Peers as treatments*

Brian Krauth
Simon Fraser University

May 28, 2024

Abstract

This paper develops a general framework for interpreting linear regression estimates of contextual peer effects under random or conditionally random peer assignment. Rather than imposing the strong assumption that peers influence individual outcomes solely and directly through specific observed characteristics, the model considers social interaction with a given peer group as a treatment with an unknown and variable treatment effect. In this setting, a wide variety of conventional peer effect regressions are informative and can be interpreted as identifying treatment effect heterogeneity along dimensions of interest to the researcher. The framework is then used to clarify the limitations of common research designs and to suggest avenues for improving 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

*Email: `bkrauth@sfu.ca`. Revisions 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.

(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 often provides little guidance on which peer characteristics to include in an empirical model, so some researchers include whatever potentially relevant peer variables are available while other researchers 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) 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 exactly the same effect on a given individual's outcome. 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 counterfactual outcomes depend 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 a natural alternative formulation in which each person has an unobserved and individual-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. A 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. Instead, estimated contextual effects can be understood as describing treatment effect heterogeneity along researcher-selected dimensions. The framework can be used to define causal peer effects in terms of explicit counterfactuals, to state conditions under which they are identified, and to provide simple regression-based estimators.

The implications of this model support and clarify many common empirical practices. The causal 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 important limitations) random cohorts or subgroups within endogenously-selected larger groups.

A straightforward and testable (if strong) separability assumption supports use of the simple linear-in-means model and allows the researcher to interpret its coefficients as predicting the effect of replacing a randomly-selected peer of one observed type with a randomly-selected peer of another observed type. Non-separability can be accommodated in using a binning approach in which coefficients from a parsimonious binned specification can be interpreted as weighted averages of richer specifications with weights that can be calculated.

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 imposes subtle but important constraints on the set of counterfactual group allocations whose impacts 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.

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 non-random 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 specify a model with only contextual effects and interpret it as the reduced form of a more general structural model that may include endogenous effects.

The empirical literature on social effects 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, effort, gender, race, ethnicity, personality, mental health, disruptive behavior, special needs, native 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. While a detailed survey of findings on classroom peer effects is beyond the scope of the current paper, 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' own achievement is similar to that of girls. In addition, using the results to predict the outcome of a proposed reallocation of students is complicated by the fact that 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 empirical 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) and the subsequent literature on treatment effects with spillovers, the relevant peer characteristics are directly 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 assignment of peer groups.

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 individuals. 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 three 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 orthogonal to 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 used here helps to separate practical issues of specification choice from core assumptions about causal mechanisms. For example, the formulation here allows us to interpret the results of two different regression specifications with different explanatory variables and functional forms in the context of the same underlying causal model.

A more general difference is this paper's emphasis on clarifying the interpretation of common empirical practices rather than proposing novel estimation methods.

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. 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 marginal effect of smaller and more feasible reallocations. Carrell et al. (2013) report the results of a field experiment that uses peer effect estimates from one cohort of students to construct presumably optimal allocations for a later cohort. Notably, this allocation yielded surprisingly poor results, providing a cautionary tale on the risks of optimizing from a potentially misspecified model.

2 Model

This section develops the basic model. The model's exposition will refer to a running example application on the effect of classroom gender[2] composition on 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 classmates, even in settings 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 $i \in \mathcal{I} \equiv \{1, 2, \ldots, I\}$. Each individual is fully characterized by an **unobserved 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 vectors $\mathbf{T} \in \mathcal{T}^I$ and $\mathbf{G} \in \mathcal{G}^I$, in the sense that all variables in the model are functions of $(\mathbf{T}, \mathbf{G})$.

The unobserved type 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 large so that individuals are unique or nearly unique. The ordering of the type space is arbitrary; nearby types are not necessarily more similar, and types cannot necessarily be ranked on some some simple "quality" index.

Group membership is determined by some **group selection mechanism** that can be described by 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 $\mathbf{G_A} \in \mathcal{G}^I$ and types $\mathbf{T_A} \in \mathcal{T}^I$.

Each individual experiences a scalar **outcome** of interest $y_i \in \mathbb{R}$ that depends on

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

both the individual's own type and that of other group members:

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

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. The model takes the outcome to be a deterministic function of types and the peer group assignment, but random post-assignment shocks could be accommodated by redefining $\mathbf{Y}(\mathbf{T}, \mathbf{G})$ as a conditional mean.

For each individual i, we can observe the peer group g_i, the outcome y_i and a vector of **observed background characteristics** $\mathbf{x}_i \equiv (x_{i1}, \ldots, x_{iK}) \in \mathbb{R}^K$. These characteristics are predetermined[3] and depend only on one's own type:

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

Note that the observed characteristics are not assumed to directly affect the outcome. However, they are incorporated into the unobserved type in the sense that $\mathbf{x}_i$ is a function of τ_i. As a result, the model is consistent with a direct causal connection from $\mathbf{X}$ to $\mathbf{Y}$, a non-causal correlation in which $\mathbf{X}$ serves as a proxy variable for other predetermined characteristics, or any combination of the two. This is a key feature of this model: the characteristics in $\mathbf{x}_i$ are not assumed to be part of some "true" causal model, but rather have been chosen by the researcher based on data availability and the research question. Another researcher might choose different characteristics in the same data set, and both choices could lead to interesting and valid causal findings.

Given observed characteristics $\mathbf{X}$ and peer group assignments $\mathbf{G}$, **peer average characteristics** for individual i can be defined as $\bar{\mathbf{x}}_i \equiv (\bar{x}_{i1}, \ldots, \bar{x}_{iK}) \in \mathbb{R}^K$ where:

$$\bar{\mathbf{x}}_i \equiv \bar{\mathbf{x}}_i(\mathbf{X}, \mathbf{G}) \equiv \frac{1}{n_{g_i} - 1} \sum_{j \neq i : g_j = g_i} \mathbf{x}_j \tag{4}$$

where n_g is the number of members of group g.

Example 1 (Classmate gender effects)**.** *A researcher has data on I students allocated*

[3]That is, they are not treatments that can be manipulated by a policy maker, as in Manski (2013) and the subsequent literature on treatment effects with spillovers.

across G classrooms and aims to measure the effect of classmate gender on test scores. In this setting, the observed variables would be:

$$y_i \equiv \textit{student } i\textit{'s test score}$$
$$g_i \equiv \textit{classroom ID for student } i$$
$$\mathbf{x}_i \equiv \textit{selected characteristics of student } i\textit{, including gender}$$
$$\bar{\mathbf{x}}_i \equiv \textit{average characteristics of student } i\textit{'s classmates}$$

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 individual-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:*

$$\Pr(\mathbf{T} = \mathbf{T_A}) = \prod_{i=1}^{I} f_\tau(\tau_i(\mathbf{T_A})) \tag{5}$$

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

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

Assumption 2 (Constant group size)**.** *Each peer group in $\mathbf{G}$ has exactly n members. That is, $\Pr(\mathbf{G} \in \mathcal{G}_n^I) = 1$ where:*

$$\mathcal{G}_n^I \equiv \left\{ (g_1, \ldots, g_I) \in \mathcal{G}^I : \sum_{i=1}^{I} \mathbb{I}\left(g_i = g\right) = n, \forall g \in \{1, \ldots, G\} \right\} \tag{6}$$

is the set of all peer group allocations that have exactly n members per group, and $n = I/G$ is an integer.

Assumption 2 is a standard assumption that simplifies exposition. Variable group size is typically addressed in applied work by imposing parametric restrictions, but can be accommodated nonparametrically in this setting 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:*

$$y_i(\mathbf{T}, \mathbf{G}) = y\left(\tau_i, \{\tau_j\}_{g_i = g_j}\right) \tag{7}$$

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

Assumption 3 implies anonymous/exchangeable spillovers within peer groups, no spillovers across peer groups, and no direct effects of group assignment itself. Direct effects of group assignment or a richer structure of spillovers can in principle be accommodated in this model, but would require additional application-specific structure and identifying assumptions that are beyond the scope of this paper.

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

$$\mathbf{S}_{\mathbf{x}} \equiv \{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_K\} \tag{8}$$

where $\mathbf{e}_k \in \{0, 1\}^K$ is the unit vector of length $K \geq 1$ containing one in column k and zero elsewhere; and its probability distribution[4] *is fully described by:*

$$\mu_k \equiv \Pr(\mathbf{x}_i = \mathbf{e}_k) \qquad \text{(for all } k \in 0, 1, \ldots, K)$$
$$\boldsymbol{\mu} \equiv E(\mathbf{x}_i) = \left[\begin{array}{ccc} \mu_1 & \cdots & \mu_K \end{array}\right] \tag{9}$$

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 4 also implies that the K-vector of peer average characteristics $\bar{\mathbf{x}}_i$ fully describes the distribution of observed characteristics within individual i's peer group.

Assumption 5 (Rank condition)**.** *Let $\mathbf{d}_i = vec(1, \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i)$. Then $E(\mathbf{d}_i'\mathbf{d}_i)$ is full rank.*

[4] Note that $\mu_0 = 1 - \sum_{k=1}^{K} \mu_k$ is not included in the vector $\boldsymbol{\mu}$ but can be expressed as a function of it.

Assumption 5 is the standard rank condition needed for identification of relevant regression (best linear predictor) coefficients from the joint distribution of observables $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$.

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

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

This choice satisfies Assumption 4 since $\mathbf{x}_i \in \{0, 1\} = \{\mathbf{e}_0, \mathbf{e}_1\}$.

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 Additional conditions

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

The first set of conditions constrains the group selection mechanism. As usual, some source of purely random variation in exposure to treatment is needed to identify causal effects. In this setting, causal inference on the effect of peer groups will require some form of random peer group selection. Simple random assignment is the most straightforward scenario, but the weaker assumption of conditional random assignment is often sufficient to identify causal effects. 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:*

$$\mathbf{G} \perp \mathbf{T} \qquad (\mathbf{RA})$$

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:*

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

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

An important difference between these two forms of random assignment is that simple random assignment does not constrain the researcher's choice of background characteristics to include in $\mathbf{x}_i$. In contrast, conditional random assignment requires $\mathbf{x}_i$ to include all characteristics that affect group assignment.

The second set of conditions constrain the outcome function to be separable. Neither form of separability is necessarily required for identification, but separability simplifies analysis and interpretation, and is shown in Section 4.1 to support commonly-used methods such as the linear-in-means model.

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

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

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

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

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

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

Outcomes that are neither own-separable nor peer-separable will be called **non-separable**. Note that both forms of separability are constraints on how the unobserved types enter into the outcome, and do not depend on the specific characteristics in $\mathbf{x}_i$.

3 Defining social effects

Causal social effects can be defined in this setting by stating an explicit potential outcome function and set of counterfactuals. Individual characteristics are predetermined (as in Graham et al. (2010)) rather than manipulable (as in Manski (2013)), so the applicable counterfactuals in this model relate to the peer group assignment, and not to the characteristics of any specific individual.

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

$$\mathbf{p}_i \equiv \mathbf{p}(i, \mathbf{G}) \equiv \{j \neq i : g_j = g_i\} \tag{10}$$

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

$$y_i(\mathbf{p}) \equiv y\left(\tau_i, \{\tau_j\}_{j \in \mathbf{p}}\right) \tag{11}$$

where $\mathbf{p}$ *is any size* $n-1$ *subset of* $\mathcal{I} \setminus \{i\}$.

That is, the observed outcome for individual i is $y_i(\mathbf{p}_i)$, and the counterfacutal outcome $y_i(\mathbf{p})$ is the outcome that would have been observed if individual i had instead been assigned the counterfactual peer group $\mathbf{p}$.

Counterfactual peer group assignments can be conceptualized in three distinct ways. We can consider the effect of changing a single peer (**peer effects**), an entire peer group (**group effects**), or the peer group assignment mechanism itself (**reallocation effects**). In addition, these effects can be defined for the average individual in the population (**average effects**), or conditioned on the observed characteristics of the treated individual (**conditional effects**).

This section will develop quantitative definitions of peer effects and group effects in terms of an individual-specific potential outcome function that describes the outcome a given individual would experience if exposed to a given set of peers. Reallocation effects are addressed as an extension in Section 5.

3.1 Peer effects

Peer effects can be defined as the average effect of replacing a single peer of one observed type with a single peer of another observed type.

Definition 6 (Average peer effect)**.** *The* **average peer effect** *(APE_ℓ) of peers of observed type ℓ is defined as:*

$$APE_\ell \equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{p}}) - y_i(\{j'\} \cup \tilde{\mathbf{p}})\middle| \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right) \tag{12}$$

where $\tilde{\mathbf{p}}$ *is a purely random draw of* $n-2$ *peers from* f_τ.

Although equation (12) looks complex, the concept is simple. Take a randomly-selected individual (i) with a randomly-constructed peer group ($\tilde{\mathbf{p}}$), and replace a randomly-selected peer (j') of the base observed type ($\mathbf{x}_{j'} = \mathbf{e}_0$) with a randomly-selected peer (j) of observed type ℓ ($\mathbf{x}_j = \mathbf{e}_\ell$). The average peer effect is the predicted change in this individual's outcome.

If we think of interaction with a given peer as a unique "treatment,", average peer effects can be interpreted as describing the heterogeneity of these treatment effects across observed types, and are thus analogous to the conditional average treatment effect estimated in the literature on heterogeneous treatment effects (Abrevaya et al., 2015). One difference from a typical treatment effects setting is that there is no natural "untreated" state, so average peer effects are defined relative to the average peer in some arbitrarily selected base population. Regardless of the base population, average peer effects can be used to make comparisons between any two observed types: the average effect of replacing the average observed type k peer with the average observed type ℓ peer is $APE_\ell - APE_k$.

Rather than averaging across all treated 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** *($CPE_{k\ell}$) of peers of observed type ℓ on treated individuals of observed type k is defined as:*

$$CPE_{k\ell} \equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{p}}) - y_i(\{j'\} \cup \tilde{\mathbf{p}}) \middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right) \tag{13}$$

where $\tilde{\mathbf{p}}$ is a purely random draw of $n-2$ peers from f_τ.

That is, $CPE_{k\ell}$ can be considered the effect on the average observed type k individual of replacing the average base-type peer with the average observed type ℓ peer.

Note that average and conditional peer effects are both well-defined under the model's maintained assumptions and do not require additional conditions like separability or random assignment. However, these additional conditions are important for identification and interpretation.

Example 3 (Peer effects by classmate gender)**.** *Continuing the classmate gender effects example, average and conditional peer effects can be defined as follows:*

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

Note that APE_1 will be a weighted average of CPE_{01} and CPE_{11}.

3.2 Group effects

Peer **group effects** can be defined as the average effect of replacing the entire peer group with another peer group that has a different distribution of observable characteristics. These effects can be distinct from average peer effects when the marginal impact of a given peer depends on the treated individual's other peers. For example, a hyperactive classmate may be more disruptive if there are other hyperactive students in the classroom, or the social dynamics of a classroom may change if girls outnumber boys. These considerations may imply nonlinear relationships between peer group composition and outcomes (Hoxby and Weingarth, 2005).

As discussed in Section 2.2, the maintained assumptions of constant group size (Assumption 2) and discrete observed characteristics (Assumption 4) imply that $\bar{\mathbf{x}}_i$ has a finite support $\mathbf{S}_{\bar{\mathbf{x}}}$ and provides a complete description of the frequency distribution of observed characteristics among i's peers. That is, person i has exactly $(n-1)\bar{x}_{i\ell}$ peers of observed type ℓ. This greatly simplifies the modeling of nonlinearity because *any* (linear or nonlinear) function of observed peer characteristics is equivalent on the support to a *linear* (affine) function of some saturated categorical vector constructed from $\bar{\mathbf{x}}_i$.

Definition 8 (Saturated peer group variable)**.** *Let the* **saturated peer group variable** $\mathbf{z}^S(\bar{\mathbf{x}}_i)$ *be defined by:*

$$\mathbf{z}^S(\bar{\mathbf{x}}_i) = \sum_{s=1}^{S} \mathbf{e}_s \mathbb{I}\left(m(\bar{\mathbf{x}}_i) = s\right) \tag{14}$$

where $S = |\mathbf{S}_{\bar{\mathbf{x}}}| - 1$*,* $m : \mathbf{S}_{\bar{\mathbf{x}}} \to \{0, 1, \ldots, S\}$ *is an arbitrary strict ordering on* $\mathbf{S}_{\bar{\mathbf{x}}}$*, and* $\mathbf{e}_s$ *is the unit vector of length* S *containing one in column* s *and zero elsewhere.*

Saturated peer group variables allow for a very general model, but the support of $\bar{\mathbf{x}}_i$ is often too large for $\mathbf{z}^S(\bar{\mathbf{x}}_i)$ to be a practical explanatory variable. In that case, the researcher may prefer a more parsimonious regression model based on binning $\bar{\mathbf{x}}_i$.

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

$$\mathbf{z}_i = \mathbf{z}(\bar{\mathbf{x}}_i) = \sum_{b=1}^{B} \mathbf{e}_b \mathbb{I}\left(\bar{\mathbf{x}}_i \in \mathbf{S}^b_{\bar{\mathbf{x}}}\right) \tag{15}$$

where $(\mathbf{S}^0_{\bar{\mathbf{x}}}, \mathbf{S}^1_{\bar{\mathbf{x}}}, \ldots, \mathbf{S}^B_{\bar{\mathbf{x}}})$ *is a partition of* $\mathbf{S}_{\bar{\mathbf{x}}}$ *(the support of* $\bar{\mathbf{x}}_i$*), and* $\mathbf{e}_b$ *is the unit vector of length* B *containing one in column* b *and zero elsewhere. Bin* b *is a* **singleton** *if*

$|\mathbf{S}^b_{\bar{\mathbf{x}}}| = 1$ *and* **pooled** *if* $|\mathbf{S}^b_{\bar{\mathbf{x}}}| > 1$.

The binned group variable $\mathbf{z}_i$ is defined by the researcher, and can include any mix of singleton and pooled bins. Note that the saturated group variable is a special case of a binned group variable whose bins are all singletons, and that any binned variable can be written as a linear function of the saturated variable.

Example 4 (Binned group variables for classmate gender)**.** *Continuing the classmate gender 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 binned group variables from* $\bar{\mathbf{x}}_i$ *including:*

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

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

- *Alll-female, all-male, and mixed* ($B = 2$)*:*

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

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

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

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

Definition 10 (Group effects)**.** *The* **average group effect** *of a bin b peer group is defined as:*

$$AGE_b \equiv E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \tag{16}$$

and the **conditional group effect** *of a bin b peer group on treated individuals of observed type k is defined as:*

$$CGE_{kb} \equiv E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \tag{17}$$

where $\tilde{\mathbf{p}}$ *is a purely random draw*[5] *of* $n-1$ *peers from* f_τ*.*

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-b 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 5 (Group effects for classmate gender)**.** *Continuing the classmate gender 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 that peer group size is odd and all peer groups are either majority-male or majority-female. Then the following average and conditional group effects can be defined:*

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

The "average majority-male peer group" here is based on the distribution of group composition that would be observed under random assignment, which places high weight on groups where boys are a slight majority and low but nonzero weight on groups that are all or almost all boys.

[5] Note that AGE_b and $CGE_{k,b}$ 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 7 in Section 5.2 shows that AGE_b and $CGE_{k,b}$ are only informative about peer group reallocations that preserve this conditional distribution (e.g., if $\mathbf{S}^b_{\bar{\mathbf{x}}}$ is a singleton). See Section 5.2 for additional details.

4 Main 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 interpreted 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

Proposition 1 below shows that simple causal effects can typically be interpreted as a weighted average of more complex effects, with weights that depend on the probability distribution of $\mathbf{x}_i$.

Proposition 1 (Aggregation). *Given Assumptions 1-5:*

1. *Conditional effects can be aggregated to yield average effects:*

$$APE_\ell = \sum_{k=0}^{K} \mu_k CPE_{k\ell} \tag{18}$$

$$AGE_b = \sum_{k=0}^{K} \mu_k CGE_{kb} \tag{19}$$

where $\mu_k = E(x_{ik}) = \Pr(\mathbf{x}_i = \mathbf{e}_k)$ as defined earlier.

2. *Saturated group effects can be aggregated to yield group effects for any other binning scheme:*

$$AGE_b = \sum_{k=0}^{K} \sum_{s=1}^{S} \mu_k w_{sb}(\boldsymbol{\mu}) CGE_{ks}^S \tag{20}$$

$$CGE_{kb} = \sum_{s=1}^{S} w_{sb}(\boldsymbol{\mu}) CGE_{ks}^S \tag{21}$$

where CGE_{ks}^S is the conditional group effect for bin s of the saturated variable

$\mathbf{z}^S(\bar{\mathbf{x}}_i)$, $w_{sb}(\boldsymbol{\mu})$ is a weighting function given by:

$$w_{sb}(\boldsymbol{\mu}) = \frac{\sum_{\bar{\mathbf{x}} \in \mathbf{S}^b_{\bar{\mathbf{x}}} : \mathbf{z}^S(\bar{\mathbf{x}}) = \mathbf{e}_s} \mathcal{M}(\bar{\mathbf{x}}, n, \boldsymbol{\mu})}{\sum_{\bar{\mathbf{x}} \in \mathbf{S}^b_{\bar{\mathbf{x}}}} \mathcal{M}(\bar{\mathbf{x}}, n, \boldsymbol{\mu})} - \frac{\sum_{\bar{\mathbf{x}} \in \mathbf{S}^0_{\bar{\mathbf{x}}} : \mathbf{z}^S(\bar{\mathbf{x}}) = \mathbf{e}_s} \mathcal{M}(\bar{\mathbf{x}}, n, \boldsymbol{\mu})}{\sum_{\bar{\mathbf{x}} \in \mathbf{S}^0_{\bar{\mathbf{x}}}} \mathcal{M}(\bar{\mathbf{x}}, n, \boldsymbol{\mu})} \tag{22}$$

and:

$$\mathcal{M}(\bar{\mathbf{x}}, n, \boldsymbol{\mu}) = \frac{(n-1)!}{\prod_{k=0}^{K}((n-1)\bar{x}_{\cdot k})!} \prod_{k=0}^{K} \mu_k^{(n-1)\bar{x}_{\cdot k}} \tag{23}$$

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

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

Proposition 2 below 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 individual-specific or pair-specific latent variables. Average and conditional peer effects can also be expressed in terms of conditional expectations of these latent variables, and Proposition 4 in the next section establishes conditions under which these conditional expectations are identified.

Proposition 2 (Separability)**.** *Given Assumptions 1-5:*

1. *If outcomes are peer-separable (PS), then each individual's potential outcome function can be expressed in the form:*

$$y_i(\mathbf{p}) = \sum_{j \in \mathbf{p}} PE_{ij} \tag{24}$$

where $PE_{ij} = PE(\tau_i, \tau_j)$ and:

$$CPE_{k\ell} = E(PE_{ij} | \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell) - E(PE_{ij} | \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0) \tag{25}$$

$$APE_\ell = E(PE_{ij} | \mathbf{x}_j = \mathbf{e}_\ell) - E(PE_{ij} | \mathbf{x}_j = \mathbf{e}_0) \tag{26}$$

for all observable categories (k, ℓ).

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

$$y_i(\mathbf{p}) = OE_i + \sum_{j \in \mathbf{p}} PE_j \tag{27}$$

where $OE_i = OE(\tau_i)$, $PE_j = PE(\tau_j)$ *and:*

$$CPE_{k\ell} = APE_\ell = E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_j|\mathbf{x}_j = \mathbf{e}_0) \tag{28}$$

for all observable categories (k, ℓ).

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)**.** *Given Assumptions 1-5:*

1. *If peers are randomly assigned conditional on observables (CRA) and outcomes are peer-separable (PS), then:*

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i) = L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \tag{29}$$

   where $L(\cdot|\cdot)$ *is the best linear predictor.*

2. *If peers are randomly assigned conditional on observables (CRA) and outcomes are peer-separable and own separable (PS, OS), then:*

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) = L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \tag{30}$$

The restrictions in equations (29) and (30) imply that certain coefficients in a linear regression are zero, which can be easily tested using standard methods. Note that separability is a property of the outcome function $y(\cdot, \cdot)$ and not of 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 any test based on these implications depends on the specific variables chosen by the researcher.

4.2 Identification

The results in this section show conditions under which the social effects defined in Section 3 are identified from the joint distribution of observable variables $(\mathbf{Y}, \mathbf{X}, \mathbf{G})$. The identification results are constructive and suggest simple estimators whose implementation is described in Section 4.3.

Proposition 4 below shows identification under a simple random assignment research design. Proposition 5 later in this section shows identification under the weaker assumption of conditional random assignment, and Proposition 9 in Section 6 shows identification under a more complex nested assignment design.

Proposition 4 (Identification with random assignment)**.** *Given Assumptions 1-5:*

1. *If peers are randomly assigned (RA) and outcomes are peer-separable (PS), then average and conditional peer effects are identified from the joint distribution of* $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$*:*

$$APE_\ell = \frac{\alpha_{2\ell}}{n-1} \tag{31}$$

$$CPE_{k\ell} = \frac{\beta_{2\ell} + \beta_{3k\ell}}{n-1} \tag{32}$$

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

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \equiv \alpha_0 + \mathbf{x}_i\boldsymbol{\alpha}_1 + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2 \tag{33}$$

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \equiv \beta_0 + \mathbf{x}_i\boldsymbol{\beta}_1 + \bar{\mathbf{x}}_i\boldsymbol{\beta}_2 + \mathbf{x}_i\boldsymbol{\beta}_3\bar{\mathbf{x}}_i' \tag{34}$$

i.e., $\alpha_{2\ell}$ *is element* ℓ *of* $\boldsymbol{\alpha}_2$*,* $\beta_{2\ell}$ *is element* ℓ *of* $\boldsymbol{\beta}_2$*,* $\beta_{3k\ell}$ *is the element in row* k *and column* ℓ *of* $\boldsymbol{\beta}_3$ *for all* $k > 0$*, and* $\beta_{30\ell} \equiv 0$ *for all* ℓ*.*

2. *If peers are randomly assigned (RA), then average and conditional group effects are identified from the joint distribution of* $(y_i, \mathbf{x}_i, \mathbf{z}_i)$*:*

$$AGE_b = \gamma_{2b} \tag{35}$$

$$CGE_{kb} = \delta_{2b} + \delta_{3kb} \tag{36}$$

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

$$L(y_i|\mathbf{x}_i, \mathbf{z}_i) \equiv \gamma_0 + \mathbf{x}_i\boldsymbol{\gamma}_1 + \mathbf{z}_i\boldsymbol{\gamma}_2 \tag{37}$$

$$L(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i) \equiv \delta_0 + \mathbf{x}_i\boldsymbol{\delta}_1 + \mathbf{z}_i\boldsymbol{\delta}_2 + \mathbf{x}_i\boldsymbol{\delta}_3\mathbf{z}_i' \tag{38}$$

i.e., γ_{2b} *is element* b *of* $\boldsymbol{\gamma}_2$*,* δ_{2b} *is element* b *of* $\boldsymbol{\delta}_2$*,* δ_{3kb} *is the element in row* k *and column* b *of* $\boldsymbol{\delta}_3$ *for all* $k > 0$*, and* $\delta_{30b} \equiv 0$ *for all* b*.*

Proposition 4 shows conditions under which each causal social effect defined in Section 3 can be expressed in terms of a linear regression model whose coefficients can be identified from the joint distribution[6] of observables.

Example 6 (Identifying peer effects for classmate gender)**.** *Continuing the classmate gender effects example, suppose two researchers have access to a data set in which students are randomly assigned to classrooms:*

[6] Note that $(y_i, \mathbf{x}_i, \mathbf{z}_i)$ can be derived from $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$, which can be derived from $(\mathbf{Y}, \mathbf{X}, \mathbf{G})$.

- *Researcher A estimates the effect of male classmates on test scores using the conventional linear-in-means model (33).*
- *Researcher B estimates the heterogeneous linear-in-means model (34) using the same data as Researcher A.*

Under the assumption of peer separability, Part 1 of Proposition 4 allows Researcher A to interpret his coefficient on $\bar{\mathbf{x}}_i$ *as the effect on the average student of replacing the average female classmate with the average male classmate, and allows Researcher B to interpret her 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. Adding the coefficient on the interaction term* $\mathbf{x}_i'\bar{\mathbf{x}}_i$ *gives Researcher B the effect on the average male 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 some 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.*
- *Researcher B's finding of heterogeneity (i.e. a nonzero coefficient on the interaction term) does not invalidate Researcher A's analysis based on equation (33) that ignores heterogeneity.*

In other words, a wide range of empirical specifications can be estimated and each provides valid and potentially useful results.

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 (Identifying group effects for classmate gender)**.** *Continuing the classmate gender effects example, suppose that Researcher C and Researcher D have the same data as Researcher A and Researcher B, but:*

- *Researcher C divides peer groups into majority-female and majority-male; i.e.,* $\mathbf{z}_i = \mathbb{I}(\bar{\mathbf{x}}_i > 0.5)$.
- *Researcher D divides peer groups into all-male, all-female, and mixed.*
- *Each researcher estimates a regression of* y_i *on* $(\mathbf{x}_i, \mathbf{z}_i)$ *for their chosen* $\mathbf{z}_i$.

Then Part 2 of Proposition 4 says that Researcher C's 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, and that Researcher D's coefficients can be similarly interpreted with respect to her chosen* $\mathbf{z}_i$*. Note that:*

- *The results in Proposition 4 apply regardless of the researcher's choice of* $\mathbf{z}_i$*.*
- *Researcher C or Researcher D's results could be used to test peer separability, and may thus invalidate Researcher A and Researcher B's results.*

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 modifications. 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)**.** *Given Assumptions 1-5, if peers are randomly assigned conditional on observable characteristics (CRA), then:*

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) = E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) \tag{39}$$

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

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

Proposition 5 (Identification with conditional random assignment)**.** *Given Assumptions 1-5:*

1. *If peers are randomly assigned conditional on observable characteristics (CRA) and outcomes are peer separable (PS), then average and conditional peer effects are identified from the joint distribution of* $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$*:*

$$APE_\ell = \sum_{k=0}^{K} \mu_k \frac{\beta_{2\ell} + \beta_{3k\ell}}{n-1} \tag{40}$$

$$CPE_{k\ell} = \frac{\beta_{2\ell} + \beta_{3k\ell}}{n-1} \tag{41}$$

   where $(\beta_{2\ell}, \beta_{3k\ell})$ *are defined as in equation (34).*

2. *If peers are randomly assigned conditional on observable characteristics (CRA), then average and conditional group effects are identified from the joint distribution*

of $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$*:*

$$AGE_b = \sum_{k=0}^{K} \sum_{s=1}^{S} \mu_k w_{sb}(\boldsymbol{\mu})(\lambda_{2s} + \lambda_{3ks}) \tag{42}$$

$$CGE_{kb} = \sum_{s=1}^{S} w_{sb}(\boldsymbol{\mu})(\lambda_{2s} + \lambda_{3ks}) \tag{43}$$

where $w_{sb}(\cdot)$ *is defined in Proposition 1 and* $\boldsymbol{\lambda} = (\lambda_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3)$ *are the coefficients from the best linear predictor:*

$$L(y_i|\mathbf{x}_i, \mathbf{z}^S(\bar{\mathbf{x}}_i), \mathbf{x}_i'\mathbf{z}^S(\bar{\mathbf{x}}_i)) \equiv \lambda_0 + \mathbf{x}_i\boldsymbol{\lambda}_1 + \mathbf{z}^S(\bar{\mathbf{x}}_i)\boldsymbol{\lambda}_2 + \mathbf{x}_i\boldsymbol{\lambda}_3\mathbf{z}^S(\bar{\mathbf{x}}_i)' \tag{44}$$

i.e., λ_{2s} *is element* s *of* $\boldsymbol{\lambda}_2$*,* λ_{3ks} *is the element in row* k *and column* s *of* $\boldsymbol{\lambda}_3$ *for all* $k > 0$*, and* $\lambda_{30s} \equiv 0$ *for all* s*.*

While Proposition 5 applies more generally than Proposition 4, this generality comes at the cost that some estimands are weighted averages of regression coefficients[7] rather than just the coefficients. The reason for this is that both peer effects and group effects are defined in terms of a hypothetical randomly-assigned peer group, so some re-weighting is required when the assignment mechanism in the data deviates from simple random assignment. As in Proposition 1, the weights in Proposition 5 can be recovered from the probability distribution of $\mathbf{x}_i$.

4.3 Estimation and inference

The identification results in Propositions 4 and 5 are constructive and suggest simple plug-in estimators that are easily implemented in standard statistical packages. This section provides an informal discussion of estimation and inference in this setting.

Suppose the researcher has a sample of N observations on $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$ from a large population that satisfies the model assumptions. Sampling models vary in the applied literature, so rather than specifying the details of the sampling scheme we simply assume it satisfies all conditions required for:

$$\sqrt{N}\left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}\right) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{45}$$

where $\boldsymbol{\psi} \equiv (\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\lambda})$ is a vector of previously-defined population means and best linear predictor coefficients, and $\hat{\boldsymbol{\psi}} \equiv (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\lambda}})$ is a consistent and asymptotically

[7] Alternatively, the regression coefficients can be interpreted as weighted averages of causal effects.

normal estimator of $\boldsymbol{\psi}$. In most applications, the researcher will have a cluster sample of size $N = nG$ constructed from data on all n members of G randomly selected groups, $\hat{\boldsymbol{\mu}}$ will be the sample average of $\mathbf{x}_i$, and $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\lambda}})$ will be the OLS regression coefficients. In other applications, the researcher may observe data for a random sample of individuals, each of whom can be linked to some aggregate data source such as census tract characteristics.

If peers are randomly assigned, Proposition 4 shows that peer and group effects correspond to best linear predictor coefficients or linear combinations of those coefficients:

$$\widehat{APE}_\ell = \frac{\hat{\alpha}_{2\ell}}{n-1} \qquad \text{if (PS, RA)} \tag{46}$$

$$\widehat{CPE}_{k\ell} = \frac{\hat{\beta}_{2\ell} + \hat{\beta}_{3k\ell}}{n-1} \qquad \text{if (PS, RA)} \tag{47}$$

$$\widehat{AGE}_b = \hat{\gamma}_{2b} \qquad \text{if (RA)} \tag{48}$$

$$\widehat{CGE}_{kb} = \hat{\delta}_{2b} + \hat{\delta}_{3kb} \qquad \text{if (RA)} \tag{49}$$

If peers are randomly assigned conditional on observables, Proposition 5 shows that peer and group effects can be expressed as linear combinations of best linear predictor coefficients or as weighted averages of those coefficients. As a result, they can be estimated by:

$$\widehat{APE}_\ell = \frac{\hat{\beta}_{2\ell} + \sum_{k=0}^{K} \hat{\mu}_k \hat{\beta}_{3k\ell}}{n-1} \qquad \text{if (PS, CRA)} \tag{50}$$

$$\widehat{CPE}_{k\ell} = \frac{\hat{\beta}_{2\ell} + \hat{\beta}_{3k\ell}}{n-1} \qquad \text{if (PS, CRA)} \tag{51}$$

$$\widehat{AGE}_b = \sum_{k=0}^{K} \sum_{s=0}^{S} \hat{\mu}_k w_{sb}(\hat{\boldsymbol{\mu}})(\hat{\lambda}_{2s} + \hat{\lambda}_{3ks}) \qquad \text{if (CRA)} \tag{52}$$

$$\widehat{CGE}_{kb} = \sum_{s=0}^{S} w_{sb}(\hat{\boldsymbol{\mu}})(\hat{\lambda}_{2s} + \hat{\lambda}_{3ks}) \qquad \text{if (CRA)} \tag{53}$$

Five of these eight estimators are just linear combinations of OLS coefficients, so the researcher can apply standard cluster-robust asymptotic inference procedures to construct standard errors and confidence intervals, or to perform hypothesis tests.

Inference is slightly more complicated for the three estimators that include weights based on $\hat{\boldsymbol{\mu}}$, as their asymptotic variance depends on the *joint* distribution of $\hat{\boldsymbol{\mu}}$ and the regression coefficients. A straightforward general approach is to define $\hat{\boldsymbol{\psi}}$ as the just-identified GMM estimator[8] for the vector of moment conditions:

$$E\left(\begin{bmatrix} \mathbf{x}_i - \boldsymbol{\mu} \\ y_i - \alpha_0 - \mathbf{x}_i\boldsymbol{\alpha}_1 - \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2 \\ \mathbf{x}_i'(y_i - \alpha_0 - \mathbf{x}_i\boldsymbol{\alpha}_1 - \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2) \\ etc. \end{bmatrix}\right) = \mathbf{0} \tag{54}$$

and $\hat{\boldsymbol{\Sigma}}$ as the associated (cluster-robust) GMM variance matrix. Under the usual GMM regularity conditions:

$$\hat{\boldsymbol{\Sigma}} \stackrel{P}{\to} \boldsymbol{\Sigma} \tag{55}$$

The parameter (vector) of interest can then be defined as $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\psi})$ for some differentiable function $\boldsymbol{\theta}(\cdot)$, and its estimator can be defined as $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\hat{\boldsymbol{\psi}})$. Then $\hat{\boldsymbol{\theta}}$ has the asymptotic distribution:

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \stackrel{D}{\to} N(\mathbf{0}, (\nabla\boldsymbol{\theta}(\boldsymbol{\psi}))\boldsymbol{\Sigma}(\nabla\boldsymbol{\theta}(\boldsymbol{\psi}))') \tag{56}$$

where $\nabla\boldsymbol{\theta}(\boldsymbol{\psi})$ is the Jacobian matrix of $\boldsymbol{\theta}(\boldsymbol{\psi})$, and the asymptotic variance can be estimated:

$$(\nabla\boldsymbol{\theta}(\hat{\boldsymbol{\psi}}))\hat{\boldsymbol{\Sigma}}(\nabla\boldsymbol{\theta}(\hat{\boldsymbol{\psi}}))' \stackrel{P}{\to} (\nabla\boldsymbol{\theta}(\boldsymbol{\psi}))\boldsymbol{\Sigma}(\nabla\boldsymbol{\theta}(\boldsymbol{\psi}))' \tag{57}$$

Similarly, a hypothesis of the form $\boldsymbol{\theta}(\boldsymbol{\psi}) = \mathbf{0}$ can be tested using the Wald statistic:

$$H_0 : \boldsymbol{\theta}(\boldsymbol{\psi}) = \mathbf{0} \qquad \Longrightarrow \qquad \boldsymbol{\theta}(\hat{\boldsymbol{\psi}})'\left((\nabla\boldsymbol{\theta}(\hat{\boldsymbol{\psi}}))\hat{\boldsymbol{\Sigma}}(\nabla\boldsymbol{\theta}(\hat{\boldsymbol{\psi}}))'\right)^{-1}\boldsymbol{\theta}(\hat{\boldsymbol{\psi}}) \stackrel{D}{\to} \chi^2(r) \tag{58}$$

where r is the number of restrictions imposed by the null. Each of these steps is standard, and can be implemented by commonly-available software (e.g., the `gmm`, `nlcom`, and `testnl` commands in Stata).

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 predetermined by the researcher. This scenario fits many applications in which the researcher has a specific research question or 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 want 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

[8] Note that the GMM estimator here is identical to the OLS estimator; the purpose of applying GMM here is to estimate the full $\boldsymbol{\Sigma}$ matrix including the asymptotic covariance of $\hat{\boldsymbol{\mu}}$ with the regression coefficients using commonly-available tools.

and Imbens (2016) or Wager and Athey (2018) can be adapted to this setting. Similarly, the additive structure of peer effects under separability can be exploited by a sieve-based method when the individual characteristic of interest is continuous. These extensions are beyond the scope of this paper and left to future research.

5 Extension: Reallocation effects

The peer and group effects defined in Section 3 predict the effect of a change in the composition of a representative individual's peer group. Given a fixed population of individuals, 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, a researcher 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 (Bhattacharya, 2009; Graham et al., 2010). This section defines both the set of reallocations that can be considered as well as the corresponding reallocation effects. It also provides results on identification and estimation of reallocation effects.

5.1 Defining reallocation effects

As with peer and group effects, the first step is to define the relevant counterfactual, which in this case is a reallocation of individuals across groups based on observed characteristics. Such a reallocation must be feasible given the frequency distribution of characteristics.

Definition 11 (Feasible reallocation)**.** *A* **feasible reallocation mechanism** *is a function* $\mathbf{G}_R : \mathbf{S}_{\mathbf{x}}^I \times [0,1] \to \mathcal{G}_n^I$*. A* **feasible reallocation** *is a random vector* $\tilde{\mathbf{G}}_R \equiv \mathbf{G}_R(\mathbf{X}, \epsilon)$ *where* $\mathbf{G}_R$ *is a feasible reallocation mechanism and* ϵ *is a random variable with probability distribution* $\epsilon | \mathbf{T} \sim U(0,1)$*.*

The random variable ϵ serves as a randomization device that can be used to select among observationally-equivalent (same value of $\mathbf{x}_i$) individuals and thus average over the conditional distribution of unobserved heterogeneity. Feasible reallocations satisfy conditional random assignment (CRA) by construction, and the actual allocation can be treated as a feasible reallocation if it satisfies conditional random assignment.

Example 8 (Feasible reallocations by gender)**.** *Continuing the classmate gender effects example, 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 of the following feasible reallocation mechanisms:*

- *Simple random assignment:*
  $\mathbf{G}_R(\mathbf{X}, \epsilon)$ *is a random draw from* $\mathcal{G}_n^I$.
- *All classes single-gender:*
  $\mathbf{G}_R(\mathbf{X}, \epsilon)$ *is a random draw from* $\left\{(g_1, \ldots, g_I) \in \mathcal{G}_n^I : \sum_{i:g_i=g} \mathbf{x}_i \in \{0, n\}\right\}$
- *All classes perfectly mixed:*
  $\mathbf{G}_R(\mathbf{X}, \epsilon)$ *is a random draw from* $\left\{(g_1, \ldots, g_I) \in \mathcal{G}_n^I : \sum_{i:g_i=g} \mathbf{x}_i = n/2\right\}$

Each feasible reallocation implies a probability distribution over outcomes, so the predicted average outcome can be compared across any two feasible reallocations.

Definition 12 (Reallocation effects)**.** *The* **average reallocation effect** *of the feasible reallocation mechanism* $\mathbf{G}_R$ *is defined as:*

$$ARE(\mathbf{G}_R) \equiv E(y_i(\mathbf{p}(i, \mathbf{G}_R(\mathbf{X}, \epsilon))) - y_i(\tilde{\mathbf{p}})) \tag{59}$$

and its **conditional reallocation effect** *on treated individuals of observed type* k *is defined as:*

$$CRE_k(\mathbf{G}_R) \equiv E(y_i(\mathbf{p}(i, \mathbf{G}_R(\mathbf{X}, \epsilon))) - y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k) \tag{60}$$

where $\tilde{\mathbf{p}}$ *is a purely random draw of* $n-1$ *peers from* f_τ.

Reallocation effects compare the predicted outcome under the proposed reallocation mechanism to the predicted outcome under a benchmark allocation of simple random assignment. The choice of benchmark is arbitrary, and predicted outcomes for any two reallocation mechanisms $\mathbf{G}_0$ and $\mathbf{G}_1$ can be compared by calculating $ARE(\mathbf{G}_1) - ARE(\mathbf{G}_0)$.

As with peer and group effects, average and conditional reallocation effects have a straightforward relationship, described in Proposition 6 below.

Proposition 6 (Aggregation for reallocation effects)**.** *Conditional reallocation effects can be aggregated to yield average reallocation effects:*

$$ARE(\mathbf{G}_R) = \sum_{k=0}^{K} \mu_k CRE_k(\mathbf{G}_R) \tag{61}$$

5.2 Identification of reallocation effects

Proposition 7 below describes how the reallocation effects defined in Section 5.1 can be described in terms of the peer and/or group effects defined in Section 3.

Proposition 7 (Identification of reallocation effects)**.** *Let* $\mathbf{G}_R$ *be a feasible reallocation mechanism. Then given Assumptions 1-5:*

1. *If* $(\mathbf{S}^1_{\bar{\mathbf{x}}}, \ldots, \mathbf{S}^B_{\bar{\mathbf{x}}})$ *are singletons, and* $\Pr(\bar{\mathbf{x}}_i(\mathbf{X}, \mathbf{G}_R(\mathbf{X}, \epsilon)) \in \mathbf{S}^0_{\bar{\mathbf{x}}}) = 0$*, then:*

$$ARE(\mathbf{G}_R) = \sum_{k=0}^{K} \sum_{b=1}^{B} \mu_k \Delta z_{kb}(\mathbf{G}_R) CGE_{kb} \tag{62}$$

$$CRE_k(\mathbf{G}_R) = \sum_{b=1}^{B} \Delta z_{kb}(\mathbf{G}_R) CGE_{kb} \tag{63}$$

where:

$$\Delta z_{kb}(\mathbf{G}_R) \equiv \Pr(\bar{\mathbf{x}}_i(\mathbf{X}, \mathbf{G}_R(\mathbf{X}, \epsilon)) \in \mathbf{S}^b_{\bar{\mathbf{x}}} | \mathbf{x}_i = \mathbf{e}_k) - \Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) \in \mathbf{S}^b_{\bar{\mathbf{x}}} | \mathbf{x}_i = \mathbf{e}_k) \tag{64}$$

2. *If outcomes are peer-separable (PS), then:*

$$ARE(\mathbf{G}_R) = (n-1) \sum_{k=0}^{K} \sum_{\ell=1}^{K} \mu_k \Delta \bar{x}_{k\ell}(\mathbf{G}_R) CPE_{k\ell} \tag{65}$$

$$CRE_k(\mathbf{G}_R) = (n-1) \sum_{\ell=1}^{K} \Delta \bar{x}_{k\ell}(\mathbf{G}_R) CPE_{k\ell} \tag{66}$$

where:

$$\Delta \bar{x}_{k\ell}(\mathbf{G}_R) \equiv E(\bar{x}_{i\ell}(\mathbf{X}, \mathbf{G}_R(\mathbf{X}, \epsilon)) | \mathbf{x}_i = \mathbf{e}_k) - \mu_\ell \tag{67}$$

3. *If outcomes are peer-separable and own-separable (PS, OS), then:*

$$ARE(\mathbf{G}_R) = 0 \tag{68}$$

$$CRE_k(\mathbf{G}_R) = (n-1) \sum_{\ell=1}^{K} \Delta \bar{x}_{k\ell}(\mathbf{G}_R) APE_\ell \tag{69}$$

Proposition 7 implies that reallocation effects can be expressed as weighted averages of peer and group effects that are identified from the joint distribution of $(\mathbf{Y}, \mathbf{X}, \mathbf{G})$ under conditions described in Propositions 4 and 5.

However, Part 1 of Proposition 7 indicates an important limitation when outcomes are not peer-separable: identifying reallocation effects from the joint distribution of $(y_i, \mathbf{x}_i, \mathbf{z}_i)$ requires that the binning scheme used to construct $\mathbf{z}_i$ must assign a unique value of $\mathbf{z}_i$ for each distinct $\bar{\mathbf{x}}_i$ in the reallocation's support. 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 $\bar{\mathbf{x}}_i$. This limitation does not affect identification from the joint distribution of $(\mathbf{Y}, \mathbf{X}, \mathbf{G})$ or $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$ since the saturated model can always be constructed from either of these joint distributions.

Example 9 (Identifying classroom reallocation effects)**.** *Continuing the classmate gender effects example, suppose the researcher estimates group effects based on five bins:*

- *All-boy* ($\bar{\mathbf{x}}_i = 1$)
- *Majority-boy* ($0.5 < \bar{\mathbf{x}}_i < 1$),
- *Exactly balanced* ($\bar{\mathbf{x}}_i = 0.5$)
- *Majority-girl* ($0.0 < \bar{\mathbf{x}}_i < 0.5$)
- *All-girl* ($\bar{\mathbf{x}}_i = 0$).

The all-boy, balanced and all-girl bins are singletons, while the majority-boy and majority-girl bins are pooled. Proposition 7 implies that

- *The researcher's 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.*
- *The researcher's 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 choose a binning scheme rich enough to identify the (conditional) average reallocation effects needed to evaluate the reallocation(s) of interest. This may not be practical for a given sample, so there will typically be a variance/bias trade-off. There may not be enough classrooms in a singleton bin for adequate statistical precision, and some pooling of similar classrooms can increase precision at a cost of bias from aggregating bins with dissimilar average effects.

5.3 Estimating reallocation effects

Proposition 7 provides a starting point for estimating reallocation effects by a plug-in method:

$$\widehat{ARE}(\mathbf{G}_R) = \begin{cases} 0 & \text{if (PS, OS)} \\ \sum_{k=0}^{K} \hat{\mu}_k \widehat{CRE}_k(\mathbf{G}_R) & \text{if (PS, CRA) or (singletons, CRA)} \end{cases} \tag{70}$$

$$\widehat{CRE}_k(\mathbf{G}_R) = \begin{cases} (n-1)\sum_{\ell=1}^{K} \Delta\bar{x}_{k\ell}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})\widehat{APE}_\ell & \text{if (PS, OS, CRA)} \\ (n-1)\sum_{\ell=1}^{K} \Delta\bar{x}_{k\ell}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})\widehat{CPE}_{k\ell} & \text{if (PS, CRA)} \\ \sum_{b=1}^{B} \Delta z_{kb}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})\widehat{CGE}_{kb} & \text{if (singletons, CRA)} \end{cases} \tag{71}$$

where:

$$\Delta\bar{x}_{k\ell}(\mathbf{G}_R, \hat{\boldsymbol{\mu}}) \equiv E\left(\bar{x}_{i\ell}(\mathbf{X}, \mathbf{G}_R) \middle| \mathbf{x}_i = \mathbf{e}_k, \frac{\sum_{j\neq i}\mathbf{x}_j}{I-1} = \hat{\boldsymbol{\mu}}\right) - \hat{\mu}_\ell \tag{72}$$

$$\begin{aligned} \Delta z_{kb}(\mathbf{G}_R, \hat{\boldsymbol{\mu}}) \equiv\ & \Pr\left(\bar{\mathbf{x}}_i(\mathbf{X}, \mathbf{G}_R(\mathbf{X}, \epsilon)) \in \mathbf{S}^b_{\bar{\mathbf{x}}} \middle| \mathbf{x}_i = \mathbf{e}_k, \frac{\sum_{j\neq i}\mathbf{x}_j}{I-1} = \hat{\boldsymbol{\mu}}\right) \\ & - \Pr\left(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) \in \mathbf{S}^b_{\bar{\mathbf{x}}} \middle| \mathbf{x}_i = \mathbf{e}_k, \frac{\sum_{j\neq i}\mathbf{x}_j}{I-1} = \hat{\boldsymbol{\mu}}\right) \end{aligned} \tag{73}$$

Both $\Delta\bar{x}_{k\ell}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})$ and $\Delta z_{kb}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})$ can be calculated by enumeration, or approximated by simulation. The asymptotic properties of the estimators defined in (70) and (71) depend on the specific reallocation mechanism $\mathbf{G}_R$ chosen by the researcher, and how the resulting value of $\Delta\bar{x}_{k\ell}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})$ and/or $\Delta z_{kb}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})$ depends on $\hat{\boldsymbol{\mu}}$. For example, the delta method can be appplied if $\Delta\bar{x}_{kb}(\mathbf{G}_R, \hat{\boldsymbol{\mu}})$ is a differentiable function of $\hat{\boldsymbol{\mu}}$.

6 Extension: Nested assignment mechanisms

Although simple random assignment is the ideal setting for studying peer effects, many empirical 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 nested 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 non-random selection into schools. The key identifying assumption of this research design is that each cohort within a school represents a random selection (due to random timing of birth) from a school-specific distribution.

This section adds a general nested assignment 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 alternative strategies for applications in which those conditions do not hold. The model is accompanied by a running example based on classroom peer effects in family income.

6.1 Maintained assumptions

The model is as defined in Section 2, with additional maintained assumptions as given below.

Assumption 6 (Group clusters)**.** *Each peer group belongs to a* **group cluster** $c \in \mathcal{C} \equiv \{1, \ldots, C\}$*:*

$$\mathbf{C} \equiv \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_I \end{bmatrix} \equiv \begin{bmatrix} c(g_1) \\ c(g_2) \\ \vdots \\ c(g_I) \end{bmatrix} \equiv \mathbf{C}(\mathbf{G}) \tag{74}$$

where c_i is the group cluster for individual i and $c : \mathcal{G} \to \mathcal{C}$ is a known function. To simplify exposition, each group cluster is assumed to include the same number ($\frac{G}{C}$) of peer groups.

Assignment to group clusters will typically depend on unobserved type.

Assumption 7 (Group clusters and types)**.** *Each individual's type is an independent draw from a type distribution that varies by group cluster:*

$$\Pr(\mathbf{T}|\mathbf{C}) = \prod_{i=1}^{I} f_{\tau|c}(\tau_i, c_i) \tag{75}$$

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

Assumption 7 allows the distribution of unobserved types, and thus the distribution of observed types and outcomes, to vary systematically across group clusters. As in Assumption 1, the arbitrary ordering of individuals makes independence a mostly innocuous assumption.

Finally, it will be convenient to assume that every observed type of individual can be found at every group cluster.

Assumption 8 (Full support)**.** *Every observed type can be observed at every group cluster:*

$$\Pr(\mathbf{x}_i = \mathbf{e}_k \cap c_i = c) > 0 \tag{76}$$

for all (k, c)*.*

Assumption 8 ensures that various conditional expectations in the analysis below are well-defined. The key results can be easily adapted to the case where some characteristic-cluster pairs have zero probability.

Example 10 (Classmate family income)**.** *Consider a simplified classroom peer effects setting in which students come from rich* ($\mathbf{x}_i = 1$) *or poor* ($\mathbf{x}_i = 0$) *families, and attend the local public* ($c_i = 1$) *or private* ($c_i = 2$) *school. Both schools have a mix of rich and poor students, but rich students are more likely to attend the private school* ($E(\mathbf{x}_i|c_i = 2) > E(\mathbf{x}_i|c_i = 1)$)*. Within each school, students are randomly assigned to cohorts/classrooms as a result of random timing of birth* ($\tau_i \perp\!\!\!\perp g_i|c_i$)*.*

6.2 Optional assumptions

The key assumption needed for any identification results in this setting is that there is random assignment to groups *within* each group cluster.

Definition 13 (Random assignment within cluster)**.** *Peer groups are* **randomly assigned within cluster (RAWC)** *if:*

$$\mathbf{G} \perp\!\!\!\perp \mathbf{T}|\mathbf{C} \tag{\textbf{RAWC}}$$

In addition, various restrictions on the outcome function $y(\cdot)$ are potentially helpful. In the interest of space and clarity, the analysis will focus on identification of peer effects under the assumption of peer separability. By Proposition 2, this will allow assumptions on the outcome function to be stated in terms of the latent variable PE_{ij}.

The first pair of assumptions restrict cross-cluster heterogeneity in a manner that is similar to standard fixed effects models. The assumption of cluster invariance allows peer effects to vary across observable types, but rules out all cross-cluster variation in peer effects.

Definition 14 (Cluster invariance)**.** *If outcomes are peer-separable (PS), the pairwise peer effects* PE_{ij} *are* **cluster invariant (CI)** *if :*

$$E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c, c_j = c') = E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c) \tag{\textbf{CI}}$$

for all (k, ℓ, c, c').

The assumption of constant shifts allows own effects to vary systematically across group clusters and across observable types, but rule out cross-cluster differences that vary by observable type.

Definition 15 (Constant shifts). *If outcomes are peer-separable (PS) and own-separable (OS), the own effects* OE_i *have* **constant shifts (CS)** *if:*

$$E(OE_i|\mathbf{x}_i = \mathbf{e}_k, c_i = c) - E(OE_i|\mathbf{x}_i = \mathbf{e}_0, c_i = c) = E(OE_i|\mathbf{x}_i = \mathbf{e}_k) - E(OE_i|\mathbf{x}_i = \mathbf{e}_0) \quad (\mathbf{CS})$$

for all (k, c).

Cluster invariance and constant shifts are both strong assumptions. Some results can be obtained for the somewhat weaker assumption of partial cluster invariance.

Definition 16 (Partial cluster invariance). *If outcomes are peer-separable (PS), the pairwise peer effects* PE_{ij} *are* **partially cluster invariant (PCI)** *if:*

$$E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0, c_i = c, c_j = c') = E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0, c_i = c) \quad (\mathbf{PCI})$$

for all (k, c, c').

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

6.3 Simple fixed effects models

Proposition 8 gives conditions under which a researcher can interpret the coefficients of a simple linear fixed effects model as measuring average peer effects.

Proposition 8 (Identification via fixed effects). *Given Assumptions 1-8, suppose that peer groups are randomly assigned within cluster (RAWC),and outcomes are peer-separable (PS) and own-separable (OS) with cluster invariance (CI) and constant shifts (CS). Then average peer effects are identified from the joint distribution of* $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i, c_i)$*:*

$$APE_\ell = CPE_{k\ell} = \alpha_{2\ell}/(n-1) \quad (77)$$

for all (k, ℓ)*, where:*

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}, c_i = c) = \alpha_0^c + \mathbf{x}\boldsymbol{\alpha}_1 + \bar{\mathbf{x}}\boldsymbol{\alpha}_2 \quad (78)$$

is a linear regression model with group cluster fixed effects.

Intuitively, fixed effects models allow group cluster to matter for the outcome, but only in ways that shift the outcome by the same amount for everyone in that group cluster, and that can be interpreted as reflecting differences in own effects and not differences in peer effects. The example below suggests how strong these assumptions would be in a typical application.

Example 11 (Assumptions needed for school fixed effects)**.** *Continuing the classmate family income example, the assumptions needed for Proposition 8 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 cluster invariance (CI).*
- *the student quality gap between rich and poor students to vary across schools. This would violate constant shifts (CS).*

Although Proposition 8 is an identification result, it has straightforward implications for estimation and inference. Equation (78) can be estimated by standard linear fixed effects methods, and standard cluster-robust asymptotic inference applies.

6.4 Heterogeneous-coefficient models

Relaxing the strong assumptions needed for Proposition 8 yields a set of heterogeneous-coefficient regression models that can be given varying causal interpretations.

For example, random assignment within cluster implies that the results of Proposition 4 can be applied on a cluster-by-cluster basis. To show this result, it is first necessary to define cluster-specific peer effects and best linear predictors.

Definition 17 (Cluster-specific peer effects)**.** *Let* **cluster-specific peer effects** *for group cluster c 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 group cluster:*

$$APE^c_\ell \equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}}) | \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0, c_i = c_j = c\right) \quad (79)$$

$$CPE^c_{k\ell} \equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}}) | \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0, \mathbf{x}_i = \mathbf{e}_k, c_i = c_j = c\right) \quad (80)$$

where $\tilde{\mathbf{q}}$ *is a purely random draw of* $(n-2)$ *peers from* f_τ.

Definition 18 (Cluster-specific best linear predictor)**.** *Let* c *be a group cluster, and let* $\mathbf{d}_i = \mathbf{d}(\mathbf{x}_i, \bar{\mathbf{x}}_i)$ *be a vector of variables (possibly including constant and interaction terms) such that* $E(\mathbf{d}_i'\mathbf{d}_i | c_i = c)$ *is nonsingular. Then the* **cluster-specific best linear predictor** *of* y_i *given* $\mathbf{d}_i$ *for group cluster* c *is:*

$$L^c(y_i|\mathbf{d}_i) \equiv \mathbf{d}_i \xi^c \quad (81)$$

where $\xi^c \equiv E(\mathbf{d}_i'\mathbf{d}_i | c_i = c)^{-1} E(\mathbf{d}_i' y_i | c_i = c)$

As shown in Part 1 of Proposition 9 below, Proposition 4 can be applied cluster-by-cluster to relate cluster-specific coefficients to the corresponding cluster-specific peer effects. These cluster-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 group cluster.

Unfortunately, within-cluster peer effects are generally insufficient to measure the effect of a reallocation *across* group clusters. This is a critical limitation that is typically not addressed in empirical work, as many reallocations of interest (e.g. Example 12 below) represent shifts across rather than within group clusters.

Results 2 and 3 in Proposition 9 below show that either form of cluster invariance can allow (non-cluster-specific) peer effects to be expressed as the average of cluster-specific effects. They are therefore identified and can be used to predict the result of a feasible reallocation across group clusters.

Proposition 9 (Identification under random assignment within cluster)**.** *Given Assumptions 1-8:*

1. *If peers are randomly assigned within cluster (RAWC) and outcomes are peer-separable (PS), then cluster-specific peer effects for each group cluster* $c \in \mathcal{C}$ *are identified from the joint distribution of* $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i, c_i)$*:*

$$CPE^c_{k\ell} = \frac{\beta^c_{2\ell} + \beta^c_{3k\ell}}{n-1} \quad (82)$$

$$APE^c_\ell = \frac{\alpha^c_{2\ell}}{n-1} \quad (83)$$

where $\boldsymbol{\alpha}^c \equiv (\alpha_0^c, \boldsymbol{\alpha}_1^c, \boldsymbol{\alpha}_2^c)$ *is the vector of coefficients from the cluster-specific best linear predictor:*

$$L^c(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \equiv \alpha_0^c + \mathbf{x}_i\boldsymbol{\alpha}_1^c + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2^c \tag{84}$$

and $\boldsymbol{\beta}^c \equiv (\beta_0^c, \boldsymbol{\beta}_1^c, \boldsymbol{\beta}_2^c, \boldsymbol{\beta}_3^c)$ *is the vector of coefficients from the cluster-specific best linear predictor:*

$$L^c(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \equiv \beta_0^c + \mathbf{x}_i\boldsymbol{\beta}_1^c + \bar{\mathbf{x}}_i\boldsymbol{\beta}_2^c + \mathbf{x}_i\boldsymbol{\beta}_3^c\bar{\mathbf{x}}_i' \tag{85}$$

2. *If peers are randomly assigned within cluster (RAWC) and outcomes are peer-separable (PS) with cluster invariance (CI), then peer effects are identified from the joint distribution of* $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i, c_i)$*:*

$$CPE_{k\ell} = \frac{E(\beta_{2\ell}^{c_i} + \beta_{3k\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_k)}{n-1} \tag{86}$$

$$APE_\ell = \frac{\sum_{k=0}^K \mu_k E(\beta_{2\ell}^{c_i} + \beta_{3k\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_k)}{n-1} \tag{87}$$

for all (k, ℓ)*.*

3. *If peers are randomly assigned within cluster (RAWC) and outcomes are peer-separable (PS) and own-separable (OS) with partial cluster invariance (PCI), then peer effects are identified from the joint distribution of* $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i, c_i)$*:*

$$CPE_{k\ell} = APE_\ell = \frac{E(\alpha_{2\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_\ell)}{n-1} \tag{88}$$

for all (k, ℓ)*.*

Designs based on random cohorts are common in the applied literature, but Propositions 8 and 9 show that they imply complications beyond those seen in a simple or conditional random assignment design. Researchers using random cohort designs have the option of imposing strong restrictions on heterogeneity (as in Proposition 8), by combining somewhat weaker restrictions with more explicit handling of heterogeneous coefficients (as in Proposition 9), or by noting that the results only apply to within-cluster comparisons and reallocations.

Example 12 (Within-school effects of rich and poor peers)**.** *Continuing the classmate family income example, suppose the researcher is unwilling to assume cluster invariance. Then within-cluster coefficients can be used to predict:*

- *The average effect of replacing a poor public school student with a rich public school student (APE^1_{21})*
- *The average effect of replacing a poor private school student with a rich private school student (APE^2_{21}).*

However, these coefficients cannot be used to predict the average effect of replacing a poor public school student with a rich private school student. Such a reallocation is never observed in the data.

The assumption of (partial) cluster invariance would make this prediction possible because it implies that poor public school students and poor private school students are equivalent as peers. Therefore, APE^2_{21} is also the the average effect of replacing a poor public school student with a rich private school student .

As with the other identification results in this paper, Proposition 9 has clear implications for estimation and inference. The estimating equations (84) and (85) fit within the framework of heterogeneous-coefficient linear panel data models. Wooldridge (2010, p. 377-381) provides a useful overview of estimation, inference procedures, and limitations for this class of models.

7 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, this information is useful in identifying and interpreting both peer effects and reallocation effects. 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 19 (Direct contextual effects)**.** *Outcomes are subject to* **direct contextual effects (DCE)** *in the* **full set of relevant 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 $u_i = u(\tau_i)$ such that:*

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

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

That is, direct contextual effects are a fixed function of a specific set of relevant characteristics. Note that the outcome is always subject to direct contextual effects in the trivial case $\mathbf{x}^*(\tau_i) = \tau_i$, so direct contextual effects are only a restrictive assumption if one specifies a specific vector as the full set of relevant characteristics.

7.1 Identification and omitted variables bias

When the outcome is subject to direct contextual effects in the *observed* variables $\mathbf{x}_i$, Proposition 10 below shows that the $h(\cdot)$ function is identified and can be estimated under the usual conditions. This function can then be used to predict various counterfactuals, including the average/conditional peer, group and reallocation effects defined earlier. In addition, $h(\cdot)$ can be used to predict the result of any reallocation based only on its effect on $\bar{\mathbf{x}}_i$, while the estimands considered in Proposition 4 only predict the result of conditionally random reallocations.

Proposition 10 also shows that direct contextual effects can overcome the limitations of research designs based on random assignment within cluster. That is, if the model exhibits direct contextual effects in $\mathbf{x}_i$, then any two individuals with the same observed characteristics have the same effect on their peers. This is a sufficient condition for cluster invariance, so the identification result in Part 2 of Proposition 9 applies.

Proposition 10 (Identification of direct contextual effects)**.** *Given Assumptions 1-8, if outcomes are subject to direct contextual effects (DCE) in $\mathbf{x}_i$, then:*

1. *If peers are conditionally randomly assigned (CRA), then $h(\cdot)$ is identified from the joint distribution of $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$:*

$$h\left(\mathbf{x}, \left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}\right) = E\left(y_i \,\middle|\, \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \frac{1}{n-1}\sum_{j=1}^{n-1}\mathbf{x}^j\right) \tag{89}$$

   for all values on the support of $(\mathbf{x}_i, \bar{\mathbf{x}}_i)$.

2. *If peers are randomly assigned within cluster (RAWC) and outcomes are peer-separable (PS), then average and conditional peer effects are identified from the joint distribution of $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i, c_i)$:*

$$CPE_{k\ell} = \frac{E(\beta_{2\ell}^{c_i} + \beta_{3k\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_k)}{n-1} \tag{90}$$

$$APE_{\ell} = \frac{\sum_{k=0}^{K}\mu_k E(\beta_{2\ell}^{c_i} + \beta_{3k\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_k)}{n-1} \tag{91}$$

   for all (k, ℓ).

In other words, the assumption that the outcome is subject to direct contextual effects in $\mathbf{x}_i$ provides additional identifying power. Unfortunately, the data requirements for that assumption are substantial - the researcher needs data on *everything* about each person that potentially affects their influence on other people - and unlikely to be met in most applications. One exception is where the researcher has a clear structural model of the causal channels by which peers influence one another, and can observe all of these channels in the available data.

When the observed characteristics form a strict subset of the relevant characteristics, estimation of the "true" $h(\cdot)$ is subject to omitted variables bias in the usual manner, and any resulting predictions may also be biased. However, the main results in this paper still apply, and average/conditional peer and group effects are identified under the usual conditions. As a result, a simple regression model with one or two peer characteristics is often preferable to a more complex model, unless the more complex model can be assumed to include the full set of relevant characteristics.

Example 13 (Peer gender, disruptive behavior, and direct contextual effects)**.** *Returning to the classmate gender effects example, consider two researchers (Researcher A and Researcher B) with data on* $(y_i, \mathbf{x}_i, \bar{\mathbf{x}}_i)$ *from a set of randomly-assigned classrooms where* $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i1}x_{i2})$ *and:*

- x_{i1} *is an indicator for whether student* i *is male.*
- x_{i2} *is an indicator for whether student* i *engages in disruptive behavior.*
- *Boys are more likely to engage in disruptive behavior*

$$E(x_{i2}|x_{i1} = 1) > E(x_{i2}|x_{i1} = 0)$$

- $\mathbf{x}_i$ *includes the full set of relevant characteristics. More specifically, peer disruptive behavior is the only relevant characteristic:*

$$h\left(\mathbf{x}, \left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}\right) = \omega_0 + \omega_1 \bar{x}_{i2}$$

- *Researcher A estimates a regression of* y_i *on own and peer gender* $(x_{i1}, \bar{x}_{i1})$*.*
- *Researcher B estimates a regression of* y_i *on own and peer gender and disruptive behavior* $(\mathbf{x}_i, \bar{\mathbf{x}}_i)$*.*

Researcher A will not recover the correct model of $h(\cdot)$*, but will still recover average and conditional peer effects for peer gender, and can make predictions about the consequences of a feasible reallocation by peer gender. In contrast, Researcher B will recover the correct model of* $h(\cdot)$*, and will correctly conclude that peer gender does not matter once one accounts for peer disruptive behavior. However, this does not mean that a reallocation* *across classrooms by gender will have no effect. Behavior varies by gender, so any change in gender composition may also change the rate of disruptive peer behavior. In the absence of information on the magnitude of this indirect effect, Researcher B's results may actually be less informative than Researcher A's on the consequences of a reallocation by gender.*

8 Conclusion

This paper has established a simple framework for thinking about contextual effects, clarifying common empirical procedures in the context of this framework, and suggesting simple and easily-implemented enhancements to those procedures. The extensions show that the framework can usefully be applied to a wide variety of settings and applications. Taken as a whole, the results 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 or average peer effects across the two categories under relatively weak assumptions. In contrast, a regression with many related peer characteristics is difficult to interpret without imposing the very strong assumptions needed to identify direct contextual effects and considering in detail the relationship between these characteristics.

A second implication is that researchers can estimate multiple distinct regression models, 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 few 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 how conditional or average peer effects vary 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. For example, 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 precise 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, a research design based on random cohorts (groups) within non-randomly assigned schools (group clusters) typically identifies only the consequences of a reallocation within the school (group cluster). Highly restrictive homogeneity assumptions are required to identify the consequences of reallocations across schools.

The results in this paper emphasize a clearer understanding of simple models and standard estimation methods rather than the development of novel or elaborate methods. Simple models and methods are central to empirical research and merit substantial attention. However, the framework developed here provides several clear avenues for further research that develops and applies novel econometric methods with empirical data. For example, the methods described here assume the researcher has predetermined a small set of discrete characteristics to use as conditioning variables. In applications where the available characteristics are continuous and/or high-dimensional, researchers may be interested in pursuing a more data-driven exploration of peer group heterogeneity. Recent advances in the use of machine learning and other tools for 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 peer and group effects from limited data. Another potential avenue of further research is a more detailed application of heterogeneous coefficients methods to measuring peer effects in cluster-based designs.

References

Abrevaya, Jason, Yu-Chin Hsu, and Robert P. Lieli, "Estimating conditional average treatment effects," *Journal of Business and Economic Statistics*, 2015, *33* (4), 485–505.

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

1. Each result can be derived from the definition. For average peer effects:

$$
\begin{aligned}
APE_\ell &= E\left(y_i(\{j\}\cup\tilde{\mathbf{p}}) - y_i(\{j'\}\cup\tilde{\mathbf{p}})\middle|\,\mathbf{x}_j=\mathbf{e}_\ell,\mathbf{x}_{j'}=\mathbf{e}_0\right) && \text{(definition)}\\
&= \sum_{k=0}^{K} E\left(y_i(\{j\}\cup\tilde{\mathbf{p}}) - y_i(\{j'\}\cup\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{x}_j=\mathbf{e}_\ell,\mathbf{x}_{j'}=\mathbf{e}_0\end{matrix}\right)\Pr(\mathbf{x}_i=\mathbf{e}_k|\mathbf{x}_j=\mathbf{e}_\ell,\mathbf{x}_{j'}=\mathbf{e}_0)\\
&= \sum_{k=0}^{K} E\left(y_i(\{j\}\cup\tilde{\mathbf{p}}) - y_i(\{j'\}\cup\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{x}_j=\mathbf{e}_\ell,\mathbf{x}_{j'}=\mathbf{e}_0\end{matrix}\right)\Pr(\mathbf{x}_i=\mathbf{e}_k)\\
& && (\text{since } \tau_i \perp\!\!\!\perp \tau_j,\tau_{j'})\\
&= \sum_{k=0}^{K} \mu_k CPE_{k\ell}
\end{aligned}
$$

For average group effects:

$$
\begin{aligned}
AGE_b &= E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0) && \text{(definition)}\\
&= \sum_{k=0}^{K} E\left(y_i(\tilde{\mathbf{p}})\,|\mathbf{x}_i=\mathbf{e}_k,\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b\right)\Pr\left(\mathbf{x}_i=\mathbf{e}_k|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b\right)\\
&\quad - \sum_{k=0}^{K} E\left(y_i(\tilde{\mathbf{p}})\,|\mathbf{x}_i=\mathbf{e}_k,\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0\right)\Pr\left(\mathbf{x}_i=\mathbf{e}_k|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0\right)\\
&= \sum_{k=0}^{K}\left(E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b\end{matrix}\right) - E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0\end{matrix}\right)\right)\Pr\left(\mathbf{x}_i=\mathbf{e}_k\right)\\
& && (\text{since } \tau_i \perp\!\!\!\perp \{\tau_j\}_{j\neq i},\tilde{\mathbf{p}})\\
&= \sum_{k=0}^{K} \mu_k CGE_{kb}
\end{aligned}
$$

2. For any k, b:

$$
\begin{aligned}
E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b\end{matrix}\right) &= \sum_{s=0}^{S} E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s\end{matrix}\right)\Pr\left(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b\end{matrix}\right)\\
&= E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_0\end{matrix}\right)\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_0|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b)\\
&\quad+\sum_{s=1}^{S} E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s\end{matrix}\right)\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b)\\
&= E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_0\end{matrix}\right)\left(1-\sum_{s=1}^{S}\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b)\right)\\
&\quad+\sum_{s=1}^{S} E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s\end{matrix}\right)\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b)\\
&= E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_0\end{matrix}\right)+\sum_{s=1}^{S} CGE_{ks}^S\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b)
\end{aligned}
\tag{92}
$$

Substituting result (92) into the definition of CGE_{kb} produces:

$$
\begin{aligned}
CGE_{kb} &= E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b\end{matrix}\right)-E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0\end{matrix}\right) && \text{(definition of } CGE_{kb})\\
&= \left(E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_0\end{matrix}\right)+\sum_{s=1}^{S} CGE_{ks}^S\Pr\left(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s\,|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b\right)\right)\\
&\quad-\left(E\left(y_i(\tilde{\mathbf{p}})\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_0\end{matrix}\right)+\sum_{s=1}^{S} CGE_{ks}^S\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0)\right) && \text{(by (92))}\\
&= \sum_{s=1}^{S} CGE_{ks}^S\begin{pmatrix}\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_b)\\ -\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}})=\mathbf{e}_s|\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{e}_0)\end{pmatrix}
\end{aligned}
\tag{93}
$$

For any (s,b):

$$\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{e}_s | \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) = \frac{\Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b \cap \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{e}_s)}{\Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b))} \tag{94}$$

$$= \frac{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^b : \mathbf{z}^S(\bar{\mathbf{x}}) = \mathbf{e}_s} \Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}})}{\sum_{\bar{\mathbf{x}} \in \mathbf{S}_{\bar{\mathbf{x}}}^b} \Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}})} \tag{95}$$

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

$$\Pr(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) = \mathcal{M}(\bar{\mathbf{x}}, n, \boldsymbol{\mu}) \tag{96}$$

The result then follows by substitution.

Proof for Proposition 2

1. Let:

$$PE(\tau_i, \tau_j) \equiv y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - \left(\frac{n-2}{n-1}\right) y(\tau_i, \{1, 1, 1, \ldots, 1\}) \tag{97}$$

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

$$\begin{aligned}
\sum_{j\in\mathbf{p}} PE_{ij} &= \sum_{j\in\mathbf{p}} y(\tau_i,\{\tau_j,1,1,\ldots,1\}) - \left(\frac{n-2}{n-1}\right) y(\tau_i,\{1,1,1,\ldots,1\}) \quad \text{(by (97))} \\
&= \left(\sum_{j\in\mathbf{p}} y(\tau_i,\{\tau_j,1,1,\ldots,1\})\right) - (n-2)\, y(\tau_i,\{1,1,1,\ldots,1\}) \\
&= y(\tau_i,\{1,1,\ldots,1\}) + \sum_{j\in\mathbf{p}} \left(y(\tau_i,\{\tau_j,1,1,\ldots,1\}) - y(\tau_i,\{1,1,1,\ldots,1\})\right) \\
&= y(\tau_i,\{1,1,\ldots,1\}) \\
&\quad + y(\tau_i,\{\tau_{\mathbf{p}(1)},1,1,\ldots,1\}) - y(\tau_i,\{1,1,1,\ldots,1\}) \\
&\quad + y(\tau_i,\{1,\tau_{\mathbf{p}(2)},1,\ldots,1\}) - y(\tau_i,\{1,1,1,\ldots,1\}) \\
&\qquad \vdots \\
&\quad + y(\tau_i,\{1,1,\ldots,\tau_{\mathbf{p}(n-1)}\}) - y(\tau_i,\{1,1,1,\ldots,1\}) \\
&= y(\tau_i,\{1,1,\ldots,1\}) \\
&\quad + y(\tau_i,\{\tau_{\mathbf{p}(1)},\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\ldots,\tau_{\mathbf{p}(n-1)}\}) - y(\tau_i,\{1,\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\ldots,\tau_{\mathbf{p}(n-1)}\}) \\
&\quad + y(\tau_i,\{1,\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\ldots,\tau_{\mathbf{p}(n-1)}\}) - y(\tau_i,\{1,1,\tau_{\mathbf{p}(3)},\ldots,\tau_{\mathbf{p}(n-1)}\}) \\
&\qquad \vdots \\
&\quad + y(\tau_i,\{1,1,\ldots,1,\tau_{\mathbf{p}(n-2)},\tau_{\mathbf{p}(n-1)}\}) - y(\tau_i,\{1,1,\ldots,1,\tau_{\mathbf{p}(n-1)}\}) \\
&\quad + y(\tau_i,\{1,1,\ldots,1,\tau_{\mathbf{p}(n-1)}\}) - y(\tau_i,\{1,1,\ldots,1\}) \\
&\hspace{20em} \text{(by PS)} \\
&= y(\tau_i,\{\tau_{\mathbf{p}(1)},\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\ldots,\tau_{\mathbf{p}(n-1)}\}) \\
&= y_i(\mathbf{p})
\end{aligned}$$

which is result (24). To prove results (25) and (26), first note that $(\tau_i,\tau_j) \perp\!\!\!\perp \tau_{j'}$ by equation (5), so:

$$(PE_{ij},\mathbf{x}_i,\mathbf{x}_j) \perp\!\!\!\perp \mathbf{x}_{j'} \tag{98}$$

Then:

$$CPE_{k\ell} = E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}})\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right)$$

(definition of $CPE_{k\ell}$)

$$= E\left(\left(PE_{ij} + \sum_{j'' \in \tilde{\mathbf{q}}} PE_{ij''}\right) - \left(PE_{ij'} + \sum_{j'' \in \tilde{\mathbf{q}}} PE_{ij''}\right)\middle| \begin{array}{l} \mathbf{x}_i = \mathbf{e}_k, \\ \mathbf{x}_j = \mathbf{e}_\ell, \\ \mathbf{x}_{j'} = \mathbf{e}_0 \end{array}\right)$$

(by (24))

$$= E\left(PE_{ij} - PE_{ij'}\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right)$$
$$= E\left(PE_{ij}\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right) - E\left(PE_{ij'}\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right)$$
$$= E\left(PE_{ij}\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right) - E\left(PE_{ij}\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_{j'} = \mathbf{e}_\ell, \mathbf{x}_j = \mathbf{e}_0\right)$$
$$= E\left(PE_{ij}\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(PE_{ij}\middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0\right) \qquad \text{(by (98))}$$

which is the result in (25) and:

$$APE_\ell = E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}})\middle| \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right) \quad \text{(definition of } APE_\ell\text{)}$$
$$= E\left(\left(PE_{ij} + \sum_{j'' \in \tilde{\mathbf{q}}} PE_{ij''}\right) - \left(PE_{ij'} + \sum_{j'' \in \tilde{\mathbf{q}}} PE_{ij''}\right)\middle| \begin{array}{l} \mathbf{x}_j = \mathbf{e}_\ell, \\ \mathbf{x}_{j'} = \mathbf{e}_0 \end{array}\right)$$

(by (24))

$$= E\left(PE_{ij} - PE_{ij'}\middle| \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right)$$
$$= E\left(PE_{ij}\middle| \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right) - E\left(PE_{ij'}\middle| \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right)$$
$$= E\left(PE_{ij}\middle| \mathbf{x}_j = \mathbf{e}_\ell, \mathbf{x}_{j'} = \mathbf{e}_0\right) - E\left(PE_{ij}\middle| \mathbf{x}_{j'} = \mathbf{e}_\ell, \mathbf{x}_j = \mathbf{e}_0\right)$$
$$= E\left(PE_{ij}\middle| \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(PE_{ij}\middle| \mathbf{x}_j = \mathbf{e}_0\right) \qquad \text{(by (98))}$$

which is the result in (26).

2. Let $PE_j \equiv y(1, \{\tau_j, 1, 1, \ldots, 1\}) - y(1, \{1, 1, 1, \ldots, 1\})$, and let $OE_i \equiv y(\tau_i, \{1, 1, 1, \ldots, 1\})$.

Then for any i:

$$\begin{aligned}
PE_{ij} &= \left(y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - \left(\frac{n-2}{n-1}\right) y(\tau_i, \{1, 1, 1, \ldots, 1\}) \right) && \text{(by (97))} \\
&= \frac{y(\tau_i, \{1, 1, 1, \ldots, 1\})}{n-1} + y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\}) \\
&= \frac{y(\tau_i, \{1, 1, 1, \ldots, 1\})}{n-1} + y(1, \{\tau_j, 1, 1, \ldots, 1\}) - y(1, \{1, 1, 1, \ldots, 1\}) \\
& && \text{(by OS)} \\
&= \frac{OE_i}{n-1} + PE_j && (99)
\end{aligned}$$

Substituting the result in (99) into (24) yields the result in (27). Substituting that same result into (25) and (26) yields:

$$\begin{aligned}
APE_\ell &= E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_0) && \text{(by (26))} \\
&= E\left(\frac{OE_i}{n-1} + PE_j \middle| \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(\frac{OE_i}{n-1} + PE_j \middle| \mathbf{x}_j = \mathbf{e}_0\right) && \text{(by (99))} \\
&= E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_j|\mathbf{x}_j = \mathbf{e}_0) + \frac{E\left(OE_i|\mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(OE_i|\mathbf{x}_j = \mathbf{e}_0\right)}{n-1} \\
&= E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_j|\mathbf{x}_j = \mathbf{e}_0) + \frac{E\left(OE_i\right) - E\left(OE_i\right)}{n-1} \\
& && \text{(by (5)} \implies OE_i \perp\!\!\!\perp \mathbf{x}_j) \\
&= E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_j|\mathbf{x}_j = \mathbf{e}_0) \\
CPE_{k\ell} &= E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell) - E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0) && \text{(by (25))} \\
&= E\left(\frac{OE_i}{n-1} + PE_j \middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(\frac{OE_i}{n-1} + PE_j \middle| \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0\right) \\
& && \text{(by (99))} \\
&= E\left(PE_j|\, \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(PE_j|\, \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0\right) && (100) \\
&\quad + \frac{E\left(OE_i|\, \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(OE_i|\, \mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0\right)}{n-1} \\
&= E\left(PE_j|\, \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(PE_j|\, \mathbf{x}_j = \mathbf{e}_0\right) + \frac{E\left(OE_i|\, \mathbf{x}_i = \mathbf{e}_k\right) - E\left(OE_i|\, \mathbf{x}_i = \mathbf{e}_k\right)}{n-1} \\
& && \text{(by (5)} \implies (\mathbf{x}_i, OE_i) \perp\!\!\!\perp (\mathbf{x}_j, PE_j)) \\
&= E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_j|\mathbf{x}_j = \mathbf{e}_0) \\
&= APE_\ell
\end{aligned}$$

which is the result in (28).

Proof for Proposition 3

1. Let $\tilde{\mathbf{G}}$ be a purely random group assignment and let $\tilde{\mathbf{p}}_i = \mathbf{p}(i, \tilde{\mathbf{G}})$. Since $\mathbf{Y}(\cdot)$ satisfies (PS) and $\tilde{\mathbf{G}}$ satisfies (RA), Part 1 of Proposition 4 applies to the joint distribution of counterfactual outcomes $(\mathbf{Y}(\tilde{\mathbf{G}}), \mathbf{X}, \bar{\mathbf{X}}(\mathbf{X}, \tilde{\mathbf{G}}))$. Since $\mathbf{G}$ satisfies (CRA), Lemma 1 applies to the joint distribution of actual outcomes $(\mathbf{Y}, \mathbf{X}, \bar{\mathbf{X}})$. Let the vector of best linear predictor coefficients $\boldsymbol{\zeta}$ be defined as in equation (103) of the proof for Proposition 4. Then:

$$
\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) && \text{(by (39) in Lemma 1)} \\
&= \zeta_0(n-1) + \mathbf{x}\boldsymbol{\zeta}_1(n-1) + \bar{\mathbf{x}}\boldsymbol{\zeta}_2(n-1) + \mathbf{x}\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}' \\
& && \text{(by (106) in the proof for Proposition 4)}
\end{aligned}
$$

Applying the law of iterated projections:

$$
\begin{aligned}
L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) &= L(E(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) && \text{(law of iterated projections)} \\
&= L\left(\begin{array}{l} \zeta_0(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1(n-1) \\ \quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' \end{array} \middle| \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i \right) \\
& && \text{(result above)} \\
&= \zeta_0(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1(n-1) && (101) \\
&\quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' \\
L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i) &= L(E(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i) \\
& && \text{(law of iterated projections)} \\
&= L\left(\begin{array}{l} \zeta_0(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1(n-1) \\ \quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' \end{array} \middle| \begin{array}{l} \mathbf{x}_i, \bar{\mathbf{x}}_i, \\ \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i \end{array} \right) \\
& && \text{(result above)} \\
&= \zeta_0(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1(n-1) && (102) \\
&\quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' \\
&= L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) && \text{(by (101) and (102))}
\end{aligned}
$$

which is result (29).

2. The assumptions here (PS, OS, CRA) imply that all results in Propositions 2

and 5 apply. Therefore:

$$\begin{aligned} APE_\ell = CPE_{k\ell} & \quad \text{for all } k && \text{(by (28) in Proposition 2)} \\ = \frac{\beta_{2\ell} + \beta_{3k\ell}}{n-1} & && \text{(by (41) in Proposition 5)} \end{aligned}$$

which can only be true if $\beta_{3k\ell} = \beta_{30\ell} = 0$ for all k, ℓ.

Proof for Proposition 4

1. By (PS), Part 1 of Proposition 2 applies. Let $\boldsymbol{\zeta} \equiv (\zeta_0, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3)$ satisfy:

$$E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell) = \zeta_0 + \mathbf{e}_k\boldsymbol{\zeta}_1 + \mathbf{e}_\ell\boldsymbol{\zeta}_2 + \mathbf{e}_k\boldsymbol{\zeta}_3\mathbf{e}_\ell' \tag{103}$$

The linear functional form in (103) is without loss of generality since $\mathbf{x}$ is categorical. The estimand $CPE_{k\ell}$ can be expressed as a function of $\boldsymbol{\zeta}$:

$$\begin{aligned} CPE_{k\ell} &= E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell) && (\text{PS} \Longrightarrow \text{ (26) in Proposition 2}) \\ &\quad - E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0) \\ &= \left(\zeta_0 + \mathbf{e}_k\boldsymbol{\zeta}_1 + \mathbf{e}_\ell\boldsymbol{\zeta}_2 + \mathbf{e}_k\boldsymbol{\zeta}_3\mathbf{e}_\ell'\right) && \text{(by (103))} \\ &\quad - \left(\zeta_0 + \mathbf{e}_k\boldsymbol{\zeta}_1 + \mathbf{e}_0\boldsymbol{\zeta}_2 + \mathbf{e}_k\boldsymbol{\zeta}_3\mathbf{e}_0'\right) \\ &= \left(\zeta_0 + \mathbf{e}_k\boldsymbol{\zeta}_1 + \mathbf{e}_\ell\boldsymbol{\zeta}_2 + \mathbf{e}_k\boldsymbol{\zeta}_3\mathbf{e}_\ell'\right) - \left(\zeta_0 + \mathbf{e}_k\boldsymbol{\zeta}_1\right) && (\text{since } \mathbf{e}_0 = \mathbf{0}) \\ &= \mathbf{e}_\ell\boldsymbol{\zeta}_2 + \mathbf{e}_k\boldsymbol{\zeta}_3\mathbf{e}_\ell' \\ &= \zeta_{2\ell} + \zeta_{3k\ell} && (104) \end{aligned}$$

The next step is to show the relationship between the coefficients in ζ and the coefficients in β:

$$
\begin{aligned}
E(y_i|\mathbf{X},\mathbf{G}) &= E\left(\sum_{j\in\mathbf{p}_i} PE_{ij}\middle|\mathbf{X},\mathbf{G}\right) && (\text{PS} \implies (24) \text{ in Proposition 2})\\
&= E\left(\sum_{j=1}^{I} PE_{ij}\mathbb{I}\left(j\in\mathbf{p}_i\right)\middle|\mathbf{X},\mathbf{G}\right)\\
& && (\text{where } \mathbb{I}\left(\cdot\right) \text{ is the indicator function})\\
&= \sum_{j=1}^{I} E(PE_{ij}|\mathbf{X},\mathbf{G})\mathbb{I}\left(j\in\mathbf{p}_i\right) && (\text{since } \mathbb{I}\left(j\in\mathbf{p}_i\right) \text{ is a function of } \mathbf{G})\\
&= \sum_{j\in\mathbf{p}_i} E(PE_{ij}|\mathbf{X},\mathbf{G})\\
&= \sum_{j\in\mathbf{p}_i} E(PE_{ij}|\mathbf{X}) && (\text{RA} \implies (PE_{ij},\mathbf{X}) \perp\!\!\!\perp (\mathbf{G},\mathbf{p}_i))\\
&= \sum_{j\in\mathbf{p}_i} E(PE_{ij}|\mathbf{x}_i,\mathbf{x}_j) && (\text{since } (5) \implies (\tau_i,\tau_j) \perp\!\!\!\perp \tau_{j'})\\
&= \sum_{j\in\mathbf{p}_i} (\zeta_0 + \mathbf{x}_i\boldsymbol{\zeta}_1 + \mathbf{x}_j\boldsymbol{\zeta}_2 + \mathbf{x}_i\boldsymbol{\zeta}_3\mathbf{x}_j') && (\text{by } (103))\\
&= \zeta_0(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1(n-1) + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' && (105)
\end{aligned}
$$

Applying the law of iterated expectations to this result:

$$
\begin{aligned}
E(y_i|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) &= E(E(y_i|\mathbf{X},\mathbf{G})|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}})\\
& && (\text{law of iterated expectations})\\
&= E\left(\begin{matrix}\zeta_0(n-1)+\mathbf{x}_i\boldsymbol{\zeta}_1(n-1)\\ +\bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1)+\mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i'\end{matrix}\middle|\begin{matrix}\mathbf{x}_i=\mathbf{x},\\ \bar{\mathbf{x}}_i=\bar{\mathbf{x}}\end{matrix}\right) && (\text{by } (105))\\
&= \zeta_0(n-1)+\mathbf{x}\boldsymbol{\zeta}_1(n-1)+\bar{\mathbf{x}}\boldsymbol{\zeta}_2(n-1)+\mathbf{x}\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}'\\
& && (106)
\end{aligned}
$$

Applying the law of iterated projections to this result:

$$\begin{aligned}
L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i\bar{\mathbf{x}}_i') &= L(E(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i\bar{\mathbf{x}}_i') && \text{(law of iterated projections)} \\
&= L\left(\begin{array}{l} \zeta_0(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1(n-1) \\ + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' \end{array} \middle| \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i\bar{\mathbf{x}}_i' \right) && \text{(by (106))} \\
&= \underbrace{\zeta_0(n-1)}_{\beta_0} + \mathbf{x}_i \underbrace{\boldsymbol{\zeta}_1(n-1)}_{\boldsymbol{\beta}_1} + \bar{\mathbf{x}}_i \underbrace{\boldsymbol{\zeta}_2(n-1)}_{\boldsymbol{\beta}_2} + \mathbf{x}_i \underbrace{\boldsymbol{\zeta}_3(n-1)}_{\boldsymbol{\beta}_3} \bar{\mathbf{x}}_i' && (107)
\end{aligned}$$

So $\boldsymbol{\beta}_2 = \boldsymbol{\zeta}_2(n-1)$, $\boldsymbol{\beta}_3 = \boldsymbol{\zeta}_3(n-1)$ and:

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

which is result (32). To get result (31), first note that:

$$\begin{aligned}
E(PE_{ij}|\mathbf{x}_j = \mathbf{x}) &= E(E(PE_{ij}|\mathbf{x}_i, \mathbf{x}_j)|\mathbf{x}_j = \mathbf{x}) && \text{(law of iterated expectations)} \\
&= E(\zeta_0 + \mathbf{x}_i\boldsymbol{\zeta}_1 + \mathbf{x}_j\boldsymbol{\zeta}_2 + \mathbf{x}_i\boldsymbol{\zeta}_3\mathbf{x}_j'|\mathbf{x}_j = \mathbf{x}) && \text{(by (103))} \\
&= \zeta_0 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{x})\boldsymbol{\zeta}_1 + \mathbf{x}\boldsymbol{\zeta}_2 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{x})\boldsymbol{\zeta}_3\mathbf{x}' \\
& && \text{(conditioning rule)} \\
&= \zeta_0 + E(\mathbf{x}_i)\boldsymbol{\zeta}_1 + \mathbf{x}\boldsymbol{\zeta}_2 + E(\mathbf{x}_i)\boldsymbol{\zeta}_3\mathbf{x}' && \text{(since (5) } \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j) \\
&= (\zeta_0 + E(\mathbf{x}_i))\,\boldsymbol{\zeta}_1 + \mathbf{x}\left(\boldsymbol{\zeta}_2 + \boldsymbol{\zeta}_3' E(\mathbf{x}_i')\right) && (108)
\end{aligned}$$

Equation (26) from Proposition 2 implies:

$$\begin{aligned}
APE_\ell &= E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_0) && \text{(PS} \implies \text{(26) in Proposition 2)} \\
&= \left((\zeta_0 + E(\mathbf{x}_i)\boldsymbol{\zeta}_1) + \mathbf{e}_\ell\left(\boldsymbol{\zeta}_2 + \boldsymbol{\zeta}_3' E(\mathbf{x}_i')\right)\right) && \text{(by (108))} \\
&\quad - \left((\zeta_0 + E(\mathbf{x}_i)\boldsymbol{\zeta}_1) + \mathbf{e}_0\left(\boldsymbol{\zeta}_2 + \boldsymbol{\zeta}_3' E(\mathbf{x}_i')\right)\right) \\
&= \left((\zeta_0 + E(\mathbf{x}_i)\boldsymbol{\zeta}_1) + \mathbf{e}_\ell\left(\boldsymbol{\zeta}_2 + \boldsymbol{\zeta}_3' E(\mathbf{x}_i')\right)\right) && \text{(since } \mathbf{e}_0 = 0) \\
&\quad - ((\zeta_0 + E(\mathbf{x}_i)\boldsymbol{\zeta}_1)) \\
&= \mathbf{e}_\ell\left(\boldsymbol{\zeta}_2 + \boldsymbol{\zeta}_3' E(\mathbf{x}_i')\right) && (109)
\end{aligned}$$

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

$$\begin{aligned}
L(y_i|\bar{\mathbf{x}}_i) &= L(L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\bar{\mathbf{x}}_i) && \text{(law of iterated projections)} \\
&= L(\alpha_0 + \mathbf{x}_i\boldsymbol{\alpha}_1 + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2|\bar{\mathbf{x}}_i) && \text{(definition of } \boldsymbol{\alpha}\text{)} \\
&= \alpha_0 + L(\mathbf{x}_i|\bar{\mathbf{x}}_i)\boldsymbol{\alpha}_1 + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2 \\
&= (\alpha_0 + E(\mathbf{x}_i)\boldsymbol{\alpha}_1) + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2 && (\text{RA} \implies \mathbf{x}_i \perp\!\!\!\perp \bar{\mathbf{x}}_i)
\end{aligned}$$

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

$$\begin{aligned}
L(y_i|\bar{\mathbf{x}}_i) &= L(L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i)|\bar{\mathbf{x}}_i) && \text{(law of iterated projections)} \\
&= L\left(\begin{array}{l} \zeta_0(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1(n-1) \\ \quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' \end{array}\middle| \bar{\mathbf{x}}_i\right) && \text{(by (107))} \\
&= \zeta_0(n-1) + L(\mathbf{x}_i|\bar{\mathbf{x}}_i)\boldsymbol{\zeta}_1(n-1) && \text{(property of linear projection)} \\
&\quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + L(\mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i'|\bar{\mathbf{x}}_i) \\
&= \zeta_0(n-1) + E(\mathbf{x}_i)\boldsymbol{\zeta}_1(n-1) && (\text{RA} \implies \mathbf{x}_i \perp\!\!\!\perp \bar{\mathbf{x}}_i) \\
&\quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2(n-1) + E(\mathbf{x}_i)\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i' \\
&= \underbrace{\zeta_0(n-1) + E(\mathbf{x}_i)\boldsymbol{\zeta}_1(n-1)}_{\alpha_0 + E(\mathbf{x}_i)\boldsymbol{\alpha}_1} + \bar{\mathbf{x}}_i\underbrace{(\boldsymbol{\zeta}_2(n-1) + \boldsymbol{\zeta}_3'E(\mathbf{x}_i')(n-1))}_{\boldsymbol{\alpha}_2} && (110)
\end{aligned}$$

So $\boldsymbol{\alpha}_2 = (\boldsymbol{\zeta}_2(n-1) + \boldsymbol{\zeta}_3'E(\mathbf{x}_i')(n-1))$ and:

$$\begin{aligned}
APE_k &= \mathbf{e}_\ell\left(\boldsymbol{\zeta}_2 + \boldsymbol{\zeta}_3'E(\mathbf{x}_i')\right) && \text{(by (109))} \\
&= \mathbf{e}_\ell\frac{\boldsymbol{\alpha}_2}{n-1} && \text{(by (110))} \\
&= \frac{\alpha_{2\ell}}{n-1}
\end{aligned}$$

which is the result in (31).

2. Let $\tilde{\mathbf{p}}$ be a purely random draw of $(n-1)$ peers from f_τ. 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:

$$\begin{aligned}
CGE_{kb} &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \quad \text{(by (17))} \\
&= E(y_i(\mathbf{p}_i)|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_b) - E(y_i(\mathbf{p}_i)|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_0) \\
&\qquad \text{(RA} \implies \text{same joint distribution)} \\
&= E(y_i|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i = \mathbf{e}_b) - E(y_i|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i = \mathbf{e}_0) \qquad (111)
\end{aligned}$$

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:

$$\begin{aligned}
E(y_i|\mathbf{x}_i, \mathbf{z}_i) &= L(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i) \\
&= \delta_0 + \mathbf{x}_i\boldsymbol{\delta}_1 + \mathbf{z}_i\boldsymbol{\delta}_2 + \mathbf{x}_i\boldsymbol{\delta}_3\mathbf{z}_i' \qquad \text{(by (38))}
\end{aligned}$$

Combining these two results produces:

$$\begin{aligned}
CGE_{kb} &= E(y_i|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i = \mathbf{e}_b) - E(y_i|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i = \mathbf{e}_0) \qquad \text{(by (111))} \\
&= \left(\delta_0 + \mathbf{e}_k\boldsymbol{\delta}_1 + \mathbf{e}_b\boldsymbol{\delta}_2 + \mathbf{e}_k\boldsymbol{\delta}_3\mathbf{e}_b'\right) - \left(\delta_0 + \mathbf{e}_k\boldsymbol{\delta}_1 + \mathbf{e}_0\boldsymbol{\delta}_2 + \mathbf{e}_k\boldsymbol{\delta}_3\mathbf{e}_0'\right) \\
&\qquad \text{(result above)} \\
&= \left(\delta_0 + \mathbf{e}_k\boldsymbol{\delta}_1 + \mathbf{e}_b\boldsymbol{\delta}_2 + \mathbf{e}_k\boldsymbol{\delta}_3\mathbf{e}_b'\right) - (\delta_0 + \mathbf{e}_k\boldsymbol{\delta}_1) \qquad \text{(since } \mathbf{e}_0 = \mathbf{0}) \\
&= \mathbf{e}_b\boldsymbol{\delta}_2 + \mathbf{e}_k\boldsymbol{\delta}_3\mathbf{e}_b' \\
&= \delta_{2b} + \delta_{3kb}
\end{aligned}$$

which is result (36). Result (35) can be established by similar reasoning:

$$\begin{aligned}
AGE_b &= E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \qquad \text{(by (16))} \\
&= E(y_i(\mathbf{p}_i)|\mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_b) - E(y_i(\mathbf{p}_i)|\mathbf{z}_i(\mathbf{p}_i) = \mathbf{e}_0) \\
&\qquad \text{(RA} \implies \text{same joint distribution)} \\
&= E(y_i|\mathbf{z}_i = \mathbf{e}_b) - E(y_i|\mathbf{z}_i = \mathbf{e}_0) \qquad (112)
\end{aligned}$$

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

$$\begin{aligned}
E(y_i|\mathbf{z}_i) &= L(y_i|\mathbf{z}_i) \\
&= L(L(y_i|\mathbf{x}_i, \mathbf{z}_i)|\mathbf{z}_i) && \text{(law of iterated projections)} \\
&= L(\gamma_0 + \mathbf{x}_i\boldsymbol{\gamma}_1 + \mathbf{z}_i\boldsymbol{\gamma}_2|\mathbf{z}_i) && \text{(by (37))} \\
&= \gamma_0 + L(\mathbf{x}_i|\mathbf{z}_i)\boldsymbol{\gamma}_1 + \mathbf{z}_i\boldsymbol{\gamma}_2 \\
&= \gamma_0 + E(\mathbf{x}_i)\boldsymbol{\gamma}_1 + \mathbf{z}_i\boldsymbol{\gamma}_2 && (\text{RA} \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{z}_i)
\end{aligned}$$

Combining these two results:

$$\begin{aligned}
AGE_b &= E(y_i|\mathbf{z}_i = \mathbf{e}_b) - E(y_i|\mathbf{z}_i = \mathbf{e}_0) && \text{(by (112))} \\
&= (\gamma_0 + E(\mathbf{x}_i)\boldsymbol{\gamma}_1 + \mathbf{e}_b\boldsymbol{\gamma}_2) - (\gamma_0 + E(\mathbf{x}_i)\boldsymbol{\gamma}_1 + \mathbf{e}_0\boldsymbol{\gamma}_2) && \text{(result above)} \\
&= (\gamma_0 + E(\mathbf{x}_i)\boldsymbol{\gamma}_1 + \mathbf{e}_b\boldsymbol{\gamma}_2) - (\gamma_0 + E(\mathbf{x}_i)\boldsymbol{\gamma}_1) && (\text{since } \mathbf{e}_0 = \mathbf{0}) \\
&= \mathbf{e}_b\boldsymbol{\gamma}_2 \\
&= \gamma_{2b}
\end{aligned}$$

which is result (35).

Proof for Lemma 1

Choose any $\mathbf{G_A} \in \mathcal{G}_n^I$ and $\mathbf{X_A} \in \mathbb{R}^{I \times K}$. Let $\mathbf{g}_i^A \equiv (i, \mathbf{p}(i, \mathbf{G_A}))$ be an n-vector identifying all individuals in group g_i including individual i, and for any matrix $\mathbf{M}$ and vector $\mathbf{v}$ let $[\mathbf{M}]_{\mathbf{v}}$ be the submatrix constructed from rows $\mathbf{v}$ in matrix $\mathbf{M}$. Then:

$$\begin{aligned}
E(y_i|\mathbf{X}=\mathbf{X_A},\mathbf{G}=\mathbf{G_A}) &= E(y(\tau_i,\{\tau_j\}_{j\in\mathbf{p}(i,\mathbf{G})})|\mathbf{X}=\mathbf{X_A},\mathbf{G}=\mathbf{G_A}) \\
&= E(y(\tau_i,\{\tau_j\}_{j\in\mathbf{p}(i,\mathbf{G_A})})|\mathbf{X}=\mathbf{X_A},\mathbf{G}=\mathbf{G_A}) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(conditioning rule)} \\
&= E(y(\tau_i,\{\tau_j\}_{j\in\mathbf{p}(i,\mathbf{G_A})})|\mathbf{X}=\mathbf{X_A}) \qquad\qquad\qquad \text{(by CRA)} \\
&= \sum_{\mathbf{T_A}\in\mathcal{T}^n} y(\mathbf{T_A})\Pr([\mathbf{T}]_{\mathbf{g}_i^A}=\mathbf{T_A}|\mathbf{X}=\mathbf{X_A}) \\
&= \sum_{\mathbf{T_A}\in\mathcal{T}^n} y(\mathbf{T_A})\prod_{j=1}^{n}\Pr\left([\mathbf{T}]_{[\mathbf{g}_i^A]_j}=[\mathbf{T_A}]_j \,\middle|\, [\mathbf{X}]_{[\mathbf{g}_i^A]_j}=[\mathbf{X_A}]_{[\mathbf{g}_i^A]_j}\right) \\
&\qquad\qquad\qquad\qquad \text{(since } (\tau_i,\mathbf{x}_i)\perp\!\!\!\perp(\tau_j,\mathbf{x}_j) \text{ for all } i\neq j) \\
&= \underbrace{\sum_{\mathbf{T_A}\in\mathcal{T}^n} y(\mathbf{T_A})\prod_{j=1}^{n}\frac{f_\tau([\mathbf{T_A}]_j)\mathbb{I}\left(\mathbf{x}([\mathbf{T_A}]_j)=[\mathbf{X_A}]_{[\mathbf{g}_i^A]_j}\right)}{\sum_{\tau\in\mathcal{T}} f_\tau(\tau)\mathbb{I}\left(\mathbf{x}(\tau)=[\mathbf{X_A}]_{[\mathbf{g}_i^A]_j}\right)}}_{\equiv\nu(\mathbf{x}_i(\mathbf{X_A}),\bar{\mathbf{x}}_i(\mathbf{X_A},\mathbf{G_A}))}
\end{aligned} \tag{113}$$

Note that the last step in equation (113) makes use of the fact that $\bar{\mathbf{x}}_i$ fully describes the frequency distribution of characteristics $\{\mathbf{x}_j\}_{j\in\mathbf{p}_i}$, and that the $\nu(\cdot)$ function depends on the type distribution $f_\tau(\cdot)$ but not on the probability distribution of $\mathbf{G}$ other than through the conditional random assignment condition.

Since equation (113) holds for any $(\mathbf{X_A},\mathbf{G_A})$ we can also say that:

$$\begin{aligned}
E(y_i|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) &= E(E(y_i|\mathbf{X},\mathbf{G})|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) \quad \text{(law of iterated expectations)} \\
&= E(\nu(\mathbf{x}_i(\mathbf{X}),\bar{\mathbf{x}}_i(\mathbf{X},\mathbf{G})|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) \qquad \text{(by (113))} \\
&= \nu(\mathbf{x},\bar{\mathbf{x}})
\end{aligned} \tag{114}$$

Since equation (114) holds with the same $\nu(\cdot)$ function for any $\mathbf{G}$ that satisfies (CRA), it also holds for purely random $\mathbf{G}$, which implies result (39).

Proof for Proposition 5

Let $\tilde{\mathbf{G}}$ be a purely random group assignment and let $\tilde{\mathbf{p}}_i=\mathbf{p}(i,\tilde{\mathbf{G}})$. Then:

1. Let $\tilde{\boldsymbol{\beta}}\equiv(\tilde{\beta}_0,\tilde{\boldsymbol{\beta}}_1,\tilde{\boldsymbol{\beta}}_2,\tilde{\boldsymbol{\beta}}_3)$ be the best linear predictor coefficients from the counter-

factual regression model:

$$L(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i), \mathbf{x}_i'\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)) = \tilde{\beta}_0 + \mathbf{x}_i\tilde{\boldsymbol{\beta}}_1 + \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)\tilde{\boldsymbol{\beta}}_2 + \mathbf{x}_i\tilde{\boldsymbol{\beta}}_3\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)' \tag{115}$$

Since outcomes are peer-separable (PS) and $\tilde{\mathbf{G}}$ satisfies (RA), Part 1 of Proposition 4 applies to the counterfactual outcomes:

$$CPE_{k\ell} = \frac{\tilde{\beta}_{2\ell} + \tilde{\beta}_{3k\ell}}{n-1} \qquad \text{(by (32) in Proposition 4)}$$

In addition, the proof for Proposition 4 shows that the counterfactual CEF is linear under these conditions:

$$\begin{aligned} E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) &= L(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}, \mathbf{x}_i'\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \mathbf{x}'\bar{\mathbf{x}}) && \text{(by (107))} \\ &= \tilde{\beta}_0 + \mathbf{x}\tilde{\boldsymbol{\beta}}_1 + \bar{\mathbf{x}}\tilde{\boldsymbol{\beta}}_2 + \mathbf{x}\tilde{\boldsymbol{\beta}}_3\bar{\mathbf{x}}' && (116) \end{aligned}$$

Since $\mathbf{G}$ satisfies (CRA), Lemma 1 applies, which also implies that the actual CEF is linear:

$$\begin{aligned} E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) && \text{(by (39) in Lemma 1)} \\ &= \tilde{\beta}_0 + \mathbf{x}\tilde{\boldsymbol{\beta}}_1 + \bar{\mathbf{x}}\tilde{\boldsymbol{\beta}}_2 + \mathbf{x}\tilde{\boldsymbol{\beta}}_3\bar{\mathbf{x}}' && (117) \end{aligned}$$

and $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$. Therefore:

$$CPE_{k\ell} = \frac{\beta_{2\ell} + \beta_{3k\ell}}{n-1} \qquad \text{(since } \beta = \tilde{\beta}\text{)}$$

which is result (41). Result (40) follows from substitution of result (41) into result (18).

2. Let $\tilde{\boldsymbol{\lambda}} \equiv (\tilde{\lambda}_0, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2, \tilde{\boldsymbol{\lambda}}_3)$ be the best linear predictor coefficients from the counterfactual regression model:

$$L(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i, \mathbf{z}_i^S(\tilde{\mathbf{p}}_i), \mathbf{x}_i'\mathbf{z}_i^S(\tilde{\mathbf{p}}_i)) = \tilde{\lambda}_0 + \mathbf{x}_i\tilde{\boldsymbol{\lambda}}_1 + \mathbf{z}_i^S(\tilde{\mathbf{p}}_i)\tilde{\boldsymbol{\lambda}}_2 + \mathbf{x}_i\tilde{\boldsymbol{\lambda}}_3\mathbf{z}_i^S(\tilde{\mathbf{p}}_i)' \tag{118}$$

Since $\tilde{\mathbf{G}}$ satisfies (RA), Part 2 of Proposition 4 applies to the counterfactual outcomes:

$$CGE_{ks}^S = \tilde{\lambda}_{2s} + \tilde{\lambda}_{3ks} \qquad \text{(by (111) in Proposition 4)}$$

The counterfactual CEF is linear since $\mathbf{z}^S(\cdot)$ is saturated:

$$E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) = L(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i, \mathbf{z}_i^S(\tilde{\mathbf{p}}_i), \mathbf{x}_i'\mathbf{z}_i^S(\tilde{\mathbf{p}}_i)) \tag{119}$$

$$= \tilde{\lambda}_0 + \mathbf{x}\tilde{\boldsymbol{\lambda}}_1 + \mathbf{z}^S(\bar{\mathbf{x}})\tilde{\boldsymbol{\lambda}}_2 + \mathbf{x}\tilde{\boldsymbol{\lambda}}_3\mathbf{z}^S(\bar{\mathbf{x}})' \tag{120}$$

Since $\mathbf{G}$ satisfies (CRA), Lemma 1 applies, which implies that:

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) = E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) \qquad \text{(by (39) in Lemma 1)}$$

$$= \tilde{\lambda}_0 + \mathbf{x}\tilde{\boldsymbol{\lambda}}_1 + \mathbf{z}^S(\bar{\mathbf{x}})\tilde{\boldsymbol{\lambda}}_2 + \mathbf{x}\tilde{\boldsymbol{\lambda}}_3\mathbf{z}^S(\bar{\mathbf{x}})' \tag{121}$$

and $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}$. Therefore:

$$CGE_{ks}^S = \lambda_{2s} + \lambda_{3ks} \qquad \text{(since } \boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}\text{)}$$

Result (43) then follows from substitution of this result into result (21). Result (42) follows from substitution of result (43) into result (19).

Proof for Proposition 6

The result follows directly from the definitions:

$$\begin{aligned} ARE(\mathbf{G}_R) &= E(y_i(\mathbf{p}(i, \mathbf{G}_R(\mathbf{X}, \epsilon))) - y_i(\tilde{\mathbf{p}})) && \text{(definition of } ARE\text{)} \\ &= \sum_{k=0}^{K} E\left(y_i(\mathbf{p}(i, \mathbf{G}_R(\mathbf{X}, \epsilon))) - y_i(\tilde{\mathbf{p}})|\, \mathbf{x}_i = \mathbf{e}_k\right) \Pr\left(\mathbf{x}_i = \mathbf{e}_k\right) \\ &= \sum_{k=0}^{K} \mu_k CRE_k(\mathbf{G}_R) && \text{(definition of } \mu \text{ and } CRE\text{)} \end{aligned}$$

which is the result in (61).

Proof for Proposition 7

For convenience, let $\tilde{\mathbf{G}}_R \equiv \mathbf{G}_R(\mathbf{X}, \epsilon)$ $y_i^R \equiv y_i(\mathbf{p}(i, \tilde{\mathbf{G}}_R))$, $\bar{\mathbf{x}}_i^R \equiv \bar{\mathbf{x}}(\mathbf{p}(i, \tilde{\mathbf{G}}_R))$, and $\mathbf{z}_i^R \equiv \mathbf{z}(\bar{\mathbf{x}}_i^R)$.

1. Since $\tilde{\mathbf{G}}_R$ satisfies (CRA), Lemma 1 applies:

$$E(y_i^R|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i^R = \bar{\mathbf{x}}) = E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) \qquad \text{(by (39) in Lemma 1)}$$

Pick any $b > 0$. By assumption, $\mathbf{S}_{\bar{\mathbf{x}}}^b = \left\{\bar{\mathbf{x}}^b\right\}$ is a singleton, and the events $\bar{\mathbf{x}}_i^R = \bar{\mathbf{x}}^b$

and $\mathbf{z}_i^R = \mathbf{e}_b$ are identical. Therefore:

$$\begin{aligned}
E(y_i^R|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i^R = \mathbf{e}_b) &= E(y_i^R|\mathbf{x}_i = \mathbf{e}_k, \bar{\mathbf{x}}_i^R = \bar{\mathbf{x}}^b) && \text{(identical events)} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}^b) && \text{(by Lemma 1)} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) && \text{(identical events)} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + CGE_{kb} && (122)
\end{aligned}$$

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

$$\begin{aligned}
E(y_i^R|\mathbf{x}_i = \mathbf{e}_k) &= \sum_{b=0}^{B} E(y_i^R|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i^R = \mathbf{e}_b) \Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
&= \sum_{b=1}^{B} E(y_i^R|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i^R = \mathbf{e}_b) \Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
& \qquad \text{(since } \Pr(\bar{\mathbf{x}}_i^R \in \mathbf{S}_{\bar{\mathbf{x}}}^0) = 0) \\
&= \sum_{b=1}^{B} (E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + CGE_{kb}) \Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
& \qquad \text{(by (122))} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \left(\underbrace{\sum_{b=1}^{B} \Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k)}_{1} \right) \\
& \quad + \sum_{b=1}^{B} CGE_{kb} \Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + \sum_{b=1}^{B} CGE_{kb} \Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \qquad (123)
\end{aligned}$$

Similarly:

$$
\begin{aligned}
E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k) &= \sum_{b=0}^{B} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0|\mathbf{x}_i = \mathbf{e}_k) \\
&\quad + \sum_{b=1}^{B} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b) \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0|\mathbf{x}_i = \mathbf{e}_k) \\
&\quad + \sum_{b=1}^{B} \left(E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + CGE_{kb}\right) \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
&\qquad \text{(by (122))} \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) \left(\underbrace{\sum_{b=0}^{B} \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k)}_{1} \right) \\
&\quad + \sum_{b=1}^{B} CGE_{kb} \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \\
&= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + \sum_{b=1}^{B} CGE_{kb} \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k)
\end{aligned} \tag{124}
$$

Combining these results yields:

$$
\begin{aligned}
CRE_k(\mathbf{G}_R) &= E(y_i(\mathbf{p}(i, \mathbf{G}_R(\mathbf{X}, \epsilon))) - y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k) \qquad \text{(definition of } CRE\text{)} \\
&= E(y_i^R|\mathbf{x}_i = \mathbf{e}_k) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k) \\
&= \left(E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + \sum_{b=1}^{B} CGE_{kb} \Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \right) \\
&\quad - \left(E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_0) + \sum_{b=1}^{B} CGE_{kb} \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \right) \\
&\qquad \text{(by (123) and (124))} \\
&= \sum_{b=1}^{B} CGE_{kb} \left(\Pr(\mathbf{z}_i^R = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) - \Pr(\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{e}_b|\mathbf{x}_i = \mathbf{e}_k) \right) \\
&= \sum_{b=1}^{B} \Delta z_{kb}(\mathbf{G}_R) CGE_{kb}
\end{aligned}
$$

which is the result in (63). Result (62) follows by substituting (63) into result (61) of Proposition 6.

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

$$
\begin{aligned}
E(y_i^R|\mathbf{X},\tilde{\mathbf{G}}_R) &= E(y_i(\mathbf{p}(i,\tilde{\mathbf{G}}_R))|\mathbf{X},\tilde{\mathbf{G}}_R) && \text{(definition)}\\
&= E\left(\sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} PE_{ij}\middle|\mathbf{X},\tilde{\mathbf{G}}_R\right) && \text{(by Proposition 2)}\\
&= \sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} E\left(PE_{ij}|\,\mathbf{X},\tilde{\mathbf{G}}_R\right)\\
&= \sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} E\left(E(PE_{ij}|\mathbf{X},\tilde{\mathbf{G}}_R,\epsilon)\middle|\,\mathbf{X},\tilde{\mathbf{G}}_R\right)\\
& && \text{(law of iterated expectations)}\\
&= \sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} E\left(E(PE_{ij}|\mathbf{X},\epsilon)\,\middle|\mathbf{X},\tilde{\mathbf{G}}_R\right)\\
& && \text{(since } \tilde{\mathbf{G}}_R \text{ is a function of } (\mathbf{X},\epsilon))\\
&= \sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} E\left(E(PE_{ij}|\mathbf{X})\,\middle|\mathbf{X},\tilde{\mathbf{G}}_R\right) && \text{(since } \epsilon \perp\!\!\!\perp \mathbf{T})\\
&= \sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} E\left(E(PE_{ij}|\mathbf{x}_i,\mathbf{x}_j)\,\middle|\mathbf{X},\tilde{\mathbf{G}}_R\right) && \text{(since } \tau_i \perp\!\!\!\perp \tau_j \text{ for } i\neq j)\\
&= \sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} E(PE_{ij}|\mathbf{x}_i,\mathbf{x}_j)\\
&= \sum_{j\in\mathbf{p}(i,\tilde{\mathbf{G}}_R)} \zeta_0+\mathbf{x}_i\boldsymbol{\zeta}_1+\mathbf{x}_j\boldsymbol{\zeta}_2+\mathbf{x}_i\boldsymbol{\zeta}_3\mathbf{x}_j'\\
& && \text{(where } \boldsymbol{\zeta} \text{ is defined as in (103))}\\
&= \zeta_0(n-1)+\mathbf{x}_i\boldsymbol{\zeta}_1(n-1)+\bar{\mathbf{x}}_i^R\boldsymbol{\zeta}_2(n-1)+\mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i^{R\prime} && (125)
\end{aligned}
$$

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

$$
\begin{aligned}
E(y_i^R|\mathbf{x}_i=\mathbf{x}) &= E(E(y_i^R|\mathbf{X},\tilde{\mathbf{G}}_R)|\mathbf{x}_i=\mathbf{x}) && \text{(Law of iterated expectations)}\\
&= E\left(\begin{array}{l}\zeta_0(n-1)+\mathbf{x}_i\boldsymbol{\zeta}_1(n-1)\\ \quad+\bar{\mathbf{x}}_i^R\boldsymbol{\zeta}_2(n-1)+\mathbf{x}_i\boldsymbol{\zeta}_3(n-1)\bar{\mathbf{x}}_i^{R\prime}\end{array}\middle|\,\mathbf{x}_i=\mathbf{x}\right) && \text{(by (125))}\\
&= \zeta_0(n-1)+\mathbf{x}\boldsymbol{\zeta}_1(n-1) && (126)\\
&\quad+E(\bar{\mathbf{x}}_i^R|\mathbf{x}_i=\mathbf{x})\boldsymbol{\zeta}_2(n-1)+\mathbf{x}\boldsymbol{\zeta}_3(n-1)E(\bar{\mathbf{x}}_i^R|\mathbf{x}_i=\mathbf{x})'
\end{aligned}
$$

This result also applies when $\tilde{\mathbf{G}}_R$ is purely random, so:

$$\begin{aligned}
E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}) &= \zeta_0(n-1) + \mathbf{x}\boldsymbol{\zeta}_1(n-1) + E(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x})\boldsymbol{\zeta}_2(n-1) \\
&\quad + \mathbf{x}\boldsymbol{\zeta}_3(n-1)E(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x})' \\
&\hspace{6cm} \text{(by (126))} \\
&= \zeta_0(n-1) + \mathbf{x}\boldsymbol{\zeta}_1(n-1) + E(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}))\boldsymbol{\zeta}_2(n-1) \\
&\quad + \mathbf{x}\boldsymbol{\zeta}_3(n-1)E(\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}))' \\
&\hspace{6cm} (\text{RA} \implies \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) \perp\!\!\!\perp \mathbf{x}_i) \\
&= \zeta_0(n-1) + \mathbf{x}\boldsymbol{\zeta}_1(n-1) + \boldsymbol{\mu}\boldsymbol{\zeta}_2(n-1) + \mathbf{x}\boldsymbol{\zeta}_3(n-1)\boldsymbol{\mu}'
\end{aligned}$$

Then

$$\begin{aligned}
CRE_k(\mathbf{G}_R) &= E(y_i^R - y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{e}_k) \\
&= \begin{pmatrix} \zeta_0(n-1) + \mathbf{e}_k\boldsymbol{\zeta}_1(n-1) + E(\bar{\mathbf{x}}_i^R|\mathbf{x}_i = \mathbf{e}_k)\boldsymbol{\zeta}_2(n-1) \\ + \mathbf{e}_k\boldsymbol{\zeta}_3(n-1)E(\bar{\mathbf{x}}_i^R|\mathbf{x}_i = \mathbf{e}_k)' \end{pmatrix} \\
&\quad - \begin{pmatrix} \zeta_0(n-1) + \mathbf{e}_k\boldsymbol{\zeta}_1(n-1) + \boldsymbol{\mu}\boldsymbol{\zeta}_2(n-1) \\ + \mathbf{e}_k\boldsymbol{\zeta}_3(n-1)\boldsymbol{\mu}' \end{pmatrix} \\
&\hspace{6cm} \text{(by (126))} \\
&= E\left(\bar{\mathbf{x}}_i^R|\mathbf{x}_i = \mathbf{e}_k) - \boldsymbol{\mu}\right)\boldsymbol{\zeta}_2(n-1) + \mathbf{e}_k\boldsymbol{\zeta}_3(n-1)E\left(\bar{\mathbf{x}}_i^R|\mathbf{x}_i = \mathbf{e}_k) - \boldsymbol{\mu}\right)' \\
&= (n-1)\sum_{\ell=1}^{K}\left(E(\bar{x}_{i\ell}^R|\mathbf{x}_i = \mathbf{e}_k) - \mu_\ell\right)(\zeta_{2\ell} + \zeta_{3k\ell}) \\
&= (n-1)\sum_{\ell=1}^{K}\Delta\bar{x}_{k\ell}(\mathbf{G}_R)CPE_{k\ell}
\end{aligned}$$

which is the result in (66). The result in (65) follows by applying the law of total probability to (66).

3. Given (PS, OS), Part 2 of Proposition 2 applies. By equation (28) in Proposition 2, $CPE_{k\ell} = APE_\ell$ and so result (69) follows from (66) by substitution. Result (68) follows from the fact that individual and peer characteristics have the same

expected value in any feasible reallocation:

$$\begin{aligned}
\sum_{k=0}^{K} \Delta\bar{x}_{k\ell}(\mathbf{G}_R)\mu_k &= \sum_{k=0}^{K} \left(E(\bar{x}^R_{i\ell}|\mathbf{x}_i = \mathbf{e}_k) - \mu_\ell\right)\mu_k && \text{(definition of } \Delta\bar{x}\text{)} \\
&= \sum_{k=0}^{K} E(\bar{x}^R_{i\ell}|\mathbf{x}_i = \mathbf{e}_k)\Pr(\mathbf{x}_i = \mathbf{e}_k) - E(x_{i\ell})\sum_{k=0}^{K}\Pr(\mathbf{x}_i = \mathbf{e}_k) \\
& && \text{(definition of } \boldsymbol{\mu}\text{)} \\
&= E(\bar{x}^R_{i\ell}) - E(x_{i\ell}) && \text{(law of total probability)} \\
&= 0 && (127)
\end{aligned}$$

Therefore:

$$\begin{aligned}
ARE(\mathbf{G}_R) &= (n-1)\sum_{k=0}^{K}\sum_{\ell=1}^{K}\mu_k\Delta\bar{x}_{k\ell}(\mathbf{G}_R)CPE_{k\ell} && \text{(by (65))} \\
&= (n-1)\sum_{k=0}^{K}\sum_{\ell=1}^{K}\mu_k\Delta\bar{x}_{k\ell}(\mathbf{G}_R)APE_\ell && \text{(by (28) in Proposition 2)} \\
&= (n-1)\sum_{\ell=1}^{K}APE_\ell\underbrace{\sum_{k=0}^{K}\Delta\bar{x}_{k\ell}(\mathbf{G}_R)\mu_k}_{=0 \text{ by } (127)} \\
&= 0
\end{aligned}$$

which is the result in (68).

Proof for Proposition 8

Given (OS,PS), Part 2 of Proposition 2 applies, and we can express each outcome y_i as a sum of own effects OE_i and peer effects PE_j. We can also show that:

$$E\left(PE_{ij}\middle|\begin{array}{l}\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell,\\ c_i = c, c_j = c'\end{array}\right) = E\left(\frac{OE_i}{n-1} + PE_j\middle|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c, c_j = c'\right) \tag{128}$$

$$= \frac{E(OE_i|\mathbf{x}_i = \mathbf{e}_k, c_i = c)}{n-1} + E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell, c_j = c') \tag{129}$$

$$E\left(PE_{ij}\middle|\begin{array}{l}\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell,\\ c_i = c\end{array}\right) = E\left(\frac{OE_i}{n-1} + PE_j\middle|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c\right) \tag{130}$$

$$= \frac{E(OE_i|\mathbf{x}_i = \mathbf{e}_k, c_i = c)}{n-1} + E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) \tag{131}$$

Let $\boldsymbol{\zeta} \equiv (\zeta_0, \boldsymbol{\zeta}_1)$ such that:

$$E(PE_j|\mathbf{x}_j) = \zeta_0 + \mathbf{x}_j\boldsymbol{\zeta}_1 \tag{132}$$

and let $\boldsymbol{\eta}^c \equiv (\eta_0^c, \boldsymbol{\eta}_1^c)$ and $\boldsymbol{\eta} \equiv (\eta_0, \boldsymbol{\eta}_1)$ such that:

$$E(OE_i|\mathbf{x}_i, c_i = c) = \eta_0^c + \mathbf{x}_i\boldsymbol{\eta}_1^c \tag{133}$$

$$E(OE_i|\mathbf{x}_i) = \eta_0 + \mathbf{x}_i\boldsymbol{\eta}_1 \tag{134}$$

Substitiuting (129) and (131) into the definition of cluster invariance (CI) produces the implication that:

$$E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell, c_j = c') = E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) \tag{135}$$

$$= \zeta_0 + \mathbf{x}_j\boldsymbol{\zeta}_1 \tag{136}$$

Constant shifts (CS) implies that $\boldsymbol{\eta}_1^c$ does not vary across group clusters:

$$\begin{aligned}\eta_{1\ell}^c &= E(OE_i|\mathbf{x}_i = \mathbf{e}_\ell, c_i = c) - E(OE_i|\mathbf{x}_i = \mathbf{e}_0, c_i = c)\\ &= E(OE_i|\mathbf{x}_i = \mathbf{e}_\ell) - E(OE_i|\mathbf{x}_i = \mathbf{e}_0) \qquad \text{(by CS)}\\ &= \eta_{1\ell}\end{aligned}$$

implying that:

$$E(OE_i|\mathbf{x}_i, c_i = c) = \eta_0^c + \mathbf{x}_i\boldsymbol{\eta}_1 \tag{137}$$

Substituting these results into equation (27) from Part 2 of Proposition 2:

$$\begin{aligned}
E(y_i|\mathbf{X},\mathbf{G},\mathbf{C}) &= E\left(OE_i + \sum_{j\in\mathbf{p}_i} PE_j \middle| \mathbf{X},\mathbf{G},\mathbf{C}\right) && \text{(by (27))}\\
&= E\left(OE_i\,|\mathbf{X},\mathbf{G},\mathbf{C}\right) + \sum_{j\in\mathbf{p}_i} E\left(PE_j\,|\mathbf{X},\mathbf{G},\mathbf{C}\right)\\
&= E\left(OE_i\,|\mathbf{x}_i, c_i\right) + \sum_{j\in\mathbf{p}_i} E\left(PE_j\,|\mathbf{x}_j, c_j\right) && \text{(by RAWC)}\\
&= E\left(OE_i\,|\mathbf{x}_i, c_i\right) + \sum_{j\in\mathbf{p}_i} E\left(PE_j\,|\mathbf{x}_j\right) && \text{(by CI)}\\
&= \eta_0^{c_i} + \mathbf{x}_i\boldsymbol{\eta}_1 + \sum_{j\in\mathbf{p}_i}(\zeta_0 + \mathbf{x}_j\boldsymbol{\zeta}_1) && \text{(by results above)}\\
&= (\eta_0^{c_i} + \zeta_0(n-1)) + \mathbf{x}_i\boldsymbol{\eta}_1 + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_1(n-1) && (138)
\end{aligned}$$

Applying the law of iterated expectations:

$$\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}, c_i = c) &= E(E(y_i|\mathbf{X},\mathbf{G},\mathbf{C})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}, c_i = c) && \text{(by LIE)}\\
&= \underbrace{(\eta_0^c + \zeta_0(n-1))}_{\alpha_0^c} + \mathbf{x}\underbrace{\boldsymbol{\eta}_1}_{\boldsymbol{\alpha}_1} + \bar{\mathbf{x}}\underbrace{\boldsymbol{\zeta}_1(n-1)}_{\boldsymbol{\alpha}_2} && (139)
\end{aligned}$$

Finally, result (28) in Part 2 of Proposition 2 implies:

$$\begin{aligned}
CPE_{k\ell} = APE_\ell &= E(PE_i|\mathbf{x}_i = \mathbf{e}_\ell) - E(PE_i|\mathbf{x}_i = \mathbf{e}_0) && \text{(by (28))}\\
&= \zeta_{1\ell} && \text{(by (132))}\\
&= \alpha_{2\ell}/(n-1) && \text{(by (139))}
\end{aligned}$$

which is the result in (77).

Proof for Proposition 9

1. The proof here is essentially the same as the proof for part two of Proposition 4, but conditioning on c_i. Given (PS), Proposition 2 implies the potential outcome function can be written as in equation (24) and within-cluster peer effects can be

written in terms of PE_{ij}:

$$\begin{aligned} CPE^c_{k\ell} &= E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c) \\ &\quad - E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c) \\ APE^c_\ell &= E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c) - E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c) \end{aligned} \tag{140}$$

Without loss of generality, let $\boldsymbol{\zeta}^{cc'} \equiv (\zeta_0^{cc'}, \boldsymbol{\zeta}_1^{cc'}, \boldsymbol{\zeta}_2^{cc'}, \boldsymbol{\zeta}_3^{cc'})$ satisfy:

$$E(PE_{ij}|\mathbf{x}_i, \mathbf{x}_j, c_i = c, c_j = c') = \zeta_0^{cc'} + \mathbf{x}_i\boldsymbol{\zeta}_1^{cc'} + \mathbf{x}_j\boldsymbol{\zeta}_2^{cc'} + \mathbf{x}_i\boldsymbol{\zeta}_3^{cc'}\mathbf{x}_j' \tag{141}$$

These two results can be combined to find $CPE^c_{k\ell}$ in terms of ζ^{cc}:

$$\begin{aligned} CPE^c_{k\ell} &= E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c) && \text{(by (140))} \\ &\quad - E(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c) \\ &= \left(\zeta_0^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_1^{cc} + \mathbf{e}_\ell\boldsymbol{\zeta}_2^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_3^{cc}\mathbf{e}_\ell'\right) && \text{(by (141))} \\ &\quad - \left(\zeta_0^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_1^{cc} + \mathbf{e}_0\boldsymbol{\zeta}_2^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_3^{cc}\mathbf{e}_0'\right) \\ &= \left(\zeta_0^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_1^{cc} + \mathbf{e}_\ell\boldsymbol{\zeta}_2^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_3^{cc}\mathbf{e}_\ell'\right) - \left(\zeta_0^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_1^{cc}\right) && \text{(since } \mathbf{e}_0 = 0\text{)} \\ &= \mathbf{e}_\ell\boldsymbol{\zeta}_2^{cc} + \mathbf{e}_k\boldsymbol{\zeta}_3^{cc}\mathbf{e}_\ell' \\ &= \zeta_{2\ell}^{cc} + \zeta_{3k\ell}^{cc} && (142) \end{aligned}$$

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

$$
\begin{aligned}
E(y_i|\mathbf{X},\mathbf{G},\mathbf{C}) &= E\left(\sum_{j\in\mathbf{p}_i} PE_{ij} \middle| \mathbf{X},\mathbf{G},\mathbf{C}\right) && \text{(by Proposition 2)} \\
&= E\left(\sum_{j=1}^{I} PE_{ij}\mathbb{I}\left(j\in\mathbf{p}_i\right) \middle| \mathbf{X},\mathbf{G},\mathbf{C}\right) \\
& && \text{(where } \mathbb{I}\left(\right) \text{ is the indicator function)} \\
&= \sum_{j=1}^{I} E(PE_{ij}\mathbb{I}\left(j\in\mathbf{p}_i\right)|\mathbf{X},\mathbf{G},\mathbf{C}) \\
&= \sum_{j=1}^{I} E(PE_{ij}|\mathbf{X},\mathbf{G},\mathbf{C})\mathbb{I}\left(j\in\mathbf{p}_i\right) \\
& && \text{(since } \mathbb{I}\left(j\in\mathbf{p}_i\right) \text{ is a function of } \mathbf{G}) \\
&= \sum_{j\in\mathbf{p}_i} E(PE_{ij}|\mathbf{X},\mathbf{G},\mathbf{C}) \\
&= \sum_{j\in\mathbf{p}_i} E(PE_{ij}|\mathbf{X},\mathbf{C}) && (\text{RAWC} \implies \mathbf{T}\perp\!\!\!\perp\mathbf{G}|\mathbf{C}) \\
&= \sum_{j\in\mathbf{p}_i} E(PE_{ij}|\mathbf{x}_i,\mathbf{x}_j,c_i,c_j=c_i) && \text{(by (75))} \\
&= \sum_{j\in\mathbf{p}_i} \left(\zeta_0^{c_ic_i} + \mathbf{x}_i\boldsymbol{\zeta}_1^{c_ic_i} + \mathbf{x}_j\boldsymbol{\zeta}_2^{c_ic_i} + \mathbf{x}_i\boldsymbol{\zeta}_3^{c_ic_i}\mathbf{x}_j'\right) && \text{(by (141))} \\
&= \zeta_0^{c_ic_i}(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1^{c_ic_i}(n-1) + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2^{c_ic_i}(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3^{c_ic_i}(n-1)\bar{\mathbf{x}}_i'
\end{aligned} \tag{143}
$$

Applying the law of iterated projections:

$$
\begin{aligned}
L^c\left(y_i \middle| \begin{matrix}\mathbf{x}_i,\bar{\mathbf{x}}_i,\\ \mathbf{x}_i'\bar{\mathbf{x}}_i\end{matrix}\right) &= L^c\left(E\left(y_i|\,\mathbf{X},\mathbf{G},\mathbf{C}\right) \middle| \begin{matrix}\mathbf{x}_i,\bar{\mathbf{x}}_i,\\ \mathbf{x}_i'\bar{\mathbf{x}}_i\end{matrix}\right) && \text{(law of iterated projections)} \\
&= L^c\left(\begin{matrix}\zeta_0^{c_i}(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1^{c_ic_i}(n-1) \\ + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2^{c_ic_i}(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3^{c_ic_i}(n-1)\bar{\mathbf{x}}_i'\end{matrix} \middle| \begin{matrix}\mathbf{x}_i,\bar{\mathbf{x}}_i,\\ \mathbf{x}_i'\bar{\mathbf{x}}_i\end{matrix}\right) && \text{(by (143))} \\
&= \underbrace{\zeta_0^{cc}(n-1)}_{\beta_0^c} + \mathbf{x}_i\underbrace{\boldsymbol{\zeta}_1^{cc}(n-1)}_{\boldsymbol{\beta}_1^c} + \bar{\mathbf{x}}_i\underbrace{\boldsymbol{\zeta}_2^{cc}(n-1)}_{\boldsymbol{\beta}_2^c} + \mathbf{x}_i\underbrace{\boldsymbol{\zeta}_3^{cc}(n-1)}_{\boldsymbol{\beta}_3^c}\bar{\mathbf{x}}_i'
\end{aligned} \tag{144}
$$

So $\boldsymbol{\beta}_2^c = \boldsymbol{\zeta}_2^{cc}(n-1)$, $\boldsymbol{\beta}_3^c = \boldsymbol{\zeta}_3^{cc}(n-1)$, and:

$$\begin{aligned} CPE_{k\ell}^c &= \zeta_{2\ell}^{cc} + \zeta_{3k\ell}^{cc} && \text{(by (142))} \\ &= \frac{\beta_{2\ell}^c + \beta_{3k\ell}^c}{n-1} && \text{(by (144))} \end{aligned}$$

which is the result in (82). The same procedure can be used to derive the result in (83). First, express APE_ℓ^c in terms of $\boldsymbol{\zeta}^{cc}$:

$$\begin{aligned} APE_\ell^c &= E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c) && \text{(by (140))} \\ &\quad - E(PE_{ij}|\mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c) \\ &= E\left(E\left(PE_{ij}|\,\mathbf{x}_i, \mathbf{x}_j, c_i = c_j = c\right)|\,\mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c\right) \\ &\quad - E\left(E\left(PE_{ij}|\,\mathbf{x}_i, \mathbf{x}_j, c_i = c_j = c\right)|\,\mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c\right) \\ &&& \text{(law of iterated expectations)} \\ &= E\left(\zeta_0^{cc} + \mathbf{x}_i\boldsymbol{\zeta}_1^{cc} + \mathbf{x}_j\boldsymbol{\zeta}_2^{cc} + \mathbf{x}_i\boldsymbol{\zeta}_3^{cc}\mathbf{x}_j' \,\middle|\, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c\right) && \text{(by 141))} \\ &\quad - E\left(\zeta_0^{cc} + \mathbf{x}_i\boldsymbol{\zeta}_1^{cc} + \mathbf{x}_j\boldsymbol{\zeta}_2^{cc} + \mathbf{x}_i\boldsymbol{\zeta}_3^{cc}\mathbf{x}_j' \,\middle|\, \mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c\right) \\ &= \zeta_0^{cc} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c)\boldsymbol{\zeta}_1^{cc} && \text{(since } \mathbf{e}_0 = \mathbf{0}) \\ &\quad + \mathbf{e}_\ell\boldsymbol{\zeta}_2^{cc} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c)\boldsymbol{\zeta}_3^{cc}\mathbf{e}_\ell' \\ &\quad - (\zeta_0^{cc} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c)\boldsymbol{\zeta}_1^{cc}) \\ &= \zeta_0^{cc} + E(\mathbf{x}_i|c_i = c_j = c)\boldsymbol{\zeta}_1^{cc} && ((75) \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j|\mathbf{C}) \\ &\quad + \mathbf{e}_\ell\boldsymbol{\zeta}_2^{cc} + E(\mathbf{x}_i|c_i = c_j = c)\boldsymbol{\zeta}_3^{cc}\mathbf{e}_\ell' \\ &\quad - (\zeta_0^{cc} + E(\mathbf{x}_i|c_i = c_j = c)\boldsymbol{\zeta}_1^{cc}) \\ &= \mathbf{e}_\ell\boldsymbol{\zeta}_2^{cc} + E(\mathbf{x}_i|c_i = c)\boldsymbol{\zeta}_3^{cc}\mathbf{e}_\ell' \\ &= \mathbf{e}_\ell\left(\boldsymbol{\zeta}_2^{cc} + (\boldsymbol{\zeta}_3^{cc})'E(\mathbf{x}_i'|c_i = c)\right) && (145) \end{aligned}$$

Then find the relationship between $\boldsymbol{\alpha}^c$ and $\boldsymbol{\zeta}^{cc}$ by expressing $L^c(y_i|\bar{\mathbf{x}}_i)$ in terms of $\boldsymbol{\alpha}^c$ and in terms of $\boldsymbol{\zeta}^{cc}$:

$$\begin{aligned}
L^c(y_i|\bar{\mathbf{x}}_i) &= L^c\left(L^c\left(y_i|\,\mathbf{x}_i, \bar{\mathbf{x}}_i\right)|\,\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)}\\
&= L^c\left(\alpha_0^c + \mathbf{x}_i\boldsymbol{\alpha}_1^c + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2^c|\,\bar{\mathbf{x}}_i\right) && \text{(by (84))}\\
&= \alpha_0^c + L^c(\mathbf{x}_i|\bar{\mathbf{x}}_i)\boldsymbol{\alpha}_1^c + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2^c\\
&= \alpha_0^c + E(\mathbf{x}_i|c_i = c)\boldsymbol{\alpha}_1^c + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2^c && \text{(by (75))}\\
&= (\alpha_0^c + E(\mathbf{x}_i|c_i = c)\boldsymbol{\alpha}_1^c) + \bar{\mathbf{x}}_i\boldsymbol{\alpha}_2^c && (146)\\
L^c(y_i|\bar{\mathbf{x}}_i) &= L^c\left(E\left(y_i|\,\mathbf{X}, \mathbf{G}, \mathbf{C}\right)|\,\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)}\\
&= L^c\left(\begin{matrix}\zeta_0^{c_i}(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_1^{c_ic_i}(n-1)\\ +\,\bar{\mathbf{x}}_i\boldsymbol{\zeta}_2^{c_ic_i}(n-1) + \mathbf{x}_i\boldsymbol{\zeta}_3^{c_ic_i}(n-1)\bar{\mathbf{x}}_i'\end{matrix}\middle|\,\bar{\mathbf{x}}_i\right) && \text{(by (144))}\\
&= \zeta_0^c(n-1) + L^c(\mathbf{x}_i|\bar{\mathbf{x}}_i)\boldsymbol{\zeta}_1^{cc}(n-1)\\
&\quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2^{cc}(n-1) + L^c(\mathbf{x}_i|\bar{\mathbf{x}}_i)\boldsymbol{\zeta}_3^{cc}(n-1)\bar{\mathbf{x}}_i'\\
&= \zeta_0^c(n-1) + E(\mathbf{x}_i|c_i = c)\boldsymbol{\zeta}_1^{cc}(n-1) && \text{(by (75))}\\
&\quad + \bar{\mathbf{x}}_i\boldsymbol{\zeta}_2^{cc}(n-1) + E(\mathbf{x}_i|c_i = c)\boldsymbol{\zeta}_3^{cc}(n-1)\bar{\mathbf{x}}_i'\\
&= \underbrace{\zeta_0^c(n-1) + E(\mathbf{x}_i|c_i = c)\boldsymbol{\zeta}_1^{cc}(n-1)}_{\alpha_0^c + E(\mathbf{x}_i|c_i = c)\boldsymbol{\alpha}_1^c} && (147)\\
&\quad + \bar{\mathbf{x}}_i\underbrace{\left(\boldsymbol{\zeta}_2^{cc}(n-1) + (\boldsymbol{\zeta}_3^{cc})'E(\mathbf{x}_i'|c_i = c)(n-1)\right)}_{\boldsymbol{\alpha}_2^c}
\end{aligned}$$

So $\boldsymbol{\alpha}_2^c = (\boldsymbol{\zeta}_2^{cc}(n-1) + (\boldsymbol{\zeta}_3^{cc})'E(\mathbf{x}_i'|c_i = c)(n-1))$ and:

$$\begin{aligned}
APE_k^c &= \mathbf{e}_\ell(\boldsymbol{\zeta}_2^{cc} + (\boldsymbol{\zeta}_3^{cc})'E(\mathbf{x}_i'|c_i = c)) && \text{(by (145))}\\
&= \mathbf{e}_\ell\frac{\boldsymbol{\alpha}_2^c}{n-1} && \text{(by (146) and (147))}\\
&= \frac{\alpha_{2\ell}^c}{n-1}
\end{aligned}$$

which is the result in (83).

2. Given peer separability (PS), Proposition 2 implies the potential outcome function can be written as in equation (24) and average and conditional peer effects can be expressed in terms of the latent variable PE_{ij}. Cluster invariance (CI) implies

that:

$$
\begin{aligned}
E\left(PE_{ij}\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{x}_j=\mathbf{e}_\ell\end{matrix}\right) &= E\left(E\left(PE_{ij}\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,c_i,c_j,\\ \mathbf{x}_j=\mathbf{e}_\ell\end{matrix}\right)\middle|\begin{matrix}\mathbf{x}_i=\mathbf{e}_k,\\ \mathbf{x}_j=\mathbf{e}_\ell\end{matrix}\right) && \text{(by LIE)}\\
&= \sum_{c=1}^{C}\sum_{c'=1}^{C}\begin{pmatrix}E\left(PE_{ij}\left|\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,c_i=c,c_j=c'\right.\right)\\ \times\Pr(c_i=c,c_j=c'|\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell)\end{pmatrix}\\
& && \text{(by LTP)}\\
&= \sum_{c=1}^{C}\sum_{c'=1}^{C}\begin{pmatrix}E\left(PE_{ij}\left|\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,c_i=c,c_j=c'\right.\right)\\ \times\Pr(c_i=c|\mathbf{x}_i=\mathbf{e}_k)\\ \times\Pr(c_j=c'|\mathbf{x}_j=\mathbf{e}_\ell)\end{pmatrix}\\
& && \text{(since } i \perp\!\!\!\perp j\text{)}\\
&= \sum_{c=1}^{C}\sum_{c'=1}^{C}\begin{pmatrix}E\left(PE_{ij}\left|\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,c_i=c\right.\right)\\ \times\Pr(c_i=c|\mathbf{x}_i=\mathbf{e}_k)\\ \times\Pr(c_j=c'|\mathbf{x}_j=\mathbf{e}_\ell)\end{pmatrix} && \text{(by (CI))}\\
&= \sum_{c=1}^{C}\begin{pmatrix}E\left(PE_{ij}\left|\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,c_i=c\right.\right)\Pr(c_i=c|\mathbf{x}_i=\mathbf{e}_k)\\ \times\underbrace{\sum_{c'=1}^{C}\Pr(c_j=c'|\mathbf{x}_j=\mathbf{e}_\ell)}_{=1}\end{pmatrix}\\
&= \sum_{c=1}^{C}E\left(PE_{ij}\left|\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,c_i=c\right.\right)\Pr(c_i=c|\mathbf{x}_i=\mathbf{e}_k)\\
& && (148)
\end{aligned}
$$

We can then substitute this result into result (25) in Proposition 2 to get:

$$
\begin{aligned}
CPE_{k\ell} &= E\left(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell\right) - E\left(PE_{ij}|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0\right) && \text{(by (25))} \\
&= \sum_{c=1}^{C} E\left(PE_{ij}\left|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c\right.\right) \Pr(c_i = c|\mathbf{x}_i = \mathbf{e}_k) && \text{(by (148))} \\
&\quad - \sum_{c=1}^{C} E\left(PE_{ij}\left|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0, c_i = c\right.\right) \Pr(c_i = c|\mathbf{x}_i = \mathbf{e}_k) \\
&= \sum_{c=1}^{C} \begin{pmatrix} E\left(PE_{ij}\left|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c\right.\right) \\ - E\left(PE_{ij}\left|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0, c_i = c\right.\right) \end{pmatrix} \Pr(c_i = c|\mathbf{x}_i = \mathbf{e}_k) \\
&= \sum_{c=1}^{C} \begin{pmatrix} E\left(PE_{ij}\left|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_\ell, c_i = c_j = c\right.\right) \\ - E\left(PE_{ij}\left|\mathbf{x}_i = \mathbf{e}_k, \mathbf{x}_j = \mathbf{e}_0, c_i = c_j = c\right.\right) \end{pmatrix} \Pr(c_i = c|\mathbf{x}_i = \mathbf{e}_k) \\
& && \text{(by (CI))} \\
&= \sum_{c=1}^{C} CPE_{k\ell}^{c} \Pr(c_i = c|\mathbf{x}_i = \mathbf{e}_k) && \text{(by (140))} \\
&= E(CPE_{k\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_k) \\
&= \frac{E(\beta_{2\ell}^{c_i} + \beta_{3k\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_k)}{n-1} && \text{(by (82))}
\end{aligned}
$$

which is the result in (86). Result (87) follows from substituition of (86) into (18).

3. Given (OS,PS), Part 2 of Proposition 2 applies, and we can express each outcome y_i as a sum of own effects OE_i and peer effects PE_j. Equations (129) and (131) in the proof for Proposition 8 apply here as well, and we can substitute these results into the definition of partial cluster invariance (PCI) to produce the implication that:

$$E(PE_j|\mathbf{x}_j = \mathbf{e}_0, c_j = c') = E(PE_j|\mathbf{x}_j = \mathbf{e}_0) \tag{149}$$

For any ℓ, the law of total probability implies that:

$$E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) = \sum_{c=1}^{C} E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell, c_j = c) \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) \tag{150}$$

In addition, we can derive:

$$
\begin{aligned}
E(PE_j|\mathbf{x}_j = \mathbf{e}_0) &= E(PE_j|\mathbf{x}_j = \mathbf{e}_0) \underbrace{\sum_{c=1}^{C} \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell)}_{=1} \\
&= \sum_{c=1}^{C} E(PE_j|\mathbf{x}_j = \mathbf{e}_0) \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) \qquad (151)
\end{aligned}
$$

Substituting in, we get:

$$
\begin{aligned}
APE_\ell &= E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell) - E(PE_j|\mathbf{x}_j = \mathbf{e}_0) && \text{(by (28))} \\
&= \sum_{c=1}^{C} E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell, c_j = c) \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) && \text{(by (150, 151))} \\
&\quad - \sum_{c=1}^{C} E(PE_j|\mathbf{x}_j = \mathbf{e}_0) \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) \\
&= \sum_{c=1}^{C} E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell, c_j = c) \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) && \text{(by (149))} \\
&\quad - \sum_{c=1}^{C} E(PE_j|\mathbf{x}_j = \mathbf{e}_0, c_j = c) \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) \\
&= \sum_{c=1}^{C} \begin{pmatrix} E(PE_j|\mathbf{x}_j = \mathbf{e}_\ell, c_j = c) \\ - E(PE_j|\mathbf{x}_j = \mathbf{e}_0, c_j = c) \end{pmatrix} \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) \\
&= \sum_{c=1}^{C} APE_\ell^c \Pr(c_j = c|\mathbf{x}_j = \mathbf{e}_\ell) && \text{(by (28, 140))} \\
&= E(APE_\ell^{c_i}|\mathbf{x}_i = \mathbf{e}_\ell) \\
&= \frac{E(\alpha_{2\ell}^{c_i}|\mathbf{x}_i = \mathbf{e}_\ell)}{n-1} && \text{(by (83))}
\end{aligned}
$$

which is the result in (88).

Proof for Proposition 10

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

$$
E\left(h(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\mathbf{p}}) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}) = \bar{\mathbf{x}} \right) = h(\mathbf{x}, \{\mathbf{x}^1, \dots, \mathbf{x}^{n-1}\}) \qquad (152)
$$

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

$$(\mathbf{x}_i^*, u_i) \perp\!\!\!\perp \left\{\mathbf{x}_j^*\right\}_{j\in\tilde{\mathbf{p}}} \tag{153}$$

which implies:

$$\begin{aligned} E(u_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) &= E(u_i|\mathbf{x}_i = \mathbf{x}) && \text{(by (153))} \\ &= 0 && (154) \end{aligned}$$

By (CRA), Lemma 1 holds and therefore:

$$\begin{aligned} E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) && \text{(by (39) in Lemma 1)} \\ &= E\left(h\left(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\tilde{\mathbf{p}}}\right) + u_i \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}\right) && \text{(by DCE)} \\ &= E\left(h\left(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\tilde{\mathbf{p}}}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}\right) \\ &\quad + E\left(u_i \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}\right) \\ &= E\left(h\left(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\tilde{\mathbf{p}}}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}\right) && \text{(by (154))} \\ &= h\left(\mathbf{x}, \left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}\right) && \text{(by (152))} \end{aligned}$$

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. (PS) implies that result (24) in Proposition 2 holds, where:

$$\begin{aligned} PE_{ij} &= y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - \frac{n-2}{n-1} y(\tau_i, \{1, 1, 1, \ldots, 1\}) && \text{(by (97))} \\ &= h(\mathbf{x}_i, \{\mathbf{x}_j, \mathbf{x}(1), \ldots, \mathbf{x}(1)\}) + u_i - \frac{n-2}{n-1}\left(h(\mathbf{x}_i, \{\mathbf{x}(1), \mathbf{x}(1), \ldots, \mathbf{x}(1)\}) + u_i\right) \\ & && \text{(by DCE)} \\ &= \underbrace{h(\mathbf{x}_i, \{\mathbf{x}_j, \mathbf{x}(1), \ldots, \mathbf{x}(1)\}) - \frac{n-2}{n-1} h(\mathbf{x}_i, \{\mathbf{x}(1), \mathbf{x}(1), \ldots, \mathbf{x}(1)\})}_{\equiv h_2(\mathbf{x}_i, \mathbf{x}_j)} + \frac{u_i}{n-1} \end{aligned} \tag{155}$$

It follows by substitution that:

$$
\begin{aligned}
E\left(PE_{ij}\,\middle|\,\begin{array}{l}\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,\\ c_i=c,c_j=c'\end{array}\right) &= E\left(h_2(\mathbf{x}_i,\mathbf{x}_j)+\frac{u_i}{n-1}\,\middle|\,\begin{array}{l}\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,\\ c_i=c,c_j=c'\end{array}\right)\\
&= h_2(\mathbf{e}_k,\mathbf{e}_\ell)+\frac{E(u_i|\mathbf{x}_i=\mathbf{e}_k,c_i=c)}{n-1}\quad (\text{since } i \perp\!\!\!\perp j)\\
E\left(PE_{ij}\,\middle|\,\begin{array}{l}\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,\\ c_i=c\end{array}\right) &= E\left(h_2(\mathbf{x}_i,\mathbf{x}_j)+\frac{u_i}{n-1}\,\middle|\,\begin{array}{l}\mathbf{x}_i=\mathbf{e}_k,\mathbf{x}_j=\mathbf{e}_\ell,\\ c_i=c\end{array}\right)\\
&= h_2(\mathbf{e}_k,\mathbf{e}_\ell)+\frac{E(u_i|\mathbf{x}_i=\mathbf{e}_k,c_i=c)}{n-1}\quad (\text{since } i \perp\!\!\!\perp j)
\end{aligned}
$$

which implies condition (CI). Therefore, Part 2 of Proposition 9 applies, and results (86) and (87) in that proposition imply results (90) and (91) here.