Peers as treatments*

Brian Krauth
Simon Fraser University

September 22, 2022

Abstract

Social interaction effects are often estimated under the strong assumption that an individual's choices are a direct function of the observed characteristics of their reference group. This paper considers a less restrictive potential outcomes framework in which interaction with a given peer or peer group is considered a treatment with an unknown and variable treatment effect. In this framework, conventional peer effect regressions can be interpreted as characterizing treatment effect heterogeneity. This framework is then used to clarify identification and interpretation of commonly-used peer effect models and to suggest avenues for improving conventional empirical practice.

JEL codes: C21, C31
Keywords: Peer effects, social interactions, contextual effects

1 Introduction

Empirical researchers often aim to measure the impact of peers or some other reference group on a person's choices or outcomes. Much of this research is based on a behavioral model, generally associated with[1] Manski (1993), in which an individual's outcome responds directly to the observed outcomes (endogenous effects) and characteristics (contextual effects) of peers. Manski's formulation has inspired an extensive literature developing methods for modeling endogenous effects and for empirically distinguishing them from both contextual effects and endogenous peer selection.

The modeling of contextual effects has seen less formal attention despite their prevalence in empirical research. Economic theory provides little guidance on which peer characteristics to include in the model, so some researchers include whatever potentially relevant peer variables are available while others include only a single variable of interest. The results of these varying ad hoc specifications are often difficult to interpret or compare across studies (Fruehwirth, 2014)

*Contact email: `bkrauth@sfu.ca`. Revised versions available at `bvkrauth.github.io/publication/peertreat`

[1]In Manski (1993), behavior responds to the conditional expectation of peer behavior and characteristics, but in most subsequent empirical work it is taken to respond to their observed values. Blume et al. (2011, p. 891-892) discuss this distinction and some of its implications.

in the absence of a unifying framework or model selection criterion. Many of these difficulties are a byproduct of interpreting contextual effects as *direct* and *constant*: in the absence of an endogenous effect, any two peer groups with the same observed characteristics are assumed to have the exact same effect on outcomes. This interpretation imposes strong data requirements and identifying assumptions in order to estimate the model and make relevant counterfactual predictions. More specifically, the estimated model must include all potentially relevant peer characteristics, and the counterfactuals must be defined in terms of their impact on all of those characteristics. These requirements are unlikely to be met in most applications, and may lead to substantial omitted variables bias.

This paper describes an alternative "peers as treatments" formulation in which each person has an unobserved and person-specific influence on peer outcomes. This influence is analogous to a standard treatment effect, but each person represents a distinct treatment whose effect on peers may vary across treated individuals and with the other group members. An person's effect on peers may be correlated with observed background characteristics, but need not be an exact function of these characteristics as in the traditional model. In this setting, estimated contextual effects are best understood as describing treatment effect heterogeneity along researcher-selected dimensions. The framework can then be used to define causal peer effects in terms of explicit counterfactuals, to state conditions under which they are identified, and to provide simple estimators.

The implications of this model support and clarify many common empirical practices. The peer effects defined in this paper can usually be estimated by linear regressions similar to those regularly used in empirical research. Simple linear models provide useful information, and researchers can use different model specifications to explore different dimensions of peer effect heterogeneity without needing to take a stand on the "true" model. Peer effects can be identified using simple random assignment of peers, random assignment based on observable characteristics, or (with some caveats) random cohorts or subgroups within endogenously-selected larger groups.

At the same time, the model implies clear recommendations and constraints for future empirical work. First, parsimonious specifications with a few binary or categorical explanatory variables are more robustly informative than the ad hoc specifications with many variables that often appear in empirical research. Second, the precise source of identifying randomness in peer group formation has subtle but important implications for the set of counterfactuals that can be credibly assessed. For example, the random cohort research design commonly used to measure classroom peer effects only identifies the impact of counterfactual student allocations *within* the school, and say little about cross-school reallocations unless the researcher is willing to impose strong assumptions. Finally, while the results here show almost any form of peer effect heterogeneity can be identified and interpreted, credible statistical inference on peer effect heterogeneity is hampered by the risk of spurious inferences resulting from unreported specification search and multiple testing. Techniques from the literature on treatment effect heterogeneity such as pre-analysis plans or more careful data-driven inference (Wager and Athey, 2018) can in principle be adapted to this setting, though the detailed implementation is left to future research.

1.1 Related literature

The contemporary economics literature on measuring social effects has been primarily aimed at addressing the challenges described by Manski (1993): distinguishing true social effects from spurious social effects due to nonrandom peer selection or unobserved common shocks, and distinguishing endogenous social effects from contextual social effects. Subsequent empirical research has addressed the first of these issues by exploiting natural experiments in which peer group assignment is affected by purely random factors, while methodological research has addressed the second issue by exploiting nonlinearity (Brock and Durlauf, 2000), exclusion restrictions (Gaviria and Raphael, 2001), or social network structure (Graham, 2008; Bramoullé et al., 2009). When endogenous and contextual effects cannot be separately identified, a common solution is to estimate a model with only contextual effects and either assume no endogenous effects or interpret the regression model as the reduced form of a more general structural model with endogenous effects.

The related empirical literature is vast, and much of it emphasizes contextual effects. For example, the classroom peer effects literature includes hundreds of papers on how student outcomes (typically but not always test scores) are affected by observed peer ability, peer effort, peer gender, peer race and ethncity, peer personality, peer mental health, disruptive peers, peers with special needs, peers speaking English as a second language, etc. Other papers (Arcidiacono et al., 2012; Isphording and Zölitz, 2020) measure the effect of a more general concept of unobserved peer "quality" as inferred from individual fixed effects. A detailed survey of findings on classroom peer effects is beyond the scope of the current paper, but several general conclusions can be drawn: peer characteristics often matter, and they can matter in ways that are not fully described by a simple one-dimensional peer quality measure. For example, several papers find that students with learning disabilities (which have a negative effect on own achievement) have a positive effect on peer achievement, and boys are regularly found to reduce peer achievement even in subjects where boys perform as well as girls. In addition, the various dimensions along which peers seem to matter are clearly related: language and ethnicity are nearly inseparable, as are gender and behavior. Changing one contextual factor through classroom assignments will tend to change other related factors, making it difficult to reach clear policy conclusions on the consequences of alternative peer group assignment mechanisms.

Much of this applied work follows Manski (1993) in treating the contextual effect as a direct and constant function of peer characteristics. As in the current paper, more recent methodological research has used a treatment effects framework to relax these assumptions and clarify the counterfactual policies that can be assessed under a given set of model assumptions. Manski (2013) and Li et al. (2019) relax the assumption that peer effects are constant across treated individuals while retaining the assumption that they depend directly on the observed peer characteristics. In Manski (2013), the relevant peer characteristics are manipulable individual-level treatments that have variable effects on both own and peer outcomes. Peer groups are fixed, peer effects are identified through random assignment to treatment, and the policy of interest is a counterfactual assignment of treatments. In Li et al. (2019), the relevant peer characteristics are non-manipulable background characteristics, and each person's observed characteristics have a direct effect on peer outcomes that varies across the treated individuals but not across peers with a given set of observed characteristics. Peer effects are identified through random assignment of individuals to peer groups, and the policy of interest is a counterfactual peer group assignment.

Graham et al. (2010) is similar to this paper in allowing peer effects to be both variable and indirect. That is, the effect of one person on another may depend on unobserved characteristics of both indviduals. In their model, observed peer characteristics do not directly affect the outcome but are imperfect proxies for unobserved peer characteristics that do. Their policy of interest is a counterfactual peer group assignment, as in Li et al. (2019) and this paper. The analysis and results in this paper are complementary to those in Graham et al. (2010), but differ in several important ways:

1. Graham et al. (2010) consider a single binary individual characteristic (e.g., race), while this paper considers a richer (categorical) characteristics space.
2. Graham et al. (2010) assume peer groups are large enough that peer group composition can be treated as a continuous variable. As a result:
   (a) Estimation is based on nonparametric kernel regressions, their derivatives and various integrals/averages of those derivatives.
   (b) As Graham et al. (2010) note, this assumption implies that "our estimands and estimators are not appropriate for situations where groups are small (e.g., college roommates)'.'

   In contrast, this paper assumes peer groups are small (finite) so that peer group composition is a discrete variable. This property facilitates the use of linear models, and fits many applications - classrooms, roommates, close friends, etc. - better than the "large groups" assumption.
3. Graham et al. (2010) model the observed characteristics as independent of unobserved heterogeneity, while this paper models the observed characteristics as a function of unobserved heterogeneity. The two formulations are substantively equivalent (one can map one model to the other by redefining variables), but the formulation here helps to separate practical issues of specification choice from core identifying assumptions about causal mechanisms.

Additional details on these differences are provided throughout the paper. A more general difference is this paper's emphasis on explaining, clarifying and improving upon current empirical practice.

Finally, this paper is among several that use estimated peer effect models to to predict the consequences of counterfactual allocations of individuals to peer groups. Bhattacharya (2009) develops algorithms to find optimal assignments from a given set of model estimates. Carrell et al. (2013) report the results of a field experiment that uses peer effect estimates from one class cohort of students to construct presumably[2] optimal allocations for a later cohort. Graham et al. (2010) note that the large changes needed to reach an optimal group assignment are typically infeasible and emphasize tools for predicting the effect of more feasible local reallocations like a small reduction in segregation.

[2]Notably, this "optimal" allocation yielded surprisingly poor results, providing a cautionary tale on the risks of large scale social engineering from restrictive models estimated with limited data.

2 Model

This section develops the basic model. Section 3 defines various causal social effects within the model environment, Section 4 establishes the main results, and Sections 5 and 6 discuss various extensions. The model's exposition will refer to a running example application to studying the effect of classroom gender[3] composition on an academic achievement as measured by test scores. This question has been investigated extensively in the empirical literature, for example by Hoxby (2000), Lavy and Schlosser (2011) and Eisenkopf et al. (2015). This research typically finds a substantial positive effect of female peers, even in settings and academic subjects where boys and girls have similar average outcomes. It is thus a natural application of this model, which does not assume that peers can be ordered in a single quality dimension.

2.1 Basic framework and notation

The model features a population of heterogeneous **individuals** arbitrarily indexed by $i \in \mathcal{N} \equiv \{1, 2, \ldots, N\}$. Each individual is fully characterized by an **unobservable type** $\tau_i \in \mathcal{T} \equiv \{1, 2, \ldots, T\}$ and membership in some social **group** $g_i \in \mathcal{G} \equiv \{1, 2, \ldots, G\}$. The population as a whole is fully characterized by the random N-vectors $\mathbf{T} \in \mathcal{T}^N$ and $\mathbf{G} \in \mathcal{G}^N$, in the sense that all random variables in the model are functions of $(\mathbf{T}, \mathbf{G})$.

An individual's type τ_i represents everything about the individual that is potentially relevant in this domain. The type space is finite to allow the use of elementary probability theory, but it can be quite large so that each type is typically unique to a particular individual. The ordering of the type space is arbitrary; nearby types are not necessarily more similar, and types do not necessarily correspond to some scalar "quality" index that is monotonically related to outcomes.

Group membership is determined by some **group selection mechanism** which is a discrete conditional PDF of the form:

$$f_{\mathbf{G}|\mathbf{T}}(\mathbf{G_A}, \mathbf{T_A}) \equiv \Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{T} = \mathbf{T_A}) \tag{1}$$

for any fixed vector of group assignments $\mathbf{G_A} \in \mathcal{G}^N$ and types $\mathbf{T_A} \in \mathcal{T}^N$.

Each individual experiences a scalar **outcome** of interest $y_i \in \mathbb{R}$ which depends on both the individual's own type and that of other group members:

$$\mathbf{Y} \equiv \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \equiv \begin{bmatrix} y_1(\mathbf{T}, \mathbf{G}) \\ \vdots \\ y_N(\mathbf{T}, \mathbf{G}) \end{bmatrix} \equiv \mathbf{Y}(\mathbf{T}, \mathbf{G}) \tag{2}$$

The model does not include a direct causal effect of peer outcomes ("endogenous effects" in the language of Manski 1993) but it can be interpreted as the reduced form of such a model.

For each individual i, the researcher observes the peer group g_i, the outcome y_i and a vector of individual background **characteristics** $\mathbf{x}_i \in \mathbb{R}^K$. These background characteristics are

[3] Following previous research and many available data sources, gender is treated as binary throughout the example.

predetermined[4] and depend only on one's own type:

$$\mathbf{X} \equiv \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \equiv \begin{bmatrix} \mathbf{x}(\tau_1) \\ \vdots \\ \mathbf{x}(\tau_N) \end{bmatrix} \equiv \mathbf{X}(\mathbf{T}) \tag{3}$$

Note that observed characteristics do not directly affect the outcome, though there will typically be a relationship between observed characteristics and outcomes via their shared dependence on the unobserved type. This is a key feature of this model: the explanatory variables are not assumed to be part of some "true" causal model, but rather have been chosen by the researcher based on data availability and researcher interest. Another researcher might choose different explanatory variables, and both choices could lead to interesting and valid causal results.

Example 1 (Gender peer effects in school)**.** *A researcher has data on N students allocated across G classrooms and aims to measure the effect of classmate gender on test scores. In this setting, the outcome and peer group variables would be:*

$$y_i \equiv \text{student } i\text{'s test score}$$
$$g_i \equiv \text{classroom ID for student } i$$

and the observed characteristics in $\mathbf{x}_i$ would include student i's gender along with any other characteristics the researcher chooses fom the available data. The unobserved type τ_i would represent everything in $\mathbf{x}_i$ along with student i's ability, past academic and nonacademic experiences, personality, family and neighborhood context, mental and physical health, special needs, and any other potentially relevant indvidual-level factors.

2.2 Maintained assumptions

This section states some basic assumptions that will be maintained throughout the analysis.

Assumption 1 (Independent types)**.** *Each individual's type is an independent draw from a common type distribution:*

$$\Pr(\mathbf{T} = \mathbf{T_A}) = \prod_{i=1}^{N} f_\tau(\tau_i(\mathbf{T_A})) \tag{4}$$

where $f_\tau : \mathcal{T} \to [0,1]$ is some unknown discrete PDF.

Assumption 1 is innocuous: the indexing of individuals is arbitrary, so unconditional independence is supported by standard exchangeability arguments. This unconditional independence does not imply independence of types conditional on observed characteristics $\mathbf{X}$ or group memberships $\mathbf{G}$.

Assumption 2 (Constant group size)**.** *Each peer group in $\mathbf{G}$ has exactly n members, which in turn implies that $N = nG$.*

[4] That is, they are not treatments that can be manipulated by a policy maker, as in Manski (2013).

Assumption 2 is a standard assumption that simplifies exposition. Variable group size can be accommodated by including group size as a conditioning/explanatory variable.

Assumption 3 (Group interactions)**.** *Given individual types and peer groups, the outcome for individual i is:*

$$y_i(\mathbf{T}, \mathbf{G}) = y\left(\tau_i, \{\tau_j\}_{g_i=g_j}\right) \tag{5}$$

where $y : \mathcal{T}^n \to \mathbb{R}$ is an unknown function.

Assumption 3 implies anonymous/exchangeable spillovers within peer groups, no spillovers across peer groups, no direct effects of group assignment itself itself, and no post-assignment random factors that might affect the outcome. Direct effects of group assignment and post-assignment random factors are common in applied work but can be incorporated into the model in various ways for specific applications.

Assumption 4 (Discrete characteristics)**.** *The support of $\mathbf{x}_i$ is:*

$$\mathbf{S_x} \equiv \{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_K\} \tag{6}$$

where $\mathbf{e}_k \in \{0,1\}^K$ is a unit vector containing one in column k and zero elsewhere; and its probability distribution[5] *is fully described by:*

$$\begin{aligned} \mu_s &\equiv \Pr(\mathbf{x}_i = \mathbf{e}_s) && \text{(for all } s \in 0, 1, \ldots, K) \\ \mu &\equiv E(\mathbf{x}_i) = \begin{bmatrix} \mu_1 & \cdots & \mu_K \end{bmatrix} && (7) \end{aligned}$$

Assumption 4 abstracts from functional form considerations by taking the observable characteristics $\mathbf{x}_i$ to be a K-vector of categorical dummy variables. If the original set of individual characteristics does not have this structure, the researcher can generate this structure by binning continuous variables, including interactions, etc.

Assumption 5 (Information available to the researcher)**.** *The researcher directly observes the joint distribution of* $(\mathbf{X}, \mathbf{Y}, \mathbf{G})$.

Assumption 5 abstracts from sampling considerations in order to focus on identification issues. Sampling designs can vary substantially in applied work on social interactions, as the presence of social interaction effects generally violates the simple random sampling assumption. The identification results in this paper are constructive, and suggest estimators based on linear regression coefficients whose statistical properties are well-understood in a wide variety of sampling designs.

Example 2 (Variable selection for gender peer effects)**.** *Continuing the gender peer effects example, suppose the researcher decides to only include a single ($K = 1$) binary gender variable:*

$$\mathbf{x}_i \equiv \begin{cases} 1 & \textit{if student } i \textit{ is male} \\ 0 & \textit{if student } i \textit{ is female} \end{cases}$$

[5] Note that $\mu_0 = 1 - \sum_{s=1}^{K} \mu_s$ is not included in the vector μ but can be expressed as a function of it, so it will be treated as if it were part of μ when convenient to do so.

This choice satisfies Assumption 4.

If the researcher also wishes to include a lagged test score in the model, Assumption 4 could be satisfied by binning the test score (e.g. into quartiles or deciles), and interacting the binned test score with gender. With b bins for the test score, $\mathbf{x}_i$ *would be a unit vector of length* $K = 2b - 1$.

2.3 Optional assumptions

This section defines several additional assumptions that are *not* maintained throughout the paper, but rather are required for particular propositions.

The first set of optional assumptions constrain the group selection mechanism. As usual, some source of purely random variation in treatment status is needed to identify causal effects. In this setting, causal inference will require some form of random group selection. Simple random assignment is the most straightforward scenario, but the weaker assumption of conditional random assignment is often sufficient for causal identification. Section 4.2 shows the role of random assignment in identification.

Definition 1 (Simple random assignment)**.** *The group selection mechanism* $f_{\mathbf{G}|\mathbf{T}}$ *satisfies* **simple random assignment (RA)** *if:*

$$\mathbf{G} \perp\!\!\!\perp \mathbf{T} \tag{\textbf{RA}}$$

i.e., peer group assignment does not depend on one's unobservable type or any other predetermined characteristics.

Definition 2 (Conditional random assignment)**.** *The group selection mechanism* $f_{\mathbf{G}|\mathbf{T}}$ *satisfies* **conditional random assignment (CRA)** *based on observed characteristics if:*

$$\mathbf{G} \perp\!\!\!\perp \mathbf{T}|\mathbf{X} \tag{\textbf{CRA}}$$

i.e., peer group assignment may depend on one's observable characteristics but does not otherwise depend on one's unobservable type.

Note that simple random assignment does not constrain the researcher's choice of background characteristics to include in $\mathbf{x}_i$, while conditional random assignment requires $\mathbf{x}_i$ to include all characteristics that affect group assignment.

The second set of optional assumptions constrain the outcome function to be separable in various ways. Separability is typically not required for identification, but it simplifies analysis and interpretation. Section 4.1 shows how separability assumptions facilitate relatively simple and easily-interpreted empirical models such as the linear in means model.

Definition 3 (Peer separability)**.** *Peer effects are* **peer-separable (PSE)** *if the effect of replacing one peer with another does not depend on one's other peers:*

$$y\left(\tau_i, \left\{\tau_j', \tau\right\}\right) - y\left(\tau_i, \left\{\tau_j, \tau\right\}\right) = y\left(\tau_i, \left\{\tau_j', \tau'\right\}\right) - y\left(\tau_i, \left\{\tau_j, \tau'\right\}\right) \tag{\textbf{PSE}}$$

for any $\tau_i, \tau_j, \tau_j' \in \mathcal{T}$ *and* $\tau, \tau' \in \mathcal{T}^{n-2}$.

Definition 4 (Own separability)**.** *Peer effects are* **own-separable (OSE)** *if the effect of replacing one peer group with another does not depend on one's own type:*

$$y(\tau_i, \{\tau'\}) - y(\tau_i, \{\tau\}) = y(\tau_i', \{\tau'\}) - y(\tau_i', \{\tau\}) \tag{\textbf{OSE}}$$

for any $\tau_i, \tau_i' \in \mathcal{T}$ *and* $\tau, \tau' \in \mathcal{T}^{n-1}$.

Peer effects that are neither own-separable nor peer-separable will be called **non-separable**. Note that separability is a constraint on how unobserved types enter into the outcome, and is not dependent on the specific characteristics in $\mathbf{x}_i$.

3 Defining social effects

Given this model, we can now define causal social effects in terms of an explicit potential outcome function and set of counterfactuals. Individual characteristics $(\mathbf{T}, \mathbf{X})$ are predetermined (as in Graham et al. (2010)) rather than manipulable (as in Manski (2013)), so the applicable counterfactuals in this model describe the peer group assignment $\mathbf{G}$.

In a fixed population of individuals, changes to one peer group imply corresponding changes to other peer groups. For example, a school can only increase the number of boys in one classroom by reducing the number of boys in another classroom. As a result, counterfactual peer group assignments can be conceptualized in either of two distinct ways:

- **Peer group effects**: The predicted response of a typical individual to changing a single peer (peer effects) or their entire peer group (group effects).
- **Reallocation effects**: The predicted response of the population to a feasible reallocation of peers.

In addition, these effects can be defined for the average individual in the population (average effects), or conditioning on observed characteristics of the treated individual (conditional effects). Combining these features yields six causal effects of potential interest that are summarized in Table 1 and defined later in this section.

3.1 Peer effects

Individual (observed and unobserved) characteristics are predetermined, so causal peer effects for a given individual are defined in terms of an individual-specific potential outcome function that treats the peer group assignment as the relevant counterfactual.

Definition 5 (Potential outcomes)**.** *Individual i's* **peer group** *is defined as:*

$$\mathbf{p}_i \equiv \mathbf{p}(i, \mathbf{G}) \equiv \{j \neq i : g_j = g_i\} \tag{8}$$

and their **potential outcome function** *is defined as:*

$$y_i(\mathbf{p}) \equiv y\left(\tau_i, \{\tau_j\}_{j \in \mathbf{p}}\right) \tag{9}$$

Estimand	Treated unit	Treatment
Average peer effect (APE)	Average person	Replace one peer
Conditional peer effect (CPE)	Person w/given characteristics	Replace one peer
Average group effect (AGE)	Average person	Replace the peer group
Conditional group effect (CGE)	Person w/given characteristics	Replace the peer group
Average reallocation effect (ARE)	Everyone	A feasible reallocation
Conditional reallocation effect (CRE)	Everyone w/given characteristics	A feasible reallocation

Table 1: Social effects defined for this model.

where the counterfactual peer group $\mathbf{p}$ *is any size* $n-1$ *subset of* $\mathcal{N} \setminus \{i\}$.

That is, the observed outcome for individual i is $y_i(\mathbf{p}_i)$, and $y_i(\mathbf{p})$ is the counterfactual outcome that would have been observed if individual i had instead been assigned the peer group $\mathbf{p}$. The potential outcome function is not directly observable but can be used to define various causal peer effects, all based on the idea of replacing randomly-selected peers with one set of observed characteristics with randomly-selected new peers who have different observed characteristics.

The first estimand to be defined is the average effect of replacing a single peer of one observable type with a single peer of another observable type.

Definition 6 (Average peer effect)**.** *The* **average peer effect** *(APE_k) of peers of observed type k relative to peers of observed type zero is defined as:*

$$APE_k \equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{p}}) - y_i(\{j'\} \cup \tilde{\mathbf{p}}) | \, \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0\right) \tag{10}$$

where $\tilde{\mathbf{p}}$ *is a purely random draw of* $n-2$ *peers from* $\mathcal{N} \setminus \{i, j, j'\}$*, and the* $1 \times K$ *matrix of such effects is defined as:*

$$\mathbf{APE} \equiv \begin{bmatrix} APE_1 & APE_2 & \cdots & APE_K \end{bmatrix} \tag{11}$$

Although equation (9) looks complex, the concept is simple. Take a randomly-selected individual with a randomly-constructed peer group, and replace a randomly-selected peer from the base group with a randomly-selected peer from group k. The average peer effect is the predicted change in this individual's outcome.

Average peer effects can be interpreted as describing the heterogeneity of peers across identifiable groups, and are thus analogous to the conditional average treatment effect estimated in the literature on heterogeneous treatment effects (e.g. Wager and Athey (2018)). One difference from that setting is that there is no natural "untreated" state, so average peer effects are defined relative to the average peer in an arbitrarily selected base category. Regardless of the base category chosen, $APE_\ell - APE_k$ can be interpreted as the average effect of replacing an average peer from category k with an average peer from category ℓ.

Rather than averaging across all individuals, researchers may also be interested in how peer effects vary with the observed characteristics of the treated individual:

Definition 7 (Conditional peer effect)**.** *The* **conditional peer effect** *($CPE_{s,k}$) of peers of observed type k on individuals of observed type s is defined as:*

$$CPE_{s,k} \equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{p}}) - y_i(\{j'\} \cup \tilde{\mathbf{p}})\right| \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0) \tag{12}$$

where $\tilde{\mathbf{p}}$ is a purely random draw of $n-2$ peers from $\mathcal{N} \setminus \{i, j, j'\}$, and the $(K+1) \times K$ matrix of such effects is defined as:

$$\mathbf{CPE} \equiv \begin{bmatrix} CPE_{0,1} & \cdots & CPE_{0,K} \\ \vdots & \ddots & \vdots \\ CPE_{K,1} & \cdots & CPE_{K,K} \end{bmatrix} \tag{13}$$

That is, $CPE_{s,k}$ can be considered the effect on the typical category-s individual of replacing a typical base-category peer with a typical category-k peer.

Note that **APE** and **CPE** are both well-defined under the model's maintained assumptions and do not require the data generating process satisfies optional assumptions such as separability or random assignment. However, these assumptions may affect identification and interpretation.

3.2 Group effects

The effect of one peer may depend on the other peer group members. For example, the effect of a disruptive student on classmates may depend on whether there are other disruptive students in the classroom. As a result, researchers may wish to analyze peer effects at the group level rather than the individual level. Such an analysis requires the researcher to construct variables describing the group as a whole, starting with the simple peer group average:

Definition 8 (Peer average characteristics)**.** *Let* **peer average characteristics** *for individual i be defined as:*

$$\bar{\mathbf{x}}_i = \bar{\mathbf{x}}_i(\mathbf{X}, \mathbf{G}) = \frac{1}{n-1} \sum_{j \neq i : g_j = g_i} \mathbf{x}_j \tag{14}$$

Empirical research on peer effects often uses the "linear in means" model, which models the outcome as a linear function of own characteristics $\mathbf{x}_i$ and peer average characteristics $\bar{\mathbf{x}}_i$. Other studies (Hoxby and Weingarth, 2005) argue that the linear in means model is too restrictive, and emphasize specifications that include nonlinear functions of peer characteristics: measures of within-group heterogeneity, threshold or "critical mass" effects, etc.

The maintained assumptions described in Section 2.2 greatly simplify the modeling of nonlinearity in this setting. Observed characteristics $\mathbf{x}_i$ are categorical by Assumption 4 and group size n is fixed by Assumption 2, so $\bar{\mathbf{x}}_i$ has a finite support and provides a complete description of the frequency distribution of observed characteristics among i's peers. As a result, any nonlinear function of observed peer characteristics can be expressed as a linear function of some categorical variable constructed by binning or otherwise dicretizing $\bar{\mathbf{x}}_i$.

Definition 9 (Binned peer group variable)**.** *Let the* **binned peer group variable** $\mathbf{z}_i \in \{0,1\}^M$ *be defined by:*

$$\mathbf{z}_i = \mathbf{z}(\bar{\mathbf{x}}_i) = \sum_{m=1}^{M} \mathbf{e}_m \mathbb{I}\left(\bar{\mathbf{x}}_i \in \mathbf{S}_{\bar{\mathbf{x}}}^m\right) \tag{15}$$

where $(\mathbf{S}_{\bar{\mathbf{x}}}^0, \mathbf{S}_{\bar{\mathbf{x}}}^1, \ldots, \mathbf{S}_{\bar{\mathbf{x}}}^M)$ *is a partition of* $\mathbf{S}_{\bar{\mathbf{x}}}$ *(the support of* $\bar{\mathbf{x}}_i$*), and* $\mathbf{e}_m$ *is the unit vector containing one in column m and zero elsewhere. Bin m is a* **singleton** *if* $|\mathbf{S}_{\bar{\mathbf{x}}}^m| = 1$ *and* **pooled** *if* $|\mathbf{S}_{\bar{\mathbf{x}}}^m| > 1$*. A partition that consists only of singleton bins is* **saturated***, as is the associated binned variable.*

The group variable $\mathbf{z}_i$ is defined by the econometrician, and can include any mix of singleton and pooled categories. For reference, it will also be useful to define a second group variable based on a specific saturated partition.

Definition 10 (Saturated peer group variable)**.** *Let the* **saturated peer group variable** $\mathbf{z}_i^{sat} \in \{0,1\}^{M^{sat}}$ *be defined*[6] *by:*

$$\mathbf{z}_i^{sat} = \mathbf{z}^{sat}(\bar{\mathbf{x}}_i) = \mathbf{e}_{(n-1)\sum_{k=1}^{K} \bar{\mathbf{x}}_{ik} n^{k-1}} \tag{16}$$

Note that the econometrician could choose $\mathbf{z}_i = \mathbf{z}_i^{sat}$, but in practice there will be a bias/variance trade off as the number of observations per category decrease in the number of bins.

Example 3 (Variables for gender peer effects)**.** *Continuing the gender peer effects example, the gender composition of student i's classroom is fully described by* $\bar{\mathbf{x}}_i \in [0,1]$*, the proportion of classmates who are male. Its support is* $\mathbf{S}_{\bar{\mathbf{x}}} = \left\{0, \frac{1}{n-1}, \ldots, 1\right\}$ *which has* $|\mathbf{S}_{\bar{\mathbf{x}}}| = n$ *elements. The researcher can construct various categorical variables from* $\bar{\mathbf{x}}_i$ *including:*

- *Majority-female or majority-male (*$M = 1$*):*

$$\mathbf{z}_i = \mathbf{z}(\bar{\mathbf{x}}_i) = \begin{cases} 0 & \text{if } 0.0 \leq \bar{\mathbf{x}}_i \leq 0.5 \\ 1 & \text{if } 0.5 < \bar{\mathbf{x}}_i \leq 1.0 \end{cases}$$

[6] Note that defining a specific variable requires choosing a specific ordering of categories, but the ordering is arbitrary. That is, there is nothing relevant about the formula $(n-1)\sum_{k=1}^{K} \bar{\mathbf{x}}_{ik} n^{k-1}$ other than it is one of many ways to assign a different bin number to each possible value of $\bar{\mathbf{x}}_i$.

- *All-female, all-male, and mixed ($M = 2$):*

$$\mathbf{z}_i = \mathbf{z}(\bar{\mathbf{x}}_i) = \begin{cases} [\ 0 \quad 0\ ] & \text{if } \bar{\mathbf{x}}_i = 0.0 \\ [\ 1 \quad 0\ ] & \text{if } 0.0 < \bar{\mathbf{x}}_i < 1.0 \\ [\ 0 \quad 1\ ] & \text{if } \bar{\mathbf{x}}_i = 1.0 \end{cases}$$

- *A saturated variable ($M = n - 1$) that nests all other options:*

$$\mathbf{z}_i = \mathbf{z}(\bar{\mathbf{x}}_i) = \begin{cases} [\ 0 \quad 0 \quad \dots \quad 0\ ] & \text{if } \bar{\mathbf{x}}_i = 0.0 \\ [\ 1 \quad 0 \quad \dots \quad 0\ ] & \text{if } \bar{\mathbf{x}}_i = \frac{1}{n-1} \\ & \vdots \\ [\ 0 \quad 0 \quad \dots \quad 1\ ] & \text{if } \bar{\mathbf{x}}_i = 1.0 \end{cases}$$

Given a researcher's choice of $\mathbf{z}_i$, one can define peer *group* effects with or without conditioning on the characteristics of the treated individual:

Definition 11 (Group effects)**.** *The* **average group effect** *of a bin m peer group (relative to a bin zero peer group) is defined as:*

$$AGE_m \equiv E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \tag{17}$$

and the **conditional group effect** *of a bin m peer group on category s individuals is defined as:*

$$CGE_{s,m} \equiv E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \tag{18}$$

where $\tilde{\mathbf{p}}$ is a purely random draw[7] of $n - 1$ peers from $\mathcal{N} \setminus \{i\}$. The M-vector **AGE** *and the $(K + 1) \times M$ matrix* **CGE** *are also defined accordingly.*

The average group effect can be interpreted as the effect on a randomly-selected individual of replacing a randomly constructed bin-zero peer group with a randomly constructed bin-m peer group, and the conditional group effect is the same quantity for a randomly-selected individual from a particular category. As with average and conditional peer effects, average and conditional group effects are well-defined under the maintained assumptions of the model, though their identification and interpretation may depend on additional conditions.

Example 4 (Peer group effects for gender peer effects)**.** *Continuing the gender peer effects example, let $\mathbf{z}_i = \mathbb{I}(\bar{\mathbf{x}}_i > 0.5)$ be an indicator for whether the peer group is majority male. For convenience, assume n is even so there are no exactly-balanced peer groups. Then the following causal effects can be defined:*

[7] Note that AGE_m and $CGE_{s,m}$ are defined in terms of a purely random draw of peers, and thus imposes a particular conditional distribution for $\Pr(\bar{\mathbf{x}}_i|\mathbf{z}_i)$. Proposition 6 in Section 4.3 shows that AGE_m and $CGE_{s,m}$ are only informative about peer group reallocations that preserve this conditional distribution (e.g., if $\mathbf{S}^m_{\bar{\mathbf{x}}}$ is a singleton). See Section 4.3 for additional details.

- APE_1 *is the effect on the average student of replacing an average female peer with an average male peer.*
- CPE_{11} *is the effect on the average male student of replacing an average female peer with an average male peer.*
- CPE_{01} *is the effect on the average female student of replacing an average female peer with an average male peer.*
- AGE_1 *is the effect on the average student of replacing the average majority-male peer group with the average majority-female peer group.*
- CGE_{01} *is the effect on the average female student of replacing the average majority-male peer group with the average majority-female peer group.*
- CGE_{11} *is the effect on the average male student of replacing the average majority-male peer group with the average majority-female peer group.*

3.3 Reallocation effects

The peer and group effects defined in Sections 3.1 and 3.2 predict the effect of a change in the composition of a representative individual's peer group. As discussed above, with a fixed population any change in the composition of one peer group implies a corresponding change in the composition of at least one other peer group. As a result, we may also be interested in the somewhat different question of reallocation effects: how average outcomes are affected by a feasible change to the entire social network **G** (Bhattacharya, 2009; Graham et al., 2010). This section defines both feasible reallocations and the corresponding reallocation effects.

As with peer group effects, the first step is to define the relevant counterfactual, which in this case is a feasible reallocation. Policymakers generally do not have information on unobserved types, so the feasible reallocations of interest will be based on the observed characteristics **X** and can also include a random component.

Definition 12 (Feasible reallocation)**.** *A* **feasible reallocation** *is a function* $\mathbf{G}_R : \mathbb{R}^{N+1} \to \mathcal{G}^N$ *such that the counterfactual peer group allocation* $\mathbf{G}_R(\mathbf{X}, \sigma)$ *satisfies Assumption 2 for a given randomization device* $\sigma|\mathbf{T} \sim U(0,1)$*.*

For example, a researcher might wish to compare single-gender, balanced, and randomly-mixed classroom assignments. The randomization device σ allows the researcher to average over a conditional probability distribution for these group assignments rather than specify a particular group assignment. Note that feasible reallocations satisfy conditional random assignment (CRA) by construction.

Definition 13 (Reallocation effects)**.** *Let* $\mathbf{G}_0 = \mathbf{G}_0(\mathbf{X}, \sigma)$ *and* $\mathbf{G}_1 = \mathbf{G}_1(\mathbf{X}, \sigma)$ *be two feasible reallocations. The* **average reallocation effect** *of a change from allocation* $\mathbf{G}_0$ *to allocation* $\mathbf{G}_1$ *is defined as:*

$$ARE(\mathbf{G}_0, \mathbf{G}_1) \equiv E(y_{i1} - y_{i0}) \tag{19}$$

and the **conditional reallocation effect** *of that same change is defined as:*

$$CRE_s(\mathbf{G}_0, \mathbf{G}_1) \equiv E(y_{i1} - y_{i0} | \mathbf{x}_i = \mathbf{e}_s) \tag{20}$$

$$\mathbf{CRE}(\mathbf{G}_0, \mathbf{G}_1) \equiv \begin{bmatrix} CRE_0 & CRE_1 & \cdots & CRE_K \end{bmatrix} \tag{21}$$

where $y_{iR} = y_i(\mathbf{p}(i, \mathbf{G}_R))$ *is the counterfactual outcome for individual* i *under reallocation* R.

Example 5 (Reallocations for gender peer effects)**.** *Suppose for convenience that* n *and* G *are even and that exactly half of students are boys. A researcher could define reallocation effects for any pair of the following feasible reallocations:*

- *Simple random assignment.*
  $\mathbf{G}_R(\mathbf{X}, \sigma)$ *is a random sample from the set of group assignments satisfying Assumption 2.*
- *All classes single-gender.*
  $\mathbf{G}_R(\mathbf{X}, \sigma)$ *is a random sample from the set of group assignments satisfying Assumption 2 such that* $\sum_{g_i=g} \mathbf{x}_i \in \{0, n\}$ *for all* g.
- *All classes perfectly mixed.*
  $\mathbf{G}_R(\mathbf{X}, \sigma)$ *is a random sample from the set of group assignments satisfying Assumption 2 such that* $\sum_{g_i=g} \mathbf{x}_i = n/2$ *for all* g.

Any other allocation that satisfies (CRA) could also be considered, including the original allocation if applicable.

4 Results

This section demonstrates the relevant properties of the model. The main result is Proposition 4, which shows conditions under which simple linear regression models can be intepreted as measuring peer effects or peer group effects as defined in Section 3. For example, peer separability and random assignment are sufficient conditions for the simple linear-in-means model to be interpreted as measuring average peer effects. Other propositions consider weaker assumptions and more complex estimands, and typically show that the effect of interest can be expressed in terms of either linear regression coefficients or a weighted average of such coefficients.

4.1 Aggregation and separability

Before discussing identification in detail, some preliminary results are helpful. Proposition 1 shows that simple causal effects can typically be interpreted as a weighted average of more complex effects, with weights that can easily be recovered from the probability distribution of $\mathbf{x}_i$:

Proposition 1 (Aggregation)**.** *1. Conditional effects can be aggregated to yield average ef-*

fects:

$$APE_k = \sum_{s=0}^{K} \Pr(\mathbf{x}_i = \mathbf{e}_s) CPE_{s,k} \tag{22}$$

$$AGE_m = \sum_{s=0}^{K} \Pr(\mathbf{x}_i = \mathbf{e}_s) CGE_{s,m} \tag{23}$$

$$ARE(\mathbf{G}_0, \mathbf{G}_1) = \sum_{s=0}^{K} \Pr(\mathbf{x}_i = \mathbf{e}_s) CRE_s(\mathbf{G}_0, \mathbf{G}_1) \tag{24}$$

2. *Conditional group effects for a saturated partition can be aggregated to yield group effects for any other partition:*

$$CGE_{s,m} = \sum_{r=1}^{M^{sat}} w_{rm}(\mu) CGE_{s,r}^{sat} \tag{25}$$

where $CGE_{s,r}^{sat}$ *is the conditional group effect for bin* r *of ithe saturated variable* $\mathbf{z}_i^{sat}$, $w_{rm}(\mu)$ *is a weighting function given by:*

$$w_{rm}(\mu) = \frac{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^m : \mathbf{z}^{sat}(\bar{\mathbf{x}}) = \mathbf{e}_r} M(\bar{\mathbf{x}}, n, \mu)}{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^m} M(\bar{\mathbf{x}}, n, \mu)} - \frac{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^0 : \mathbf{z}^{sat}(\bar{\mathbf{x}}) = \mathbf{e}_r} M(\bar{\mathbf{x}}, n, \mu)}{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^0} M(\bar{\mathbf{x}}, n, \mu)} \tag{26}$$

and:

$$M(\bar{\mathbf{x}}, n, \mu) = \frac{(n-1)!}{\prod_{s=0}^{K} ((n-1)\bar{\mathbf{x}}_{\cdot s})!} \prod_{s=0}^{K} \mu_s^{(n-1)\bar{\mathbf{x}}_{\cdot s}} \tag{27}$$

is the probability of drawing $(n-1)\bar{\mathbf{x}}$ *from a multinomial distribution with* $(n-1)$ *trials and categorical probability vector* μ*.*

The results in Proposition 1 are not particularly surprising, but are useful to keep in mind when choosing and comparing model specifications.

Proposition 2 shows how separability assumptions can be employed to simplify the analysis. In particular, a peer-separable potential outcome function can always be written as the sum of a set of latent variables: an individual-specific own effect and a set of individual-specific or pair-specific peer effects. Average and conditional peer effects **APE** and **CPE** can also be expressed in terms of conditional expectations of these latent variables. While neither the own effect nor the peer effect is directly observable for a given individual, Proposition 4 in the next section establishes conditions under which average peer effects are identified.

Proposition 2 (Separability)**.** *1. If peer effects are peer-separable (PSE), then each individual's potential outcome function can be expressed in the form:*

$$y_i(\mathbf{p}) = o_i + \sum_{j \in \mathbf{p}} p_{ij} \tag{28}$$

where $o_i = o(\tau_i)$, $p_{ij} = p(\tau_i, \tau_j)$ *and:*

$$CPE_{s,k} = E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) - E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) \tag{29}$$

$$APE_k = E(p_{ij}|\mathbf{x}_j = \mathbf{e}_k) - E(p_{ij}|\mathbf{x}_j = \mathbf{e}_0) \tag{30}$$

for all observable categories (s, k).

2. *If peer effects are peer-separable and own-separable (PSE, OSE), then each individual's potential outcome function can be expressed in the form:*

$$y_i(\mathbf{p}) = o_i + \sum_{j \in \mathbf{p}} p_j \tag{31}$$

where $o_i = o(\tau_i)$, $p_j = p(\tau_j)$ *and:*

$$CPE_{s,k} = APE_k = E(p_j|\mathbf{x}_j = \mathbf{e}_k) - E(p_j|\mathbf{x}_j = \mathbf{e}_0) \tag{32}$$

for all observable categories (s, k).

Separability assumptions are convenient but not necessarily correct. Fortunately, they have testable implications as shown in Proposition 3 below.

Proposition 3 (Testable implications of separability). 1. *If peers are randomly assigned conditional on observables (CRA) and peer effects are peer separable (PSE), then:*

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i) = L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \tag{33}$$

2. *If peers are randomly assigned conditional on observables (CRA) and peer effects are peer separable and own separable (PSE, OSE), then:*

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) = L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \tag{34}$$

or equivalently $\beta_3 = 0$.

Note that separability is a property of the outcome function $y(\cdot, \cdot)$ and not the particular explanatory variables $(\mathbf{x}_i, \mathbf{z}_i)$ chosen by the researcher. As a result, the implications in Proposition 3 hold for any $(\mathbf{x}_i, \mathbf{z}_i)$, though the power of a test based on these implications depends on the choice.

4.2 Identification of peer and group effects

Proposition 4 below shows identification under a simple random assignment research design. The identification analysis also suggests some simple estimators. Proposition 5 later in this section shows identification under conditional random assignment, and Proposition 8 in Section 5 shows identification under a more complex two-stage assignment design.

Proposition 4 (Identification with random assignment). 1. *If peers are randomly assigned (RA) and peer effects are peer separable (PSE), then peer effects* **APE** *and* **CPE** *are* *identified:*

$$APE_k = \frac{\alpha_{2k}}{n-1} \tag{35}$$

$$CPE_{s,k} = \frac{\beta_{2k} + \beta_{3sk}}{n-1} \tag{36}$$

where $(\alpha_{2k}, \beta_{2k}, \beta_{3sk})$ *are coefficients from the best linear predictors:*

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \equiv \alpha_0 + \mathbf{x}_i\alpha_1 + \bar{\mathbf{x}}_i\alpha_2 \tag{37}$$

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \equiv \beta_0 + \mathbf{x}_i\beta_1 + \bar{\mathbf{x}}_i\beta_2 + \mathbf{x}_i\beta_3\bar{\mathbf{x}}_i' \tag{38}$$

i.e., α_{2k} *is the* k*th element of* α_2*,* β_{2k} *is the* k*th element of* β_2*,* β_{3sk} *is the element in row* s *and column* k *of* β_3 *for all* $s > 0$*, and* $\beta_{30k} = 0$*.*

2. *If peers are randomly assigned (RA), then group effects* **AGE** *and* **CGE** *are identified:*

$$AGE_m = \gamma_{2m} \tag{39}$$

$$CGE_{s,m} = \delta_{2m} + \delta_{3sm} \tag{40}$$

where $(\gamma_{2m}, \delta_{2m}, \delta_{3sm})$ *are coefficients from the best linear predictors:*

$$L(y_i|\mathbf{x}_i, \mathbf{z}_i) \equiv \gamma_0 + \mathbf{x}_i\gamma_1 + \mathbf{z}_i\gamma_2 \tag{41}$$

$$L(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i) \equiv \delta_0 + \mathbf{x}_i\delta_1 + \mathbf{z}_i\delta_2 + \mathbf{x}_i\delta_3\mathbf{z}_i' \tag{42}$$

i.e., γ_{2m} *is element* m *of* γ_2*,* δ_{2m} *is element* m *of* δ_2*,* δ_{3sm} *is element* (s, m) *of* δ_3 *for all* $s > 0$*, and* $\delta_{30m} = 0$ *for all* m*.*

Proposition 4 shows conditions under which each causal peer group effect defined in Section 3 can be expressed in terms of a linear regression model.

Example 6 (Peer-separable gender effects)**.** *Using the data described in the previous examples, suppose Researcher A estimates the effect of male classmates on test scores using the conventional linear-in-means model (37). Under the assumption of peer separability, Part 2 of Proposition 4 allows Researcher A to interpret the coefficient on* $\bar{\mathbf{x}}_i$ *as the effect on the average student of replacing the average female classmate with the average male classmate. Note that there are no other control variables, and gender does not appear in the underlying structural model; instead this analysis is interpreted as an analysis of heterogeneity. Another researcher with the same data but other* $\mathbf{x}_i$ *variables - race, ethnicity, language spoken at home, immigration status, etc. - could explore those other aspects of heterogeneity either separately or in any combination.*

Now suppose that Researcher B has the same data, but estimates the heterogeneous linear-in-means model (38). Part 2 of Proposition 4 also shows that the assumption of peer-separability allows Researcher B to interpret the coefficient on $\bar{\mathbf{x}}_i$ *as the effect on the average female student of replacing the average female classmate with the average male classmate, and the . Adding the coefficient on the interaction term* $\mathbf{x}_i\bar{\mathbf{x}}_i$ *gives the effect on the average male student of replacing the average female classmate with the average male classmate. Note that a finding of heterogeneity (i.e. a nonzero coefficient on the interaction term) by Researcher B does not* *invalidate Researcher A's analysis based on equation (37), as that analysis can still be interpreted as averaging these heterogeneous effects across all treatment units. Both specifications are valid, in the sense of recovering an estimand of interest.*

Although the assumption of peer separability provides a simple interpretation of linear-in-means results, empirical researchers have shown increasing interest in contextual effects that go beyond the linear-in-means model, and have repeatedly found evidence for such nonlinearities.

Example 7 (Nonlinear gender effects)**.** *Suppose that Researcher C has the same data as Researcher A, but divides peer groups into majority-male and majority female; i.e.,* $\mathbf{z}_i = 1$ *if* $\bar{\mathbf{x}}_i > 0.5$ *and* $\mathbf{z}_i = 0$ *if* $\bar{\mathbf{x}}_i \leq 0.5$*, and estimates a regression of* y_i *on* $(\mathbf{x}_i, \mathbf{z}_i)$*. Then part 1 of Proposition 4 says that the coefficient on* $\mathbf{z}_i$ *can be interpreted as the effect on the average student of replacing the average (randomly constructed) majority-female peer group with the average (randomly constructed) majority-male peer group. In addition, these averages can be identified separately for male and female students from a regression of* y_i *on* $(\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i)$*. Again, the results in Proposition 4 apply regardless of the researcher's choice of how to construct* $\mathbf{z}_i$*. Researcher C could compare majority-male versus majority-female peer groups, or could compare all-male, all-female, and mixed peer groups. Note that a finding of nonlinearity by Researcher C could invalidate the assumption of peer separability and thus invalidate the results of Resesarchers A and B.*

Although identification and interpretation are simplest with random assignment, many of the results in Proposition 4 also hold under conditional random assignment while others require minor modfications. To show this it is first necessary to show (in Lemma 1 below) that the conditional expectation function is the same under random assignment and conditional random assignment.

Lemma 1 (Conditional random assignment)**.** *If peers are randomly assigned conditional on observable characteristics (CRA), then:*

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) = E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) \tag{43}$$

where $\tilde{\mathbf{p}}$ *is a purely random draw of* $(n-1)$ *peers from* $\mathcal{N} \setminus \{i\}$*.*

Proposition 5, which shows identification under conditional random assignment, then follows.

Proposition 5 (Identification with conditional random assignment)**.** *1. If peers are randomly assigned conditional on observable characteristics (CRA) and peer effects are peer separable (PSE), then peer effects* **APE** *and* **CPE** *are identified:*

$$APE_k = \sum_{s=0}^{K} \mu_s \frac{\beta_{2k} + \beta_{3sk}}{n-1} \tag{44}$$

$$CPE_{s,k} = \frac{\beta_{2k} + \beta_{3sk}}{n-1} \tag{45}$$

where $(\beta_{2k}, \beta_{3sk})$ *are defined as in equation (38).*

2. *If peers are randomly assigned conditional on observable characteristics (CRA), then group effects* **AGE** *and* **CGE** *are identified:*

$$AGE_m = \sum_{s=0}^{K} \sum_{r=1}^{M^{sat}} \mu_s w_{rm}(\mu)(\delta_{2r}^{sat} + \delta_{3sr}^{sat}) \tag{46}$$

$$CGE_{s,m} = \sum_{r=1}^{M^{sat}} w_{rm}(\mu)(\delta_{2r}^{sat} + \delta_{3sr}^{sat}) \tag{47}$$

where $\delta^{sat} = (\delta_0^{sat}, \delta_1^{sat}, \delta_2^{sat}, \delta_3^{sat})$ *are the coefficients from estimating equation (42) with saturated group variable* $\mathbf{z}_i^{sat}$.

While Proposition 5 is more general than Proposition 4, this generality comes at the cost that some estimands are weighted averages of regression coefficients rather than just the coefficients. The reason this is the case is that both peer effects and group effects are defined in terms of a hypothetical randomly-assigned peer group, so some reweighting is required when the observed peer group is randomly assigned based on observed characteristics. As in Proposition 1, the probability weights in Proposition 5 can be recovered either directly from the probability distribution of $\mathbf{x}_i$ or by simulating random draws from that distribution.

4.3 Identification of reallocation effects

Proposition 6 below describes how the reallocation effects defined in Section 3.3 can be described in terms of the estimands defined in Section 3.1.

Proposition 6 (Reallocation effects)**.** *Let* $\mathbf{G}_0, \mathbf{G}_1$ *be two feasible reallocations, and let* $\bar{\mathbf{x}}_{iR} = \bar{\mathbf{x}}_i(\mathbf{T}, \mathbf{G}_R)$ *and* $\mathbf{z}_{iR} = \mathbf{z}(\bar{\mathbf{x}}_{iR})$ *be the counterfactual peer group composition of individual i under reallocation* $\mathbf{G}_R$. *Then:*

1. *If* $(\mathbf{S}_{\bar{\mathbf{x}}}^1, \ldots, \mathbf{S}_{\bar{\mathbf{x}}}^M)$ *are singletons, and* $\Pr(\bar{\mathbf{x}}_{i0} \in \mathbf{S}_{\bar{\mathbf{x}}}^0) = \Pr(\bar{\mathbf{x}}_{i1} \in \mathbf{S}_{\bar{\mathbf{x}}}^0) = 0$, *then:*

$$ARE(\mathbf{G}_0, \mathbf{G}_1) = \sum_{s=0}^{K} \mu_s E(\mathbf{z}_{i1} - \mathbf{z}_{i0} | \mathbf{x}_i = \mathbf{e}_s)\mathbf{CGE}_s' \tag{48}$$

$$CRE_s(\mathbf{G}_0, \mathbf{G}_1) = E(\mathbf{z}_{i1} - \mathbf{z}_{i0} | \mathbf{x}_i = \mathbf{e}_s)\mathbf{CGE}_s' \tag{49}$$

where $\mathbf{CGE}_s$ *is row s of the matrix* **CGE**.

2. *If peer effects are peer separable (PSE), then:*

$$ARE(\mathbf{G}_0, \mathbf{G}_1) = (n-1) \sum_{s=0}^{K} \mu_s E(\bar{\mathbf{x}}_{i1} - \bar{\mathbf{x}}_{i0} | \mathbf{x}_i = \mathbf{e}_s)\mathbf{CPE}_s' \tag{50}$$

$$CRE_s(\mathbf{G}_0, \mathbf{G}_1) = (n-1) E(\bar{\mathbf{x}}_{i1} - \bar{\mathbf{x}}_{i0} | \mathbf{x}_i = \mathbf{e}_s)\mathbf{CPE}_s' \tag{51}$$

where $\mathbf{CPE}_s$ *is row s of the matrix* **CPE**.

3. If peer effects are peer separable and own separable (PSE, OSE), then:

$$ARE(\mathbf{G}_0, \mathbf{G}_1) = 0 \tag{52}$$

$$CRE_s(\mathbf{G}_0, \mathbf{G}_1) = (n-1)E(\bar{\mathbf{x}}_{i1} - \bar{\mathbf{x}}_{i0}|\mathbf{x}_i = \mathbf{e}_s)\mathbf{APE}' \tag{53}$$

Parts 2 and 3 of Proposition 6 show that separability allows the effect of an alternative allocation to be inferred straightforwardly from average or conditional peer effects. In contrast, Part 1 of Proposition 6 indicates an important limitation on the use of nonlinear peer effect regressions to predict the results of a reallocation: the partition of $\bar{\mathbf{x}}_i$ must assign a unique value of $\mathbf{z}_i$ for each distinct value of $\bar{\mathbf{x}}_i$ in the support of either reallocation. Values of $\bar{\mathbf{x}}_i$ outside of that support can be pooled. The intuition here is that within a pooled category, the distribution of $\mathbf{z}_i$ does not pin down the distribution of $\bar{\mathbf{x}}_i$, so two allocation rules may have the same distribution of $\mathbf{z}_i$ but not the same distribution of $\bar{\mathbf{x}}_i$.

Example 8 (The effects of a classroom reallocation by gender)**.** *Suppose the researcher estimates a nonlinear model using five categories in constructing $\mathbf{z}_i$: all-boy ($\bar{\mathbf{x}}_i = 1$), majority-boy ($0.5 < \bar{\mathbf{x}}_i < 1$), balanced ($\bar{\mathbf{x}}_i = 0.5$), majority-girl ($0.0 < \bar{\mathbf{x}}_i < 0.5$), and all-girl ($\bar{\mathbf{x}}_i = 0$). The all-boy, balanced and all-girl categories are singletons, while the majority-boy and majority-girl categories are pooled. Proposition 6 implies those results can be used to predict the result of a change from balanced to gender-segregated classrooms, or from the baseline random allocation to a balanced or gender-segregated allocation. However, the results cannot be used to predict the effect of a change from balanced to majority-boy and majority-girl classrooms. The natural solution to this issue is to remember that the researcher chooses the partition, and can in principle always choose a partition rich enough to identify (conditional) average reallocation effects. For example, one might redefine categories to include singleton categories for $\bar{\mathbf{x}}_i = 0.5$, $\bar{\mathbf{x}}_i = 0.25$ and $\bar{\mathbf{x}}_i = 0.75$ and then use the results to compare a balanced or random allocation to one in which roughly half of classrooms are 75% boys and the other half are 75% girls. In practice, the usual bias/variance trade off applies to the choice of partitions. There may not be enough exactly-balanced classrooms for adequate statistical precision, and some pooling of nearly-balanced classrooms can increase precision at a cost of bias from aggregating categories with dissimilar average effects.*

4.4 Estimation

Although the primary focus of this paper is on identification, the identification results above are constructive and suggest simple plug-in estimators that will be consistent and asymptotically normal in a wide variety of settings.

Propositions 4 and 5 express peer and group effects in terms of best linear predictors $(\alpha, \beta, \gamma, \delta, \delta^{sat})$, or weighted averages whose weights depend on $\mu = E(\mathbf{x}_i)$. Let $\theta \equiv (\mu, \alpha, \beta, \gamma, \delta, \delta^{sat})$ and let $\hat{\theta} \equiv (\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \hat{\delta}^{sat})$ be a consistent and asymptotically normal estimator constructed from a sample of N individuals:

$$\sqrt{N}\left(\hat{\theta} - \theta\right) \to^D N(0, \Sigma) \tag{54}$$

with consistent covariance matrix estimator $\hat{\Sigma} \to^P \Sigma$. In a typical application, the data will be constructed from a random sample of G groups, $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$ will be the sample average, the remaining elements of $\hat{\theta}$ will be linear regression coefficients, and $\hat{\Sigma}$ will be robust to group-level clustering.

Propositions 4 and 5 suggest a natural set of plug-in estimators based on the elements of $\hat{\theta}$:

$$\widehat{APE}_k = \begin{cases} \frac{\hat{\alpha}_{2k}}{n-1} & \text{if (PSE, RA)} \\ \sum_{s=0}^{K}\hat{\mu}_s\widehat{CPE}_{sk} & \text{if (PSE, CRA)} \end{cases}$$

$$\widehat{CPE}_{sk} = \left\{ \frac{\hat{\beta}_{2k}+\hat{\beta}_{3sk}}{n-1} \quad \text{if (PSE, CRA)} \right.$$

$$\widehat{AGE}_m = \begin{cases} \hat{\gamma}_{2m} & \text{if (RA)} \\ \sum_{s=0}^{K}\hat{\mu}_s\widehat{CGE}_{sm} & \text{if (CRA)} \end{cases}$$

$$\widehat{CGE}_{sm} = \begin{cases} \hat{\delta}_{2m} + \hat{\delta}_{3sm} & \text{if (RA)} \\ \sum_{r=0}^{M^{sat}} w_{rm}(\hat{\mu})(\hat{\delta}_{2r}^{sat} + \hat{\delta}_{3sr}^{sat}) & \text{if (CRA)} \end{cases}$$

Since these estimators are differentiable functions of $\hat{\theta}$, they will be consistent and asymptotically normal, with an asymptotic covariance matrix that can be derived using the delta method.

Similarly, Proposition 6 provides a starting point for estimating reallocation effects by a plug-in method. Estimating reallocation effects is complicated by the fact that the set of feasible reallocations in the population depends on μ. However, since all feasible reallocations are defined as functions of $\mathbf{X}$ and the randomization device σ, the set of feasible reallocations in a given sample depends only on $\hat{\mu}$.

$$\widehat{ARE}(\mathbf{G}_0, \mathbf{G}_1) = \begin{cases} 0 & \text{if (PSE, OSE)} \\ \sum_{s=0}^{K}\hat{\mu}_s\widehat{CRE}_s(\mathbf{G}_0, \mathbf{G}_1) & \text{if (PSE, CRA) or (singletons, CRA)} \end{cases}$$

$$\widehat{CRE}_s(\mathbf{G}_0, \mathbf{G}_1) = \begin{cases} (n-1)\Delta\bar{x}_s(\mathbf{G}_0, \mathbf{G}_1, \hat{\mu})\widehat{\mathbf{APE}}' & \text{if (PSE, OSE, CRA)} \\ (n-1)\Delta\bar{x}_s(\mathbf{G}_0, \mathbf{G}_1, \hat{\mu})\widehat{\mathbf{CPE}}_s' & \text{if (PSE, CRA)} \\ \Delta\mathbf{z}_s(\mathbf{G}_0, \mathbf{G}_1, \hat{\mu})\widehat{\mathbf{CGE}}_s' & \text{if (singletons, CRA)} \end{cases}$$

where:

$$\Delta\bar{x}_s(\mathbf{G}_0, \mathbf{G}_1, \hat{\mu}) \equiv E(\bar{\mathbf{x}}_{i1} - \bar{\mathbf{x}}_{i0}|\mathbf{x}_i = \mathbf{e}_s, \hat{\mu}) \tag{55}$$

$$\Delta\mathbf{z}_s(\mathbf{G}_0, \mathbf{G}_1, \hat{\mu}) \equiv E(\mathbf{z}_{i1} - \mathbf{z}_{i0}|\mathbf{x}_i = \mathbf{e}_s, \hat{\mu}) \tag{56}$$

The asymptotic properties of these estimators depend on the reallocations chosen by the researcher, because $\Delta\bar{x}_s(\mathbf{G}_0, \mathbf{G}_1, \hat{\mu})$ and $\Delta\mathbf{z}_s(\mathbf{G}_0, \mathbf{G}_1, \hat{\mu})$ are not necessarily continuous or differentiable in $\hat{\mu}$. If they are, these estimators will be consistent and asymptotically normal.

Note that these estimators have been defined in terms of a set of categories for $\mathbf{x}_i$ and bins for $\mathbf{z}_i$ (if applicable) that have been determined by the researcher in advance of the data. This scenario fits many applications in which the researcher has a specific research question (ideally in a pre-analysis plan) and the individual characteristics relevant to that question are naturally discrete. When individual characteristics of interest are continuous or high-dimensional, or when peer groups are large (so that there are many possible bins for defining $\mathbf{z}_i$), researchers may need to construct categories and bins in a data-driven manner that balances model flexibility with statistical precision. Tree-based methods for estimating conditional average treatment effects such as those developed in Athey and Imbens (2016) or Wager and Athey (2018) can be adapted to this setting, though this adaptation is beyond the scope of this paper and left to future research.

5 Extension: Multi-stage assignment

Although simple random assignment is the ideal setting for studying peer effects, many peer effect studies are based on a more complex research design in which individuals are non-randomly assigned to large groups and then randomly assigned to smaller groups within those large groups. For example, classroom peer effects are typically estimated using a research design associated with Hoxby (2000): panel data with multiple grade cohorts within multiple schools is used in combination with linear fixed effects regression models to account for nonrandom selection into schools. The key identifying assumption of this research design is that selection into a given grade cohort within a school is random (due to random timing of birth) conditional on the nonrandom selection of school. However, the linear fixed effects implementation is quite restrictive, treating the causal effect of interest as a fixed parameter and imposing other coefficient homogeneity restrictions that may or may not be reasonable in a given setting.

This section adds a general multi-stage selection design to the potential outcomes framework developed in Section 2, demonstrates conditions under which the linear fixed effects regression model will recover causal peer effects, and proposes alterative strategies for applications in which those conditions do not hold.

Example 9 (Public and private schools)**.** *Consider a simplified classorom peer effects setting in which students are rich ($\mathbf{x}_i = 1$) or poor ($\mathbf{x}_i = 0$) and attend the local public ($\ell = 1$) or private ($\ell = 2$) school. Both schools have a mix of rich and poor students, but rich students are more likely than poor students to attend the private school. Within each school, students are randomly assigned to cohorts/classrooms as a result of random timing of birth.*

5.1 Model

The model is as defined in Section 2 with additional assumptions and definitions as given below. In the interest of space and clarity, the analysis will focus on the case where peer separability holds.

Each peer group belongs to a **location** $\ell \in \mathcal{L} \equiv \{1, \ldots, L\}$:

$$\mathbf{L} \equiv \begin{bmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_N \end{bmatrix} \equiv \begin{bmatrix} \ell(g_1) \\ \ell(g_2) \\ \vdots \\ \ell(g_N) \end{bmatrix} \equiv \mathbf{L}(\mathbf{G}) \tag{57}$$

where ℓ_i is the location for individual i and $\ell : \mathcal{G} \to \mathcal{L}$ is a known function. To simplify exposition, each location is assumed to include r peer groups, which implies that $G = rL$ and $N = nG = nrL$.

Assignment to location will typically depend on unobserved type:

Assumption 6 (Locations and types)**.** *Each individual's type is an independent draw from a type distribution that varies by location:*

$$\Pr(\mathbf{T}|\mathbf{L}) = \prod_{i=1}^{N} f_{\tau|\ell}(\tau_i, \ell_i) \tag{58}$$

where $f_{\tau|\ell} : \mathcal{T} \times \mathcal{L} \to [0,1]$ is some unknown discrete conditional PDF.

Assumption 6 allows the distribution of uobserved types to vary across locations. As in Assumption 1, the arbitrary ordering makes independence a mostly innocuous assumption.

Definition 14 (Random assignment by location)**.** *Peer groups are* **randomly assigned by location (RAL)** *if:*

$$\mathbf{G} \perp\!\!\!\perp \mathbf{T}|\mathbf{L} \tag{\textbf{RAL}}$$

Random assignment by location is stated here as an optional assumption but will be needed for all results in Section 5

5.2 Identification with simple fixed effects models

Most empirical research in this setting uses linear models with constant coefficients and a location fixed effect, so a natural starting point is to establish conditions under which that procedure will recover conditional average peer effects. Intiutitively, fixed effects models allow location to matter for the outcome, but only in ways that shift the outcome by the same amount for everyone at that location.

Definition 15 (Constant shifts)**.** *Suppose that peer separability (PSE) and own separability (OSE) both hold. Then own effects have* **constant shifts (CS)** *if:*

$$E(o_i|\mathbf{x}_i = \mathbf{e}_k, \ell_i = \ell) - E(o_i|\mathbf{x}_i = \mathbf{e}_0, \ell_i = \ell) = E(o_i|\mathbf{x}_i = \mathbf{e}_k) - E(o_i|\mathbf{x}_i = \mathbf{e}_0) \tag{\textbf{CS}}$$

for all (k, ℓ).

In addition, the location fixed effect is taken to reflect exclusively differences in own effects, and not systematic differences in peer effects.

Definition 16 (Location invariance)**.** *Suppose that peer separability (PSE) holds. Peer effects are* **location invariant (LI)** *if:*

$$E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell, \ell_j = \ell') = E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) \tag{\textbf{LI}}$$

for all (s, k, ℓ, ℓ').

These two assumptions can then be used to support the use of simple fixed effects models to estimate peer effects in this setting.

Proposition 7 (Identification via fixed effects regression)**.** *Suppose that peer groups are randomly assigned by location (RAL), and the outcome function satisfies peer separability (PSE), own separability (OSE), location invariance (LI) and constant shifts (CS). Then:*

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}, \ell_i = \ell) = \alpha_0^\ell + \mathbf{x}\alpha_1 + \bar{\mathbf{x}}\alpha_2 \tag{59}$$

and:

$$APE_k = CPE_{sk} = \alpha_{2k}/(n-1) \tag{60}$$

for all (s, k)*.*

Proposition 7 gives conditions under which a researcher can estimate a standard linear fixed effects model and interpret the coefficients on average peer characteristics as reflecting an average peer effect. The example below suggests how strong these assumptions would be in a typical application.

Example 10. *Continuing the two-school example, the assumptions needed for Proposition 7 would allow:*

- *Rich students to be systematically better/worse students (own effect) than poor students.*
- *Rich students to be systematically better/worse peers (peer effect) than poor students.*
- *Private school students to be systematically better/worse students than public school students*

but would not allow:

- *Private school students to be systematically better/worse peers than public school students. This would violate location invariance (LI).*
- *The student quality gap between rich and poor students to vary across the two schools. This would violate constant shifts (CS).*

5.3 Identification with heterogeneous-coefficient models

Relaxing the assumptions in Proposition 7 yields a set of heterogeneous-coefficient regression models that can be given varying causal intepretations. For example, random assignment by location implies that peer effects can be measured separately within each location even without any restrictions on heterogeneity across locations.

Definition 17 (Location-specific regression models)**.** *The coefficients of the* **location-specific linear in means models** *are defined as the vectors* $\alpha^\ell \equiv (\alpha_0^\ell, \alpha_1^\ell, \alpha_2^\ell)$*, and* $\beta^\ell \equiv (\beta_0^\ell, \beta_1^\ell, \beta_2^\ell, \beta_3^\ell)$ *such that:*

$$L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \equiv \alpha_0^\ell + \mathbf{x}_i\alpha_1^\ell + \bar{\mathbf{x}}_i\alpha_2^\ell \tag{61}$$

$$L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \equiv \beta_0^\ell + \mathbf{x}_i\beta_1^\ell + \bar{\mathbf{x}}_i\beta_2^\ell + \mathbf{x}_i\beta_3^\ell\bar{\mathbf{x}}_i' \tag{62}$$

where $L^\ell(\cdot|\cdot)$ *is the best linear predictor conditional on* $\ell_i = \ell$*, i.e.*

$$\alpha^\ell = E(\mathbf{w}_i'\mathbf{w}_i|\ell_i = \ell)^{-1}E(\mathbf{w}_i'y_i|\ell_i = \ell) \text{ where } \mathbf{w}_i = (1, \mathbf{x}_i, \bar{\mathbf{x}}_i) \tag{63}$$

$$\beta^\ell = E(\mathbf{w}_i'\mathbf{w}_i|\ell_i = \ell)^{-1}E(\mathbf{w}_i'y_i|\ell_i = \ell) \text{ where } \mathbf{w}_i = (1, \mathbf{x}_i, \bar{\mathbf{x}}_i, vec(\mathbf{x}_i'\bar{\mathbf{x}}_i))$$

Equations (61) and (62) correspond to standard heterogeneous-coefficient linear panel data models. Wooldridge (2010, p. 377-381) describes estimation and testing procedures for this class of models. For the purpose of discussing identification, these location-specific coefficients will be taken as given. In a given sample, there will be a bias/variance trade-off that may make it more practical to estimate a more restrictive regression model as an approximation to the fully heterogeneous model.

Definition 18 (Location-specific peer effects)**.** *Let* **location-specific peer effects** *for location ℓ be defined as the average effect of replacing a randomly-selected peer from the base category with a randomly-selected peer from another category and in the same location:*

$$\begin{aligned} APE_k^{\ell} &\equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}}) | \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0, \ell_i = \ell_j = \ell\right) \\ CPE_{s,k}^{\ell} &\equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}}) | \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0, \mathbf{x}_i = \mathbf{e}_s, \ell_i = \ell_j = \ell\right) \end{aligned} \tag{64}$$

where $\tilde{\mathbf{q}}$ is a purely random draw of $(n-2)$ peers from $\mathcal{N} \setminus \{i, j\}$.

As shown in result 1 of Proposition 8 below, Proposition 4 can be applied location-by-location to relate location-specific coefficients to the corresponding location-specific peer effects. These location-specific peer effects can then be used to find reallocation effects for any feasible reallocation across peer groups that keeps every individual within the same location. Unfortunately, within-location peer effects are generally insufficient to measure the effect of any reallocation across locations. This is a critical limitation that is typically not addressed in empirical work, as many reallocations of interest represent shifts across rather than within locations.

Example 11. *The within-location coefficients can be used to predict the effect of replacing a rich private school student with a poor private school student, or the effect of replacing a rich public school student with a poor public school student. But they cannot predict the effect of replacing a rich private school student with a poor public school student, or of replacing the average rich student in the population with the average poor student in the population.*

In order to predict the effects of reallocations across locations, the variability of peer effects across locations can be restricted using either the previously-defined concept of location invariance or the slightly weaker restriction of partial location invariance.

Definition 19 (Partial location invariance)**.** *Peer effects are* **partially location invariant (PLI)** *if they are peer separable and:*

$$E(p_{ij} | \mathbf{x}_j = \mathbf{e}_0, \ell_j = \ell) = E(p_{ij} | \mathbf{x}_j = \mathbf{e}_0) \tag{\textbf{PLI}}$$

for all ℓ.

Partial location invariance essentially requires that location invariance applies to at least one oberved type. For convenience, this observed type is taken to be the base type.

Results 2 and 3 in Proposition 8 below show that either form of location invariance can allow (non-location-specific) peer effects to be expressed as a weighted average of location-specific effects. They are therefore identified and can be used to predict the result of a feasible reallocation across locations.

Proposition 8 (Identification under within-location random assignment). *1. If peers are randomly assigned by location (RAL) and peer effects are peer separable (PSE), then location specific peer effects are identified:*

$$CPE^{\ell}_{s,k} = \frac{\beta^{\ell}_{2k} + \beta^{\ell}_{3sk}}{n-1} \tag{65}$$

$$APE^{\ell}_{k} = \frac{\alpha^{\ell}_{2k}}{n-1} \tag{66}$$

for each location $\ell \in \mathcal{L}$.

2. *If peers are randomly assigned by location (RAL) and peer effects are peer separable and location invariant (PSE, LI), then peer effects are identified:*

$$CPE_{s,k} = \frac{E(\beta^{\ell_i}_{2k}) + E(\beta^{\ell_i}_{3sk})}{n-1} \tag{67}$$

$$APE_{k} = \frac{E(\alpha^{\ell_i}_{2k})}{n-1} \tag{68}$$

for all (s,k).

3. *If peers are randomly assigned by location (RAL) and peer effects are peer separable, own separable and partially location invariant (PSE, OSE, PLI), then peer effects are identified:*

$$CPE_{s,k} = APE_{k} = \frac{E(\alpha^{\ell_i}_{2k} | \mathbf{x}_i = \mathbf{e}_k)}{n-1} \tag{69}$$

for all (s,k).

To summarize the results in this section, designs based on random cohorts are common in the applied literature but imply complications beyond those in a simple or even conditional random assignment design. Researchers using such designs have the option of imposing strong restrictions on heterogeneity (as in Proposition 7), by combining somewhat weaker restrictions with more explicit handling of hetereogeneous coefficients (as in Proposition 8), or by noting that the results only apply to within-location comparisons and reallocations.

6 Extension: Direct contextual effects

As discussed in the introduction, much of the applied literature treats contextual effects as if they were *direct* and *constant*. That is, the effect peers have on a particular individual is a parametric function of a limited set of own and peer characteristics. If the researcher has access to the correctly-specified model and full set of relevant characteristics, identification and interpretation of both peer effects and reallocation effects are dramatically simplified. Otherwise, the results may be subject to substantial omitted variables bias. This section considers and analyzes direct contextual effects as a special case of the model developed in previous sections.

Definition 20 (Direct contextual effects). *Peer effects are said to be* **direct contextual effects (DCE)** *in the vector of characteristics $\mathbf{x}^*_i = \mathbf{x}^*(\tau_i) \in \mathbb{R}^{K^*}$ if there exists an unknown function*

$h : \mathbb{R}^{nK^*} \to \mathbb{R}$ and scalar $\epsilon_i = \epsilon(\tau_i)$ such that:

$$y\left(\tau_i, \{\tau_j\}_{g_j=g_i}\right) = h\left(\mathbf{x}^*(\tau_i), \{\mathbf{x}^*(\tau_j)\}_{g_j=g_i}\right) + \epsilon(\tau_i) \tag{\textbf{DCE}}$$

and $E(\epsilon_i|\mathbf{x}_i^*) = 0$.

The key restriction in (DCE) is that the peer effect is a fixed function of a specific set of characteristics. The relevant characteristics can be taken as discrete for tractability. In isolation, the assumption of direct contextual effects can be made trivially true by defining $\mathbf{x}_i^*$ as a unit vector of $K^* = (T-1)$ indicator variables for each of the T unobserved types. The content of the assumption therefore depends on the specific characteristics in $\mathbf{x}_i^*$. In particular, the vector $\mathbf{x}_i^*$ of relevant variables and the vector $\mathbf{x}_i$ of available variables do not necessarily coincide.

Definition 21 (No omitted variables)**.** *A researcher is said to have* **no omitted variables (NOV)** *if the outcome has direct contextual effects in the observed characteristics $\mathbf{x}_i$:*

$$\mathbf{x}_i = \mathbf{x}_i^* \tag{\textbf{NOV}}$$

The assumption of no omitted variables is very strong, as it requires that the researcher includes every potentially relevant peer characteristic in the regression model.

6.1 Separability and functional form

Assumption (DCE) does not restrict the functional form for $h(\cdot,\cdot)$, but Proposition 9 below shows other assumptions may pin down a convenient functional form.

Proposition 9 (Separability and direct contextual effects)**.**

1. *If direct contextual effects are peer separable (DCE, PSE), then each individual's potential outcome function can be expressed in the form:*

$$\begin{aligned} y_i(\mathbf{p}) &= o_i + \sum_{j\in\mathbf{p}} p_{ij} && (70)\\ &= h_1(\mathbf{x}_i^*) + \epsilon_i + \sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) \\ &= \psi_0 + \mathbf{x}_i^*\psi_1 + \bar{\mathbf{x}}_{\mathbf{p}}^*\psi_2 + \mathbf{x}_i^*\psi_3\bar{\mathbf{x}}^*(\mathbf{p})' + \epsilon_i && \text{(if } \mathbf{x}^* \text{ is categorical)}\end{aligned}$$

   where $h_1 : \mathbb{R}^{K^*} \to \mathbb{R}$ and $h_2 : \mathbb{R}^{2K^*} \to \mathbb{R}$ are functions and $\psi \equiv (\psi_0, \psi_1, \psi_2, vec(\psi_3))$ is some vector of coefficients.

2. *If direct contextual effects are peer separable and own separable (DCE, PSE, OSE), then each individual's potential outcome function can be expressed in the form:*

$$\begin{aligned} y_i(\mathbf{p}) &= o_i + \sum_{j\in\mathbf{p}} p_j && (71)\\ &= h_1(\mathbf{x}_i^*) + \epsilon_i + \sum_{j\in\mathbf{p}} h_3(\mathbf{x}_j^*) \\ &= \psi_0 + \mathbf{x}_i^*\psi_1 + \bar{\mathbf{x}}^*(\mathbf{p})\psi_2 + \epsilon_i && \text{(if } \mathbf{x}^* \text{ is categorical)}\end{aligned}$$

where $h_1 : \mathbb{R}^{K^*} \to \mathbb{R}$ and $h_3 : \mathbb{R}^{K^*} \to \mathbb{R}$ are functions and $\psi \equiv (\psi_0, \psi_1, \psi_2)$ is some vector of coefficients.

When peer separability and own separability are assumed, direct contextual effects can be written in terms of group-level averages of individual (if both forms of separability are assumed) or pairwise (if only peer separability is assumed) characteristics. If these separability assumptions are combined with a categorical specification of $\mathbf{x}^*$, the model reduces to the standard linear-in-means contextual effects model.

6.2 Identification and omitted variables bias

Direct contextual effects do not violate the conditions for Propositions 4, 5, or 8, so heterogeneous and conditional average effects continue to be identified under the conditions described in those propositions. Proposition 10 below shows additional identification results that apply when contextual effects are direct.

Proposition 10 (Identification of direct contextual effects)**.**

1. *If peers are conditionally randomly assigned (CRA) and peer effects are direct contextual effects with no omitted variables (DCE, NOV), then $h(\cdot, \cdot)$ is identified:*

$$h\left(\mathbf{x}, \left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}\right) = E\left(y_i \,|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}\right) \tag{72}$$

for all values of $\mathbf{x}$ and $\bar{\mathbf{x}} \equiv \left(\sum_{j=1}^{n-1} \mathbf{x}^j\right)/(n-1)$ on the support of $(\mathbf{x}_i, \bar{\mathbf{x}}_i)$.

2. *If peers are randomly assigned by location (RAL) and peer effects are peer-separable direct contextual effects with no omitted variables (PSE, DCE, NOV), then* **CPE** *and* **APE** *are identified:*

$$CPE_{s,k} = \frac{E(\beta_{2k}^{\ell_i}) + E(\beta_{3sk}^{\ell_i})}{n-1} \tag{73}$$

$$APE_k = \frac{E(\alpha_{2k}^{\ell_i})}{n-1} \tag{74}$$

The first result in Proposition 10 shows that the "structural" parameters of the direct contextual effects model are identified. The second result shows that direct contextual effects are helpful in securing identification with random assignment by location, essentially because they imply location invariance.

Unfortunately, the data requirements for Proposition 10 are extremely demanding: the (NOV) assumption requires data on *everything* about each person that affects their influence on other people. In the more likely case in which the researcher has data on some subset of those characteristics, conventional omitted variables concepts apply. To simplify the discussion, suppose that the direct contextual effects take the linear in means form:

$$h(\mathbf{x}_i^*, \left\{\mathbf{x}_j^*\right\}_{j\in\mathbf{p}}) = \psi_0 + \mathbf{x}_i^*\psi_1 + \bar{\mathbf{x}}_i^*(\mathbf{p})\psi_2 \tag{75}$$

that peer groups are randomly assigned (RA) and that $\mathbf{x}_i^* = (\mathbf{x}_i, \mathbf{u}_i)$, where $\mathbf{u}_i$ is some vector of omitted individual-level variables. Then:

$$y_i = \psi_0 + \mathbf{x}_i\psi_{11} + \mathbf{u}_i\psi_{12} + \bar{\mathbf{x}}_i\psi_{21} + \bar{\mathbf{u}}_i\psi_{22} + \epsilon_i \tag{76}$$

Let $L(\mathbf{u}_i|\mathbf{x}_i) = \pi_0 + \mathbf{x}_i\pi_1$. Then by (RA), $L(\mathbf{u}_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) = \pi_0 + \mathbf{x}_i\pi_1$, $L(\bar{\mathbf{u}}_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) = \pi_0 + \bar{\mathbf{x}}_i\pi_1$, and:

$$\begin{aligned} L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) &= L(\psi_0 + \mathbf{x}_i\psi_{11} + \mathbf{u}_i\psi_{12} + \bar{\mathbf{x}}_i\psi_{21} + \bar{\mathbf{u}}_i\psi_{22} + \epsilon_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \\ &= \psi_0 + \mathbf{x}_i\psi_{11} + (\pi_0 + \mathbf{x}_i\pi_1)\psi_{12} + \bar{\mathbf{x}}_i\psi_{21} + (\pi_0 + \bar{\mathbf{x}}_i\pi_1)\psi_{22} \\ &= \underbrace{(\psi_0 + \pi_0\psi_{12} + \pi_0\psi_{21})}_{\alpha_0} + \mathbf{x}_i \underbrace{(\psi_{11} + \pi_1\psi_{12})}_{\alpha_1} + \bar{\mathbf{x}}_i \underbrace{(\psi_{21} + \pi_1\psi_{22})}_{\alpha_2} \end{aligned} \tag{77}$$

In other words, ψ_{21} (the direct effect of $\bar{\mathbf{x}}_i$) is identified from $(\mathbf{X}, \mathbf{Y}, \mathbf{G})$ and can be estimated by the conventional OLS regression coefficient α only if all omitted peer characteristics are either irrelevant ($\psi_{22} = 0$) or uncorrelated with the included characteristics ($\pi_1 = 0$). Otherwise the conventional regression yields a biased estimate of ψ_{21}. This result is a simple variation on the textbook omitted variables problem, but bears some emphasis: random assignment of peers does not imply random assignment of individual peer characteristics, and there is typically substantial correlation among an individual's characteristics.

6.3 Peer group reallocations

If the data requirements for Proposition 10 are met, the assumption of direct contextual effects has the potential advantage of allowing the direct comparison of predicted outcomes across any two specific allocations. That is, for any individuals i, j, j' and set of peers $\mathbf{q}$, the effect of replacing peer j with peer j' is:

$$y_i(j \cup \mathbf{q}) - y_i(j' \cup \mathbf{q}) = h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_j^*, \mathbf{x}_r^*\right\}_{r\in\mathbf{q}}\right) - h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{j'}^*, \mathbf{x}_r^*\right\}_{r\in\mathbf{q}}\right) \tag{78}$$

which depends only on $\mathbf{X}$ and $\mathbf{q}$. For example, if ψ_2 is the direct causal effect of $\bar{\mathbf{x}}_i^*$ on y_i, then any reallocation of peers that changes $\bar{\mathbf{x}}_i^*$ by Δ units changes y_i by $\Delta\psi_2$ units. The direct contextual effects model can thus be used to predict exact outcome differences between any two alternative allocations or allocation rules.

In contrast, knowledge of heterogeneous and conditional average peer effects only allows for comparisons within the class of alternative allocation mechanisms described in Section 4.3: those in which groups are assigned randomly conditional on the observable characteristics in the data. Returning to the classroom gender effects example, the conditional average peer effect **APE** can be interpreted as the *average* effect of replacing a *randomly selected* male peer with a *randomly selected* female peer. In contrast, if direct contextual effects are assumed in which $\mathbf{x}_i^*$ is a gender indicator, then the contextual effects parameter ψ_2 is the effect of replacing *any* male peer with *any* female peer.

However, just as the requirement of complete data on $\mathbf{x}_i^*$ represents a barrier to identification of direct contextual effects, it also represents a potential barrier to using the results for policy analysis. For example, suppose that a student's academic achievement is a known function of peer gender and the number of peers with an attention-deficit/hyperactivity disorder (ADHD). School policymakers cannot directly change a student's gender or ADHD status, so any change in peer characteristics can only be implemented by reallocating individuals. Since measured ADHD rates are much higher for boys, a reallocation by gender alone will induce a corresponding reallocation by ADHD. Therefore, policymakers must know the consequences of a candidate reallocation for all elements of $\mathbf{x}_i^*$ in order to use direct contextual effect parameters in predicting the consequences of that reallocation for outcomes. Even if researchers are able to collect detailed data on peer characteristics sufficient to identify direct contextual effects, those results are only usable by policymakers with similarly detailed data.

7 Conclusion

The results established here have several implications for empirical research on contextual peer effects, and on their potential application to policy.

The first implication is that simple model specifications based on categorical explanatory variables will often be more informative than "kitchen sink" regressions that attempt to incorporate every potentially relevant peer characteristic available in the data. A simple specification that uses a single binary peer characteristic (high/low income, black/white, male/female, etc.) can be interpreted as measuring the difference in conditional average peer effects across the two categories under relatively weak assumptions. In constrast, a regression with many related peer characteristics is difficult to interpret without imposing the very strong assumptions needed to identify direct contextual effects: i.e., that there are no relevant omitted peer characteristics that are correlated with observed peer characteristics.

A second implication is that researchers can estimate multiple distinct but logically consistent regressions, with each providing information on a different comparison. This is particularly relevant in a literature heavily focused on estimating a variety of specifications using a few key data sets such as the Add Health survey or the longitudinal student records of those U.S. states and Canadian provinces that make such data available. For example, one researcher might estimate a regression with peer parental income as the explanatory variable, while another estimates a similar regression with the same data using peer parental education as the explanatory variable. If the researchers' goal is to measure direct contextual effects, at least one of these models is misspecified. In contrast, if the researchers' goal is to measure conditional average peer effects across identifiable groups, each of these models is informative and any apparent conflict between their results can be reconciled by estimating a third regression that includes both peer variables and their interaction.

A third implication is that the dimension and mechanism of randomization is important in ways that are not often appreciated. Conditional average peer effects describe the effect of replacing a randomly selected peer from one category with a randomly selected peer from another category. This corresponds to the effect of replacing any peer from one category with any peer from the other category only if peer effects are homogeneous within categories, i.e., the researcher has estimated a direct contextual effect. Similarly, research designs based on random cohorts within nonrandomly assigned schools (locations) only identify the consequences of a reallocation within the school (location). Consequences of reallocations across schools are only identified under quite restrictive homogeneity assumptions.

Finally, reinterpreting conventional peer effects as measuring heterogeneity in treatment effects opens up several opportunities for further research. In particular, the analysis in this paper takes the construction of categories as a given choice of the researcher. Recent advances in the use of machine learning and other tools for more systematically analyzing treatment effect heterogeneity (Wager and Athey, 2018) may be adapted to this setting, and open up the possibility of identifying robust predictors of productive peers and peer groups from data.

References

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler, "Estimating Spillovers using Panel Data, with an Application to the Classroom," *Quantitative Economics*, 2012, *3* (3), 421–470.

Athey, Susan and Guido Imbens, "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 2016, *113* (27), 7353–7360.

Bhattacharya, Debopam, "Inferring optimal peer assignment from experimental data," *Journal of the American Statistical Association*, 2009, *104* (486), 486–500.

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides, "Identification of Social Interactions," in Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, eds., *Handbook of Social Economics Volume 1B*, Elsevier, 2011.

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

Brock, William A. and Steven N. Durlauf, "Interactions-based models," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics, Volume 5*, North-Holland, 2000.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica*, 2013, *81* (3), 855–882.

Eisenkopf, Gerald, Zohal Hessami, Urs Fischbacher, and Heinrich W. Ursprung, "Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland," *Journal of Economic Behavior and Organization*, 2015, *115*, 123–143.

Fruehwirth, Jane Cooley, "Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box," *Review of Economics and Statistics*, 2014, *96* (3), 514–523.

Gaviria, Alejandro and Steven Raphael, "School-based peer effects and juvenile behavior," *Review of Economics and Statistics*, 2001, *83* (2), 257–268.

Graham, Bryan S., "Identifying social interactions through conditional variance restrictions," *Econometrica*, 2008, *76* (3), 643–660.

— **, Guido W. Imbens, and Geert Ridder**, "Measuring the effects of segregation in the presence of social spillovers: A nonparametric approach," Working paper 2010.

Hoxby, Caroline M., "Peer effects in the classroom: Learning from gender and race variation," Working Paper 7867, NBER 2000.

— **and Gretchen Weingarth**, "Taking race out of the equation: School reassignment and the structure of peer effects," Working Paper 2005.

Isphording, Ingo E. and Ulf Zölitz, "The value of a peer," Working paper 342, University of Zurich 2020.

Lavy, Victor and Analía Schlosser, "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, 2011, *3* (1), 1–33.

Li, Xinran, Peng Ding, Qian Lin, Dawei Yang, and Jun S. Liu, "Randomization Inference for Peer Effects," *Journal of the American Statistical Association*, 2019, *114* (528), 1651–1664.

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

— , "Identification of treatment response with social interactions," *The Econometrics Journal*, 2013, *16*, S1–S23.

Wager, Stefan and Susan Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data, Second Edition*, MIT Press, 2010.

Not-for-publication Appendix

Proof for Proposition 1

These results follow from the definition of each item. Part 1 of the proposition can be derived as follows:

$$\sum_{s=0}^{K} CPE_{s,k}\Pr(\mathbf{x}_i=\mathbf{e}_s) = \sum_{s=0}^{K} E\left(y_i(\{j\}\cup\tilde{\mathbf{p}}) - y_i(\{j'\}\cup\tilde{\mathbf{p}})\middle|\,\mathbf{x}_i=\mathbf{e}_s,\mathbf{x}_j=\mathbf{e}_k,\mathbf{x}_{j'}=\mathbf{e}_0\right)\Pr(\mathbf{x}_i=\mathbf{e}_s) \tag{79}$$

$$= E\left(y_i(\{j\}\cup\tilde{\mathbf{p}}) - y_i(\{j'\}\cup\tilde{\mathbf{p}})\middle|\,\mathbf{x}_j=\mathbf{e}_k,\mathbf{x}_{j'}=\mathbf{e}_0\right) \tag{80}$$

$$= APE_k \tag{81}$$

$$\sum_{s=0}^{K} CGE_{s,m}\Pr(\mathbf{x}_i=\mathbf{e}_s) = \sum_{s=0}^{K}\left(E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i=\mathbf{e}_s,\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i=\mathbf{e}_s,\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0)\right)\Pr(\mathbf{x}_i=\mathbf{e}_s) \tag{82}$$

$$= E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0) \tag{83}$$

$$= AGE_m \tag{84}$$

Part 2 of the proposition can be derived as follows. First note that for any m:

$$E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{array}{r}\mathbf{x}_i=\mathbf{e}_s,\\ \mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m\end{array}\right) = \sum_{r=0}^{M^{sat}} E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{array}{r}\mathbf{x}_i=\mathbf{e}_s,\\ \mathbf{z}_i^{sat}=\mathbf{e}_r\end{array}\right)\Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}})=\mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m) \tag{85}$$

$$= E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{array}{r}\mathbf{x}_i=\mathbf{e}_s,\\ \mathbf{z}_i^{sat}=\mathbf{e}_0\end{array}\right)\Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}})=\mathbf{e}_0|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m) \tag{86}$$

$$+ \sum_{r=1}^{M^{sat}} E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{array}{r}\mathbf{x}_i=\mathbf{e}_s,\\ \mathbf{z}_i^{sat}=\mathbf{e}_r\end{array}\right)\Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}})=\mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m)$$

$$= E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{array}{r}\mathbf{x}_i=\mathbf{e}_s,\\ \mathbf{z}_i^{sat}=\mathbf{e}_0\end{array}\right)\left(1-\sum_{r=1}^{M^{sat}}\Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}})=\mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m)\right) \tag{87}$$

$$+ \sum_{r=1}^{M^{sat}} E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{array}{r}\mathbf{x}_i=\mathbf{e}_s,\\ \mathbf{z}_i^{sat}=\mathbf{e}_r\end{array}\right)\Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}})=\mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m)$$

$$= E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{array}{r}\mathbf{x}_i=\mathbf{e}_s,\\ \mathbf{z}_i^{sat}=\mathbf{e}_0\end{array}\right) + \sum_{r=1}^{M^{sat}} CGE_{s,r}^{sat}\Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}})=\mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_m) \tag{88}$$

Substituting this result into the definition of $CGE_{s,m}$ produces:

$$
\begin{aligned}
CGE_{s,m} &= E\left(y_i(\tilde{\mathbf{p}}) \middle| \begin{matrix} \mathbf{x}_i = \mathbf{e}_s, \\ \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m \end{matrix}\right) - E\left(y_i(\tilde{\mathbf{p}}) \middle| \begin{matrix} \mathbf{x}_i = \mathbf{e}_s, \\ \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0 \end{matrix}\right) && (89) \\
&= \left(E\left(y_i(\tilde{\mathbf{p}}) \middle| \begin{matrix} \mathbf{x}_i = \mathbf{e}_s, \\ \mathbf{z}_i^{sat} = \mathbf{e}_0 \end{matrix}\right) + \sum_{r=1}^{M^{sat}} CGE_{s,r}^{sat} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r | \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m)\right) && (90) \\
&- \left(E\left(y_i(\tilde{\mathbf{p}}) \middle| \begin{matrix} \mathbf{x}_i = \mathbf{e}_s, \\ \mathbf{z}_i^{sat} = \mathbf{e}_0 \end{matrix}\right) + \sum_{r=1}^{M^{sat}} CGE_{s,r}^{sat} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r | \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0)\right) \\
&= \sum_{r=1}^{M^{sat}} CGE_{s,r}^{sat} \begin{pmatrix} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r | \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r | \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix} && (91)
\end{aligned}
$$

For any (r, m):

$$
\begin{aligned}
\Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r | \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) &= \frac{\Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m \cap \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r)}{\Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m))} && (92) \\
&= \frac{\Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) \in \bar{\mathbf{x}} : \mathbf{z}(\bar{\mathbf{x}}) = \mathbf{e}_m \cap \mathbf{z}^{sat}(\bar{\mathbf{x}}) = \mathbf{e}_r)}{\Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) \in \bar{\mathbf{x}} : \mathbf{z}(\bar{\mathbf{x}}) = \mathbf{e}_m)} && (93) \\
&= \frac{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^m : \mathbf{z}^{sat}(\bar{\mathbf{x}}) = \mathbf{e}_r} \Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}})}{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^m} \Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}})} && (94)
\end{aligned}
$$

By construction, $(n-1)\bar{\mathbf{x}}_i(\tilde{\mathbf{p}})$ is a random draw from the multinomial distribution, so:

$$\Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) = M(\bar{\mathbf{x}}, n, \mu) \quad (95)$$

The result then follows by substitution.

Proof for Proposition 2

1. For every i, let:

$$
\begin{aligned}
o_i &\equiv y(\tau_i, \{1, 1, \ldots, 1\}) && (96) \\
p_{ij} &\equiv y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\})
\end{aligned}
$$

where unobserved type 1 has been chosen as an arbitrary reference type. Then:

$$
\begin{aligned}
o_i + \sum_{j \in \mathbf{p}} p_{ij} &= y(\tau_i, \{1, 1, \ldots, 1\}) && \text{(by (96))} \\
&\quad + \sum_{j \in \mathbf{p}} \left(y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\}) \right) \\
&= y(\tau_i, \{1, 1, \ldots, 1\}) \\
&\quad + y(\tau_i, \{\tau_{\mathbf{p}(1)}, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\}) \\
&\qquad \vdots \\
&\quad + y(\tau_i, \{\tau_{\mathbf{p}(n-1)}, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\}) \\
& && \text{(expansion of summation)} \\
&= y(\tau_i, \{1, 1, \ldots, 1\}) \\
&\quad + y(\tau_i, \{\tau_{\mathbf{p}(1)}, \tau_{\mathbf{p}(2)}, \tau_{\mathbf{p}(3)}, \ldots, \tau_{\mathbf{p}(n-1)}\}) - y(\tau_i, \{1, , \tau_{\mathbf{p}(2)}, \tau_{\mathbf{p}(3)}, \ldots, \tau_{\mathbf{p}(n-1)}\}) \\
&\quad + y(\tau_i, \{\tau_{\mathbf{p}(2)}, \tau_{\mathbf{p}(3)}, \ldots, \tau_{\mathbf{p}(n-1)}, 1\}) - y(\tau_i, \{1, \tau_{\mathbf{p}(3)}, \ldots, \tau_{\mathbf{p}(n-1)}, 1\}) \\
&\qquad \vdots \\
&\quad + y(\tau_i, \{\tau_{\mathbf{p}(n-2)}, \tau_{\mathbf{p}(n-1)}, 1, \ldots, 1\}) - y(\tau_i, \{1, \tau_{\mathbf{p}(n-1)}, 1, \ldots, 1\}) \\
&\quad + y(\tau_i, \{\tau_{\mathbf{p}(n-1)}, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\}) \\
& && \text{(by PSE)} \\
&= y(\tau_i, \{\tau_{\mathbf{p}(1)}, \tau_{\mathbf{p}(2)}, \tau_{\mathbf{p}(3)}, \ldots, \tau_{\mathbf{p}(n-1)}\}) \\
&= y_i(\mathbf{p}) && \text{(by (9))}
\end{aligned}
$$

which is result (28). To prove (29) and (30), first note that $(\tau_i, \tau_j) \perp\!\!\!\perp \tau_{j'}$ by equation (4), so:

$$
(p_{ij}, \mathbf{x}_j) \perp\!\!\!\perp \mathbf{x}_{j'} \tag{97}
$$

Then:

$$
\begin{aligned}
CPE_{s,k} &= E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}}) \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0 \right) && \text{(by (12))} \\
&= E\left(\left(o_i + p_{ij} + \sum_{j'' \in \tilde{\mathbf{q}}} p_{ij''} \right) - \left(o_i + p_{ij'} + \sum_{j'' \in \tilde{\mathbf{q}}} p_{ij''} \right) \middle| \begin{matrix} \mathbf{x}_i = \mathbf{e}_s, \\ \mathbf{x}_j = \mathbf{e}_k, \\ \mathbf{x}_{j'} = \mathbf{e}_0 \end{matrix} \right) && \text{(by (28))} \\
&= E\left(p_{ij} - p_{ij'} \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0 \right) \\
&= E\left(p_{ij} \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0 \right) - E\left(p_{ij'} \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0 \right) \\
&= E\left(p_{ij} \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0 \right) - E\left(p_{ij} \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_{j'} = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0 \right) \\
&= E\left(p_{ij} \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k \right) - E\left(p_{ij} \middle| \, \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0 \right) && \text{(by (97))}
\end{aligned}
$$

which is the result in (29) and:

$$
\begin{aligned}
APE_k &= E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}}) \middle| \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0\right) && \text{(by (10))} \\
&= E\left(\left(o_i + p_{ij} + \sum_{j'' \in \tilde{\mathbf{q}}} p_{ij''} \right) - \left(o_i + p_{ij'} + \sum_{j'' \in \tilde{\mathbf{q}}} p_{ij''} \right) \middle| \begin{matrix} \mathbf{x}_j = \mathbf{e}_k, \\ \mathbf{x}_{j'} = \mathbf{e}_0 \end{matrix} \right) && \text{(by (28))} \\
&= E\left(p_{ij} - p_{ij'} \middle| \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0\right) \\
&= E\left(p_{ij} \middle| \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0\right) - E\left(p_{ij'} \middle| \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0\right) \\
&= E\left(p_{ij} \middle| \mathbf{x}_j = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_0\right) - E\left(p_{ij} \middle| \mathbf{x}_{j'} = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0\right) \\
&= E\left(p_{ij} \middle| \mathbf{x}_j = \mathbf{e}_k\right) - E\left(p_{ij} \middle| \mathbf{x}_j = \mathbf{e}_0\right) && \text{(by (97))}
\end{aligned}
$$

which is the result in (30).

2. Let $p_j \equiv p_{1j}$. Then for any i:

$$
\begin{aligned}
p_{ij} &= (y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\})) && \text{(by (96))} \\
&= y(\tau_1, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_1, \{1, 1, 1, \ldots, 1\}) && \text{(by OSE)} \\
&= p_{1j} && \text{(by (96))} \\
&= p_j && (98)
\end{aligned}
$$

Substituting the result in (98) into (28) yields the result in (31). Substituting that same result into (29) and (30) yields:

$$
\begin{aligned}
APE_k &= E(p_{ij}|\mathbf{x}_j = \mathbf{e}_k) - E(p_{ij}|\mathbf{x}_j = \mathbf{e}_0) && \text{(by (30))} \\
&= E(p_j|\mathbf{x}_j = \mathbf{e}_k) - E(p_j|\mathbf{x}_j = \mathbf{e}_0) && \text{(by (98))} \\
CPE_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) - E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) && \text{(by (29))} \\
&= E(p_j|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) - E(p_j|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) && \text{(by (98))} \\
&= E(p_j|\mathbf{x}_j = \mathbf{e}_k) - E(p_j|\mathbf{x}_j = \mathbf{e}_0) && \text{(since (4) } \implies \mathbf{x}_i \perp\!\!\!\perp (\mathbf{x}_j, p_j)) \\
&= APE_k
\end{aligned}
$$

which are the results in (32).

Proof for Proposition 3

The conditions for both parts of Proposition 4 are met here, so its results apply.

1. Let $\tilde{\mathbf{G}}$ be a purely random group assignment and let $\tilde{\mathbf{p}}_i = \mathbf{p}(i, \tilde{\mathbf{G}})$. Part two of Proposition 4 applies to the joint distribution of $(\mathbf{X}, \mathbf{Y}(\tilde{\mathbf{G}}), \bar{\mathbf{X}}(\mathbf{X}, \tilde{\mathbf{G}}))$ since $\mathbf{Y}(\cdot)$ satisfies (PSE) and $\tilde{\mathbf{G}}$ satisfies (RA). By (CRA), Lemma 1 applies to the joint distribution of $(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}})$ Let (λ, η)

be defined as in equation (103) of the proof for Proposition 4. Then:

$$
\begin{aligned}
E(y_i|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)=\bar{\mathbf{x}}) && \text{(by (43) in Lemma 1)}\\
&= (\lambda_0+\eta_0(n-1))\\
&\quad +\mathbf{x}(\lambda_1+\eta_1(n-1))\\
&\quad +\bar{\mathbf{x}}\eta_2(n-1)\\
&\quad +\mathbf{x}\eta_3(n-1)\bar{\mathbf{x}}'\\
& && \text{(by result (106) in the proof for Proposition 4)}
\end{aligned}
$$

Applying the law of iterated projections:

$$
\begin{aligned}
L(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i) &= L(E(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i)|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i) && \text{(law of iterated projections)}\\
&= L\left(\begin{array}{l}(\lambda_0+\eta_0(n-1))\\ \;+\mathbf{x}_i(\lambda_1+\eta_1(n-1))\\ \;+\bar{\mathbf{x}}_i\eta_2(n-1)\\ \;+\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\end{array}\middle|\,\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i\right) && \text{(result above)}\\
&= (\lambda_0+\eta_0(n-1)) && (99)\\
&\quad +\mathbf{x}_i(\lambda_1+\eta_1(n-1))\\
&\quad +\bar{\mathbf{x}}_i\eta_2(n-1)\\
&\quad +\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\\
L(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i,\mathbf{z}_i) &= L(E(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i)|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i,\mathbf{z}_i) && \text{(law of iterated projections)}\\
&= L\left(\begin{array}{l}(\lambda_0+\eta_0(n-1))\\ \;+\mathbf{x}_i(\lambda_1+\eta_1(n-1))\\ \;+\bar{\mathbf{x}}_i\eta_2(n-1)\\ \;+\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\end{array}\middle|\,\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i,\mathbf{z}_i\right) && \text{(result above)}\\
&= (\lambda_0+\eta_0(n-1)) && (100)\\
&\quad +\mathbf{x}_i(\lambda_1+\eta_1(n-1))\\
&\quad +\bar{\mathbf{x}}_i\eta_2(n-1)\\
&\quad +\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\\
&= L(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i) && \text{(by (99) and (100))}
\end{aligned}
$$

which is result (33).

2. The assumptions here (PSE, OSE, CRA) imply that all results in Propositions 2 and 5 apply. Therefore:

$$
\begin{aligned}
APE_k &= CPE_{s,k} \qquad \text{for all } s && \text{(by (32) in Proposition 2)}\\
&= \frac{\beta_{2k}+\beta_{3sk}}{n-1} && \text{(by (45) in Proposition 5)}
\end{aligned}
$$

which can only be true if $\beta_{3sk} = 0$ for all s, k.

Proof for Proposition 4

1. Let $\tilde{\mathbf{p}}$ be a purely random draw of $(n-1)$ peers from $\mathcal{N} \setminus \{i\}$. By (RA), the actual peer group $\mathbf{p}_i$ is also a purely random draw from this set, so its joint distribution with $(y_i(\cdot), \mathbf{X})$ is identical to the joint distribution of $\tilde{\mathbf{p}}$ with $(y_i(\cdot), \mathbf{X})$. Then:

$$\begin{aligned}
CGE_{s,m} &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) && \text{(by (18))} \\
&= E(y_i(\mathbf{p}_i)|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_m) - E(y_i(\mathbf{p}_i)|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_0) \\
& && \text{(RA} \implies \text{same joint distribution)} \\
&= E(y_i|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i = \mathbf{e}_m) - E(y_i|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i = \mathbf{e}_0) && (101)
\end{aligned}$$

Since $\mathbf{x}_i$ and $\mathbf{z}_i$ are categorical, $E(y_i|\mathbf{x}_i, \mathbf{z}_i)$ is trivially linear in $(\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i)$. Therefore:

$$\begin{aligned}
E(y_i|\mathbf{x}_i, \mathbf{z}_i) &= L(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i) \\
&= \delta_0 + \mathbf{x}_i\delta_1 + \mathbf{z}_i\delta_2 + \mathbf{x}_i\delta_3\mathbf{z}_i' && \text{(by (42))}
\end{aligned}$$

Combining these two results produces:

$$\begin{aligned}
CGE_{s,m} &= E(y_i|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i = \mathbf{e}_m) - E(y_i|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i = \mathbf{e}_0) && \text{(by (101))} \\
&= (\delta_0 + \mathbf{e}_s\delta_1 + \mathbf{e}_m\delta_2 + \mathbf{e}_s\delta_3\mathbf{e}_m') - (\delta_0 + \mathbf{e}_s\delta_1 + \mathbf{e}_0\delta_2 + \mathbf{e}_s\delta_3\mathbf{e}_0') && \text{(result above)} \\
&= (\delta_0 + \mathbf{e}_s\delta_1 + \mathbf{e}_m\delta_2 + \mathbf{e}_s\delta_3\mathbf{e}_m') - (\delta_0 + \mathbf{e}_s\delta_1) && \text{(since } \mathbf{e}_0 = 0) \\
&= \mathbf{e}_m\delta_2 + \mathbf{e}_s\delta_3\mathbf{e}_m' \\
&= \delta_{2m} + \delta_{3sm}
\end{aligned}$$

which is the result in (40). Result (39) can be established by similar reasoning:

$$\begin{aligned}
AGE_m &= E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) && \text{(by (17))} \\
&= E(y_i(\mathbf{p}_i)|\mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_m) - E(y_i(\mathbf{p}_i)|\mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_0) \\
& && \text{(RA} \implies \text{same joint distribution)} \\
&= E(y_i|\mathbf{z}_i = \mathbf{e}_m) - E(y_i|\mathbf{z}_i = \mathbf{e}_0) && (102)
\end{aligned}$$

Since $\mathbf{z}_i$ is categorical, $E(y_i|\mathbf{z}_i)$ is trivially linear in $\mathbf{z}_i$. Therefore:

$$\begin{aligned}
E(y_i|\mathbf{z}_i) &= L(y_i|\mathbf{z}_i) \\
&= L(L(y_i|\mathbf{x}_i, \mathbf{z}_i)|\mathbf{z}_i) && \text{(law of iterated projections)} \\
&= L(\gamma_0 + \mathbf{x}_i\gamma_1 + \mathbf{z}_i\gamma_2|\mathbf{z}_i) && \text{(by (41))} \\
&= \gamma_0 + L(\mathbf{x}_i|\mathbf{z}_i)\gamma_1 + \mathbf{z}_i\gamma_2 \\
&= \gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{z}_i\gamma_2 && (\text{RA} \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{z}_i)
\end{aligned}$$

Combining these two results:

$$\begin{aligned}
AGE_m &= E(y_i|\mathbf{z}_i = \mathbf{e}_m) - E(y_i|\mathbf{z}_i = \mathbf{e}_0) && \text{(by (102))} \\
&= (\gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{e}_m\gamma_2) - (\gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{e}_0\gamma_2) && \text{(result above)} \\
&= (\gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{e}_m\gamma_2) - (\gamma_0 + E(\mathbf{x}_i)\gamma_1) && \text{(since } \mathbf{e}_0 = 0) \\
&= \mathbf{e}_m\gamma_2 \\
&= \gamma_{2m}
\end{aligned}$$

which is result (39).

2. By (PSE), Proposition 2 applies. Let $\lambda \equiv (\lambda_0, \lambda_1)$ and $\eta \equiv (\eta_1, \eta_2, \eta_3)$ satisfy:

$$\begin{aligned}
E(o_i|\mathbf{x}_i = \mathbf{e}_s) &= \lambda_0 + \mathbf{e}_s\lambda_1 \\
E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) &= \eta_0 + \mathbf{e}_s\eta_1 + \mathbf{e}_k\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_k'
\end{aligned} \tag{103}$$

The linear functional forms in (103) are without loss of generality since $\mathbf{x}$ is categorical. The estimand $CPE_{s,k}$ can be expressed as a function of η:

$$\begin{aligned}
CPE_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) && \text{(by (30) in Proposition 2)} \\
&\quad - E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) \\
&= (\eta_0 + \mathbf{e}_s\eta_1 + \mathbf{e}_k\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_k') && \text{(by (103))} \\
&\quad - (\eta_0 + \mathbf{e}_s\eta_1 + \mathbf{e}_0\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_0') \\
&= (\eta_0 + \mathbf{e}_s\eta_1 + \mathbf{e}_k\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_k') - (\eta_0 + \mathbf{e}_s\eta_1) && \text{(since } \mathbf{e}_0 = 0) \\
&= \mathbf{e}_k\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_k' \\
&= \eta_{2k} + \eta_{3sk} && (104)
\end{aligned}$$

The next step is to show the relationship between the coefficients in (λ, η) and the coefficients in β:

$$\begin{aligned}
E(y_i|\mathbf{X},\mathbf{G}) &= E\left(o_i + \sum_{j\in\mathbf{p}_i} p_{ij}\,\middle|\,\mathbf{X},\mathbf{G}\right) && \text{(by (28) in Proposition 2)}\\
&= E\left(o_i + \sum_{j=1}^{N} p_{ij}\mathbb{I}\left(j\in\mathbf{p}_i\right)\,\middle|\,\mathbf{X},\mathbf{G}\right) && \text{(where } \mathbb{I}\left(\cdot\right) \text{ is the indicator function)}\\
&= E(o_i|\mathbf{X},\mathbf{G}) + \sum_{j=1}^{N} E(p_{ij}|\mathbf{X},\mathbf{G})\mathbb{I}\left(j\in\mathbf{p}_i\right)\\
& && \text{(since } \mathbb{I}\left(j\in\mathbf{p}_i\right) \text{ is a function of } \mathbf{G})\\
&= E(o_i|\mathbf{X},\mathbf{G}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X},\mathbf{G})\\
&= E(o_i|\mathbf{X}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X}) && \text{(since RA} \implies (o_i, p_{ij}, \mathbf{X}) \perp\!\!\!\perp (\mathbf{G},\mathbf{p}_i))\\
&= E(o_i|\mathbf{x}_i) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j) && \text{(since (4)} \implies (\tau_i,\tau_j) \perp\!\!\!\perp \tau_{j'})\\
&= \lambda_0 + \mathbf{x}_i\lambda_1 + \sum_{j\in\mathbf{p}_i}(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}_j') && \text{(by (103))}\\
&= \lambda_0 + \mathbf{x}_i\lambda_1 + \eta_0(n-1) + \mathbf{x}_i\eta_1(n-1) + \bar{\mathbf{x}}_i\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\\
&= (\lambda_0 + \eta_0(n-1)) + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) + \bar{\mathbf{x}}_i\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'
\end{aligned} \tag{105}$$

Applying the law of iterated expectations to this result:

$$\begin{aligned}
E(y_i|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) &= E(E(y_i|\mathbf{X},\mathbf{G})|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) && \text{(law of iterated expectations)}\\
&= E\left(\begin{array}{l}(\lambda_0+\eta_0(n-1)) + \mathbf{x}_i(\lambda_1+\eta_1(n-1))\\ +\,\bar{\mathbf{x}}_i\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\end{array}\,\middle|\,\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}\right)\\
& && \text{(by (105))}\\
&= (\lambda_0+\eta_0(n-1)) + \mathbf{x}(\lambda_1+\eta_1(n-1)) + \bar{\mathbf{x}}\eta_2(n-1) + \mathbf{x}\eta_3(n-1)\bar{\mathbf{x}}'
\end{aligned} \tag{106}$$

Applying the law of iterated projections to this result:

$$
\begin{aligned}
L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i\bar{\mathbf{x}}_i') &= L(E(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i\bar{\mathbf{x}}_i') && \text{(law of iterated projections)} \\
&= L\left(\left. \begin{array}{l} (\lambda_0 + \eta_0(n-1)) \\ + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) \\ + \bar{\mathbf{x}}_i\eta_2(n-1) \\ + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i' \end{array} \right| \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i\bar{\mathbf{x}}_i' \right) && \text{(by (106))} \\
&= \underbrace{(\lambda_0 + \eta_0(n-1))}_{\beta_0} + \mathbf{x}_i \underbrace{(\lambda_1 + \eta_1(n-1))}_{\beta_1} + \bar{\mathbf{x}}_i \underbrace{\eta_2(n-1)}_{\beta_2} + \mathbf{x}_i \underbrace{\eta_3(n-1)}_{\beta_3} \bar{\mathbf{x}}_i'
\end{aligned} \tag{107}
$$

So $\beta_2 = \eta_2(n-1)$, $\beta_3 = \eta_3(n-1)$ and:

$$
\begin{aligned}
CPE_{s,k} &= \eta_{2k} + \eta_{3sk} && \text{(by (104))} \\
&= \frac{\beta_{2k} + \beta_{3sk}}{n-1} && \text{(by (107))}
\end{aligned}
$$

which is the result in (36). To get result (35), first note that:

$$
\begin{aligned}
E(p_{ij}|\mathbf{x}_j = \mathbf{x}) &= E(E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j)|\mathbf{x}_j = \mathbf{x}) && \text{(law of iterated expectations)} \\
&= E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}_j'|\mathbf{x}_j = \mathbf{x}) && \text{(by (103))} \\
&= \eta_0 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{x})\eta_1 + \mathbf{x}\eta_2 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{x})\eta_3\mathbf{x}' && \text{(conditioning rule)} \\
&= \eta_0 + E(\mathbf{x}_i)\eta_1 + \mathbf{x}\eta_2 + E(\mathbf{x}_i)\eta_3\mathbf{x}' && \text{(since (4)} \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j) \\
&= (\eta_0 + E(\mathbf{x}_i))\,\eta_1 + \mathbf{x}\,(\eta_2 + \eta_3' E(\mathbf{x}_i')) && (108)
\end{aligned}
$$

Equation (30) from Proposition 2 implies:

$$
\begin{aligned}
APE_k &= E(p_{ij}|\mathbf{x}_j = \mathbf{e}_k) - E(p_{ij}|\mathbf{x}_j = \mathbf{e}_0) && \text{(by (30) in Proposition 2)} \\
&= \ ((\eta_0 + E(\mathbf{x}_i)\eta_1) + \mathbf{e}_k\,(\eta_2 + \eta_3' E(\mathbf{x}_i'))) && \text{(by (108))} \\
&\quad - ((\eta_0 + E(\mathbf{x}_i)\eta_1) + \mathbf{e}_0\,(\eta_2 + \eta_3' E(\mathbf{x}_i'))) \\
&= \ ((\eta_0 + E(\mathbf{x}_i)\eta_1) + \mathbf{e}_k\,(\eta_2 + \eta_3' E(\mathbf{x}_i'))) && \text{(since } \mathbf{e}_0 = 0) \\
&\quad - ((\eta_0 + E(\mathbf{x}_i)\eta_1)) \\
&= \mathbf{e}_k\,(\eta_2 + \eta_3' E(\mathbf{x}_i')) && (109)
\end{aligned}
$$

Assumption (RA) implies that $\mathbf{x}_i \perp\!\!\!\perp \bar{\mathbf{x}}_i$, so:

$$
\begin{aligned}
L(y_i|\bar{\mathbf{x}}_i) &= L(L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\bar{\mathbf{x}}_i) && \text{(law of iterated projections)} \\
&= L(\alpha_0 + \mathbf{x}_i\alpha_1 + \bar{\mathbf{x}}_i\alpha_2|\bar{\mathbf{x}}_i) && \text{(definition of } \alpha) \\
&= \alpha_0 + L(\mathbf{x}_i|\bar{\mathbf{x}}_i)\alpha_1 + \bar{\mathbf{x}}_i\alpha_2 \\
&= (\alpha_0 + E(\mathbf{x}_i)\alpha_1) + \bar{\mathbf{x}}_i\alpha_2 && (\text{RA} \implies \mathbf{x}_i \perp\!\!\!\perp \bar{\mathbf{x}}_i)
\end{aligned}
$$

Having expressed $L(y_i|\bar{\mathbf{x}}_i)$ in terms of the the coefficients in α, it can also be expressed in terms of the coefficients in (λ, η):

$$\begin{aligned}
L(y_i|\bar{\mathbf{x}}_i) &= L(L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i)|\bar{\mathbf{x}}_i) && \text{(law of iterated projections)} \\
&= L\left(\begin{array}{l} (\lambda_0 + \eta_0(n-1)) + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) \\ + \bar{\mathbf{x}}_i\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i' \end{array} \middle| \bar{\mathbf{x}}_i \right) && \text{(by (107))} \\
&= (\lambda_0 + \eta_0(n-1)) + L(\mathbf{x}_i|\bar{\mathbf{x}}_i)(\lambda_1 + \eta_1(n-1)) && \text{(property of linear projection)} \\
&\quad + \bar{\mathbf{x}}_i\eta_2(n-1) + L(\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'|\bar{\mathbf{x}}_i) \\
&= (\lambda_0 + \eta_0(n-1)) + E(\mathbf{x}_i)(\lambda_1 + \eta_1(n-1)) && (\text{RA} \implies \mathbf{x}_i \perp\!\!\!\perp \bar{\mathbf{x}}_i) \\
&\quad + \bar{\mathbf{x}}_i\eta_2(n-1) + E(\mathbf{x}_i)\eta_3(n-1)\bar{\mathbf{x}}_i' \\
&= \underbrace{(\lambda_0 + \eta_0(n-1) + E(\mathbf{x}_i)(\lambda_1 + \eta_1(n-1)))}_{\alpha_0 + E(\mathbf{x}_i)\alpha_1} + \bar{\mathbf{x}}_i \underbrace{(\eta_2(n-1) + \eta_3' E(\mathbf{x}_i')(n-1))}_{\alpha_2}
\end{aligned} \tag{110}$$

So $\alpha_2 = (\eta_2(n-1) + \eta_3' E(\mathbf{x}_i')(n-1))$ and:

$$\begin{aligned}
APE_k &= \mathbf{e}_k \left(\eta_2 + \eta_3' E(\mathbf{x}_i')\right) && \text{(by (109))} \\
&= \mathbf{e}_k \frac{\alpha_2}{n-1} && \text{(by (110))} \\
&= \frac{\alpha_{2k}}{n-1}
\end{aligned}$$

which is the result in (35).

Lemma 1

Choose any $\mathbf{T_A} \in \mathcal{T}^N$, $\mathbf{G_A} \in \mathcal{G}^N$ and $\mathbf{X_A} \in \mathbb{R}^{NK}$, let $(\tau_i(\mathbf{T_A}), \mathbf{x}_i(\mathbf{X_A}))$ represent row i in $(\mathbf{T_A}, \mathbf{X_A})$, and let $\mathbf{p}_i(\mathbf{G_A}) = \mathbf{p}(i, \mathbf{G_A})$. Assumption (CRA) implies that:

$$\begin{aligned}
\Pr(\mathbf{T} = \mathbf{T_A}|\mathbf{G} = \mathbf{G_A}, \mathbf{X} = \mathbf{X_A}) &= \Pr(\mathbf{T} = \mathbf{T_A}|\mathbf{X} = \mathbf{X_A}) && \text{(by } \textbf{(CRA)}\text{)} \\
&= \prod_{i=1}^N \Pr(\tau_i = \tau_i(\mathbf{T_A})|\mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A})) && \text{(by (4))} \\
&= \prod_{i=1}^N \frac{\Pr(\tau_i = \tau_i(\mathbf{T_A}) \cap \mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A}))}{\Pr(\mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A}))} \\
&= \prod_{i=1}^N \mathbb{I}\left(\mathbf{x}_i(\mathbf{X_A}) = \mathbf{x}(\tau_i(\mathbf{T_A}))\right) \frac{\Pr(\tau_i = \tau_i(\mathbf{T_A}))}{\Pr(\mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A}))}
\end{aligned} \tag{111}$$

Therefore:

$$
\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E\left(y_i(\mathbf{p}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}_i) = \bar{\mathbf{x}}\right) && \text{(by (5) and (9))} \\
&= E\left(y\left(\tau_i, \{\tau_j\}_{j\in\mathbf{p}_i}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}_i) = \bar{\mathbf{x}}\right) \\
&= \sum_{\mathbf{T_A}\in\mathcal{T}^N} \sum_{\mathbf{G_A}\in\mathcal{G}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_A})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} \cap \mathbf{G} = \mathbf{G_A} | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}_i(\mathbf{G_A})) = \bar{\mathbf{x}}) \end{pmatrix}
\end{aligned}
$$

Let $\chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}}) \equiv \{\mathbf{X} : \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}(i, \mathbf{G_A})) = \bar{\mathbf{x}}\}$. Then:

$$
\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= \sum_{\mathbf{T_A}\in\mathcal{T}^N} \sum_{\mathbf{G_A}\in\mathcal{G}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_A})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} \cap \mathbf{G} = \mathbf{G_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \end{pmatrix} \\
& && \text{(equivalent events)} \\
&= \sum_{\mathbf{T_A}\in\mathcal{T}^N} \sum_{\mathbf{G_A}\in\mathcal{G}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_A})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \\ * \Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \end{pmatrix} && \text{(by (CRA))} \\
&= \sum_{\mathbf{G_A}\in\mathcal{G}^N} \Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \left(\sum_{\mathbf{T_A}\in\mathcal{T}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_A})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \end{pmatrix} \right)
\end{aligned}
$$

Let $\mathbf{G_B} \equiv (1, 1, \ldots, 1, 2, 2, \ldots, 2, \ldots, N)$. By the exchangeability/independence of the rows in $(\mathbf{T}, \mathbf{X})$, $\mathbf{G_A}$ can be replaced with $\mathbf{G_B}$ in the second part of the expression above:

$$
\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= \sum_{\mathbf{G_A}\in\mathcal{G}^N} \Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \left(\sum_{\mathbf{T_A}\in\mathcal{T}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_B})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}})) \end{pmatrix} \right) \\
&= \sum_{\mathbf{T_A}\in\mathcal{T}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_B})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}})) \end{pmatrix} \overbrace{\left(\sum_{\mathbf{G_A}\in\mathcal{G}^N} \Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \right)}^{=1} \\
&= \sum_{\mathbf{T_A}\in\mathcal{T}^N} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_B})}\right) \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}}))
\end{aligned}
\tag{112}
$$

Since equation (112) applies for all $\mathbf{G}$ that satisfy (CRA) it also applies for purely random $\mathbf{G}$. Let $\tilde{\mathbf{p}}$ be a purely random draw of $(n-1)$ peers from $\mathcal{N} \setminus \{i\}$. Then

$$
\begin{aligned}
E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}(\tilde{\mathbf{p}})) &= \sum_{\mathbf{T_A}\in\mathcal{T}^N} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j\in\mathbf{p}_i(\mathbf{G_B})}\right) \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}})) \\
& && \text{(by (112))} \\
&= E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) && \text{(also by (112))}
\end{aligned}
$$

which is the result in (43).

Proof for Proposition 5

Let $\tilde{\mathbf{G}}$ be a purely random group assignment and let $\tilde{\mathbf{p}}_i = \mathbf{p}(i, \tilde{\mathbf{G}})$. Since $\tilde{\mathbf{G}}$ satisfies (RA) and $\mathbf{G}$ satisfies (CRA), Lemma 1 applies:

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) = E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) \qquad \text{(by (43) in Lemma 1)}$$

Then:

1. Let $\delta^{\tilde{sat}}$ be the coefficients from estimating equation (42) with counterfactual data on outcomes $y_i(\tilde{\mathbf{p}}_i)$ and peer group composition $\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}_i)$. Since $\mathbf{z}_i^{sat}$ is saturated, the events $\mathbf{z}_i^{sat} = \mathbf{e}_m$ and $\{\bar{\mathbf{x}}_i\} = \mathbf{S}_{\bar{\mathbf{x}}}^m$ are identical, implying that:

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \mathbf{z}_i^{sat} = \mathbf{z}) = E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}_i) = \mathbf{z}) \qquad \text{for all } \mathbf{x}, \mathbf{z}$$

   which implies that:

$$\delta^{sat} = \delta^{\tilde{sat}} \tag{113}$$

   Since $\tilde{\mathbf{G}}$ satisfies (RA), part one of Proposition 4 applies to the counterfactiual $y_i(\tilde{\mathbf{p}}_i)$ and $\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}_i)$. Therefore:

$$\begin{aligned} CGE_{s,m}^{sat} &= \delta_{2m}^{\tilde{sat}} + \delta_{3sm}^{\tilde{sat}} && \text{(by (40) in Proposition 4)} \\ &= \delta_{2m}^{sat} + \delta_{3sm}^{sat} && \text{(by (113))} \end{aligned}$$

   Result (47) then follows from substitution of this result into result (25). Result (46) follows from substitution of result (47) into result (23).

2. Let $\tilde{\beta}$ be the coefficients from estimating equation (38) from with counterfactual outcomes $y_i(\tilde{\mathbf{p}}_i)$ and peer group composition $\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)$. Result (43) in Lemma 1 implies that:

$$\beta = \tilde{\beta} \tag{114}$$

   Since peer effects satistfy (PSE) and $\tilde{\mathbf{G}}$ satisfies (RA), part two of Proposition 4 applies to the counterfactiual $y_i(\tilde{\mathbf{p}}_i)$ and $\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)$. Therefore:

$$\begin{aligned} CPE_{s,k} &= \frac{\tilde{\beta}_{2k} + \tilde{\beta}_{3sk}}{n-1} && \text{(by (36) in Proposition 4)} \\ &= \frac{\beta_{2k} + \beta_{3sk}}{n-1} && \text{(by (114))} \end{aligned}$$

   which is result (45). Result (44) follows from substitution of result (45) into result (22).

Proof for Proposition 6

1. Let $\tilde{\mathbf{p}}$ be a purely random draw of $(n-1)$ peers from $\mathcal{N} \setminus \{i\}$. By construction, both $\mathbf{G}_0$ and $\mathbf{G}_1$ satisfy (CRA) and Lemma 1 applies. Therefore:

$$
\begin{aligned}
E(y_{i0}|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_{i0} = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) && \text{(by (43) in Lemma 1)} \\
&= E(y_{i1}|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_{i1} = \bar{\mathbf{x}}) && (115)
\end{aligned}
$$

For any $m > 0$, $\mathbf{S}^m_{\bar{\mathbf{x}}}$ is a singleton $\{\bar{\mathbf{x}}^m\}$ so the events $\bar{\mathbf{x}}_{i1} = \bar{\mathbf{x}}^m$ and $\mathbf{z}_{i1} = \mathbf{e}_m$ are identical. Therefore, for any $m > 0$:

$$
\begin{aligned}
E(y_{i1}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_{i1} = \mathbf{e}_m) &= E(y_{i1}|\mathbf{x}_i = \mathbf{e}_s, \bar{\mathbf{x}}_{i1} = \bar{\mathbf{x}}^m) && \text{(identical events)} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}^m) && \text{(by (115))} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) && \text{(identical events)} \\
&= CGE_{s,m} + E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) && (116)
\end{aligned}
$$

Averaging over all values of $\mathbf{z}$:

$$
\begin{aligned}
E(y_{i1}|\mathbf{x}_i = \mathbf{e}_s) &= \sum_{m=0}^{M} E(y_{i1}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_{i1} = \mathbf{e}_m) \Pr(\mathbf{z}_{i1} = \mathbf{e}_m|\mathbf{x}_i = \mathbf{e}_s) \\
&= \sum_{m=1}^{M} E(y_{i1}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_{i1} = \mathbf{e}_m) \Pr(\mathbf{z}_{i1} = \mathbf{e}_m|\mathbf{x}_i = \mathbf{e}_s) \\
& \qquad \text{(since } \Pr(\bar{\mathbf{x}}_{i1} \in \mathbf{S}^0_{\bar{\mathbf{x}}}) = 0) \\
&= \sum_{m=1}^{M} (CGE_{s,m} + E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0)) \Pr(\mathbf{z}_{i1} = \mathbf{e}_m|\mathbf{x}_i = \mathbf{e}_s) \\
& \qquad \text{(by (116))} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \left(\underbrace{\sum_{m=1}^{M} \Pr(\mathbf{z}_{i1} = \mathbf{e}_m|\mathbf{x}_i = \mathbf{e}_s)}_{1} \right) \\
& \quad + \sum_{m=1}^{M} CGE_{s,m} \Pr(\mathbf{z}_{i1} = \mathbf{e}_m|\mathbf{x}_i = \mathbf{e}_s) \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + \sum_{m=1}^{M} CGE_{s,m} \Pr(\mathbf{z}_{i1} = \mathbf{e}_m|\mathbf{x}_i = \mathbf{e}_s) \\
& \qquad (117)
\end{aligned}
$$

This result applies to y_{i0} as well as to y_{i1}, so:

$$\begin{aligned}
CRE_s(\mathbf{G}_0, \mathbf{G}_1) &= E(y_{i1} - y_{i0} | \mathbf{x}_i = \mathbf{e}_s) \\
&= E(y_{i1} | \mathbf{x}_i = \mathbf{e}_s) - E(y_{i0} | \mathbf{x}_i = \mathbf{e}_s) \\
&= \left(E(y_i(\tilde{\mathbf{p}}) | \mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + \sum_{m=1}^{M} CGE_{s,m} \Pr(\mathbf{z}_{i1} = \mathbf{e}_m | \mathbf{x}_i = \mathbf{e}_s) \right) \\
&\quad - \left(E(y_i(\tilde{\mathbf{p}}) | \mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + \sum_{m=1}^{M} CGE_{s,m} \Pr(\mathbf{z}_{i0} = \mathbf{e}_m | \mathbf{x}_i = \mathbf{e}_s) \right) \\
&\qquad \text{(by (117))} \\
&= \sum_{m=1}^{M} CGE_{s,m} \left(\Pr(\mathbf{z}_{i1} = \mathbf{e}_m | \mathbf{x}_i = \mathbf{e}_s) - \Pr(\mathbf{z}_{i0} = \mathbf{e}_m | \mathbf{x}_i = \mathbf{e}_s) \right) \\
&= \mathbf{CGE}_s E(\mathbf{z}'_{i1} - \mathbf{z}'_{i0} | \mathbf{x}_i = \mathbf{e}_s)
\end{aligned}$$

which is the result in (49).

2. Given (PSE), part 1 of Proposition 2 applies:

$$\begin{aligned}
E(y_{i1} | \mathbf{X}, \mathbf{G}_1) &= E(y_i(\mathbf{p}_{i1}) | \mathbf{X}, \mathbf{G}_1) && \text{(definition)} \\
&= E\left(o_i + \sum_{j \in \mathbf{p}_{i1}} p_{ij} \,\middle|\, \mathbf{X}, \mathbf{G}_1 \right) && \text{(by Proposition 2)} \\
&= E\left(o_i | \, \mathbf{X}, \mathbf{G}_1 \right) + \sum_{j \in \mathbf{p}_{i1}} E\left(p_{ij} | \, \mathbf{X}, \mathbf{G}_1 \right) \\
&= E\left(E(o_i | \mathbf{X}, \mathbf{G}_1, \sigma) | \, \mathbf{X}, \mathbf{G}_1 \right) + \sum_{j \in \mathbf{p}_{i1}} E\left(E(p_{ij} | \mathbf{X}, \mathbf{G}_1, \sigma) | \, \mathbf{X}, \mathbf{G}_1 \right) \\
& && \text{(law of iterated expectations)} \\
&= E\left(E(o_i | \mathbf{X}, \sigma) | \, \mathbf{X}, \mathbf{G}_1 \right) + \sum_{j \in \mathbf{p}_{i1}} E\left(E(p_{ij} | \mathbf{X}, \sigma) | \, \mathbf{X}, \mathbf{G}_1 \right) \\
& && \text{(since } \mathbf{G}_1 \text{ is a function of } (\mathbf{X}, \sigma)) \\
&= E\left(E(o_i | \mathbf{x}_i) | \, \mathbf{X}, \mathbf{G}_1 \right) + \sum_{j \in \mathbf{p}_{i1}} E\left(E(p_{ij} | \mathbf{x}_i, \mathbf{x}_j) | \, \mathbf{X}, \mathbf{G}_1 \right) \\
&= E(o_i | \mathbf{x}_i) + \sum_{j \in \mathbf{p}_{i1}} E(p_{ij} | \mathbf{x}_i, \mathbf{x}_j) \\
&= \lambda_0 + \mathbf{x}_i \lambda_1 + \sum_{j \in \mathbf{p}_{i1}} \eta_0 + \mathbf{x}_i \eta_1 + \mathbf{x}_j \eta_2 + \mathbf{x}_i \eta_3 \mathbf{x}'_j \\
& && \text{(where } (\lambda, \eta) \text{ are defined as in (103))} \\
&= (\lambda_0 + \eta_0(n-1)) + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) + \bar{\mathbf{x}}_{i1} \eta_2 (n-1) + \mathbf{x}_i \eta_3 (n-1) \bar{\mathbf{x}}'_{i1} && (118)
\end{aligned}$$

Averaging over values of $\bar{\mathbf{x}}$:

$$\begin{aligned}
E(y_{i1}|\mathbf{x}_i = \mathbf{x}) &= E(E(y_{i1}|\mathbf{X}, \mathbf{G}_1)|\mathbf{x}_i = \mathbf{x}) && \text{(Law of iterated expectations)} \\
&= E\left(\begin{array}{l} (\lambda_0 + \eta_0(n-1)) + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) \\ + \bar{\mathbf{x}}_{i1}\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}'_{i1} \end{array} \middle| \mathbf{x}_i = \mathbf{x} \right) && \text{(by (118))} \\
&= (\lambda_0 + \eta_0(n-1)) + \mathbf{x}(\lambda_1 + \eta_1(n-1)) && (119) \\
&\quad + E(\bar{\mathbf{x}}_{i1}|\mathbf{x}_i = \mathbf{x})\eta_2(n-1) + \mathbf{x}\eta_3(n-1)E(\bar{\mathbf{x}}'_{i1}|\mathbf{x}_i = \mathbf{x})
\end{aligned}$$

This result also applies to $\mathbf{G}_0$, so:

$$\begin{aligned}
CRE_s(\mathbf{G}_0, \mathbf{G}_1) &= E(y_{i1} - y_{i0}|\mathbf{x}_i = \mathbf{e}_s) \\
&= \left(\begin{array}{l} (\lambda_0 + \eta_0(n-1)) + \mathbf{e}_s(\lambda_1 + \eta_1(n-1)) \\ + E(\bar{\mathbf{x}}_{i1}|\mathbf{x}_i = \mathbf{e}_s)\eta_2(n-1) + \mathbf{e}_s\eta_3(n-1)E(\bar{\mathbf{x}}'_{i1}|\mathbf{x}_i = \mathbf{e}_s) \end{array} \right) \\
&\quad - \left(\begin{array}{l} (\lambda_0 + \eta_0(n-1)) + \mathbf{e}_s(\lambda_1 + \eta_1(n-1)) \\ + E(\bar{\mathbf{x}}_{i0}|\mathbf{x}_i = \mathbf{e}_s)\eta_2(n-1) + \mathbf{e}_s\eta_3(n-1)E(\bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{e}_s) \end{array} \right) \\
& && \text{(by (119))} \\
&= (\eta'_2(n-1) + \mathbf{e}_s\eta_3(n-1))\left(E(\bar{\mathbf{x}}'_{i1}|\mathbf{x}_i = \mathbf{e}_s) - E(\bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{e}_s)\right) \\
&= \mathbf{CPE}_s E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{e}_s)(n-1)
\end{aligned}$$

which is the result in (51).

3. Given (PSE, OSE), part two of Proposition 2 applies. By equation (32) in Proposition 2, $\mathbf{CPE}_s = \mathbf{APE}$ and so the the result in (53) follows from (51) by substitution. The second result follows from:

$$\begin{aligned}
E(y_{i1} - y_{i0}) &= \sum_{s=0}^{K} E(y_{i1} - y_{i0}|\mathbf{x}_i = \mathbf{e}_s)\Pr(\mathbf{x}_i = \mathbf{e}_s) && \text{(law of total probability)} \\
&= \sum_{s=0}^{K} CRE_s(\mathbf{G}_0, \mathbf{G}_1)\Pr(\mathbf{x}_i = \mathbf{e}_s) && \text{(definition of } CRE_s) \\
&= \sum_{s=0}^{K} \mathbf{APE}\, E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{e}_s)(n-1)\Pr(\mathbf{x}_i = \mathbf{e}_s) && \text{(by result (53))} \\
&= \mathbf{APE}(n-1)\sum_{s=0}^{K} E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{e}_s)\Pr(\mathbf{x}_i = \mathbf{e}_s) \\
&= \mathbf{APE}(n-1)\underbrace{E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0})}_{0} \\
&= 0
\end{aligned}$$

Proof for Proposition 7

Given (OSE,PSE), result 2 in Proposition 2 applies, and we can express each outcome y_i as a sum of own effects o_i and peer effects p_j. Let $\eta \equiv (\eta_0, \eta_1)$ such that:

$$E(p_i|\mathbf{x}_i) = \eta_0 + \mathbf{x}_i\eta_1 \tag{120}$$

and let $\lambda^\ell \equiv (\lambda_0^\ell, \lambda_1^\ell)$ and $\lambda \equiv (\lambda_0, \lambda_1)$ such that:

$$E(o_i|\mathbf{x}_i, \ell_i = \ell) = \lambda_0^\ell + \mathbf{x}_i\lambda_1^\ell \tag{121}$$

$$E(o_i|\mathbf{x}_i) = \lambda_0 + \mathbf{x}_i\lambda_1 \tag{122}$$

By constant shifts (CS), λ_1^ℓ does not vary across locations:

$$\begin{aligned}
\lambda_{1k}^\ell &= E(o_i|\mathbf{x}_i = \mathbf{e}_k, \ell_i = \ell) - E(o_i = \mathbf{e}_0|\mathbf{x}_i, \ell_i = \ell) \\
&= E(o_i|\mathbf{x}_i = \mathbf{e}_k) - E(o_i = \mathbf{e}_0|\mathbf{x}_i) && \text{(by CS)} \\
&= \lambda_{1k}
\end{aligned}$$

implying that:

$$E(o_i|\mathbf{x}_i, \ell_i = \ell) = \lambda_0^\ell + \mathbf{x}_i\lambda_1 \tag{123}$$

Substituting these results into equation (31):

$$\begin{aligned}
E(y_i|\mathbf{X}, \mathbf{G}, \mathbf{L}) &= E\left(o_i + \sum_{j \in \mathbf{p}_i} p_j \,\middle|\, \mathbf{X}, \mathbf{G}, \mathbf{L} \right) && \text{(by (31))} \\
&= E\left(o_i \,|\mathbf{X}, \mathbf{G}, \mathbf{L}\right) + \sum_{j \in \mathbf{p}_i} E\left(p_j \,|\mathbf{X}, \mathbf{G}, \mathbf{L}\right) \\
&= E\left(o_i \,|\mathbf{x}_i, \ell_i\right) + \sum_{j \in \mathbf{p}_i} E\left(p_j \,|\mathbf{x}_j, \ell_j\right) && \text{(by RAL)} \\
&= E\left(o_i \,|\mathbf{x}_i, \ell_i\right) + \sum_{j \in \mathbf{p}_i} E\left(p_j \,|\mathbf{x}_j\right) && \text{(by LI)} \\
&= \lambda_0^{\ell_i} + \lambda_1\mathbf{x}_i + \sum_{j \in \mathbf{p}_i} (\eta_0 + \mathbf{x}_j\eta_1) && \text{(by results above)} \\
&= (\lambda_0^{\ell_i} + \eta_0(n-1)) + \lambda_1\mathbf{x}_i + \bar{\mathbf{x}}_i\eta_1(n-1) && (124)
\end{aligned}$$

Applying the law of iterated expectations:

$$\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}, \ell_i = \ell) &= E(E(y_i|\mathbf{X}, \mathbf{G}, \mathbf{L})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}, \ell_i = \ell) && \text{(by LIE)} \\
&= \underbrace{(\lambda_0^\ell + \eta_0(n-1))}_{\alpha_0^\ell} + \mathbf{x} \underbrace{\lambda_1}_{\alpha_1} + \bar{\mathbf{x}} \underbrace{\eta_1(n-1)}_{\alpha_2} && (125)
\end{aligned}$$

Finally, result 2 in Proposition 2 implies:

$$\begin{aligned} CPE_{sk} = APE_k &= E(p_i|\mathbf{x}_i = \mathbf{e}_k) - E(p_i|\mathbf{x}_i = \mathbf{e}_0) && (126)\\ &= \eta_{1k} && (127)\\ &= \alpha_{2k}/(n-1) && (128) \end{aligned}$$

Proof for Proposition 8

1. The proof here is essentially the same as the proof for part two of Proposition 4, but conditioning on ℓ_i. Given (PSE), Proposition 2 implies the potential outcome function can be written as in equation (28) and within-location conditional average peer effects can be written in terms of p_{ij}:

$$\begin{aligned} CPE^{\ell}_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell) && (129)\\ &\quad - E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0, \ell_i = \ell_j = \ell)\\ APE^{\ell}_k &= E(p_{ij}|\mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell) - E(p_{ij}|\mathbf{x}_j = \mathbf{e}_0, \ell_i = \ell_j = \ell) \end{aligned}$$

Without loss of generality, let $\lambda^{\ell} \equiv (\lambda^{\ell}_0, \lambda^{\ell}_1)$ and $\eta^{\ell\ell'} \equiv (\eta^{\ell\ell'}_1, \eta^{\ell\ell'}_2, \eta^{\ell\ell'}_3)$ satisfy:

$$\begin{aligned} E(o_i|\mathbf{x}_i, \ell_i = \ell) &= \lambda^{\ell}_0 + \mathbf{x}_i\lambda^{\ell}_1 && (130)\\ E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell, \ell_j = \ell') &= \eta^{\ell\ell'}_0 + \mathbf{x}_i\eta^{\ell\ell'}_1 + \mathbf{x}_j\eta^{\ell\ell'}_2 + \mathbf{x}_i\eta^{\ell\ell'}_3\mathbf{x}'_j \end{aligned}$$

These two results can be combined to find $\mathbf{CPE}^{\ell}$ in terms of $\eta^{\ell\ell}$:

$$\begin{aligned} CPE^{\ell}_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell) && \text{(by (129))}\\ &\quad - E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0, \ell_i = \ell_j = \ell)\\ &= \left(\eta^{\ell\ell}_0 + \mathbf{e}_s\eta^{\ell\ell}_1 + \mathbf{e}_k\eta^{\ell\ell}_2 + \mathbf{e}_s\eta^{\ell\ell}_3\mathbf{e}'_k\right) && \text{(by (130))}\\ &\quad - \left(\eta^{\ell\ell}_0 + \mathbf{e}_s\eta^{\ell\ell}_1 + \mathbf{e}_0\eta^{\ell\ell}_2 + \mathbf{e}_s\eta^{\ell\ell}_3\mathbf{e}'_0\right)\\ &= \left(\eta^{\ell\ell}_0 + \mathbf{e}_s\eta^{\ell\ell}_1 + \mathbf{e}_k\eta^{\ell\ell}_2 + \mathbf{e}_s\eta^{\ell\ell}_3\mathbf{e}'_k\right) - \left(\eta^{\ell\ell}_0 + \mathbf{e}_s\eta^{\ell\ell}_1\right) && \text{(since } \mathbf{e}_0 = 0)\\ &= \mathbf{e}_k\eta^{\ell\ell}_2 + \mathbf{e}_s\eta^{\ell\ell}_3\mathbf{e}'_k\\ &= \eta^{\ell\ell}_{2k} + \eta^{\ell\ell}_{3sk} && (131) \end{aligned}$$

The next step is to find the relationship between $\eta^{\ell\ell}$ and β^{ℓ} by finding the best linear predictor $L^{\ell}(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i)$ in terms of $\eta^{\ell\ell}$:

$$
\begin{aligned}
E(y_i|\mathbf{X},\mathbf{G},\mathbf{L}) &= E\left(o_i + \sum_{j\in\mathbf{p}_i} p_{ij} \middle| \mathbf{X},\mathbf{G},\mathbf{L}\right) && \text{(by Proposition 2)}\\
&= E\left(o_i + \sum_{j=1}^{N} p_{ij}\mathbb{I}\left(j\in\mathbf{p}_i\right) \middle| \mathbf{X},\mathbf{G},\mathbf{L}\right) \\
& && \text{(where } \mathbb{I}\left(\right) \text{ is the indicator function)}\\
&= E(o_i|\mathbf{X},\mathbf{G},\mathbf{L}) + \sum_{j=1}^{N} E(p_{ij}\mathbb{I}\left(j\in\mathbf{p}_i\right)|\mathbf{X},\mathbf{G},\mathbf{L})\\
&= E(o_i|\mathbf{X},\mathbf{G},\mathbf{L}) + \sum_{j=1}^{N} E(p_{ij}|\mathbf{X},\mathbf{G},\mathbf{L})\mathbb{I}\left(j\in\mathbf{p}_i\right)\\
& && \text{(since } \mathbb{I}\left(j\in\mathbf{p}_i\right) \text{ is a function of } \mathbf{G})\\
&= E(o_i|\mathbf{X},\mathbf{G},\mathbf{L}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X},\mathbf{G},\mathbf{L})\\
&= E(o_i|\mathbf{X},\mathbf{L}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X},\mathbf{L}) && (\text{RAL} \implies \mathbf{T}\perp\!\!\!\perp\mathbf{G}|\mathbf{L})\\
&= E(o_i|\mathbf{x}_i,\ell_i) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j,\ell_i,\ell_j=\ell_i) && \text{(by (58))}\\
&= \lambda_0^{\ell_i} + \mathbf{x}_i\lambda_1^{\ell_i} + \sum_{j\in\mathbf{p}_i}\left(\eta_0^{\ell_i\ell_i} + \mathbf{x}_i\eta_1^{\ell_i\ell_i} + \mathbf{x}_j\eta_2^{\ell_i\ell_i} + \mathbf{x}_i\eta_3^{\ell_i\ell_i}\mathbf{x}_j'\right) && \text{(by (130))}\\
&= \left(\lambda_0^{\ell_i} + \eta_0^{\ell_i\ell_i}(n-1)\right) + \mathbf{x}_i\left(\lambda_1^{\ell_i\ell_i} + \eta_1^{\ell_i\ell_i}(n-1)\right) && (132)\\
&\quad + \bar{\mathbf{x}}_i\eta_2^{\ell_i\ell_i}(n-1) + \mathbf{x}_i\eta_3^{\ell_i\ell_i}(n-1)\bar{\mathbf{x}}_i'
\end{aligned}
$$

Applying the law of iterated projections:

$$
\begin{aligned}
L^{\ell}\left(y_i \middle| \begin{matrix}\mathbf{x}_i,\bar{\mathbf{x}}_i,\\ \mathbf{x}_i'\bar{\mathbf{x}}_i\end{matrix}\right) &= L^{\ell}\left(E\left(y_i|\,\mathbf{X},\mathbf{G},\mathbf{L}\right)|\,\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)}\\
&= L^{\ell}\left(\begin{matrix}(\lambda_0^{\ell_i} + \eta_0^{\ell_i}(n-1))\\ + \mathbf{x}_i(\lambda_1^{\ell_i\ell_i} + \eta_1^{\ell_i\ell_i}(n-1))\\ + \bar{\mathbf{x}}_i\eta_2^{\ell_i\ell_i}(n-1)\\ + \mathbf{x}_i\eta_3^{\ell_i\ell_i}(n-1)\bar{\mathbf{x}}_i'\end{matrix} \middle| \mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i\right) && \text{(by (132))}\\
&= \underbrace{(\lambda_0^{\ell} + \eta_0^{\ell\ell}(n-1))}_{\beta_0^{\ell}} + \mathbf{x}_i\underbrace{(\lambda_1^{\ell} + \eta_1^{\ell\ell}(n-1))}_{\beta_1^{\ell}} && (133)\\
&\quad + \bar{\mathbf{x}}_i\underbrace{\eta_2^{\ell\ell}(n-1)}_{\beta_2^{\ell}} + \mathbf{x}_i\underbrace{\eta_3^{\ell\ell}(n-1)}_{\beta_3^{\ell}}\bar{\mathbf{x}}_i'
\end{aligned}
$$

So $\beta_2^\ell = \eta_2^{\ell\ell}(n-1)$, $\beta_3^\ell = \eta_3^{\ell\ell}(n-1)$, and:

$$\begin{aligned} CPE_{s,k}^\ell &= \eta_{2k}^{\ell\ell} + \eta_{3sk}^{\ell\ell} && \text{(by (131))} \\ &= \frac{\beta_{2k}^\ell + \beta_{3sk}^\ell}{n-1} && \text{(by (133))} \end{aligned}$$

which is the result in (65). The same procedure can be used to derive the result in (66). First, express $\mathbf{APE}^\ell$ in terms of $\eta^{\ell\ell}$:

$$\begin{aligned} APE_k^\ell &= E(p_{ij}|\mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell) - E(p_{ij}|\mathbf{x}_j = \mathbf{e}_0, \ell_i = \ell_j = \ell) && \text{(by (129))} \\ &= E\left(E\left(p_{ij}|\, \mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell_j = \ell\right)|\, \mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell\right) \\ &\quad - E\left(E\left(p_{ij}|\, \mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell_j = \ell\right)|\, \mathbf{x}_j = \mathbf{e}_0, \ell_i = \ell_j = \ell\right) \\ & && \text{(law of iterated expectations)} \\ &= E\left(\eta_0^{\ell\ell} + \mathbf{x}_i\eta_1^{\ell\ell} + \mathbf{x}_j\eta_2^{\ell\ell} + \mathbf{x}_i\eta_3^{\ell\ell}\mathbf{x}_j' \middle|\, \mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell\right) && \text{(by 130))} \\ &\quad - E\left(\eta_0^{\ell\ell} + \mathbf{x}_i\eta_1^{\ell\ell} + \mathbf{x}_j\eta_2^{\ell\ell} + \mathbf{x}_i\eta_3^{\ell\ell}\mathbf{x}_j' \middle|\, \mathbf{x}_j = \mathbf{e}_0, \ell_i = \ell_j = \ell\right) \\ &= \eta_0^{\ell\ell} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell)\eta_1^{\ell\ell} && \text{(since } \mathbf{e}_0 = 0) \\ &\quad + \mathbf{e}_k\eta_2^{\ell\ell} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_k, \ell_i = \ell_j = \ell)\eta_3^{\ell\ell}\mathbf{e}_k' \\ &\quad - \left(\eta_0^{\ell\ell} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_0, \ell_i = \ell_j = \ell)\eta_1^{\ell\ell}\right) \\ &= \eta_0^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell_j = \ell)\eta_1^{\ell\ell} && ((58) \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j|\mathbf{L}) \\ &\quad + \mathbf{e}_k\eta_2^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell_j = \ell)\eta_3^{\ell\ell}\mathbf{e}_k' \\ &\quad - \left(\eta_0^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell_j = \ell)\eta_1^{\ell\ell}\right) \\ &= \mathbf{e}_k\eta_2^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell)\eta_3^{\ell\ell}\mathbf{e}_k' \\ &= \mathbf{e}_k\left(\eta_2^{\ell\ell} + (\eta_3^{\ell\ell})'E(\mathbf{x}_i'|\ell_i = \ell)\right) && (134) \end{aligned}$$

Then find the relationship between α^ℓ and $\eta^{\ell\ell}$ by expressing $L^\ell(y_i|\bar{\mathbf{x}}_i)$ in terms of α^ℓ and in terms of $\eta^{\ell\ell}$:

$$\begin{aligned}
L^\ell(y_i|\bar{\mathbf{x}}_i) &= L^\ell\left(L^\ell\left(y_i|\,\mathbf{x}_i,\bar{\mathbf{x}}_i\right)\big|\,\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)}\\
&= L^\ell\left(\alpha_0^\ell + \mathbf{x}_i\alpha_1^\ell + \bar{\mathbf{x}}_i\alpha_2^\ell\big|\,\bar{\mathbf{x}}_i\right) && \text{(by (61))}\\
&= \alpha_0^\ell + L^\ell(\mathbf{x}_i|\bar{\mathbf{x}}_i)\alpha_1^\ell + \bar{\mathbf{x}}_i\alpha_2^\ell\\
&= \alpha_0^\ell + E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell + \bar{\mathbf{x}}_i\alpha_2^\ell && \text{(by (58))}\\
&= \left(\alpha_0^\ell + E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell\right) + \bar{\mathbf{x}}_i\alpha_2^\ell && \text{(135)}\\
L^\ell(y_i|\bar{\mathbf{x}}_i) &= L^\ell\left(E\left(y_i|\,\mathbf{X},\mathbf{G},\mathbf{L}\right)|\,\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)}\\
&= L^\ell\left(\begin{array}{l|}(\lambda_0^{\ell_i}+\eta_0^{\ell_i}(n-1))\\ \;+\mathbf{x}_i(\lambda_1^{\ell_i\ell_i}+\eta_1^{\ell_i\ell_i}(n-1))\\ \;+\bar{\mathbf{x}}_i\eta_2^{\ell_i\ell_i}(n-1)\\ \;+\mathbf{x}_i\eta_3^{\ell_i\ell_i}(n-1)\bar{\mathbf{x}}_i'\end{array}\;\bar{\mathbf{x}}_i\right) && \text{(by (133))}\\
&= (\lambda_0^\ell+\eta_0^\ell(n-1)) + L^\ell(\mathbf{x}_i|\bar{\mathbf{x}}_i)(\lambda_1^{\ell\ell}+\eta_1^{\ell\ell}(n-1))\\
&\quad + \bar{\mathbf{x}}_i\eta_2^{\ell\ell}(n-1) + L^\ell(\mathbf{x}_i|\bar{\mathbf{x}}_i)\eta_3^{\ell\ell}(n-1)\bar{\mathbf{x}}_i'\\
&= (\lambda_0^\ell+\eta_0^\ell(n-1)) + E(\mathbf{x}_i|\ell_i=\ell)(\lambda_1^{\ell\ell}+\eta_1^{\ell\ell}(n-1)) && \text{(by (58))}\\
&\quad + \bar{\mathbf{x}}_i\eta_2^{\ell\ell}(n-1) + E(\mathbf{x}_i|\ell_i=\ell)\eta_3^{\ell\ell}(n-1)\bar{\mathbf{x}}_i'\\
&= \underbrace{(\lambda_0^\ell+\eta_0^\ell(n-1)) + E(\mathbf{x}_i|\ell_i=\ell)(\lambda_1^{\ell\ell}+\eta_1^{\ell\ell}(n-1))}_{\alpha_0^\ell+E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell} && \text{(136)}\\
&\quad + \bar{\mathbf{x}}_i\underbrace{\left(\eta_2^{\ell\ell}(n-1) + (\eta_3^{\ell\ell})'E(\mathbf{x}_i'|\ell_i=\ell)(n-1)\right)}_{\alpha_2^\ell}
\end{aligned}$$

So $\alpha_2^\ell = (\eta_2^{\ell\ell}(n-1) + (\eta_3^{\ell\ell})'E(\mathbf{x}_i'|\ell_i=\ell)(n-1))$ and:

$$\begin{aligned}
APE_k^\ell &= \mathbf{e}_k(\eta_2^{\ell\ell} + (\eta_3^{\ell\ell})'E(\mathbf{x}_i'|\ell_i=\ell)) && \text{(by (134))}\\
&= \mathbf{e}_k\frac{\alpha_2^\ell}{n-1} && \text{(by (135) and (136))}\\
&= \frac{\alpha_{2k}^\ell}{n-1}
\end{aligned}$$

which is the result in (66).

2. First note that:

$$\begin{aligned}
CPE_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) - E(p_{ij}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) && \text{(by Proposition 2)} \\
&= E(E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i, \ell_j)|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) && \text{(law of iterated expectations)} \\
&\quad - E(E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i, \ell_j)|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) \\
&= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) && \text{(by (130))} \\
&\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_i} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) \\
&= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) && (137) \\
&\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j}|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0)
\end{aligned}$$

Without assumption (LI), **CPE** is not identified since $\eta^{\ell_i\ell_j}$ is only identified when $\ell_i = \ell_j$. However, (LI) implies that there exists a constant vector $\eta = (\eta_0, \eta_1, \eta_2, \eta_3)$ such that:

$$\eta^{\ell\ell'} = \eta \qquad \text{for all } \ell, \ell' \tag{138}$$

Applying (138) to (137) allows $CPE_{s,k}$ to be expressed in terms of η:

$$\begin{aligned}
CPE_{s,k} &= E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}_j'|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) && \text{(by (138))} \\
&\quad - E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}_j'|\mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_0) \\
&= (\eta_0 + \mathbf{e}_s\eta_1 + \mathbf{e}_k\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_k') - (\eta_0 + \mathbf{e}_s\eta_1 + \mathbf{e}_0\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_0') \\
&= \mathbf{e}_k\eta_2 + \mathbf{e}_s\eta_3\mathbf{e}_k' \\
&= \eta_{2k} + \eta_{3sk} && (139)
\end{aligned}$$

Applying (138) to (133) allows $L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i)$ to be expressed in terms of η:

$$L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) = \underbrace{(\lambda_0^\ell + \eta_0(n-1))}_{\beta_0^\ell} + \mathbf{x}_i \underbrace{(\lambda_1^\ell + \eta_1(n-1))}_{\beta_1^\ell} + \bar{\mathbf{x}}_i \underbrace{\eta_2(n-1)}_{\beta_2^\ell} + \mathbf{x}_i \underbrace{\eta_3(n-1)}_{\beta_3^\ell} \bar{\mathbf{x}}_i' \tag{140}$$

which implies that:

$$\begin{aligned}
CPE_{s,k} &= \eta_{2k} + \eta_{3sk} && \text{(by (139))} \\
&= \frac{E(\beta_{2k}^{\ell_i}) + E(\beta_{3sk}^{\ell_i})}{n-1} && \text{(by (140))}
\end{aligned}$$

which is the result in (67).[8] To prove the result in (68), first note that:

$$
\begin{aligned}
APE_k &= E(p_{ij}|\mathbf{x}_j = \mathbf{e}_k) - E(p_{ij}|\mathbf{x}_j = \mathbf{e}_0) && \text{(by Proposition 2)} \\
&= \quad E(E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i, \ell_j)|\mathbf{x}_j = \mathbf{e}_k) && \text{(law of iterated expectations)} \\
&\quad - E(E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i, \ell_j)|\mathbf{x}_j = \mathbf{e}_0) \\
&= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_j = \mathbf{e}_k) && \text{(by (130))} \\
&\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_i} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_j = \mathbf{e}_0) \\
&= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_j = \mathbf{e}_k) && (141) \\
&\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j}|\mathbf{x}_j = \mathbf{e}_0)
\end{aligned}
$$

Applying (138) to (141) allows APE_k to be expressed in terms of η:

$$
\begin{aligned}
APE_k &= E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}_j'|\mathbf{x}_j = \mathbf{e}_k) && \text{(by (138))} \\
&\quad - E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}_j'|\mathbf{x}_j = \mathbf{e}_0) \\
&= \ (\eta_0 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_k)\eta_1 + \mathbf{e}_k\eta_2 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_k)\eta_3\mathbf{e}_k') \\
&\quad - (\eta_0 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_0)\eta_1 + \mathbf{e}_0\eta_2 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_0)\eta_3\mathbf{e}_0') \\
&= \ (\eta_0 + E(\mathbf{x}_i)\eta_1 + \mathbf{e}_k\eta_2 + E(\mathbf{x}_i)\eta_3\mathbf{e}_k') && (\text{since } \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j) \\
&\quad - (\eta_0 + E(\mathbf{x}_i)\eta_1 + \mathbf{e}_0\eta_2 + E(\mathbf{x}_i)\eta_3\mathbf{e}_0') \\
&= \mathbf{e}_k\eta_2 + E(\mathbf{x}_i)\eta_3\mathbf{e}_k' && (\text{since } \mathbf{e}_0 = 0) \\
&= \mathbf{e}_k(\eta_2 + \eta_3' E(\mathbf{x}_i')) && (142)
\end{aligned}
$$

Applying (138) to (136) allows $L^\ell(y_i|\bar{\mathbf{x}}_i)$ to be expressed in terms of η:

$$
\begin{aligned}
L^\ell(y_i|\bar{\mathbf{x}}_i) = \ & \underbrace{(\lambda_0^\ell + \eta_0(n-1)) + E(\mathbf{x}_i|\ell_i = \ell)(\lambda_1^{\ell\ell} + \eta_1(n-1))}_{\alpha_0^\ell + E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell} && (143) \\
& + \bar{\mathbf{x}}_i \underbrace{(\eta_2(n-1) + \eta_3' E(\mathbf{x}_i'|\ell_i = \ell)(n-1))}_{\alpha_2^\ell}
\end{aligned}
$$

which implies that:

$$
\begin{aligned}
E(\alpha_2^{\ell_i}) &= E\left(\eta_2(n-1) + \eta_3' E(\mathbf{x}_i'|\ell_i)(n-1)\right) && \text{(by (143))} \\
&= \left(\eta_2 + \eta_3' E(E(\mathbf{x}_i'|\ell_i))\right)(n-1) \\
&= \left(\eta_2 + \eta_3' E(\mathbf{x}_i')\right)(n-1) && (144)
\end{aligned}
$$

[8] Note that β_2^ℓ, β_3^ℓ, and α_2^ℓ do not vary by ℓ, so it is not strictly necessary to average across locations in (67) and (68) rather than simply choosing an arbitrary location. In a finite sample, an average of noisy estimators would typically outperform any one of those estimators.

and therefore:

$$\begin{aligned}
APE_k &= \mathbf{e}_k(\eta_2 + \eta_3' E(\mathbf{x}_i')) && \text{(by (142))} \\
&= \frac{\mathbf{e}_k E(\alpha_2^{\ell_i})}{n-1} && \text{(by (144))} \\
&= \frac{E(\alpha_{2k}^{\ell_i})}{n-1}
\end{aligned}$$

which is the result in (68).

3. Assumption (OSE) implies that:

$$\begin{aligned}
E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell, \ell_j = \ell') &= E(p_j|\mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell, \ell_j = \ell') && \text{(by OSE)} \\
&= E(p_j|\mathbf{x}_j, \ell_j = \ell') && \text{(by (58))} \\
\implies (\eta_0^{\ell\ell'}, \eta_1^{\ell\ell'}, \eta_2^{\ell\ell'}, \eta_3^{\ell\ell'}) &= (\eta_0^{\ell'}, 0, \eta_2^{\ell'}, 0) && (145)
\end{aligned}$$

Applying (145) to (136) produces:

$$L^\ell(y_i|\bar{\mathbf{x}}_i) = \underbrace{(\lambda_0^\ell + \eta_0^\ell(n-1)) + E(\mathbf{x}_i|\ell_i = \ell)\lambda_1^\ell}_{\alpha_0^\ell + E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell} + \bar{\mathbf{x}}_i \underbrace{\eta_2^\ell(n-1)}_{\alpha_2^\ell} \quad (146)$$

Applying (145) to (141) produces:

$$\begin{aligned}
APE_k &= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_j = \mathbf{e}_k) && \text{(by (141))} \\
&\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j}|\mathbf{x}_j = \mathbf{e}_0) \\
&= E(\eta_0^{\ell_j} + \mathbf{x}_j\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{e}_k) && \text{(by (145))} \\
&\quad - E(\eta_0^{\ell_j}|\mathbf{x}_j = \mathbf{e}_0) \\
&= E(\eta_0^{\ell_j}|\mathbf{x}_j = \mathbf{e}_k) - E(\eta_0^{\ell_j}|\mathbf{x}_j = \mathbf{e}_0) + \mathbf{e}_k E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{e}_k) && (147)
\end{aligned}$$

The third term in this expression is identified, but the first two terms are not, because η_0^ℓ cannot be distinguished from λ_0^ℓ. However, assumption (PLI) implies that $\eta_0^{\ell'} = \eta_0$, so:

$$\begin{aligned}
APE_k &= E(\eta_0|\mathbf{x}_j = \mathbf{e}_k) - E(\eta_0|\mathbf{x}_j = \mathbf{e}_0) + \mathbf{e}_k E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{e}_k) && \text{(by PLI)} \\
&= \eta_0 - \eta_0 + \mathbf{e}_k E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{e}_k) \\
&= \mathbf{e}_k E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{e}_k) \\
&= \mathbf{e}_k E(\eta_2^{\ell_i}|\mathbf{x}_i = \mathbf{e}_k) && \text{(by exchangeability)} \\
&= \frac{\mathbf{e}_k E(\alpha_2^{\ell_i}|\mathbf{x}_i = \mathbf{e}_k)}{n-1} && \text{(by (146))} \\
&= \frac{E(\alpha_{2k}^{\ell_i}|\mathbf{x}_i = \mathbf{e}_k)}{n-1}
\end{aligned}$$

which is the result in (69).

Proof for Proposition 9

1. Let $\mathbf{q}$ be an arbitrary group of $(n-2)$ peers, and let $\mathbf{q}_0$ be a group of $(n-2)$ peers for whom $\mathbf{x}^* = 0$. Then (DCE) implies:

$$\begin{aligned} y_i(j \cup \mathbf{q}) - y_i(j' \cup \mathbf{q}) &= \Big(h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) + \epsilon_i\Big) - \Big(h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) + \epsilon_i\Big) \\ &= h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) - h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) \qquad (148) \end{aligned}$$

and (PSE) implies:

$$\begin{aligned} y_i(j \cup \mathbf{q}) - y_i(j' \cup \mathbf{q}) &= y_i(j \cup \mathbf{q}_0) - y_i(j' \cup \mathbf{q}_0) && \text{(by (PSE))} \\ &= h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}_0}) - h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}_0}) && \text{(by (148))} \\ &= h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{0,0,\ldots,0\}) - h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{0,0,\ldots,0\}) && (149) \end{aligned}$$

Let:

$$\begin{aligned} h_1(\mathbf{x}_i^*) &\equiv h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \qquad (150) \\ h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) &\equiv h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_j^*, \ldots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \end{aligned}$$

Then:

$$\begin{aligned} y_i(\mathbf{p}) &= h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \ldots, \mathbf{x}_{\mathbf{p}((n-1))}^*\right\}\right) + \epsilon_i \qquad \text{(by DCE)} \\ &= \begin{pmatrix} h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \\ +h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, 0, \ldots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \\ +h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \mathbf{x}_{\mathbf{p}(2)}^*, 0, \ldots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, 0, \ldots, 0\right\}\right) \\ + \quad \vdots \\ +h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \ldots, \mathbf{x}_{\mathbf{p}((n-1))}^*\right\}\right) - h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \ldots, \mathbf{x}_{\mathbf{p}((n-1)-1)}^*, 0\right\}\right) \end{pmatrix} + \epsilon_i \\ &= \begin{pmatrix} h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \\ +h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, 0, \ldots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \\ +h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(2)}^*, 0, \ldots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \\ + \quad \vdots \\ +h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}((n-1))}^*, 0, \ldots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) \end{pmatrix} + \epsilon_i \quad \text{(by (148) and (149))} \\ &= h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) + \sum_{j\in\mathbf{p}} h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_j^*, 0, \ldots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0,\ldots,0\}\right) + \epsilon_i \\ &= h_1(\mathbf{x}_i^*) + \sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i \end{aligned}$$

which is the first result in equation (70). If $\mathbf{x}^*$ is categorical, then these functions can be written in the form:

$$\begin{aligned} h_1(\mathbf{x}_i^*) &= \psi_0 + \mathbf{x}_i^*\psi_1 \\ h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) &= \phi_0 + \mathbf{x}_i^*\phi_1 + \mathbf{x}_j^*\phi_2 + \mathbf{x}_i^*\phi_3(\mathbf{x}_j^*)' \end{aligned} \tag{151}$$

Note that $h_2(\mathbf{x}_i^*, 0) = 0$ which implies $\phi_0 = \phi_1 = 0$ and:

$$\begin{aligned} y_i(\mathbf{p}) &= h_1(\mathbf{x}_i^*) + \sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i && \text{(by (70))} \\ &= \psi_0 + \mathbf{x}_i^*\psi_1 + \sum_{j\in\mathbf{p}} \left(\mathbf{x}_j^*\phi_2 + \mathbf{x}_i^*\phi_3(\mathbf{x}_j^*)'\right) + \epsilon_i && \text{(by (151))} \\ &= \psi_0 + \mathbf{x}_i^*\psi_1 + \left(\sum_{j\in\mathbf{p}} \mathbf{x}_j^*\right)\phi_2 + \mathbf{x}_i^*\phi_3\left(\sum_{j\in\mathbf{p}} \mathbf{x}_j^*\right)' + \epsilon_i \\ &= \psi_0 + \mathbf{x}_i^*\psi_1 + \bar{\mathbf{x}}_i^*(\mathbf{p})\underbrace{\phi_2(n-1)}_{\psi_2} + \mathbf{x}_i^*\underbrace{\phi_3(n-1)}_{\psi_2}\bar{\mathbf{x}}_i^*(\mathbf{p})' + \epsilon_i \end{aligned}$$

which is the last result in equation (70).

2. First note that:

$$\begin{aligned} y_i(\mathbf{p}) - y_i(\mathbf{p}') &= \left(h_1(\mathbf{x}_i^*) + \sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i\right) - \left(h_1(\mathbf{x}_i^*) + \sum_{j\in\mathbf{p}'} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i\right) && \text{(by (70))} \\ &= \left(\sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*)\right) - \left(\sum_{j\in\mathbf{p}'} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*)\right) && (152) \end{aligned}$$

Choose any individual i' such that $\mathbf{x}_{i'}^* = 0$, and let $h_3(\mathbf{x}_j^*) \equiv h_2(0, \mathbf{x}_j^*)$. Then:

$$\begin{aligned} y_i(\mathbf{p}) - y_i(\mathbf{p}') &= y_{i'}(\mathbf{p}) - y_{i'}(\mathbf{p}) && \text{(by (OSE))} \\ &= \left(\sum_{j\in\mathbf{p}} h_2(0, \mathbf{x}_j^*)\right) - \left(\sum_{j\in\mathbf{p}'} h_2(0, \mathbf{x}_j^*)\right) && \text{(by (152))} \\ &= \left(\sum_{j\in\mathbf{p}} h_3(\mathbf{x}_j^*)\right) - \left(\sum_{j\in\mathbf{p}'} h_3(\mathbf{x}_j^*)\right) \end{aligned}$$

The results in equation (71) follow by substitution into the results in equation (70) .

Proof for Proposition 10

1. Since $\mathbf{x}, \mathbf{x}^1, \ldots, \mathbf{x}^{n-1}$ are categorical, $\bar{\mathbf{x}}$ fully describes $\left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}$ and:

$$E\left(h(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\mathbf{p}})\middle|\, \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}) = \bar{\mathbf{x}}\right) = h(\mathbf{x}, \left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}) \tag{153}$$

for all $\mathbf{p}$. Let $\tilde{\mathbf{p}}$ be a purely random draw of $(n-1)$ peers from $\mathcal{N} \setminus \{i\}$. Then:

$$(\mathbf{x}_i^*, \epsilon_i) \perp\!\!\!\perp \left\{\mathbf{x}_j^*\right\}_{j \in \tilde{\mathbf{p}}} \tag{154}$$

which implies:

$$\begin{aligned} E(\epsilon_i | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) &= E(\epsilon_i | \mathbf{x}_i = \mathbf{x}) && \text{(by (154))} \\ &= 0 && (155) \end{aligned}$$

By (CRA), Lemma 1 holds and therefore:

$$\begin{aligned} E(y_i | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}}) | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) && \text{(by (43) in Lemma 1)} \\ &= E\left(h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_j^*\right\}_{j \in \tilde{\mathbf{p}}}\right) + \epsilon_i \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) && \text{(by DCE)} \\ &= E\left(h\left(\mathbf{x}_i, \left\{\mathbf{x}_j\right\}_{j \in \tilde{\mathbf{p}}}\right) + \epsilon_i \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) && \text{(by NOV)} \\ &= E\left(h\left(\mathbf{x}_i, \left\{\mathbf{x}_j\right\}_{j \in \tilde{\mathbf{p}}}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) \\ &\quad + E\left(\epsilon_i \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) \\ &= E\left(h\left(\mathbf{x}_i, \left\{\mathbf{x}_j\right\}_{j \in \tilde{\mathbf{p}}}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) && \text{(by (155))} \\ &= h\left(\mathbf{x}, \left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}\right) && \text{(by (153))} \end{aligned}$$

Since the left side of this equation is identified for all $(\mathbf{x}, \bar{\mathbf{x}})$ on the support of $(\mathbf{x}_i, \bar{\mathbf{x}}_i)$, so is the right side.

2. (PSE,DCE) implies that the first result in Proposition 9 applies:

$$\begin{aligned} p_{ij} &= h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) && \text{(by Proposition 9)} \\ &= h_2(\mathbf{x}_i, \mathbf{x}_j) && \text{(by (NOV))} \end{aligned}$$

It follows by substitution that:

$$\begin{aligned} E(p_{ij} | \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) &= E(h_2(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) && \text{(result above)} \\ &= h_2(\mathbf{e}_s, \mathbf{e}_k) && \text{(conditioning rule)} \\ E\left(p_{ij} \middle| \begin{array}{l} \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \\ \ell_i = \ell, \ell_j = \ell' \end{array} \right) &= E\left(h_2(\mathbf{x}_i, \mathbf{x}_j) \middle| \begin{array}{l} \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k, \\ \ell_i = \ell, \ell_j = \ell' \end{array} \right) && \text{(result above)} \\ &= h_2(\mathbf{e}_s, \mathbf{e}_k) && \text{(conditioning rule)} \\ &= E(p_{ij} | \mathbf{x}_i = \mathbf{e}_s, \mathbf{x}_j = \mathbf{e}_k) && (156) \end{aligned}$$

which is condition (LI). Therefore the second result in Proposition 8 applies.

Proof for Equation (25)

$$CGE_{s,m} = E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \qquad \text{(by (18))}$$

$$= \sum_{r=0}^{M^{sat}} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r) \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m)$$
$$- \sum_{r=0}^{M^{sat}} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r) \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0)$$

(law of total probability)

$$= \sum_{r=0}^{M^{sat}} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r) \begin{pmatrix} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix}.$$

(independence of $\mathbf{x}_i$ and $(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}), \mathbf{z}_i(\tilde{\mathbf{p}}))$)

$$= \sum_{r=0}^{M^{sat}} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r) \begin{pmatrix} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix}$$
$$- \sum_{r=0}^{M^{sat}} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_0) \begin{pmatrix} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix}$$
$$+ \sum_{r=0}^{M^{sat}} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_0) \begin{pmatrix} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix}$$

$$= \sum_{r=0}^{M^{sat}} \overbrace{\begin{pmatrix} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r) \\ - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix}}^{= 0 \text{ for } r=0,\ CGE_{s,r} \text{ for } r>0} \begin{pmatrix} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix}$$
$$+ E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_s, \mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_0) \underbrace{\begin{pmatrix} \sum_{r=0}^{M^{sat}} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \sum_{r=0}^{M^{sat}} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix}}_{=(1=1)=0}$$

$$= \sum_{r=1}^{M^{sat}} CGE_{s,r}^{sat} \begin{pmatrix} \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_m) \\ - \Pr(\mathbf{z}_i^{sat}(\tilde{\mathbf{p}}) = \mathbf{e}_r|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \end{pmatrix} \qquad \text{(see (25))}$$