.52	0.07
N = 1000	30	8.82	1.14	0.53	0.05	16.23	2.46	0.52	0.05	5.74	0.62	5.72	0.61
	100	2.15	0.06	0.10	0.00	4.16	0.13	0.14	0.01	1.51	0.05	1.51	0.04
	200	0.73	0.06	0.03	0.01	1.40	0.09	0.05	0.00	0.54	0.02	0.54	0.02
N = 5000	30	8.81	0.60	0.53	0.02	16.21	1.45	0.51	0.03	5.71	0.51	5.73	0.52
	100	2.17	0.02	0.10	0.00	4.19	0.05	0.15	0.00	1.52	0.01	1.52	0.01
	200	0.72	0.04	0.03	0.00	1.39	0.05	0.05	0.00	0.55	0.00	0.55	0.00

Notes. MSEs computed over 500 simulation draws. Specification A applies the estimator from Theorem 3*(ii)*, setting $K = 1$ and $\alpha_n = W_n'd$ (the same data generating process as in Figure 1). Specification B applies the estimator from Theorem 3*(i)*, setting $K = 2$ and treating α_n as fully unobserved. All networks have equal size. I define X_n using $2/3$ of the network sample, and I define Z_n with the remaining $1/3$ of the sample, setting $P = 1$. I assume $\varepsilon_i | k \sim \text{Logistic}(0, 1)$. See the Appendix for more specification details.

4. EMPIRICAL APPLICATIONS

I present two empirical applications of the model. First, I analyze Project STAR—an education experiment that randomized students to classrooms—to study how peer effects differ by gender. Second, I analyze the Primary School Deworming Project (Miguel and Kremer, 2004) to study how spillovers in treatment uptake differ by age and gender. In both studies, allowing for nonuniform interaction effects yields substantively new economic conclusions.

4.1. *Project STAR: Gender Differences in Peer Effects*

The first application examines the Project STAR class size reduction experiment, previously studied by Boozer and Cacciola (2001) and Graham (2008). This program was launched in

1985 among kindergarteners in 79 public schools in Tennessee. In each school, students and teachers were randomly assigned to one of three classroom types: small (13–17 students), large (22–25), and large with a teacher's aide. At the end of the year, all students took math and reading exams. In total, the experiment covered 6,325 students over 325 classrooms.[23]

To learn about gender differences in peer effects, I study a model with two student types: male (m) and female (f). The Project STAR sample is well-suited for this decomposition, as boys and girls are both well-represented in every classroom. As outcomes, I use students' math and reading test scores, discretized to indicate whether a student scored in the top 50% statewide.[24] Since students take tests simultaneously, it is reasonable to assume they cannot perfectly predict each others' scores, and therefore they act under incomplete information.

Outcomes are given by $\omega_i = \mathbb{1}\{e_i \geq \varepsilon_i\}$, where e_i is student i's effort and ε_i is a threshold reflecting individual ability, with distribution $F_{\varepsilon|k}$. Students choose e_i to maximize utility:

$$U_i(e_i|k,n) = \Big(h_k + \alpha_n + J_{kk}\ \mathrm{E}_i(\bar{\omega}^k_{n,-i}) + J_{k\ell}\ \mathrm{E}_i(\bar{\omega}^\ell_n)\Big)\, e_i - \frac{1}{2}e_i^2. \tag{17}$$

This equation contains a gender fixed effect, h_k, which captures any systematic differences in cognitive development and prior socialization. It also contains a classroom fixed effect, α_n, that accounts for all contextual factors like teacher quality, class size, and composition. Finally, it includes gender-specific peer effects, $J_{k\ell}$, which measure how effort depends on expected performance of male and female peers. An equilibrium is defined by equation (9).

Under the experimental protocols, students at a given school cannot self-select into classrooms. So, Assumption B.1 holds conditional on school. To address potential non-random selection into schools, I include school fixed effects in the estimation.[25] When implementing the estimators, I assume ε_i has a logistic distribution, $F_{\varepsilon|k}(z) = [1 + \exp(-z)]^{-1}$, thus avoiding the practical challenges of nonparametric estimation with insufficiently rich data.

Table 3 shows IV estimates of $h_m - h_f$, $J_{mm} - J_{fm}$, and $J_{ff} - J_{mf}$ using Project STAR data. These estimates measure gender differences in private effort costs and in peer effects, while fully controlling for all classroom-level contextual factors. The results show no significant difference between h_m and h_f, suggesting that boys and girls perform similarly in absence of peer effects. However, I find that peer effects differ significantly by gender, with

[23] The data does not have classroom identifiers. So, following Boozer and Cacciola (2001) and Graham (2008), I uniquely assign students to classes by matching on class observables. I recover a sample of 6,248 students (5,801 with non-missing test scores) over 321 classrooms. For further details about the program, see Word et al. (1990).

[24] In the Appendix, I show that the estimates are also similar for alternative thresholds, such as 25% and 75%.

[25] Following Graham (2008), I also report estimates for a subsample of schools that had only one classroom of each type. This addresses the unlikely possibility of nonrandom assignment to classrooms within classroom types.

$J_{mm} > J_{fm}$ and $J_{ff} > J_{mf}$, indicating that students are much more likely to conform to their own gender than to the other gender. These differences appear slightly larger for reading tests than for math tests, although the differences between outcomes are not statistically significant. In addition, I find no evidence to reject the hypothesis that $J_{mm} - J_{fm}$ equals $J_{ff} - J_{mf}$, which could suggest that peer effect differences are symmetric across genders.

To illustrate the importance of allowing for generalized interactions, I consider, in Table A.1 of the Appendix, how the estimates change if I impose uniform peer effects ($J_{k\ell} = J$). I find that this restriction yields substantively different conclusions, such as estimating that that girls have lower private effort costs than boys ($h_f > h_m$), though this difference becomes statistically insignificant after including school fixed effects. Moreover, this homogeneity restriction overlooks key distortions that determine the impacts of classroom-level policies. Since $J_{mm} > J_{fm}$ and $J_{ff} > J_{mf}$, a policy that boosts the achievement of boys would have outsized effects on other boys but limited effects on girls. Therefore, the overall effects of a classroom intervention can vary greatly depending on which students are directly targeted.

To further validate my approach, I perform two robustness checks. First, I test whether the type-specific parameters $(h_f, h_m, J_{ff}, J_{fm}, J_{mf}, J_{mm})$ depend on observed classroom characteristics, such as the poverty rate, minority share, location, or teacher education and experience. Such dependence would imply the model is misspecified. I find no evidence that these parameters vary with classroom factors. Second, I assess the sensitivity of the IV estimates to different classroom partitioning rules, plotting histograms of estimates across alternative partitions. I find that the estimates are relatively stable across partitioning rules.

TABLE III

PROJECT STAR: IV ESTIMATES OF PEER EFFECT DIFFERENCES BY GENDER

	Top 50% Math Score				Top 50% Reading Score			
	(1)		(2)		(1)		(2)	
$h_m - h_f$	-0.112	(0.16)	0.082	(0.20)	-0.029	(0.18)	0.019	(0.29)
$J_{mm} - J_{fm}$	4.229***	(0.86)	4.755***	(0.76)	5.339***	(0.82)	4.975***	(0.55)
$J_{ff} - J_{mf}$	-4.127***	(1.01)	-4.789***	(1.07)	-5.223***	(1.09)	-4.686***	(0.81)
Observations	5,798		5,798		5,718		5,718	
School FE	No		Yes		No		Yes	
1st Stage F ($\bar{\omega}_n^m$)	29.40		30.06		25.52		25.73	
1st Stage F ($\bar{\omega}_n^f$)	19.77		19.72		13.37		12.98	

Notes. This table reports estimates of $\hat{\beta}_{f,m}$, corresponding to the estimand in Thm. 3*(i)*. Estimates reported with and without school FE. Standard errors based on the limiting distribution of $\hat{\beta}_{f,m}$ as $N \to \infty$, derived in the Appendix as $\sqrt{N}(\hat{\beta}_{f,m} - \beta_{f,m}) \xrightarrow{d} \mathcal{N}(0, Q\Omega Q')$ where $Q = \mathrm{E}(Z_n X_n')^{-1}$ and $\Omega = \mathrm{Var}(Z_n \sum_\ell (J_{f\ell} - J_{m\ell}) u_{n,a}^\ell)$. I define X_n using $2/3$—and Z_n using $1/3$—of the sample, and I set $P = 1$. *p<0.1; **p<0.05; ***p<0.01.

4.2. *Primary School Deworming: Treatment Uptake with Nonuniform Spillovers*

The second application revisits Miguel and Kremer's (2004) seminal study of the Primary School Deworming Project, conducted in Busia, Kenya during 1998-2001. In this study, 75 schools were randomly assigned to three equal-sized groups: Group 1 began receiving free deworming treatment in 1998, Group 2 in 1999, and Group 3 not until 2001. A key feature of the intervention was that individuals could opt out of treatment. Indeed, among the nearly 15,000 eligible students in 1998-1999, only about 80% received treatment in a given year. This partial compliance provides a natural setting to analyze spillovers in treatment uptake.

Following the original paper, I focus on the years 1998 and 1999, and I restrict the sample to school-aged children (ages 6-18) eligible for treatment. Eligibility covered all students at treatment schools, except girls older than 13 and, in 1999, anyone without parental consent.

I model each school as a social network where interactions are global, and I allow interactions to vary by age and gender. Given that girls older than 13 were ineligible for treatment, I specify a model with three types of students: (1) girls under 13, (2) boys under 13, and (3) boys 13 and above. Since most schools were large (averaging over 450 students), I assume students cannot track everyone else's choice and therefore act with incomplete information.

In each year t, a student i of type k in school n decides whether to accept treatment based on private benefits and social interaction effects. A student's treatment decision is given by:

$$\omega_{i,t} = \mathbb{1}\left\{ h_{k,t} + \alpha_{n,t} + X_{i,t}'c + J_{kk}\,\mathrm{E}_i(\bar{\omega}^k_{n,-i,t}) + \sum_{\ell \neq k} J_{k\ell}\,\mathrm{E}_i(\bar{\omega}^\ell_{n,t}) + \varepsilon_{i,t} \geq 0 \right\}. \tag{18}$$

This equation contains a time-varying type fixed effect, $h_{k,t}$, which captures any systematic differences by age and gender in health risk or other factors affecting treatment propensity. It also contains a time-varying school fixed effect, $\alpha_{n,t}$, which accounts for all cross-school externalities (as specified by Miguel and Kremer, 2004), as well as school-level differences in treatment provision, baseline infection rates, and other factors.[26] In addition, it includes student-level controls, $X_{i,t}$, for one's prior treatment uptake and prior school attendance, as well as an idiosyncratic taste $\varepsilon_{i,t}$ with distribution $F_{\varepsilon|k,t}$.[27] Lastly, I allow for nonuniform interaction effects, $J_{k\ell}$, which measure how uptake depends on the expected treatment rates of different types of peers. Since deworming medication was both a new technology and a

[26] In the paper, cross-school spillovers operate through the number of students living near the school who attend other treatment schools. As this channel is school×time-specific, it is fully accounted for by the fixed effect $\alpha_{n,t}$.

[27] I control for prior uptake to isolate contemporaneous interactions from pre-existing factors. In the Appendix, I also report estimates restricted to first-time eligible students. I control for attendance to ensure spillovers operate through uptake rather than attendance, given that absence on the treatment day was a key source of noncompliance. Note that only prior attendance is used, as current attendance might be endogenously affected by treatment uptake.

public good (by reducing infection rates among untreated individuals), social interactions in this setting reflect a combination of peer effects, social learning, and free-riding motives.[28]

Under the experimental design, students could not self-select into schools based on treatment, and most schools had similar observable characteristics (Miguel and Kremer, 2004). Nevertheless, to address any potential self-selection, I condition on a variety of school-level factors such as geographic zone, treatment group, local infection rate, population density, and number of nearby students attending other treatment schools. As in the previous application, I assume that the idiosyncratic tastes, $\varepsilon_{i,t}$, are logistically distributed in each period.

Table 4 presents IV estimates of $h_{k_1,t} - h_{k_2,t}$ and $J_{k_1\ell} - J_{k_2\ell}$ using data from Miguel and Kremer (2004). These estimates measure age and gender differences in private payoffs and interaction effects, while controlling for all school-level determinants of treatment uptake. I report results across four specifications, with and without year, student, and school controls. The results show no significant differences in private payoffs. Yet, I find strong differences in interaction effects, with $J_{kk} > J_{k\ell}$ for $k \neq \ell$, indicating that students are more likely to conform to peers of their own type. Also, the degree of own-type conformity, measured by $\delta_{k_1k_2} = J_{k_1k_1} + J_{k_2k_2} - J_{k_1k_2} - J_{k_2k_1}$, is highest for girls and boys under 13 ($k_1, k_2 \in \{1, 2\}$). This suggests that gender-based homophily is particularly strong among younger students.

The finding that students prefer to conform to their own type helps clarify the underlying mechanisms driving spillovers in this setting. Specifically, stronger own-type conformity is difficult to reconcile with a pure free-riding motive, which would predict uniform spillovers or spillovers proportional to the epidemiological externality imposed by each type. Rather, it is more consistent with an imitation motive, where peer effects depend on social identity.

This finding also provides insight into the distributional impacts of school-based policies. To see why, note that, within schools, there tend to be large differences in treatment uptake rates by age and gender; yet, as shown in Figure A.2 of the Appendix, the size and direction of these differences vary substantially from one school to another. This suggests that certain factors at the school-level contribute to disparities in treatment uptake. By analyzing these patterns under the generalized interactions framework, I can isolate the role of nonuniform peer effects from other school-level factors. My results show that nonuniform peer effects

[28] Kremer and Miguel (2007) also study peer effects in this setting, but their analysis differs from mine in several key ways. First, they study parents' consent decisions, while I study student uptake after parent consent is granted. Second, they study local interactions among close social contacts for a small sample of parents, where spillovers operate via the *exogenous* number of links to treatment schools. Instead, I study *endogenous* interactions in large networks (schools), controlling for the number of nearby students at treatment schools. Third, they focus on later years of the program (post-2000), after schools charged deworming fees, whereas I study uptake before such fees existed. My analysis more closely follows Bhattacharya et al. (2023), who study mosquito net adoption in Kenya.

strongly influence which students in a school are more likely to take treatment. Accounting for these differences can allow policymakers to target access to treatment more effectively.

As in the previous application, I conduct two robustness checks. First, I test whether the parameters $h_{k,t}$ and $J_{k\ell}$ depend on observed school characteristics, such as treatment group or local infection rates. I find no evidence of such dependence. Second, I plot the estimates across alternative network partitioning rules, and I show the results remain relatively stable.

TABLE IV

DEWORMING IN SCHOOLS: IV ESTIMATES OF SPILLOVER DIFFERENCES BY AGE AND GENDER

	(1)		(2)		(3)		(4)	
Girls < 13 and Boys < 13								
$h_1 - h_2$	0.928	(0.69)	0.857	(0.79)	1.431	(1.67)	1.256	(1.99)
$J_{11} - J_{21}$	8.293***	(2.53)	8.679***	(2.74)	11.428***	(4.38)	12.273**	(4.97)
$J_{12} - J_{22}$	-9.382***	(2.93)	-10.029***	(3.26)	-13.827***	(4.25)	-15.598***	(5.17)
$J_{13} - J_{23}$	0.071	(0.71)	0.397	(0.86)	0.378	(1.16)	0.600	(1.13)
Girls < 13 and Boys $\geq$ 13								
$h_1 - h_3$	-0.309	(0.89)	-0.555	(0.97)	0.873	(1.41)	1.266	(1.17)
$J_{11} - J_{31}$	5.551**	(2.78)	6.892**	(3.02)	7.401*	(4.13)	7.824**	(3.88)
$J_{12} - J_{32}$	-0.132	(2.38)	-2.382	(2.95)	-5.055	(4.45)	-6.311	(3.91)
$J_{13} - J_{33}$	-4.844***	(0.57)	-3.707***	(0.93)	-2.633**	(1.23)	-2.480*	(1.43)
Boys < 13 and Boys $\geq$ 13								
$h_2 - h_3$	-1.237	(1.09)	-1.413	(0.86)	-0.559	(1.25)	0.010	(1.46)
$J_{21} - J_{31}$	-2.743	(3.22)	-1.787	(3.02)	-4.027	(4.49)	-4.449	(5.38)
$J_{22} - J_{32}$	9.250***	(3.06)	7.647**	(3.20)	8.772	(5.43)	9.287	(6.51)
$J_{23} - J_{33}$	-4.914***	(0.47)	-4.104***	(0.61)	-3.012***	(1.00)	-3.080***	(1.02)
Year Fixed Effects	No		Yes		Yes		Yes	
Student Controls	No		No		Yes		Yes	
School Controls	No		No		No		Yes	
1st Stage F ($\bar{\omega}_n^1$)	14.10		11.20		14.35		13.09	
1st Stage F ($\bar{\omega}_n^2$)	12.05		9.97		11.72		10.40	
1st Stage F ($\bar{\omega}_n^3$)	35.60		16.69		18.56		17.88	

Notes. This table reports estimates of $\hat{\beta}_{k_1,k_2}$, corresponding to the estimand in Thm. 3*(i)*. Student controls are prior eligibility, prior uptake, and prior attendance. School controls are local infection rate, local pop. density, and number of nearby pupils at other treatment schools. Zone FE included. Standard errors based on the limiting distribution of $\hat{\beta}_{k_1,k_2}$ as $N \to \infty$, derived in the Appendix as $\sqrt{N}(\hat{\beta}_{k_1,k_2} - \beta_{k_1,k_2}) \xrightarrow{d} \mathcal{N}(0, Q\Omega Q')$, where $Q = \mathrm{E}(Z_n X_n')^{-1}$ and $\Omega = \mathrm{Var}(Z_n \sum_\ell (J_{k_1\ell} - J_{k_2\ell}) u_{n,a}^\ell)$. In all cases, $h_k - h_\ell$ denotes $\frac{1}{T}\sum_{t=1}^T (h_{k,t} - h_{\ell,t})$. I define X_n using $2/3$—and Z_n using $1/3$—of the sample, setting $P = 1$. *p<0.1; **p<0.05; ***p<0.01.

5. CONCLUSION

This paper has developed empirical tools to measure and interpret social interaction effects for a generalized class of discrete choice models, where individuals interact differently with different network members, conforming to some while distinguishing from others. First, I characterize how behavioral outcomes depend on the signs and magnitudes of endogenous interaction effects. Then, I show how to draw inference about endogenous interactions from data on individual choices, while overcoming two key identification challenges: unobserved contextual effects and unobserved agent expectations in finite networks. By resolving these challenges, I establish empirical tractability of the model in any finite network setting. I apply my method to data from two large-scale educational experiments: Project STAR and the Primary School Deworming Project. Estimates show that nonuniform interactions play an important role in both applications. Note that while this paper provides an initial analysis of nonuniform and mixed-sign interactions in discrete choice models, there is much scope for future work. In particular, one valuable extension would be to explore endogenous network formation, where individuals can choose their level of exposure to different types of people.

APPENDIX

PROOF OF PROPERTY 1. Define $\mathcal{Q}:[0,1]^K \to [0,1]^K$ and $\mathcal{H}:[0,1]^K \to \mathbb{R}^K$ such that:

$$\mathcal{Q}_k(m) = F_{\varepsilon|k}\big(h_k + \sum_{\ell=1}^{K} J_{k\ell} m^\ell\big) \quad \text{and} \quad \mathcal{H}_k(m) = m^k - \mathcal{Q}_k(m), \quad \text{for } k = 1,\dots,K.$$

Since no root of $\mathcal{H}$ lies on the boundary of $[0,1]^K$, I restrict attention to the interior $(0,1)^K$.[1] Letting $\mathbf{D}_{\mathcal{H}}$ denote the Jacobian of $\mathcal{H}$, define $\mathcal{C} = \{m \in (0,1)^K : \det(\mathbf{D}_{\mathcal{H}}(m)) = 0\}$. By Sard's Theorem, $\mathcal{H}(\mathcal{C})$ has Lebesgue measure zero. Also, for any given $y = \mathcal{H}(m)$, almost all $\{F_{\varepsilon|k}\}_k$ ensure that $y \notin \mathcal{H}(\mathcal{C})$. In particular, for $y = \mathbf{0}_K$, it follows that $\mathbf{D}_{\mathcal{H}}(m)$ is almost always nonsingular at every root of $\mathcal{H}$. By the inverse function theorem, each root of $\mathcal{H}$ is isolated. Since the equilibrium set is compact and has only isolated points, it must be finite.

To prove that the number of equilibria is odd, I apply the Poincaré-Hopf Index Theorem (Milnor, 1965, Ch. 6). To do so, first note that the set of equilibria $\mathcal{M}$ lies in $(0,1)^K$, which is diffeomorphic to an open disk. Also, the vector field $\mathcal{H}$ is smooth and points outward on the boundary, since $\lim_{m_k \to 0} \mathcal{H}(m) < 0$ and $\lim_{m_k \to 1} \mathcal{H}(m) > 0$ for all k. Moreover, for almost all $\{F_{\varepsilon|k}\}_k$, $\mathcal{H}$ has finitely-many isolated roots m, each with $\det(\mathbf{D}_{\mathcal{H}}(m)) \neq 0$. By the index theorem: $\sum_{m \in \mathcal{M}} \text{index}_m(\mathcal{H}) = 1$ where $\text{index}_m(\mathcal{H}) = 2\mathbb{1}\{\det(\mathbf{D}_{\mathcal{H}}(m)) > 0\} - 1$.

Finally, suppose there exist d_s equilibria where $\rho(\mathbf{D}(m)) < 1$. At each of these equilibria:

$$\det\big(\mathbf{D}_{\mathcal{H}}(m)\big) = \det\big(I - \mathbf{D}(m)\big) = \prod_{k=1}^{K} \big(1 - \lambda_k(m)\big) > 0,$$

where $\{\lambda_k(m)\}_{k=1}^K$ are the eigenvalues of $\mathbf{D}(m)$. By the index theorem, there must also be at least $d_s - 1$ equilibria m at which $\det(\mathbf{D}_{\mathcal{H}}(m)) < 0$, each one satisfying $\rho(\mathbf{D}(m)) > 1$.

Q.E.D.

[1] Since each $F_{\varepsilon|k}$ has positive density everywhere, $\mathcal{H}_k(m) < 0$ when $m_k = 0$ and $\mathcal{H}_k(m) > 0$ when $m_k = 1$.

PROOF OF LEMMA 1. Horn and Johnson (1985), Thm. 5.6.9-5.6.10, show that $\rho(\mathbf{D}(m))$ is the greatest lower bound for the set of all matrix norms of $\mathbf{D}(m)$. In particular, they show:

(a) $\rho(\mathbf{D}(m)) \leq ||\mathbf{D}(m)||$ for any matrix norm $||\cdot||$.

(b) For any $\epsilon > 0$, there is a matrix norm $||\cdot||$ where $||\mathbf{D}(m)|| \leq \rho(\mathbf{D}(m)) + \epsilon$.

Let $\rho(\mathbf{D}(m)) < 1$. Setting $\epsilon < 1 - \rho(\mathbf{D}(m))$, it follows from *(b)* that $||\mathbf{D}(m)|| < 1$ for some matrix norm $||\cdot||$. The system (5) is a contraction map at m under this norm, which means:

$$||m_t - m|| = ||\mathcal{Q}(m_{t-1}) - \mathcal{Q}(m)|| \leq \kappa ||m_{t-1} - m||,$$

where $\kappa = [0,1)$ for any vector m_{t-1} that lies in a sufficiently small neighborhood of m. Iterating on this inequality implies that $||m_t - m|| \leq \kappa^t ||m_0 - m||$, where $\lim_{t\to\infty} \kappa^t = 0$.

Next, let $\rho(\mathbf{D}(m)) > 1$. By *(a)*, $||\mathbf{D}(m)|| > 1$ for any matrix norm $||\cdot||$. By Henry (1981), Thm. 5.1.5., there exists a scalar $u > 0$ such that, for any $\delta > 0$, there is an initial iterate m_0 where $||m_0 - m|| < \delta$ and where some future iterate m_t, $t \geq 1$, satisfies $||m_t - m|| \geq u$.

Q.E.D.

PROOF OF PROPERTY 2. First, suppose $\mathbf{J}$—and therefore $\mathbf{D}(m)$ for all $m \in [0,1]^K$—is a non-negative matrix. Also, assume there is an equilibrium $m \in (0,1)^K$ with $\rho(\mathbf{D}(m)) > 1$. If $\mathbf{D}(m)$ is irreducible, the Perron-Frobenius theorem ensures $\mathbf{D}(m)x = \rho(\mathbf{D}(m))x > x$ for some strictly positive vector $x \in \mathbb{R}^K_{++}$. It follows that $\mathbf{D}(m)\delta x > \delta x$ for any scalar $\delta > 0$. Taking the first-order Taylor approximation of $\mathcal{Q}(m + \delta x)$ about m, I obtain the equation:

$$\mathcal{Q}(m+\delta x) = \underbrace{\mathcal{Q}(m)}_{=m} + \underbrace{\mathbf{D}(m)\delta x}_{>\delta x} + \mathrm{h}_1(m+\delta x)\delta x, \quad \text{where } \lim_{\delta\to 0} \mathrm{h}_1(m+\delta x) = 0.$$

For a small enough δ, the vectors $a = m - \delta x$ and $b = m + \delta x$, where $a, b \in (0,1)^K$, satisfy $\mathbf{0}_K < \mathcal{Q}(\mathbf{0}_K) < \mathcal{Q}(a) < a < m$ and $m < b < \mathcal{Q}(b) < \mathcal{Q}(\mathbf{1}_K) < \mathbf{1}_K$. Brouwer's fixed point theorem thus ensures that $\mathcal{Q}$ has two more fixed points, $\underline{m}$ and $\overline{m}$, satisfying $\mathbf{0}_K < \underline{m} < m$ and $m < \overline{m} < \mathbf{1}_K$. If $\underline{m}$ (or $\overline{m}$) is unstable, then by the same arguments, there would be two more equilibria: one between m and $\underline{m}$ ($\overline{m}$) and one between $\underline{m}$ ($\overline{m}$) and the boundary. As the equilibrium set is almost surely finite, there are more stable than unstable equilibria; and, by Property 1, it follows that there is exactly one more stable than unstable equilibrium.

Next, suppose $\mathbf{D}(m)$ is reducible. If $\rho(\mathbf{D}(m)) > 1$, then $\rho(B) > 1$ for some irreducible block B (with index set $\mathcal{B}$) of $\mathbf{D}(m)$. By Perron–Frobenius, there exists $x \in \mathbb{R}^K$ with $x_\ell > 0$ for $\ell \in \mathcal{B}$ and $x_\ell = 0$ otherwise, where $\mathcal{Q}(m + \delta x) > m + \delta x$ for a sufficiently small $\delta > 0$. By the arguments above, there must be exactly one more stable than unstable equilibrium.

I now show that the result extends to any setting where Assumption A holds. In particular, suppose there exists an invertible matrix $\mathbf{B}$ where $\mathbf{BJB}^{-1}$ is non-negative. For $\mathcal{I} = [0,1]^K$, define $\hat{\mathcal{Q}} : \mathbf{B}\mathcal{I} \to \mathbf{B}\mathcal{I}$ such that $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}m)$. The Jacobian matrix of $\hat{\mathcal{Q}}$ is defined by $\mathbf{D}_{\hat{\mathcal{Q}}}(m) = \mathbf{B}\mathbf{D}(\mathbf{B}^{-1}m)\mathbf{B}^{-1}$. Note that $\phi : \mathbb{R}^K \to \mathbb{R}^K$, $\phi(m) = \mathbf{B}m$, is a homeomorphism, so it preserves interior points. It follows that $\mathbf{B}\mathbf{D}(m)\mathbf{B}^{-1}$ is non-negative on int$(\mathcal{I})$ if and only if $\mathbf{D}_{\hat{\mathcal{Q}}}(m) = \mathbf{B}\mathbf{D}(\mathbf{B}^{-1}m)\mathbf{B}^{-1}$ is non-negative on $\mathbf{B}$int$(\mathcal{I})$, which is equal to int$(\mathbf{B}\mathcal{I})$.

This property implies that under Assumption A, the mapping $\hat{\mathcal{Q}}$ is monotone on int$(\mathbf{B}\mathcal{I})$, since $\mathbf{D}_{\hat{\mathcal{Q}}}(m)$ is a non-negative matrix for all $m \in$ int$(\mathbf{B}\mathcal{I})$. Moreover, m is a fixed point of $\mathcal{Q}$ if and only if $\mathbf{B}m$ is a fixed point of $\hat{\mathcal{Q}}$, since $\hat{\mathcal{Q}}(\mathbf{B}m) = \mathbf{B}\mathcal{Q}(m) = \mathbf{B}m$. Furthermore, since $\mathbf{D}(m)$ and $\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}m)$ are similar matrices, they share the same eigenvalues and thus the same spectral radius: $\rho(\mathbf{D}_{\mathcal{Q}}(m)) = \rho(\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}m))$ for all $m \in$ int$(\mathcal{I})$. This allows uniqueness and local stability of fixed points for $\mathcal{Q}$ to be characterized by the monotone map $\hat{\mathcal{Q}}$. While the ordering of $\mathcal{Q}$ and $\hat{\mathcal{Q}}$'s fixed points can differ, their number and stability coincide.

Q.E.D.

PROOF OF PROPERTY 3. If actions take values in the set $\{-1, 1\}$, then $\mathcal{W}_k(m)$ equals:

$$\mathcal{W}_k(m) = \mathrm{E}\left(\left|h_k + \sum_{\ell=1}^{K} J_{k\ell}m^\ell + \varepsilon_i\right|\right) + \eta_k + \mathrm{E}(\xi_i|k).$$

Since each $F_{\varepsilon|k}$ is symmetric about zero, $\mathcal{W}_k(m)$ is strictly increasing in $|h_k + \sum_{\ell=1}^{K} J_{k\ell}m^\ell|$. Also, in equilibrium $|\,\mathrm{E}(\bar{\omega}^k)|$ is strictly increasing in $|h_k + \sum_{\ell=1}^{K} J_{k\ell}m^\ell|$. Therefore, it must be that $\mathrm{argmax}_{m\in\mathcal{M}}\, \mathcal{W}_k(m) = \mathrm{argmax}_{m\in\mathcal{M}}\, |\,\mathrm{E}(\bar{\omega}^k)|$. The second part of Property 3 follows directly from the observations that $\lim_{h_k\to\infty} \mathrm{E}(\bar{\omega}^k) = 1$ and $\lim_{h_k\to-\infty} \mathrm{E}(\bar{\omega}^k) = -1$.

Q.E.D.

PROOF OF PROPERTY 4. Suppose $J_{kj_1}J_{j_1j_2}\cdots J_{j_Mk} \geq 0$ for any types $k, j_1, j_2, \ldots, j_M$. By Lemma A.1 of the Online Appendix, there exists a diagonal matrix $\mathbf{B} = \mathrm{diag}[\gamma_1^k, \ldots, \gamma_K^k]$, where $\gamma_\ell^k = 2\mathbb{1}\{J_{kj_1}J_{j_1j_2}\ldots J_{j_M\ell} \geq 0$ for some types $j_1, j_2, \ldots, j_M\} - 1$, for which $\mathbf{B}\mathbf{J}\mathbf{B}^{-1}$ is non-negative. As shown in the proof of Property 2, $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}m)$ is monotone on $\mathrm{int}(\mathbf{B}\mathcal{I})$. Since $(\mathbf{B}\mathcal{I}, \leq)$ is a complete lattice, Tarski's fixed point theorem ensures that the set of fixed points of $\hat{\mathcal{Q}}$ is a complete lattice and, consequently, contains both a greatest and least fixed point. If $\mathbf{B}\overline{m}$ is the greatest fixed point of $\hat{\mathcal{Q}}(m)$, then $\overline{m} = \mathrm{argmax}_{m\in\mathcal{M}}\, \mathrm{E}(\bar{\omega}^\ell)$ if $\gamma_\ell^k = 1$ and $\overline{m} = \mathrm{argmin}_{m\in\mathcal{M}}\, \mathrm{E}(\bar{\omega}^\ell)$ if $\gamma_\ell^k = -1$. Alternatively, if $\mathbf{B}\underline{m}$ is the least fixed point of $\hat{\mathcal{Q}}(m)$, then $\underline{m} = \mathrm{argmin}_{m\in\mathcal{M}}\, \mathrm{E}(\bar{\omega}^\ell)$ if $\gamma_\ell^k = 1$ and $\underline{m} = \mathrm{argmax}_{m\in\mathcal{M}}\, \mathrm{E}(\bar{\omega}^\ell)$ if $\gamma_\ell^k = -1$.

Q.E.D.

PROOF OF THEOREM 1. Fix k_1 and k_2. Since $F_{\varepsilon|k_1}$ is strictly increasing, it follows that:

$$m_n^{k_1} = F_{\varepsilon|k_1}\left(h_{k_1} - h_{k_2} + \sum_{\ell=1}^{K}(J_{k_1\ell} - J_{k_2\ell})m_n^\ell + F_{\varepsilon|k_2}^{-1}(m_n^{k_2})\right)$$

$$= F_{\varepsilon|k_1}\left(\overline{h_{k_1} - h_{k_2}} + \sum_{\ell=1}^{K} (\overline{J_{k_1\ell} - J_{k_2\ell}}) m_n^\ell + F_{\varepsilon|k_2}^{-1}(m_n^{k_2})\right)$$

if and only if $(h_{k_1} - h_{k_2}) - (\overline{h_{k_1} - h_{k_2}}) = \sum_{\ell=1}^{K} \left[(\overline{J_{k_1\ell} - J_{k_2\ell}}) - (J_{k_1\ell} - J_{k_2\ell}) \right] m_n^\ell$. This property holds for all networks n. Since each $F_{\varepsilon|k}$ is nonlinear, the vectors $\{m_n\}_n$ are not collinear across networks. Sufficient variation in α_n therefore implies that $h_{k_1} - h_{k_2} = \overline{h_{k_1} - h_{k_2}}$ and $J_{k_1\ell} - J_{k_2\ell} = \overline{J_{k_1\ell} - J_{k_2\ell}}$ for any type $\ell \in \{1, \ldots, K\}$. Since k_1 and k_2 are chosen arbitrarily, this result extends to all $k_1, k_2 \in \{1, \ldots, K\}$. The proof in the $\alpha_n = W_n' d$ case is analogous.

Q.E.D.

PROOF OF THEOREM 2. I demonstrate that $\{F_{\varepsilon|k}\}_k$ and c are identified up-to-scale, and the remainder follows as in the proof of Theorem 1. Fix any type k. By Assumption B.3, there is some element x_j of X with continuous variation on $\mathbb{R}$. Without loss of generality, let $x_j = x_1$, and normalize the coefficient c_1 to one. Also, fix some network n, and define $\zeta_n^k = h_k + \alpha_n + \sum_{\ell=1}^{K} J_{k\ell} m_n^\ell$. For any agent i in type k and network n with covariates X_i, the expected choice $\mathrm{E}(\omega_i | X_i, k, \alpha_n, \{m_n^\ell\}_\ell)$ equals $F_{\varepsilon|k}(\zeta_n^k + X_i' c)$. So, to recover $(c, F_{\varepsilon|k})$, I must show that $F_{\varepsilon|k}(\zeta_n^k + X_i' c) = \overline{F}_{\varepsilon|k}(\overline{\zeta}_n^k + X_i' \overline{c})$ implies $c = \overline{c}$ and $F_{\varepsilon|k} = \overline{F}_{\varepsilon|k}$ for any $X_i \in \operatorname{supp}(X|k)$. This property is shown by Corollary 5 of Proposition 2 in Manski (1988). Also, since this holds for all k, I thus conclude that c and $\{F_{\varepsilon|k}\}_k$ are identified up-to-scale.

Q.E.D.

PROOF OF THEOREM 3. I prove part (ii) only, as the proof of part (i) is analogous. As a first step, note that for any pair of types k and ℓ: (1) $\mathrm{E}(\bar{\omega}_{n,j}^k) = \mathrm{E}[\mathrm{E}(m_n^k + u_{n,j}^k | n)] = \mathrm{E}(m_n^k)$ for each $j \in \{a, b\}$; (2) $\mathrm{E}(\bar{\omega}_{n,a}^k \bar{\omega}_{n,b}^\ell) = \mathrm{E}[\mathrm{E}((m_n^k + u_{n,a}^k)(m_n^\ell + u_{n,b}^\ell)|n)] = \mathrm{E}(m_n^k m_n^\ell)$; and (3) $\mathrm{E}(\omega_{n,b}^\ell F_{\epsilon|k}^{-1}(\bar{\omega}_{n,a}^k)) = \mathrm{E}[\mathrm{E}((m_n^\ell + u_{n,b}^k) F_{\epsilon|k}^{-1}(\bar{\omega}_{n,a}^k)|n)] = \mathrm{E}(m_n^\ell F_{\epsilon|k}^{-1}(\bar{\omega}_{n,a}^k))$. Hence, I can write:

$$
\begin{aligned}
\mathrm{E}\left(Z_n X_n'\right)^{-1} \mathrm{E}\left(Z_n F_{\epsilon|k}^{-1}(\bar{\omega}_{n,a}^k)\right) &= \mathrm{E}\left(X_n^*(X_n^*)'\right)^{-1} \mathrm{E}\left(Z_n F_{\epsilon|k}^{-1}(\bar{\omega}_{n,a}^k)\right) && \text{by (1) and (2)} \\
&= \mathrm{E}\left(X_n^*(X_n^*)'\right)^{-1} \mathrm{E}\left(X_n^* F_{\epsilon|k}^{-1}(\bar{\omega}_{n,a}^k)\right) && \text{by (3)} \\
&= \mathrm{E}\left(X_n^*(X_n^*)'\right)^{-1} \mathrm{E}\left(X_n^*(F_{\varepsilon|k}^{-1}(m_n^k) + v_{n,a}^k)\right) \\
&= \underbrace{\mathrm{E}\left(X_n^*(X_n^*)'\right)^{-1} \mathrm{E}\left(X_n^* F_{\varepsilon|k}^{-1}(m_n^k)\right)}_{=\beta_k} + \mathrm{E}\left(X_n^*(X_n^*)'\right)^{-1} \mathrm{E}\left(X_n^* v_{n,a}^k\right),
\end{aligned}
$$

where $X_n^* = (1, W_n', m_n^1, \ldots, m_n^K)'$. As a notational shorthand, let $q_k = F_{\varepsilon|k}^{-1}$. Since q_k is real analytic on $(0, 1)$, it satisfies uniqueness of analytic continuation (Rudin, 1976, Thm. 8.5). I expand $q_k(\bar{\omega}_{n,a}^k)$ in a Taylor series about m_n^k to yield the following expression for $\mathrm{E}(v_{n,a}^k | n)$:

$$\mathrm{E}(v_{n,a}^k | n) = \mathrm{E}\left(q_k(\bar{\omega}_{n,a}^k) - q_k(m_n^k) \middle| n\right)$$

$$
\begin{aligned}
&= \mathrm{E}\left(\sum_{r=0}^{\infty}\frac{q_k^{(r)}(m_n^k)}{r!}(\bar{\omega}_{n,a}^k - m_n^k)^r - q_k(m_n^k)\middle|n\right)\\
&= q_k(m_n^k)\,\mathrm{E}\left(\bar{\omega}_{n,a}^k - m_n^k\middle|n\right) + \frac{q_k''(m_n^k)}{2!}\,\mathrm{E}\left((\bar{\omega}_{n,a}^k - m_n^k)^2\middle|n\right) + \frac{q_k'''(m_n^k)}{3!}\,\mathrm{E}\left((\bar{\omega}_{n,a}^k - m_n^k)^3\middle|n\right) + \dots\\
&= \frac{q_k''(m_n^k)}{2!}\times(I_{n,a}^k)^{-1}\,\mathrm{Var}\,(\omega_i|k,n) + \frac{q_k'''(m_n^k)}{3!}\times(I_{n,a}^k)^{-2}\,\mathrm{E}\left((\omega_i - m_n^k)^3|k,n\right) + \dots\\
&= \sum_{r=2}^{\infty}\frac{q_k^{(r)}(m_n^k)}{r!}B_r\left(0, \frac{\kappa_2(\omega|k,n)}{I_{n,a}^k}, \frac{\kappa_3(\omega|k,n)}{(I_{n,a}^k)^2}, \dots, \frac{\kappa_r(\omega|k,n)}{(I_{n,a}^k)^{r-1}}\right),
\end{aligned}
$$

where B_r is the rth complete Bell polynomial, and $\{\kappa_i(\omega|k,n)\}_i$ are cumulants of ω_i given k and n. Each $B_r(0, \frac{\kappa_2(\omega|k,n)}{I_{n,a}^k}, \frac{\kappa_3(\omega|k,n)}{(I_{n,a}^k)^2}, \dots, \frac{\kappa_r(\omega|k,n)}{(I_{n,a}^k)^{r-1}})$ is therefore a polynomial in $(I_{n,a}^k)^{-1}$ with no constant term and coefficients depending only on cumulants of ω_i given k and n. It follows that $\mathrm{E}(v_{n,a}^k|n) = \sum_{s=1}^{\infty} a_{n,s}^k/(I_{n,a}^k)^s$ for some terms $a_{n,s}^k$ that depend only on the moments of $\omega_i|(k,n)$, and not the subgroup size $I_{n,a}^k$. Noting that $\mathrm{E}(X_n^* v_{n,a}^k) = \mathrm{E}(X_n^*\,\mathrm{E}(v_{n,a}^k|n))$, I write:

$$
\begin{aligned}
\mathrm{E}\left(Z_n X_n'\right)^{-1}\mathrm{E}\left(Z_n F_{\epsilon|k}^{-1}(\bar{\omega}_{n,a}^k)\right) &= \beta_k + \mathrm{E}\left(X_n^*(X_n^*)'\right)^{-1}\mathrm{E}\left(X_n^*\sum_{s=1}^{\infty}\frac{a_{n,s}^k}{(I_{n,a}^k)^s}\right)\\
&= \beta_k + \sum_{s=1}^{\infty} C_{k,s}\,\mathrm{E}\left((I_{n,a}^k)^{-s}\right),
\end{aligned}
$$

where $C_{k,s} = \mathrm{E}\left(X_n^*(X_n^*)'\right)^{-1}\mathrm{E}\left(X_n^* a_{n,s}^k\right)$ are constant vectors that do not depend on $\{I_{n,a}^k\}_n$. A continuum of distributions of $\{I_{n,\kappa(p,n)}^k\}_n$ can be obtained by varying the partitioning rule p, that is, by selecting different (a, a') across networks. Each p yields a distinct sequence $\{\mathrm{E}((I_{n,\kappa(p,n)}^k)^{-s})\}_s$, and thus, by continuity, a distinct value of $\mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n F_{\epsilon|k}^{-1}(\bar{\omega}_{n,\kappa(p,n)}^k))$. Continuous variation of this latter quantity with $\{\mathrm{E}((I_{n,\kappa(p,n)}^k)^{-s})\}_s$ identifies β_k and $\{C_{k,s}\}_s$.

Finally, defining the matrix $\mathbf{E}_k \in \mathbb{R}_{++}^{P\times P}$ such that $[\mathbf{E}_k]_{i,j} = \mathrm{E}[(I_{n,\kappa(j,n)}^k)^{1-i}]$, I can write:

$$
\left[\mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}\left(Z_n F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,\kappa(1,n)}^k)\right), \dots, \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}\left(Z_n F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,\kappa(P,n)}^k)\right)\right] = [\beta_k, C_{k,1}, \dots, C_{k,P-1}]\mathbf{E}_k.
$$

Since $\mathbf{E}_k$ is nonsingular, I can use this property to define a linear IV estimand β_k^{IV} for β_k:

$$
\begin{aligned}
\beta_k^{IV} &= \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n Y_n^k)\\
&= \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n[\Omega_{n,a}^k F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,a}^k) + \Omega_{n,a'}^k F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,a'}^k)])\\
&= \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}\left(Z_n\sum_p\left[\mathbf{E}_k^{-1}\right]_{p,1}\left[\mathbb{1}\{I_{n,\kappa(p,n)}^k = I_{n,a}^k\}F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,a}^k) + \mathbb{1}\{I_{n,\kappa(p,n)}^k = I_{n,a'}^k\}F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,a'}^k)\right]\right)\\
&= \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}\left(Z_n\sum_p\left[\mathbf{E}_k^{-1}\right]_{p,1}F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,\kappa(p,n)}^k)\right)\\
&= \left[\mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}\left(Z_n F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,\kappa(1,n)}^k)\right), \dots, \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}\left(Z_n F_{\varepsilon|k}^{-1}(\bar{\omega}_{n,\kappa(P,n)}^k)\right)\right]\left[\mathbf{E}_k^{-1}\right]_{\cdot,1} = \beta_k.
\end{aligned}
$$

Q.E.D.

REFERENCES

AKERLOF, GEORGE A. AND RACHEL E. KRANTON (2000): “Economics and Identity,” *The Quarterly Journal of Economics*, 115 (3), 715–753. [6, 7]

BERNHEIM, B. DOUGLAS (1994): “A Theory of Conformity,” *Journal of Political Economy*, 102, 841–77. [4]

BHATTACHARYA, DEBOPAM, PASCALINE DUPAS, AND SHIN KANAYA (2023): “Demand and Welfare Analysis in Discrete Choice Models with Social Interactions,” *The Review of Economic Studies*, 91 (2), 748–784. [2, 3, 11, 14, 25]

BONHOMME, STÉPHANE (2012): “Functional Differencing,” *Econometrica*, 80 (4), 1337–1385. [15]

BONHOMME, STÉPHANE AND KEVIN DANO (2023): “Functional Differencing in Networks,” Working Paper. [15]

BOOZER, MICHAEL A. AND STEPHEN E. CACCIOLA (2001): “Inside the ’Black Box’ of Project STAR: Estimation of Peer Effects Using Experimental Data,” Center Discussion Paper 832. [21, 22]

BOXELL, LEVI, MATTHEW GENTZKOW, AND JESSE M. SHAPIRO (2022): “Cross-Country Trends in Affective Polarization,” *The Review of Economics and Statistics*, 1–60. [6]

BRAMOULLÉ, YANN, HABIBA DJEBBARI, AND BERNARD FORTIN (2009): “Identification of peer effects through social networks,” *Journal of Econometrics*, 150 (1), 41–55. [16]

BRAMOULLÉ, YANN AND RACHEL KRANTON (2007): “Public goods in networks,” *Journal of Economic Theory*, 135 (1), 478–494. [6]

BROCK, WILLIAM AND STEVEN DURLAUF (2001): “Discrete Choice with Social Interactions,” *The Review of Economic Studies*, 68 (2), 235–260. [2, 4, 5, 16, 18]

——— (2007): “Identification of Binary Choice Models with Social Interactions,” *Journal of Econometrics*, 140 (1), 52–75. [2, 14, 15, 16]

CHAMBERLAIN, GARY (1980): “Analysis of Covariance with Qualitative Data,” *The Review of Economic Studies*, 47 (1), 225–238. [15]

——— (1984): “Chapter 22 Panel data,” Elsevier, vol. 2 of *Handbook of Econometrics*, 1247–1318. [15]

DANO, KEVIN (2025): “Transition Probabilities and Moment Restrictions in Dynamic Fixed Effects Logit Models,” Working Paper. [15]

ELLISON, GLENN (1993): “Learning, Local Interaction, and Coordination,” *Econometrica*, 61, 1047–1071. [5]

GRAHAM, BRYAN S. (2008): “Identifying Social Interactions through Conditional Variance Restrictions,” *Econometrica*, 76 (3), 643–660. [21, 22]

——— (2015): “Methods of Identification in Social Networks,” *Annual Review of Economics*, 7, 465–485. [2]

GUERRA, JOSÉ-ALBERTO AND MYRA MOHNEN (2022): “Multinomial Choice with Social Interactions: Occupations in Victorian London,” *The Review of Economics and Statistics*, 104 (4), 736–747. [3]

HENRY, DANIEL (1981): *Geometric Theory of Semilinear Parabolic Equations.*, vol. 840 of *Lecture Notes in Mathematics*, Berlin: Springer-Verlag. [28]

HORN, ROGER A. AND CHARLES R. JOHNSON (1985): *Matrix Analysis*, Cambridge University Press. [28]

HOROWITZ, JOEL (2009): *Semiparametric and Nonparametric Methods in Econometrics.*, vol. 692. [13, 16]

HOXBY, CAROLINE M. (2000): “The Effects of Class Size on Student Achievement: New Evidence from Population Variation,” *The Quarterly Journal of Economics*, 115 (4), 1239–1285. [3]

JACKSON, MATTHEW O. AND YVES ZENOU (2015): "Games on Networks," Elsevier, vol. 4 of *Handbook of Game Theory with Economic Applications*, 95–163. [2]

KLINE, BRENDAN AND ELIE TAMER (2020): "Econometric Analysis of Models with Social Interactions," in *Econometric Analysis of Network Data*, ed. by Bryan Graham and Aureo de Paula, Academic Press, 149–181. [2]

KREMER, MICHAEL AND EDWARD MIGUEL (2007): "The Illusion of Sustainability," *The Quarterly Journal of Economics*, 122 (3), 1007–1065. [25]

LEE, LUNGFEI, JI LI, AND XU LIN (2014): "Binary Choice Models with Social Network Under Heterogeneous Rational Expectations," *The Review of Economics and Statistics*, 96 (3), 402–417. [3]

MANSKI, CHARLES F. (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83 (403), 729–738. [13, 15, 17, 30]

——— (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60 (3), 531–542. [16]

MIGUEL, EDWARD AND MICHAEL KREMER (2004): "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72 (1), 159–217. [3, 21, 24, 25]

MILGROM, PAUL AND JOHN ROBERTS (1990): "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *The American Economic Review*, 80 (3), 511–528. [9]

MILNOR, JOHN (1965): *Topology from the Differentiable Viewpoint*, The University Press of Virginia. [27]

PAULA, AUREO DE (2017): *Econometrics of Network Models*, Cambridge University Press, vol. 1 of *Econometric Society Monographs*, 268–323. [2]

RUDIN, WALTER (1976): *Principles of Mathematical Analysis*, New York: McGraw-Hill, 3rd ed. [30]

SALOP, STEVEN (1979): "Monopolistic Competition With Outside Goods," *Bell Journal of Economics*, 10, 141–156. [5]

SCHELLING, THOMAS C. (1971): "Dynamic models of segregation†," *The Journal of Mathematical Sociology*, 1 (2), 143–186. [6]

TOPKIS, D.M. (1998): *Supermodularity and Complementarity*, Frontiers of Economic Research, Princeton University Press. [9]

VIVES, XAVIER (1990): "Nash equilibrium with strategic complementarities," *Journal of Mathematical Economics*, 19 (3), 305–321. [9]

WOOLDRIDGE, J.M. (2013): *Introductory Econometrics: A Modern Approach*, Cengage Learning. [3]

WORD, ELIZABETH R., JOHN JOHNSTON, HELEN PATE BAIN, B. DEWAYNE FULTON, JAYNE B. ZAHARIAS, CHARLES M. ACHILLES, MARTHA NANNETTE LINTZ, JOHN FOLGER, AND CAROLYN BREDA (1990): "The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report," . [22]