Peers as treatments[*]

Brian Krauth
Simon Fraser University

August 25, 2020

Abstract

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

1 Introduction

Empirical research on social interaction effects aims to measure the impact of peers or some other reference group on an individual's behavior, choices or related outcomes. Much of the research is based on a behavioral model, generally associated with[1] Manski (1993), in which the individual outcome responds directly to both the observed behavior (endogenous effects) and the observed characteristics (exogenous or contextual effects) of peers. Manski's formulation has inspired a large literature developing econometric 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 prominence in empirical research. Existing theory provides little guidance as to which background characteristics of peers should be included in the model, so researchers will typically include the peer group averages of whatever variables are available and potentially influential. The ad hoc nature of these specifications leads to difficulties both in comparing results across studies and in interpreting the results of a given study (Fruehwirth, 2014). Many of these difficulties are a byproduct of interpreting contextual effects as the direct effect of peer characteristics on one's outcomes. That is, researchers often analyze and discuss the data as if they were trying to measure the direct effect of the race, gender, ability, parental education, or family income of peers. As will be shown below, this framework imposes strong data requirements and identifying assumptions, both for estimating the effects themselves and in using empirical results to make useful counterfactual predictions.

This paper describes an alternative formulation in which people are influenced not by the observed characteristics of their peers but by the peers themselves. In other words, each person

[*]Contact email: bkrauth@sfu.ca. Revised versions available at http://www.sfu.ca/∼bkrauth/research.htm

[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 realized values. Blume et al. (2011, p. 891-892) discuss this distinction and some of its implications.

has a direct, individual-specific influence (analogous to a treatment effect) on the choices of his or her peers. This influence may be statistically related to observed background characteristics, but is conceptually distinct from them. This framework allows for unobserved heterogeneity in peer influence and so is more flexible than the traditional treatment of contextual effects, which can be interpreted as a special case in this model. Analysis within this framework suggests estimands for peer effects that are well-defined and identified in a wide variety of settings. The peer effects defined in this paper can usually be estimated without elaborate or novel econometric procedures.

As will be shown below, many aspects of conventional practice are entirely consistent with the peers-as-treatments model. Simple linear regression models provide useful information about peer effect heterogeneity in a wide variety of settings, and different researchers can use the same data to explore different aspects of peer effect heterogeneity without needing to converge on a common model specification that includes all relevant dimensions of this heterogeneity. Peer effects can be identified using any of several sources of randomness in peer group formation, including simple random assignment, random assignment based on observables, and random cohorts or subgroups within larger groups that are not randomly assigned. At the same time, the peers-as-treatments model implies several areas in which empirical practice can be improved. First, simpler specifications with a few binary or categorical peer variables are typically more robustly informative than those with many variables. Second, the precise source of identifying randomness in peer group formation has subtle but important implications for identification, estimation and interpretation that are typically obscured in commonly-used parametric models that ignore unobserved peer heterogeneity.

Many of the results here have been previously noted informally in various applied papers, and are well-understood by experienced practitioners. However, there is a clear benefit to expressing them more precisely in the context of a general model.

1.1 Related literature

The contemporary economics literature on measuring social effects has been primarily aimed at addressing the challenges described by Manski (1993). Manski's analysis emphasizes two related but distinct identification problems in previous research: distinguishing true social effects from spurious social effects due to nonrandom peer selection or unobserved common shocks, and distinguishing endogenous social effects from contextual social effects. Manski shows nonrandom peer selection makes it difficult to identify true social effects at all. He also shows that when social effects take a "linear-in-means" form and social groups are very large, contextual and endogneous effects cannot be distinguished. Subsequent empirical research has addressed the selection problem by exploiting natural experiments in which peer group assignment is plausibly exogenous conditional on observable factors. Methodological research on distinguishing endogenous from contextual effects has generally worked by exploiting nonlinearity (Brock and Durlauf, 2000), exclusion restrictions (Gaviria and Raphael, 2001), or aspects of network structure (Bramoullé et al., 2009). More recent research addresses new issues outside of Manski's original framework in several ways.

One relevant branch of this research addresses unobserved heterogeneity in the effect of an individual on his or her peers, as is done in the current paper. Graham (2008) shows how endogeneous effects can be inferred in the presence of unobserved heterogeneity by comparing outcome variances across randomly-assigned groups of varying size. Fruehwirth (2014) emphasizes that many outcomes in the applied literature (e.g., test scores) represent outputs of a joint production process unlike Manski's framework in which the outcomes are individual choices. In her model, unobserved peer effort/input enters into that production function and peer outcome is treated as a proxy for unobserved peer input rather than as having a direct effect. Arcidiacono et al. (2012), Burke and Sass (2013), and Isphording and Zölitz (2020) use panel data with repeated observations of the same individuals in multiple peer groups to estimate each individual's unobserved peer input.

Another relevant branch addresses the use of estimated peer effect models to predict the consequences of counterfactual allocations of individuals to peer groups. Bhattacharya (2009) develops algorithms to find optimal assignments, while Graham et al. (2010) predict the effect of local reallocation such as a small reduction in segregation. Carrell et al. (2013) report the results of a field experiment that provides a cautionary tale on the risks of using reduced-form models for this purpose. In their study, the "optimal" allocation actually performs worse than random assignment when implemented in the field.

The approach in this paper is most similar to that in Graham et al. (2010), which also uses a potential outcomes framework in which the ultimate estimand of interest is the effect of some counterfactual reallocation. An important difference is that Graham et al. (2010) propose plug-in estimators based on nonparametric kernel regressions that impose limited restrictions but require a great deal of data to estimate. In contrast, the emphasis in this paper is on clarifying conditions under which commonly-used simple estimators will be informative, and on improving implementation and interpretation of these common methods.

2 Model

This section develops the basic model. Definitions and maintained assumptions of the model are presented in numbered equations, while optional assumptions are labeled by name. Matrices are written as upper-case and boldface (e.g., $\mathbf{X}$), vectors as lower-case and boldface ($\mathbf{x}$), and scalars as lower-case without boldface (x). For ease of exposition, the same notation will occasionally be used to represent a random variable, the function defining it, and its realization.

The model's exposition will refer to an example application in which a researcher is studying the effect of classroom gender composition on an academic achievement as measured by test scores. This question has been investigated extensively in the empirical literature, for example by Hoxby (2000), Lavy and Schlosser (2011) and Eisenkopf et al. (2015). This research typically finds a postive effect of female peers, even in settings and academic subjects where boys and girls have similar average outcomes. It is thus a natural application of this model, which allows an individual's behavior to affect own and peer outcomes differently.

2.1 Basic framework and notation

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

An individual's type τ_i represents every characteristic of that individual that is potentially relevant in this domain, so the type space is finite but may be quite large. Since the indexing of individuals is arbitrary, each individual's type can be taken as an independent draw from a common type distribution:

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

where $f_\tau : \mathcal{T} \to [0, 1]$ is some unknown discrete PDF. Note that unconditional independence does not necessarily imply independence conditional on $\mathbf{G}$.

Peer groups may form randomly, or may depend on $\mathbf{T}$:

$$\Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{T} = \mathbf{T_A}) = f_{\mathbf{G}|\mathbf{T}}(\mathbf{G_A}, \mathbf{T_A}) \tag{2}$$

To simplify exposition, all peer groups are assumed to have identical size n, which in turn implies that $N = nG$. Let:

$$\mathbf{p}_i \equiv \mathbf{p}(i, \mathbf{G}) \equiv \{j \neq i : g_j = g_i\} \tag{3}$$

be the set of individual i's peers, and let $\mathcal{P}_i^S$ be the collection of all size S subsets of $\mathcal{N} \setminus \{i\}$, so that $\mathbf{p}_i \in \mathcal{P}_i^{(n-1)}$.

Given individual types and peer groups, social interactions produce the outcome:

$$\mathbf{Y} \equiv \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \equiv \begin{bmatrix} y\left(\tau_1, \{\tau_j\}_{j \in \mathbf{p}_1}\right) \\ \vdots \\ y\left(\tau_N, \{\tau_j\}_{j \in \mathbf{p}_N}\right) \end{bmatrix} \equiv \mathbf{Y}(\mathbf{T}, \mathbf{G}) \tag{4}$$

according to the unknown social process $y : \mathcal{T}^n \to \mathbb{R}$. Note that peers enter into the outcome function as an unordered (multi)set because indexing is arbitary.

As written, the social interaction process has several characteristics. First, each individual's outcome is determined by the individual's own unobserved type and the unobserved type of the individual's peers. This rules out spillovers between peer groups. It also abstracts from any direct effects of the group assignment itself (the implications of which are discussed in Fruehwirth (2014)) and post-assignment random factors that might affect the outcome.

In addition, while the outcome can be interpreted as the equilibrium of some game, the structural mechanism by which peers matter is not modeled. In particular, there is no distinction made here between peer effects that arise through peer behavior (endogenous effects, in the terminology of Manski (1993)) and those that arise through predetermined peer characteristics (exogenous or contextual effects). This reduced-form approach means that the framework is potentially useful for analyzing the effects of counterfactual allocations of individuals across groups, but not necessarily of other policy interventions that change individual incentives.

Finally, one important difference between this paper and some others in the literature is that peer type is not necessarily interpreted as a scalar peer "quality". For example, Arcidiacono et al. (2012), Burke and Sass (2013) and Fruehwirth (2014) all consider models of classroom peer effects in which good (high unobserved ability/effort) students are assumed to be good peers (i.e., the treated individual's outcome is an increasing function of peer unobserved ability/effort). In contrast, the model here allows unobserved personality traits such as agreeableness or unobserved behavior such as competition for teacher attention to affect own and peer outcomes in the different directions, or to have positive effects on some peers and negative effects on others.

2.2 Data and regression models

The researcher estimates one or more regression models from the observed data $\mathbf{D} \equiv (\mathbf{X}, \mathbf{Y}, \mathbf{G})$ where:

$$\mathbf{X} \equiv \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \equiv \begin{bmatrix} \mathbf{x}(\tau_1) \\ \vdots \\ \mathbf{x}(\tau_N) \end{bmatrix} \equiv \mathbf{X}(\mathbf{T}) \tag{5}$$

is an $N \times K$ matrix of predetermined[2] observable background characteristics. This section describes those regression models, while Section 2.3 describes some related nonparametric estimands. The focus in this paper is on identification from the joint distribution of observables $\mathbf{D}$ for a fixed number of individuals N. However, the identification results will be constructive and suggest natural analog estimators (e.g. OLS) with well-understood statistical properties.

[2]Throughout the paper, these background characteristics are predetermined and are not subject to manipulation by a policy maker. Manski (2013) provides an analysis of treatment response under social interactions.

To abstract from functional form considerations, the individual characteristics $\mathbf{x}_i$ are taken to be a K-vector of dummy variables indicating membership in one of $K+1$ mutually exclusive and exhaustive categories, i.e.

$$\mathbf{x}_i \in \mathbb{C}_K \equiv \{\mathbf{c}_{0K}, \mathbf{c}_{1K}, \ldots, \mathbf{c}_{KK}\} \tag{6}$$

where $\mathbf{c}_{kK}$ is a K-vector that contains one in column k and zero elsewhere, and all categories are represented in the population:

$$\Pr(\mathbf{x}_i = \mathbf{c}_{kK}) > 0 \qquad \text{for all } k = 0, 1, \ldots, K \tag{7}$$

If the original set of individual characteristics does not have this structure, the researcher can easily generate this structure by binning continuous variables, including interactions, etc.

The researcher also uses the observed data to construct a set of variables describing each individual's peer group. Let:

$$\bar{\mathbf{X}} \equiv \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \vdots \\ \bar{\mathbf{x}}_N \end{bmatrix} \equiv \begin{bmatrix} \bar{\mathbf{x}}(\mathbf{p}_1) \\ \vdots \\ \bar{\mathbf{x}}(\mathbf{p}_N) \end{bmatrix} \equiv \bar{\mathbf{X}}(\mathbf{X}, \mathbf{G}) \tag{8}$$

where $\bar{\mathbf{x}}(\mathbf{p}) = \left(\sum_{j \in \mathbf{p}} \mathbf{x}_j\right)/(n-1)$ is the peer group average of $\mathbf{x}$ for peer group $\mathbf{p}$. Since $\mathbf{x}_i$ is categorical, $\bar{\mathbf{x}}_i$ completely describes the frequency distribution of $\mathbf{x}_j$ in i's peer group, and:

$$\bar{\mathbf{x}}_i \in \bar{\mathbb{X}} \equiv \left\{ \bar{\mathbf{x}}_i \in \left\{0, \frac{1}{n-1}, \ldots, 1\right\}^K : \sum_{k=1}^{K} \bar{\mathbf{x}}_{ik} \leq 1 \right\} \tag{9}$$

Since K and n are finite, $\bar{\mathbb{X}}$ is also finite with cardinality $\left|\bar{\mathbb{X}}\right| = \frac{(K+n-1)!}{(n-1)!K!}$.

In the classroom gender effects example, the researcher has student-level data on classroom g_i, gender $\mathbf{x}_i = \mathbb{I}\left(male\right)$, and some academic outcome of interest y_i for each student i. Then $\bar{\mathbf{x}}_i \in \bar{\mathbb{X}} = \left\{0, \frac{1}{n-1}, \ldots, 1\right\}$ is the proportion male in student i's classroom, and $|\bar{\mathbb{X}}| = n$.

Although many applied papers estimate peer effect models that are linear in $\bar{\mathbf{x}}_i$, many others allow various forms of nonlinearity. To model that case, suppose that the researcher partitions $\bar{\mathbb{X}}$ into $M+1$ non-empty, mutually exclusive, and exhaustive categories $(\bar{\mathbb{X}}^0, \bar{\mathbb{X}}^1, \ldots, \bar{\mathbb{X}}^M)$ and constructs the $N \times M$ matrix:

$$\mathbf{Z} \equiv \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_N \end{bmatrix} \equiv \begin{bmatrix} \mathbf{z}(\bar{\mathbf{x}}_1) \\ \mathbf{z}(\bar{\mathbf{x}}_2) \\ \vdots \\ \mathbf{z}(\bar{\mathbf{x}}_N) \end{bmatrix} \equiv \mathbf{Z}(\bar{\mathbf{X}}) \tag{10}$$

where $\mathbf{z} : \bar{\mathbb{X}} \to \mathbb{C}_M$ is an M-vector of dummy variables indicating membership in one of these categories:

$$\mathbf{z}_i = \mathbf{z}(\bar{\mathbf{x}}_i) \equiv \sum_{m=1}^{M} \mathbf{c}_{mM} \mathbb{I}\left(\bar{\mathbf{x}}_i \in \bar{\mathbb{X}}^m\right)$$

Category m is a **singleton** if $|\bar{\mathbb{X}}^m| = 1$ and **pooled** if $|\bar{\mathbb{X}}^m| > 1$. A partition that consists only of singleton categories is **saturated**.

Returning to the classroom gender effects example, a researcher might divide classrooms into all-female, mixed, and all-male categories:

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

while another researcher might divide into majority-female and majority-male categories:

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

and a third researcher might construct a saturated $\mathbf{z}_i$:

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

Given these variables, the researcher estimates one or more of the following linear regression models:

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \equiv \alpha_0 + \mathbf{x}_i\alpha_1 + \bar{\mathbf{x}}_i\alpha_2 \tag{11}$$

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \equiv \beta_0 + \mathbf{x}_i\beta_1 + \bar{\mathbf{x}}_i\beta_2 + \mathbf{x}_i\beta_3\bar{\mathbf{x}}_i' \tag{12}$$

$$L(y_i|\mathbf{x}_i, \mathbf{z}_i) \equiv \gamma_0 + \mathbf{x}_i\gamma_1 + \mathbf{z}_i\gamma_2 \tag{13}$$

$$L(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i) \equiv \delta_0 + \mathbf{x}_i\delta_1 + \mathbf{z}_i\delta_2 + \mathbf{x}_i\delta_3\mathbf{z}_i' \tag{14}$$

where $L(\cdot|\cdot)$ is the best linear predictor and $\alpha \equiv (\alpha_0, \alpha_1, \alpha_2)$, $\beta \equiv (\beta_0, \beta_1, \beta_2, vec(\beta_3))$, $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ and $\delta = (\delta_0, \delta_1, \delta_2, vec(\delta_3))$ are vectors of coefficients. Individual elements of these matrices are referred to in the usual manner, e.g. δ_{3sm} refers to the element in the s row and m column of δ_3. For convenience, let $\delta_{30m} \equiv 0$ for all m and $\beta_{30k} \equiv 0$ for all k.

The coefficients in equations (11)-(14) are all identified from $\mathbf{D}$ under the usual assumption that:

$$E(\mathbf{w}_i'\mathbf{w}_i) \text{ is positive definite} \tag{15}$$

where $\mathbf{w}_i$ is defined as needed; e.g., $\mathbf{w}_i = (1, \mathbf{x}_i, \bar{\mathbf{x}}_i)$ for estimating equation (11). Assumption (7) guarantees the needed variation in $\mathbf{x}_i$, while the peer group assignment mechanism determines whether the required variation in $\bar{\mathbf{x}}_i$ or $\mathbf{z}_i$ is present.

In this paper, equations (11)-(14) are reduced form estimating equations and their coefficients are not directly given causal or structural interpretations. Both the individual-level variables in $\mathbf{x}_i$ and the peer variables in $\mathbf{z}_i$ are chosen by the researcher, and each potential choice of explanatory variables implies a distinct set of regression coefficients, all of which are identified from $\mathbf{D}$ provided that condition (15) holds.

At the same time, these four specifications correspond to commonly-estimated regressions in the empirical literature. Specification (11) coresponds to the simple "linear in means" model that is most commonly estimated in applied work. In the classroom gender effects example, this model would predict an outcome based on a linear function of the student's own gender and the gender composition of the student's classroom.

The other specifications represent extensions to the simple model that are found in the literature, for example, in Hoxby and Weingarth (2005):

- Specification (12) is typically interpreted as allowing "heterogeneous effects" through the interaction term between own and peer characteristics. For example, male classmates may have a different effect on male versus female students.
- Specification (13) is typically interpreted as allowing "nonlinear effects", i.e. features of the peer group composition other than the mean can matter. For example, replacing a girl with a boy may have a different effect in a classroom that is mostly girls than it would have in a classroom that is mostly boys.

- Specification (14) is the most general and incorporates both heterogeneous and nonlinear effects.

Section 2.3 below will explicitly define several related nonparametric causal estimands, and Section 3.3 will show conditions under which the estimating equations defined in this section can be related to these causal estimands.

2.3 Estimands

This section defines several estimands of potential interest. These estimands are defined in terms of the explanatory variables described in Section 2.2 but are defined nonparametrically and impose no restrictions on the underlying model.

Individual i's **potential outcome function** is defined as:

$$y_i(\mathbf{p}) \equiv y\left(\tau_i, \{\tau_j\}_{j\in\mathbf{p}}\right) \tag{16}$$

That is, $y_i(\mathbf{p})$ is the outcome that would have been observed for person i if he or she had been assigned the peers in $\mathbf{p}$. This outcome depends on her own type τ_i and the types of her peers $\{\tau_j\}_{j\in\mathbf{p}}$. The observed outcome is $y_i = y_i(\mathbf{p}_i)$ and $y_i(\mathbf{p})$ is a counterfactual outcome for all $\mathbf{p} \neq \mathbf{p}_i$.

The **conditional average peer effect** ($CAPE_k$) of peers of observed type k relative to peers of observed type zero can be defined as:

$$CAPE_k \equiv E\left(y_i(\{j\}\cup\tilde{\mathbf{q}}) - y_i(\{j'\}\cup\tilde{\mathbf{q}})\right|\mathbf{x}_j = \mathbf{c}_{kK}, \mathbf{x}_{j'} = \mathbf{c}_{0K}) \tag{17}$$

where $\tilde{\mathbf{q}}$ is a purely random draw from $\mathcal{P}_i^{(n-2)}$, and let:

$$\mathbf{CAPE} \equiv \left[\begin{array}{cccc} CAPE_1 & CAPE_2 & \cdots & CAPE_K \end{array}\right] \tag{18}$$

The conditional average peer effect is analogous to the conditional average treatment effect estimated in the literature on heterogeneous treatment effects (e.g. Wager and Athey (2018)). The basic idea is that exposure to a particular peer j can be considered a treatment, whose effect is not directly observed but may be predicted in a useful way. Peer effect estimates can be interpreted as characterizing the heterogeneity of those effects.

This setting has several key differences from a standard treatment effects analysis. In addition to the usual unobserved heterogeneity across treated units (i), there is unobserved heterogeneity in the treatment itself (j) and possible spillovers from other treatments ($\tilde{\mathbf{q}}$). This heterogeneity is addressed by averaging over the (conditional or unconditional) distribution of unobservables. In addition, there is no natural "untreated" state, so conditional average peer effects are defined relative to the average peer in an arbitrarily-selected base category. Note that there is no meaningful analogue in this setting to the average treatment effect, as the effect of replacing the average peer with the average peer would be zero.

The **heterogeneous peer effect** ($HPE_{s,k}$) of peers of observed type k on individuals of observed type s can be defined as:

$$HPE_{s,k} \equiv E\left(y_i(\{j\}\cup\tilde{\mathbf{q}}) - y_i(\{j'\}\cup\tilde{\mathbf{q}})\right|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}, \mathbf{x}_{j'} = \mathbf{c}_{0K}) \tag{19}$$

where $\tilde{\mathbf{q}}$ is a purely random draw from $\mathcal{P}_i^{(n-2)} \cap \mathcal{P}_j^{(n-2)}$. It will also be convenient to define the $(K+1)\times K$ matrix **HPE** as:

$$\mathbf{HPE} \equiv \begin{bmatrix} HPE_{0,1} & \cdots & HPE_{0,K} \\ \vdots & \ddots & \vdots \\ HPE_{K,1} & \cdots & HPE_{K,K} \end{bmatrix} \tag{20}$$

Heterogeneous peer effects allow for analysis of heterogeneity across both treatments and treated units.

A researcher may also be interested in characterizing heterogeneity across entire peer groups rather than across individual peers. In the classoom gender effects example, the researcher may want in predict the effect of moving from a majority-male classroom to a majority-female classroom. To facilitate that type of analysis, the **conditional average group effect** of a type m peer group (relative to a type zero peer group) is defined as:

$$CAGE_m \equiv E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \tag{21}$$

where $\tilde{\mathbf{p}}$ is a purely random draw from $\mathcal{P}_i$. The **heterogeneous group effect** of a type m peer group on type s individuals is defined as:

$$HGE_{s,m} \equiv E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \tag{22}$$

where $\tilde{\mathbf{p}}$ is a purely random draw[3] from $\mathcal{P}_i$. It will also be convenient to define the M-vector **CAGE** and the $(K+1) \times M$ matrix **HGE**.

In the classroom gender effects example, suppose that $\mathbf{z}_i = \mathbb{I}(\bar{\mathbf{x}}_i > 0.5)$ is an indicator for whether the peer group is majority male. Then:

- $CAPE_1$ is the effect on the average student of replacing an average female peer with an average male peer.
- HPE_{11} is the effect on the average male student of replacing an average female peer with an average male peer.
- HPE_{01} is the effect on the average female student of replacing an average female peer with an average male peer.
- $CAGE_1$ is the effect on the average student of the average majority-male peer group relative to the average balanced-or-majority-female peer group.
- HGE_{01} is the effect on the average female student of the average majority-male peer group relative to the average balanced-or-majority-female peer group.
- HGE_{11} is the effect on the average male student of the average majority-male peer group relative to the average balanced-or-majority-female peer group.

These are all well-defined and potentially interesting estimands, and the remainder of the paper establishes conditions under which they are identified and useful for prediction.

2.4 Additional identifying assumptions

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

As is generally the case, some form of randomness in peer selection will be needed to identify social effects. This can take the form of **simple random assignment (RA)** of peers:

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

i.e., peer group assignment does not depend on one's unobservable type or any other predetermined characteristics. Assumption (RA) implies that types are IID conditional on $\mathbf{G}$.

[3] Note that $CAGE_m$ and $HGE_{s,m}$ are defined in terms of a purely random draw of peers, and thus imposes a particular conditional distribution for $\Pr(\bar{\mathbf{x}}_i|\mathbf{z}_i)$. Proposition 5 in Section 3.4 shows that $CAGE_m$ and $HGE_{s,m}$ are only informative about peer group reallocations that preserve this conditional distribution (e.g., if $\bar{\mathbb{X}}^m$ is a singleton). See Section 3.4 for additional details.

An alternative to simple random assignment is **conditional random assignment (CRA)** based on observed characteristics:

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

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

Peer effects are said to be **peer-separable (PSE)** if the effect of replacing one peer with another does not depend on the identity of any other peers. That is, for all $i \in \mathcal{N}$, $\mathbf{q}, \mathbf{q}' \in \mathcal{P}_i^{(n-2)}$ and $j, j' \notin (\mathbf{q} \cup \mathbf{q}')$, replacing peer j with peer j' would have the same effect on individual i regardless of whether his other peers are $\mathbf{q}$ or $\mathbf{q}'$:

$$y_i(\{j\} \cup \mathbf{q}) - y_i(\{j'\} \cup \mathbf{q}) = y_i(\{j\} \cup \mathbf{q}') - y_i(\{j'\} \cup \mathbf{q}') \tag{\textbf{PSE}}$$

Peer effects are said to be **own-separable (OSE)** if the effect of replacing one peer group with another is the same for everyone. That is, for all $i, i' \in \mathcal{N}$ and $\mathbf{p}, \mathbf{p}' \in \mathcal{P}_i^{(n-1)} \cap \mathcal{P}_{i'}^{(n-1)}$ replacing peer group $\mathbf{p}$ with peer group $\mathbf{p}'$ would have the same effect on individual i as it does on individual i':

$$y_i(\mathbf{p}) - y_i(\mathbf{p}') = y_{i'}(\mathbf{p}) - y_{i'}(\mathbf{p}') \tag{\textbf{OSE}}$$

Peer effects that are neither own-separable nor peer-separable will be called **non-separable**.

3 Results

3.1 Separability and its implications

The first set of results relate to the assumption of separability. Separable effects turn out to be straightforward to interpret and estimate. In particular, Proposition 1 below shows that a peer-separable potential outcome function can always be written as the sum of an individual-specific "own effect" and a set of individual-specific or pair-specific "peer effects," and that **HPE** and **CAPE** can be expressed in terms of conditional expectations of these latent variables. While neither the own effect nor the peer effect is directly observable, Proposition 3 in the next section establishes some conditions under which **HPE** and **CAPE** are identified.

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

$$y_i(\mathbf{p}) = o_i + \sum_{j \in \mathbf{p}} p_{ij} \tag{23}$$

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

$$HPE_{s,k} = E(p_{ij} | \mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) - E(p_{ij} | \mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{0K}) \tag{24}$$

$$CAPE_k = E(p_{ij} | \mathbf{x}_j = \mathbf{c}_{kK}) - E(p_{ij} | \mathbf{x}_j = \mathbf{c}_{0K}) \tag{25}$$

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

$$y_i(\mathbf{p}) = o_i + \sum_{j \in \mathbf{p}} p_j \tag{26}$$

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

$$HPE_{s,k} = CAPE_k = E(p_j | \mathbf{x}_j = \mathbf{c}_{kK}) - E(p_j | \mathbf{x}_j = \mathbf{c}_{0K}) \tag{27}$$

The latent variable o_i can be interpreted as person i's "own effect", as it is the outcome person i would experience in a peer group whose members were all drawn from some arbitrarily-chosen reference type. The latent variable p_j or p_{ij} can be interpreted as person j's "peer effect", as it is the change in outcome a person i would experience if a peer from the reference type were replaced by person j.

While separability assumptions are convenient, they may not be correct. Proposition 2 below shows that separability can be tested by a simple joint significance test of some linear regression coefficients.

Proposition 2 (Testable implications of separability)**.** *1. If peers are randomly assigned conditional on observables (CRA) and peer effects are peer separable (PSE), then:*

$$L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i) = L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \tag{28}$$

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

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

or equivalently $\beta_3 = 0$.

Note that separability is a property of the outcome function $y(\cdot, \cdot)$ and not the particular explanatory variables chosen by the researcher. However, the power of a given test based on Proposition 2 will depend on the choice of explanatory variables. For example, if $\mathbf{X}$ is completely unrelated to $\mathbf{Y}$, then (28) and (29) hold trivially even in the absence of separability.

3.2 Heterogeneity and aggregation

Heterogeneous effects can be aggregated to yield conditional average effects:

$$CAPE_k = \sum_{s=0}^{K} HPE_{s,k} \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) \tag{30}$$

$$CAGE_m = \sum_{s=0}^{K} HGE_{s,m} \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) \tag{31}$$

In addition, heterogeneous group effects for a given partition can be obtained by aggregating heterogeneous group effects for a saturated partition. That is, let $\mathbf{z}_i^S$ be a vector of $M^S \equiv (|\bar{\mathbb{X}}|-1)$ indicator variables representing a saturated partition of $\bar{\mathbb{X}}$, and let $\mathbf{HGE}^S$ be the corresponding $(K+1) \times M^S$ matrix of heterogenous group effects:

$$HGE_{s,m}^S \equiv E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{mM^S}) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{0M^S}) \tag{32}$$

Then:

$$HGE_{s,m} = \sum_{r=1}^{M^S} HGE_{s,r}^S \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ -\Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \tag{33}$$

These results imply that **CAPE** can be recovered from **HPE**, and both **CAGE** and **HGE** can be recovered from $\mathbf{HGE}^S$.

3.3 Identification

Proposition 3 below shows identification under a simple random assignment research design. The identification analysis also suggests some simple estimators. Proposition 4 later in this section shows identification under random assignment conditional on observables, and Proposition 6 in Section 4 shows identification under a more complex two-stage assignment design.

Proposition 3 (Identification with random assignment)**.** *1. If peers are randomly assigned (RA), then* **HGE** *and* **CAGE** *are identified:*

$$HGE_{s,m} = \delta_{2m} + \delta_{3sm} \tag{34}$$

$$CAGE_m = \gamma_{2m} \tag{35}$$

2. If peers are randomly assigned (RA) and peer effects are peer separable (PSE), then **HPE** *and* **CAPE** *are identified:*

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

$$CAPE_k = \frac{\alpha_{2k}}{n-1} \tag{37}$$

To gain a clearer understanding of Proposition 3, return to the classroom gender effects example. First, suppose the researcher estimates the conventional linear-in-means peer effects model described in equation (11), i.e., a linear regression of y_i on $(\mathbf{x}_i, \bar{\mathbf{x}}_i)$ where $\mathbf{x}_i = 1$ for boys and $\mathbf{x}_i = 0$ for girls. Part 2 of Proposition 3 shows that the assumption of peer-separability allows the researcher to interpret that coefficient on $\bar{\mathbf{x}}_i$ as the effect on the average student of replacing the average female classmate with the average male classmate. Note that there are no other control variables, and gender does not appear in the underlying structural model; instead this analysis is interpreted as an analysis of heterogeneity. Another researcher with the same data but other $\mathbf{x}_i$ variables - race, ethnicity, language spoken at home, immigration status, etc. - could explore those other aspects of heterogeneity either separately or in any combination.

Next, suppose the researcher wishes to consider heterogeneity across treated units by estimating equation (12), i.e., a linear regression of y_i on $(\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i\bar{\mathbf{x}}_i)$. Part 2 of Proposition 3 shows that the assumption of peer-separability allows the researcher to interpret the coefficient on $\bar{\mathbf{x}}_i$ as the effect on the average *female* student of replacing the average female classmate with the average male classmate. Adding the coefficient on the interaction term $\mathbf{x}_i\bar{\mathbf{x}}_i$ gives the effect on the average *male* student of replacing the average female classmate with the average male classmate.

Note that a finding of heterogeneity (i.e. a nonzero coefficient on the interaction term) does not invalidate the analysis based on equation (11), as that analysis can still be interpreted as averaging these heterogeneous effects across all treatment units. Both specifications are valid, in the sense of recovering an estimand of interest.

Although the assumption of peer separability provides a simple interpretation of linear-in-means results, empirical researchers have shown increasing interest in contextual effects that go beyond the linear-in-means model, and have repeatedly found evidence for such nonlinearities. Returning to the classroom gender effects example, suppose the researcher divides peer groups into majority-male and majority female; i.e., $\mathbf{z}_i = 1$ if $\bar{\mathbf{x}}_i > 0.5$ and $\mathbf{z}_i = 0$ if $\bar{\mathbf{x}}_i \leq 0.5$, and estimates a regression of y_i on $(\mathbf{x}_i, \mathbf{z}_i)$. Then part 1 of Proposition 3 says that the coefficient on $\mathbf{z}_i$ can be interpreted as the effect on the average student of replacing the average (randomly constructed) majority-female peer group with the average (randomly constructed) majority-male peer group. In addition, these averages can be identified separately for male and female students from a regression of y_i on $(\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i)$. Again, the results in Proposition 3 apply regardless of the researcher's choice of how to construct $\mathbf{z}_i$. The researcher could compare majority-male versus majority-female peer groups, or could compare all-male, all-female, and mixed peer groups.

Although identification and interpretation are simplest with random assignment, many of the results in Proposition 3 also hold under conditional random assignment. To show this it is first necessary to show (in Lemma 1 below) that the conditional expectation function is the same under random assignment and conditional random assignment.

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

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) = E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) \tag{38}$$

where $\tilde{\mathbf{p}}$ *is a purely random draw from* $\mathcal{P}_i$*.*

Proposition 4, which shows identification under conditional random assignment, then follows.

Proposition 4 (Identification with conditional random assignment)**.** *1. If peers are randomly assigned conditional on observable characteristics (CRA), then* **HGE** *and* **CAGE** *are identified:*

$$HGE_{s,m} = \sum_{r=1}^{M^S} (\delta_{2r}^S + \delta_{3sr}^S) \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \tag{39}$$

$$CAGE_m = \sum_{s=0}^{K} \sum_{r=1}^{M^S} \Pr(\mathbf{x}_i = \mathbf{c}_{sK})(\delta_{2r}^S + \delta_{3sr}^S) \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \tag{40}$$

where $\delta^S = (\delta_0^S, \delta_1^S, \delta_2^S, \delta_3^S)$ *are the coefficients from estimating equation (14) with saturated group variable* $\mathbf{z}_i^S$*.*

2. *If peers are randomly assigned conditional on observable characteristics (CRA) and peer effects are peer separable (PSE), then* **HPE** *and* **CAPE** *are identified:*

$$HPE_{s,k} = \frac{\beta_{2k} + \beta_{3sk}}{n-1} \tag{41}$$

$$CAPE_k = \sum_{s=0}^{K} \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) \frac{\beta_{2k} + \beta_{3sk}}{n-1} \tag{42}$$

While Proposition 4 is more general than Proposition 3, this generality comes at the minor cost that some estimands (**CAPE**, **CAGE**, **HGE**) are weighted averages of regression coefficients rather than just the coefficients.

3.4 Peer group reallocations

The estimands defined in Section 2.3 predict the effect of a change in the composition of a representative individual's peer group. However, with a fixed population, any change in the composition of one peer group implies a corresponding change in the composition of at least one other peer group. As a result, it is important to consider the implications of alternative allocation *policies* (Bhattacharya, 2009; Graham et al., 2010).

In order to evaluate allocation policies, it is necessary to define some additional estimands and establish conditions for their identification. Let $\mathbf{G}_0$ and $\mathbf{G}_1$ be peer group allocations defined by:

$$\mathbf{G}_0 \equiv \mathbf{G}_0(\mathbf{X}, \sigma) \qquad\qquad \mathbf{G}_1 \equiv \mathbf{G}_1(\mathbf{X}, \sigma) \tag{43}$$

where the functions $\mathbf{G}_0 : \mathbb{R}^{NK+1} \rightarrow \mathcal{G}$ and $\mathbf{G}_1 : \mathbb{R}^{NK+1} \rightarrow \mathcal{G}$ are allocation rules or policies chosen by the researcher and $\sigma \in \mathbb{R}$ is a randomization device such that:

$$\sigma|\mathbf{T} \sim U(0,1) \tag{44}$$

For example, the researcher might wish to compare single-gender to randomly-mixed classroom assignments. The randomization device σ allows the researcher to specify any conditional probability distribution over group assignments (e.g., simple random assignment). Together, (43) and (44) imply that both $\mathbf{G}_0$ and $\mathbf{G}_1$ satisfy the assumption of conditional random assignment (CRA):

$$\mathbf{G}_0 \perp\!\!\!\perp \mathbf{T}|\mathbf{X} \qquad \mathbf{G}_1 \perp\!\!\!\perp \mathbf{T}|\mathbf{X} \tag{45}$$

Let:

$$\mathbf{Y}_0 \equiv \mathbf{Y}(\mathbf{T}, \mathbf{G}_0) \qquad \mathbf{Y}_1 \equiv \mathbf{Y}(\mathbf{T}, \mathbf{G}_1) \tag{46}$$
$$\bar{\mathbf{X}}_0 \equiv \bar{\mathbf{X}}(\mathbf{X}, \mathbf{G}_0) \qquad \bar{\mathbf{X}}_1 \equiv \bar{\mathbf{X}}(\mathbf{X}, \mathbf{G}_1) \tag{47}$$
$$\mathbf{Z}_0 \equiv \mathbf{Z}(\bar{\mathbf{X}}_0) \qquad \mathbf{Z}_1 \equiv \mathbf{Z}(\bar{\mathbf{X}}_1) \tag{48}$$

be the potential outcomes and peer characteristics of interest. Finally, let $(y_{i0}, \bar{\mathbf{x}}_{i0}, \mathbf{z}_{i0})$ be arbitrary elements of $(\mathbf{Y}_0, \bar{\mathbf{X}}_0, \mathbf{Z}_0)$, and let $(y_{i1}, \bar{\mathbf{x}}_{i1}, \mathbf{z}_{i1})$ be arbitrary elements of $(\mathbf{Y}_1, \bar{\mathbf{X}}_1, \mathbf{Z}_1)$.

The **conditional average reallocation effect** of a change from allocation rule $\mathbf{G}_0$ to allocation rule $\mathbf{G}_1$ is defined as:

$$CARE_s(\mathbf{G}_0, \mathbf{G}_1) \equiv E(y_{i1} - y_{i0}|\mathbf{x}_i = \mathbf{c}_{sK}) \tag{49}$$
$$\mathbf{CARE}(\mathbf{G}_0, \mathbf{G}_1) \equiv \begin{bmatrix} CARE_0 & CARE_1 & \cdots & CARE_K \end{bmatrix} \tag{50}$$

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

$$ARE(\mathbf{G}_0, \mathbf{G}_1) \equiv E(y_{i1} - y_{i0}) \tag{51}$$

The average reallocation effect is easily calculated from the conditional average reallocation effect.

$$ARE(\mathbf{G}_0, \mathbf{G}_1) = \sum_{s=0}^{K} CARE_s(\mathbf{G}_0, \mathbf{G}_1) \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) \tag{52}$$

Proposition 5 below describes how these reallocation effects can be described in terms of the estimands defined in Section 2.3.

Proposition 5 (Reallocation effects). *1. If $(\bar{\mathbb{X}}^1, \ldots, \bar{\mathbb{X}}^M)$ are singletons, and $\Pr(\bar{\mathbf{x}}_{i0} \in \bar{\mathbb{X}}^0) = \Pr(\bar{\mathbf{x}}_{i1} \in \bar{\mathbb{X}}^0) = 0$, then:*

$$CARE_s(\mathbf{G}_0, \mathbf{G}_1) = \mathbf{HGE}_s E(\mathbf{z}'_{i1} - \mathbf{z}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK}) \tag{53}$$

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

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

$$CARE_s(\mathbf{G}_0, \mathbf{G}_1) = \mathbf{HPE}_s E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK})(n-1) \tag{54}$$

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

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

$$CARE_s(\mathbf{G}_0, \mathbf{G}_1) = \mathbf{CAPE}\, E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK})(n-1) \tag{55}$$

and the average reallocation effect is zero.

Parts two and three of Proposition 5 show that separability allows the effect of an alternative allocation to be inferred straightforwardly from the estimands defined in Section 2.3. In contrast, part one of Proposition 5 indicates an important limitation on the use of non-separable/nonlinear peer effect regressions to predict the results of a change in the allocation rule: the partition of $\bar{\mathbf{x}}_i$ must assign a unique value of $\mathbf{z}_i$ for each distinct value of $\bar{\mathbf{x}}_i$ in the support of the proposed allocation rules. 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 $\bar{\mathbf{x}}_i$.

Returning to the classroom gender effects example, suppose the researcher has constructed $\mathbf{z}_i$ using five categories: all-boy ($\bar{\mathbf{x}}_i = 1$), majority-boy ($0.5 < \bar{\mathbf{x}}_i < 1$), balanced ($\bar{\mathbf{x}}_i = 0.5$), majority-girl ($0.0 < \bar{\mathbf{x}}_i < 0.5$), and all-girl ($\bar{\mathbf{x}}_i = 0$). The all-boy, balanced and all-girl categories are singletons, while the majority-boy and majority-girl categories are pooled. Proposition 5 implies those results can be used to predict the result of a change from balanced to gender-segregated classrooms, or from the baseline random allocation to a balanced or gender-segregated allocation. However, the results cannot be used to predict the effect of a change from balanced to majority-boy and majority-girl classrooms. The natural solution to this issue is to remember that the researcher chooses the partition, and can in principle[4] always choose a partition rich enough to identify (conditional) average reallocation effects. For example, one might redefine categories to include singleton categories for $\bar{\mathbf{x}}_i = 0.5$, $\bar{\mathbf{x}}_i = 0.25$ and $\bar{\mathbf{x}}_i = 0.75$ and then use the results to compare a balanced or random allocation to one in which roughly half of classrooms are 75% boys and the other half are 75% girls.

4 Extension: Multi-stage assignment processes

Although simple random assignment is the ideal research design for studying peer effects, many peer effect studies are based on a more complex research design in which individuals are non-randomly assigned to large groups and then randomly assigned to smaller groups within those large groups. For example, peer effects in education are typically estimated using panel data with multiple grade cohorts within multiple schools, and grade cohort composition (due to timing of birth) is treated as random conditional on a school fixed effect (which hopefully accounts for nonrandom selection into schools). This section considers this research design within this paper's framework by embedding peer groups within **locations**.

4.1 Model

The model is as defined in Section 2 with additional assumptions and definitions as given below.

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

$$\mathbf{L} \equiv \begin{bmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_N \end{bmatrix} \equiv \begin{bmatrix} \ell(g_1) \\ \ell(g_2) \\ \vdots \\ \ell(g_N) \end{bmatrix} \equiv \mathbf{L}(\mathbf{G}) \tag{56}$$

where $\ell : \mathcal{G} \to \mathcal{L}$ is a known function. A simple one-stage selection mechanism can be treated as a special case with a single location. To simplify exposition, each location is assumed to include r peer groups, which implies that $G = rL$ and $N = nG = nrL$.

[4]In practice, the usual bias/variance tradeoff applies to the choice of partitions: more categories reduces the bias from aggregating categories with dissimilar average effects, but also reduces statistical precision.

Each individual's type is an independent draw from a type distribution that varies by location:

$$\Pr(\mathbf{T}|\mathbf{L}) = \prod_{i=1}^{N} f_{\tau|\ell}(\tau_i, \ell_i) \tag{57}$$

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

4.1.1 Regression models

Let $\alpha^\ell \equiv (\alpha_0^\ell, \alpha_1^\ell, \alpha_2^\ell)$, and $\beta^\ell \equiv (\beta_0^\ell, \beta_1^\ell, \beta_2^\ell, \beta_3^\ell)$ be the location-specific best linear predictors for location ℓ:

$$L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \equiv \alpha_0^\ell + \mathbf{x}_i\alpha_1^\ell + \bar{\mathbf{x}}_i\alpha_2^\ell \tag{58}$$

$$L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) \equiv \beta_0^\ell + \mathbf{x}_i\beta_1^\ell + \bar{\mathbf{x}}_i\beta_2^\ell + \mathbf{x}_i\beta_3^\ell\bar{\mathbf{x}}_i' \tag{59}$$

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

$$\begin{aligned}\alpha^\ell &= E(\mathbf{w}_i'\mathbf{w}_i|\ell_i = \ell)^{-1}E(\mathbf{w}_i'y_i|\ell_i = \ell) \text{ where } \mathbf{w}_i = (1, \mathbf{x}_i, \bar{\mathbf{x}}_i) \\ \beta^\ell &= E(\mathbf{w}_i'\mathbf{w}_i|\ell_i = \ell)^{-1}E(\mathbf{w}_i'y_i|\ell_i = \ell) \text{ where } \mathbf{w}_i = (1, \mathbf{x}_i, \bar{\mathbf{x}}_i, vec(\mathbf{x}_i'\bar{\mathbf{x}}_i))\end{aligned} \tag{60}$$

These coefficients are identified from the probability distribution for $\mathbf{D}$ under the assumption:

$$E(\mathbf{w}_i'\mathbf{w}_i|\ell_i = \ell) \text{ is positive definite for all } \ell \in \mathcal{L} \tag{61}$$

where $\mathbf{w}_i$ is defined as in (60).

The estimating equations (58) and (59) correspond to standard heterogeneous-coefficient linear panel data models, and fit into the framework of Wooldridge (2010, p. 377-381). Estimators and testing procedures can easily be adapted from that literature, or additional homogeneity restrictions can be imposed to simpler and/or more efficient estimators. For example, Section 4.3 describes a set of restrictions that yield the conventional fixed effects model.

4.1.2 Estimands

Conditional average peer effects and other estimands defined in Section 2.3 remain well-defined and are identified under some conditions described in Section 4.2 below. In addition, let **within-location conditional average peer effects** and **within-location heterogeneous peer effects** for location ℓ be defined as:

$$\begin{aligned}CAPE_k^\ell &\equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}})|\mathbf{x}_j = \mathbf{c}_{kK}, \mathbf{x}_{j'} = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell\right) \\ HPE_{s,k}^\ell &\equiv E\left(y_i(\{j\} \cup \tilde{\mathbf{q}}) - y_i(\{j'\} \cup \tilde{\mathbf{q}})|\mathbf{x}_j = \mathbf{c}_{kK}, \mathbf{x}_{j'} = \mathbf{c}_{0K}, \mathbf{x}_i = \mathbf{c}_{sK}, \ell_i = \ell_j = \ell\right)\end{aligned} \tag{62}$$

where $\tilde{\mathbf{q}}$ is a purely random draw from $\mathcal{P}_i^{(n-2)} \cap \mathcal{P}_j^{(n-2)}$, and let the matrices $\mathbf{CAPE}^\ell$ and $\mathbf{HPE}^\ell$ be defined accordingly. These within-location peer effects describe the average effect of replacing a randomly chosen peer from the base category with a randomly chosen peer from another category *at the same location.*

4.1.3 Additional identifying assumptions

As will be shown in Proposition 6 below, identification in this setting will require restrictions on both group assignment and cross-location heterogeneity. This section defines the relevant restrictions. To simplify exposition, the analysis here will focus on the case of peer separability, so that these restrictions can be described in terms of the own effect o_i and the peer effect p_{ij}.

Peer groups are **randomly assigned by location (RAL)** if:

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

In other words, location assignment may depend on type, but assignment to peer groups within a given location does not.

Peer-separable peer effects are **location invariant (LI)** if:

$$E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell, \ell_j = \ell') = E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) \quad \text{for all } s, k, \ell, \ell' \tag{LI}$$

Location invariance will be the main identifying assumption in this section. In the classroom gender effects example, location invariance would allow schools that systematically have more boys to have systematically better or worse students (conditional on the student's gender) but would not allow them to have systematically better or worse peers (conditional on the peer's gender). This is a strong assumption and may not be a reasonable one in some applications.

Peer-separable peer effects are **partially location invariant (PLI)** if:

$$E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{0K}, \ell_j = \ell) = E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{0K}) \quad \text{for all } \ell \tag{PLI}$$

This slightly weaker restriction implies that there is at least one identifiable category of individuals whose average peer effects do not vary across locations. Without loss of generality, that category is taken as the base category. In the classroom gender effects example, this would allow (for example) schools with more boys to have systematically better or worse male peers, as long as the female peers were not systematically different across schools.

4.2 Identification

Proposition 6 shows that identification of conditional average peer effects in a within-location random assignment design requires strong restrictions on heterogeneity across locations.

Proposition 6 (Identification under within-location random assignment)**.**

1. *If peers are randomly assigned by location (RAL) and peer effects are peer separable (PSE), then* $\mathbf{HPE}^{\ell}$ *and* $\mathbf{CAPE}^{\ell}$ *are identified:*

$$HPE^{\ell}_{s,k} = \frac{\beta^{\ell}_{2k} + \beta^{\ell}_{3sk}}{n-1} \tag{63}$$

$$CAPE^{\ell}_{k} = \frac{\alpha^{\ell}_{2k}}{n-1} \tag{64}$$

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

2. *If peers are randomly assigned by location (RAL) and peer effects are peer separable and location invariant (PSE, LI), then* $\mathbf{HPE}$ *and* $\mathbf{CAPE}$ *are identified:*

$$HPE_{s,k} = \frac{E(\beta^{\ell_i}_{2k}) + E(\beta^{\ell_i}_{3sk})}{n-1} \tag{65}$$

$$CAPE_{k} = \frac{E(\alpha^{\ell_i}_{2k})}{n-1} \tag{66}$$

3. *If peers are randomly assigned by location (RAL) and peer effects are peer separable, own separable and partially location invariant (PSE, OSE, PLI), then* $\mathbf{HPE}$ *and* $\mathbf{CAPE}$ *are identified:*

$$HPE_{s,k} = CAPE_{k} = \frac{E(\alpha^{\ell_i}_{2k}|\mathbf{x}_i = \mathbf{c}_{kK})}{n-1} \tag{67}$$

For example, consider a standard cohort-based research design for measuring educational peer effects, where ℓ denotes nonrandomly assigned schools and g denotes randomly assigned cohorts within schools. In this setting, there are likely to be systematic differences across schools, even after conditioning on observable characteristics. In the absence of further restrictions, it is not possible to distinguish a school whose students are less productive (low $E(o_i|\ell_i = \ell)$) from a school whose students are disruptive to their peers (low $E(p_{ij}|\ell_i = \ell_j = \ell)$). Both factors end up in the same school-level fixed effect. As a result, it is only possible to characterize within-school differences in conditional average peer effects.

The location invariance (LI) assumption allows identification of **CAPE** and **HPE** by shutting down this possibility: with location invariance there are no systematic differences in peer effects across schools other than through the observed student characteristics. The partial location invariance (PLI) assumption places slightly weaker restrictions on cross-school variation but requres the additional restriction of own-separability (OSE) for the peer effects themselves.

4.3 Coefficient heterogeneity and overidentifying restrictions

Although the within-location coefficients $(\alpha^\ell, \beta^\ell)$ are defined to allow arbitrary variation across locations, some substantive assumptions also imply equality restrictions on these coefficients that can be tested or imposed on estimation.

In particular, the combination of random within-location assignment (RAL) peer separability (PSE) and location invariance (LI) implies the coefficient restriction:

$$(\beta_2^\ell, \beta_3^\ell) = (\beta_2^0, \beta_3^0) \tag{68}$$

where (β_2^0 and β_3^0) are vectors of constants. This restriction can then be imposed in estimation, or tested. Although purely random coefficient variation cannot be distinguished from heteroskedasticity, the null of homogeneity can be tested against the alternative of coefficient heterogeneity that is systematically correlated with observed variables (Wooldridge 2010, p. 385).

If peer effects are also assumed to be own separable (OSE) and the difference in conditional average own-effects does not vary systematically across schools:

$$\begin{aligned} &E(o_i|\mathbf{x}_i = \mathbf{c}_{kK}, \ell_i = \ell) - E(o_i|\mathbf{x}_i = \mathbf{c}_{0K}, \ell_i = \ell) \\ &\qquad\qquad = E(o_i|\mathbf{x}_i = \mathbf{c}_{kK}) - E(o_i|\mathbf{x}_i = \mathbf{c}_{0K}) \quad \text{for all } \ell \qquad (\mathbf{FE}) \end{aligned}$$

then the within-location estimating equation takes on the linear fixed effects form commonly used in applied work:

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}, \ell_i = \ell) = \underbrace{(\lambda_0^\ell + \eta_0(n-1))}_{\alpha_0^\ell} + \mathbf{x} \underbrace{\lambda_1}_{\alpha_1^\ell} + \bar{\mathbf{x}} \underbrace{\eta_2(n-1)}_{\alpha_2^\ell} \tag{69}$$

which can be consistently estimated by the standard "within" estimator. As with the previous result, the coefficient homogeneity implied by these assumptions is in principle testable.

4.4 Peer group reallocations

Random assignment by location complicates predictions about the results of a given peer group reallocation. If **HPE** is identified, the results in Proposition 5 apply and it is possible to calculate reallocation effects for any reallocation. In contrast, if only location-specific peer effects $\mathbf{HPE}^\ell$ are identified, then it is only possible to calculate reallocation effects for reallocations within locations.

5 Extension: Direct contextual effects

Much of the applied literature treats contextual peer effects as if they were *direct*. That is, a specific subset of peer characteristics are assumed to have a direct impact on outcomes, and any impact of changing peers operates through its impact on these characteristics. For example, if contextual effects are based on the distribution of race and gender within the group, then any peers of the same race and gender are interchangeable.

Peer effects are said to be **direct contextual effects (DCE)** if:

$$y_i(\mathbf{p}) = h\left(\mathbf{x}_i^*, \{\mathbf{x}_j^*\}_{j\in\mathbf{p}}\right) + \epsilon_i \quad \text{where } \epsilon_i = \epsilon(\tau_i) \text{ and } E(\epsilon_i|\mathbf{x}_i^*) = 0 \qquad (\mathbf{DCE})$$

where $\mathbf{x}_i^* = \mathbf{x}^*(\tau_i)$ is some (potentially unknown) K^*-vector of characteristics and $h : \mathbb{R}^{nK^*} \to \mathbb{R}$. is some unknown function. The condition $E(\epsilon_i|\mathbf{x}_i^*) = 0$ is without loss of generality since $h(\cdot,\cdot)$ can be renormalized to make it true. The key restriction in (DCE) is that the peer effect is a fixed (but unknown) function of observable (but potentially omitted) characteristics. That is, for any individuals i, j, j' and set of peers $\mathbf{q}$, the effect of replacing peer j with peer j' is:

$$y_i(j\cup\mathbf{q}) - y_i(j'\cup\mathbf{q}) = h\left(\mathbf{x}_i^*, \{\mathbf{x}_j^*, \mathbf{x}_r^*\}_{r\in\mathbf{q}}\right) - h\left(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*, \mathbf{x}_r^*\}_{r\in\mathbf{q}}\right) \qquad (70)$$

which depends only on $\mathbf{X}$ and $\mathbf{q}$.

The vector $\mathbf{x}_i^*$ of relevant variables and the vector $\mathbf{x}_i$ of variables available to the researcher do not necessarily coincide. When they do, the researcher is said to have **no omitted variables (NOV)**:

$$\mathbf{x}_i = \mathbf{x}_i^* \qquad (\mathbf{NOV})$$

Note that the general framework does not require this assumption for identification of estimands such as **CAPE** and **HPE**.

5.1 Separability and functional form

Assumption (DCE) does not restrict the functional form for $h(\cdot,\cdot)$, but Proposition 7 below shows other assumptions such as separability may pin down a convenient functional form.

Proposition 7 (Separability and direct contextual effects)**.** *1. If direct contextual effects are peer separable (DCE, PSE), then each individual's potential outcome function can be expressed in the form:*

$$\begin{aligned} y_i(\mathbf{p}) &= o_i + \sum_{j\in\mathbf{p}} p_{ij} && (71)\\ &= h_1(\mathbf{x}_i^*) + \epsilon_i + \sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) \\ &= \theta_0 + \mathbf{x}_i^*\theta_1 + \bar{\mathbf{x}}_{\mathbf{p}}^*\theta_2 + \mathbf{x}_i^*\theta_3\bar{\mathbf{x}}^*(\mathbf{p})' + \epsilon_i && (\text{if } \mathbf{x}^* \text{ is categorical}) \end{aligned}$$

where $h_1 : \mathbb{R}^{K^*} \to \mathbb{R}$ and $h_2 : \mathbb{R}^{2K^*} \to \mathbb{R}$ are functions and $\theta \equiv (\theta_0, \theta_1, \theta_2, vec(\theta_3))$ is some vector of coefficients.

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

$$\begin{aligned} y_i(\mathbf{p}) &= o_i + \sum_{j\in\mathbf{p}} p_j && (72)\\ &= h_1(\mathbf{x}_i^*) + \epsilon_i + \sum_{j\in\mathbf{p}} h_3(\mathbf{x}_j^*) \\ &= \theta_0 + \mathbf{x}_i^*\theta_1 + \bar{\mathbf{x}}^*(\mathbf{p})\theta_2 + \epsilon_i && (\text{if } \mathbf{x}^* \text{ is categorical}) \end{aligned}$$

where $h_1 : \mathbb{R}^{K^*} \to \mathbb{R}$ and $h_3 : \mathbb{R}^{K^*} \to \mathbb{R}$ are functions and $\theta \equiv (\theta_0, \theta_1, \theta_2)$ is some vector of coefficients.

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

5.2 Identification and omitted variables bias

Direct contextual effects do not violate the conditions for Propositions 3, 4, or 6, so heterogenous and conditional average effects continue to be indentified under the conditions described in those propositions. Proposition 8 below shows additional identification results that apply when contextual effects are direct.

Proposition 8 (Identification of direct contextual effects)**.**

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

$$h\left(\mathbf{x}, \left\{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\right\}\right) = E\left(y_i \left|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}\right.\right) \tag{73}$$

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

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

$$HPE_{s,k} = \frac{E(\beta_{2k}^{\ell_i}) + E(\beta_{3sk}^{\ell_i})}{n-1} \tag{74}$$

$$CAPE_k = \frac{E(\alpha_{2k}^{\ell_i})}{n-1} \tag{75}$$

The first result in Proposition 8 shows that the "structural" parameters of the direct contextual effects model are identified. The second result shows that direct contextual effects are helpful in securing identification with random assignment by location, essentially because they imply location invariance.

Unfortunately, the data requirements for Proposition 8 are extremely demanding: the (NOV) assumption requires data on *everything* about each person that affects their influence on other people. In the more likely case in which the researcher has data on some subset of those characteristics, conventional omitted variables concepts apply. To simplify the discussion, suppose that the direct contextual effects take the linear in means form:

$$h(\mathbf{x}_i^*, \left\{\mathbf{x}_j^*\right\}_{j\in\mathbf{p}}) = \theta_0 + \mathbf{x}_i^*\theta_1 + \bar{\mathbf{x}}_i^*(\mathbf{p})\theta_2 \tag{76}$$

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

$$y_i = \theta_0 + \mathbf{x}_i\theta_{11} + \mathbf{u}_i\theta_{12} + \bar{\mathbf{x}}_i\theta_{21} + \bar{\mathbf{u}}_i\theta_{22} + \epsilon_i \tag{77}$$

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

$$\begin{aligned} L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) &= L(\theta_0 + \mathbf{x}_i\theta_{11} + \mathbf{u}_i\theta_{12} + \bar{\mathbf{x}}_i\theta_{21} + \bar{\mathbf{u}}_i\theta_{22} + \epsilon_i|\mathbf{x}_i, \bar{\mathbf{x}}_i) \\ &= \theta_0 + \mathbf{x}_i\theta_{11} + (\pi_0 + \mathbf{x}_i\pi_1)\theta_{12} + \bar{\mathbf{x}}_i\theta_{21} + (\pi_0 + \bar{\mathbf{x}}_i\pi_1)\theta_{22} \\ &= \underbrace{(\theta_0 + \pi_0\theta_{12} + \pi_0\theta_{21})}_{\alpha_0} + \mathbf{x}_i\underbrace{(\theta_{11} + \pi_1\theta_{12})}_{\alpha_1} + \bar{\mathbf{x}}_i\underbrace{(\theta_{21} + \pi_1\theta_{22})}_{\alpha_2} \end{aligned} \tag{78}$$

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

5.3 Peer group reallocations

If the data requirements for Proposition 8 are met, the assumption of direct contextual effects has the potential advantage of allowing the direct comparison of predicted outcomes across any two specific allocations. That is, if θ_2 is the direct causal effect of $\bar{\mathbf{x}}_i^*$ on y_i, then any reallocation of peers that changes $\bar{\mathbf{x}}_i^*$ by Δ units changes y_i by $\Delta\theta_2$ units. The direct contextual effects model can thus be used to predict exact outcome differences between any two alternative allocations or allocation rules.

In contrast, knowledge of heterogeneous and conditional average peer effects only allows for comparisons within the class of alternative allocation mechanisms described in Section 3.4: those in which groups are assigned randomly conditional on the observable characteristics in the data. Returning to the classroom gender effects example, the conditional average peer effect **CAPE** can be interpreted as the *average* effect of replacing a *randomly selected* male peer with a *randomly selected* female peer. In contrast, if direct contextual effects are assumed in which $\mathbf{x}_i^*$ is a gender indicator, then the contextual effects parameter θ_2 is the effect of replacing *any* male peer with *any* female peer.

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

6 Conclusion

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

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

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

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

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

References

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

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.

Burke, Mary A. and Tim R. Sass, "Classroom peer effects and student achievement," *Journal of Labor Economics*, 2013, *31* (1), 51–82.

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.

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.

A Proofs

Proposition 1

1. For every i, let:

$$
\begin{aligned}
o_i &\equiv y(\tau_i, \{1, 1, \ldots, 1\}) \\
p_{ij} &\equiv y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\})
\end{aligned} \tag{79}
$$

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

$$\begin{aligned}
o_i + \sum_{j\in\mathbf{p}} p_{ij} &= y(\tau_i, \{1,1,\dots,1\}) && \text{(by (79))}\\
&\quad + \sum_{j\in\mathbf{p}} \left(y(\tau_i,\{\tau_j,1,1,\dots,1\}) - y(\tau_i,\{1,1,1,\dots,1\})\right)\\
&= y(\tau_i,\{1,1,\dots,1\})\\
&\quad + y(\tau_i,\left\{\tau_{\mathbf{p}(1)},1,1,\dots,1\right\}) - y(\tau_i,\{1,1,1,\dots,1\})\\
&\qquad \vdots\\
&\quad + y(\tau_i,\left\{\tau_{\mathbf{p}(n-1)},1,1,\dots,1\right\}) - y(\tau_i,\{1,1,1,\dots,1\})\\
& && \text{(expansion of summation)}\\
&= y(\tau_i,\{1,1,\dots,1\})\\
&\quad + y(\tau_i,\left\{\tau_{\mathbf{p}(1)},\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\dots,\tau_{\mathbf{p}(n-1)}\right\}) - y(\tau_i,\left\{1,,\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\dots,\tau_{\mathbf{p}(n-1)}\right\})\\
&\quad + y(\tau_i,\left\{\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\dots,\tau_{\mathbf{p}(n-1)},1\right\}) - y(\tau_i,\left\{1,\tau_{\mathbf{p}(3)},\dots,\tau_{\mathbf{p}(n-1)},1\right\})\\
&\qquad \vdots\\
&\quad + y(\tau_i,\left\{\tau_{\mathbf{p}(n-2)},\tau_{\mathbf{p}(n-1)},1,\dots,1\right\}) - y(\tau_i,\left\{1,\tau_{\mathbf{p}(n-1)},1,\dots,1\right\})\\
&\quad + y(\tau_i,\left\{\tau_{\mathbf{p}(n-1)},1,1,\dots,1\right\}) - y(\tau_i,\{1,1,1,\dots,1\})\\
& && \text{(by PSE)}\\
&= y(\tau_i,\left\{\tau_{\mathbf{p}(1)},\tau_{\mathbf{p}(2)},\tau_{\mathbf{p}(3)},\dots,\tau_{\mathbf{p}(n-1)}\right\})\\
&= y_i(\mathbf{p}) && \text{(by (16))}
\end{aligned}$$

which is result (23). To prove (24) and (25), first note that $(\tau_i,\tau_j) \perp\!\!\!\perp \tau_{j'}$ by equation (1), so:

$$(p_{ij},\mathbf{x}_j) \perp\!\!\!\perp \mathbf{x}_{j'} \tag{80}$$

Then:

$$\begin{aligned}
HPE_{s,k} &= E\left(y_i(\{j\}\cup\tilde{\mathbf{q}}) - y_i(\{j'\}\cup\tilde{\mathbf{q}}) \middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right) && \text{(by (19))}\\
&= E\left(\left(o_i + p_{ij} + \sum_{j''\in\tilde{\mathbf{q}}} p_{ij''}\right) - \left(o_i + p_{ij'} + \sum_{j''\in\tilde{\mathbf{q}}} p_{ij''}\right)\middle| \begin{matrix}\mathbf{x}_i=\mathbf{c}_{sK},\\ \mathbf{x}_j=\mathbf{c}_{kK},\\ \mathbf{x}_{j'}=\mathbf{c}_{0K}\end{matrix}\right) && \text{(by (23))}\\
&= E\left(p_{ij}-p_{ij'}\middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right)\\
&= E\left(p_{ij}\middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right) - E\left(p_{ij'}\middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right)\\
&= E\left(p_{ij}\middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right) - E\left(p_{ij}\middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_{j'}=\mathbf{c}_{kK},\mathbf{x}_j=\mathbf{c}_{0K}\right)\\
&= E\left(p_{ij}\middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK}\right) - E\left(p_{ij}\middle| \mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{0K}\right) && \text{(by (80))}
\end{aligned}$$

which is the result in (24) and:

$$\begin{aligned}
CAPE_k &= E\left(y_i(\{j\}\cup\tilde{\mathbf{q}}) - y_i(\{j'\}\cup\tilde{\mathbf{q}}) \middle| \mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right) && \text{(by (17))}\\
&= E\left(\left(o_i + p_{ij} + \sum_{j''\in\tilde{\mathbf{q}}} p_{ij''}\right) - \left(o_i + p_{ij'} + \sum_{j''\in\tilde{\mathbf{q}}} p_{ij''}\right)\middle| \begin{matrix}\mathbf{x}_j=\mathbf{c}_{kK},\\ \mathbf{x}_{j'}=\mathbf{c}_{0K}\end{matrix}\right) && \text{(by (23))}\\
&= E\left(p_{ij}-p_{ij'}\middle| \mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right)\\
&= E\left(p_{ij}\middle| \mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right) - E\left(p_{ij'}\middle| \mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right)\\
&= E\left(p_{ij}\middle| \mathbf{x}_j=\mathbf{c}_{kK},\mathbf{x}_{j'}=\mathbf{c}_{0K}\right) - E\left(p_{ij}\middle| \mathbf{x}_{j'}=\mathbf{c}_{kK},\mathbf{x}_j=\mathbf{c}_{0K}\right)\\
&= E\left(p_{ij}\middle| \mathbf{x}_j=\mathbf{c}_{kK}\right) - E\left(p_{ij}\middle| \mathbf{x}_j=\mathbf{c}_{0K}\right) && \text{(by (80))}
\end{aligned}$$

which is the result in (25).

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

$$\begin{aligned}
p_{ij} &= (y(\tau_i, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_i, \{1, 1, 1, \ldots, 1\})) && \text{(by (79))} \\
&= y(\tau_1, \{\tau_j, 1, 1, \ldots, 1\}) - y(\tau_1, \{1, 1, 1, \ldots, 1\}) && \text{(by OSE)} \\
&= p_{1j} && \text{(by (79))} \\
&= p_j && (81)
\end{aligned}$$

Substituting the result in (81) into (23) yields the result in (26). Substituting that same result into (24) and (25) yields:

$$\begin{aligned}
CAPE_k &= E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{kK}) - E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{0K}) && \text{(by (25))} \\
&= E(p_j|\mathbf{x}_j = \mathbf{c}_{kK}) - E(p_j|\mathbf{x}_j = \mathbf{c}_{0K}) && \text{(by (81))} \\
HPE_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) - E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{0K}) && \text{(by (24))} \\
&= E(p_j|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) - E(p_j|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{0K}) && \text{(by (81))} \\
&= E(p_j|\mathbf{x}_j = \mathbf{c}_{kK}) - E(p_j|\mathbf{x}_j = \mathbf{c}_{0K}) && \text{(since (1)} \implies \mathbf{x}_i \perp\!\!\!\perp (\mathbf{x}_j, p_j)) \\
&= CAPE_k
\end{aligned}$$

which are the results in (27).

Proposition 2

The conditions for both parts of Proposition 3 are met here, so its results apply.

1. Let $\tilde{\mathbf{G}}$ be a purely random group assignment and let $\tilde{\mathbf{p}}_i = \mathbf{p}(i, \tilde{\mathbf{G}})$. Part two of Proposition 3 applies to the joint distribution of $(\mathbf{X}, \mathbf{Y}(\tilde{\mathbf{G}}), \bar{\mathbf{X}}(\mathbf{X}, \tilde{\mathbf{G}}))$ since $\mathbf{Y}(\cdot)$ satisfies (PSE) and $\tilde{\mathbf{G}}$ satisfies (RA). By (CRA), Lemma 1 applies to the joint distribution of $(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}})$ Let (λ, η) be defined as in equation (86) of the proof for Proposition 3. Then:

$$\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) && \text{(by (38) in Lemma 1)} \\
&= (\lambda_0 + \eta_0(n-1)) \\
&\quad + \mathbf{x}(\lambda_1 + \eta_1(n-1)) \\
&\quad + \bar{\mathbf{x}}\eta_2(n-1) \\
&\quad + \mathbf{x}\eta_3(n-1)\bar{\mathbf{x}}' \\
& && \text{(by result (89) in the proof for Proposition 3)}
\end{aligned}$$

Applying the law of iterated projections:

$$
\begin{aligned}
L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) &= L(E(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) && \text{(law of iterated projections)} \\
&= L\left(\begin{array}{l} (\lambda_0 + \eta_0(n-1)) \\ \quad + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) \\ \quad + \bar{\mathbf{x}}_i\eta_2(n-1) \\ \quad + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i' \end{array} \middle| \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i\right) && \text{(result above)} \\
&= (\lambda_0 + \eta_0(n-1)) && (82) \\
&\quad + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) \\
&\quad + \bar{\mathbf{x}}_i\eta_2(n-1) \\
&\quad + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i' \\
L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i) &= L(E(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i)|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i) && \text{(law of iterated projections)} \\
&= L\left(\begin{array}{l} (\lambda_0 + \eta_0(n-1)) \\ \quad + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) \\ \quad + \bar{\mathbf{x}}_i\eta_2(n-1) \\ \quad + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i' \end{array} \middle| \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i, \mathbf{z}_i\right) && \text{(result above)} \\
&= (\lambda_0 + \eta_0(n-1)) && (83) \\
&\quad + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) \\
&\quad + \bar{\mathbf{x}}_i\eta_2(n-1) \\
&\quad + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i' \\
&= L(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i) && \text{(by (82) and (83))}
\end{aligned}
$$

which is result (28).

2. The assumptions here (PSE, OSE, CRA) imply that all results in Propositions 1 and 4 apply. Therefore:

$$
\begin{aligned}
CAPE_k &= HPE_{s,k} \qquad \text{for all } s && \text{(by (27) in Proposition 1)} \\
&= \frac{\beta_{2k} + \beta_{3sk}}{n-1} && \text{(by (41) in Proposition 4)}
\end{aligned}
$$

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

Proposition 3

1. Let $\tilde{\mathbf{p}}$ be a purely random draw from $\mathcal{P}_i^{(n-1)}$. By (RA), the actual peer group $\mathbf{p}_i$ is also a purely random draw from $\mathcal{P}_i^{(n-1)}$, 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:

$$
\begin{aligned}
HGE_{s,m} &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \\
&&& \text{(by (22))} \\
&= E(y_i(\mathbf{p}_i)|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\mathbf{p}_i) = \mathbf{c}_{mM}) - E(y_i(\mathbf{p}_i)|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\mathbf{p}_i) = \mathbf{c}_{0M}) \\
&&& \text{(RA} \implies \text{same joint distribution)} \\
&= E(y_i|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i = \mathbf{c}_{mM}) - E(y_i|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i = \mathbf{c}_{0M}) && (84)
\end{aligned}
$$

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

$$
\begin{aligned}
E(y_i|\mathbf{x}_i, \mathbf{z}_i) &= L(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i'\mathbf{z}_i) \\
&= \delta_0 + \mathbf{x}_i\delta_1 + \mathbf{z}_i\delta_2 + \mathbf{x}_i\delta_3\mathbf{z}_i' && \text{(by (14))}
\end{aligned}
$$

Combining these two results produces:

$$
\begin{aligned}
HGE_{s,m} &= E(y_i|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i = \mathbf{c}_{mM}) - E(y_i|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i = \mathbf{c}_{0M}) && \text{(by (84))} \\
&= (\delta_0 + \mathbf{c}_{sK}\delta_1 + \mathbf{c}_{mM}\delta_2 + \mathbf{c}_{sK}\delta_3\mathbf{c}'_{mM}) - (\delta_0 + \mathbf{c}_{sK}\delta_1 + \mathbf{c}_{0M}\delta_2 + \mathbf{c}_{sK}\delta_3\mathbf{c}'_{0M}) \\
& && \text{(result above)} \\
&= (\delta_0 + \mathbf{c}_{sK}\delta_1 + \mathbf{c}_{mM}\delta_2 + \mathbf{c}_{sK}\delta_3\mathbf{c}'_{mM}) - (\delta_0 + \mathbf{c}_{sK}\delta_1) && \text{(since } \mathbf{c}_{0M} = 0\text{)} \\
&= \mathbf{c}_{mM}\delta_2 + \mathbf{c}_{sK}\delta_3\mathbf{c}'_{mM} \\
&= \delta_{2m} + \delta_{3sm}
\end{aligned}
$$

which is the result in (34). Result (35) can be established by similar reasoning:

$$
\begin{aligned}
CAGE_m &= E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) - E(y_i(\tilde{\mathbf{p}})|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) && \text{(by (21))} \\
&= E(y_i(\mathbf{p}_i)|\mathbf{z}_i(\mathbf{p}_i) = \mathbf{c}_{mM}) - E(y_i(\mathbf{p}_i)|\mathbf{z}_i(\mathbf{p}_i) = \mathbf{c}_{0M}) \\
& && \text{(RA} \implies \text{same joint distribution)} \\
&= E(y_i|\mathbf{z}_i = \mathbf{c}_{mM}) - E(y_i|\mathbf{z}_i = \mathbf{c}_{0M}) && (85)
\end{aligned}
$$

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

$$
\begin{aligned}
E(y_i|\mathbf{z}_i) &= L(y_i|\mathbf{z}_i) \\
&= L(L(y_i|\mathbf{x}_i, \mathbf{z}_i)|\mathbf{z}_i) && \text{(law of iterated projections)} \\
&= L(\gamma_0 + \mathbf{x}_i\gamma_1 + \mathbf{z}_i\gamma_2|\mathbf{z}_i) && \text{(by (13))} \\
&= \gamma_0 + L(\mathbf{x}_i|\mathbf{z}_i)\gamma_1 + \mathbf{z}_i\gamma_2 \\
&= \gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{z}_i\gamma_2 && \text{(RA} \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{z}_i\text{)}
\end{aligned}
$$

Combining these two results:

$$
\begin{aligned}
CAGE_m &= E(y_i|\mathbf{z}_i = \mathbf{c}_{mM}) - E(y_i|\mathbf{z}_i = \mathbf{c}_{0M}) && \text{(by (85))} \\
&= (\gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{c}_{mM}\gamma_2) - (\gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{c}_{0M}\gamma_2) && \text{(result above)} \\
&= (\gamma_0 + E(\mathbf{x}_i)\gamma_1 + \mathbf{c}_{mM}\gamma_2) - (\gamma_0 + E(\mathbf{x}_i)\gamma_1) && \text{(since } \mathbf{c}_{0M} = 0\text{)} \\
&= \mathbf{c}_{mM}\gamma_2 \\
&= \gamma_{2m}
\end{aligned}
$$

which is result (35).

2. By (PSE), Proposition 1 applies. Let $\lambda \equiv (\lambda_0, \lambda_1)$ and $\eta \equiv (\eta_1, \eta_2, \eta_3)$ satisfy:

$$
\begin{aligned}
E(o_i|\mathbf{x}_i = \mathbf{c}_{sK}) &= \lambda_0 + \mathbf{c}_{sK}\lambda_1 && (86) \\
E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) &= \eta_0 + \mathbf{c}_{sK}\eta_1 + \mathbf{c}_{kK}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{kK}
\end{aligned}
$$

The linear functional forms in (86) are without loss of generality since $\mathbf{x}$ is categorical. The estimand $HPE_{s,k}$ can be expressed as a function of η:

$$
\begin{aligned}
HPE_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) && \text{(by (25) in Proposition 1)} \\
&\quad - E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{0K}) \\
&= (\eta_0 + \mathbf{c}_{sK}\eta_1 + \mathbf{c}_{kK}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{kK}) && \text{(by (86))} \\
&\quad - (\eta_0 + \mathbf{c}_{sK}\eta_1 + \mathbf{c}_{0K}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{0K}) \\
&= (\eta_0 + \mathbf{c}_{sK}\eta_1 + \mathbf{c}_{kK}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{kK}) - (\eta_0 + \mathbf{c}_{sK}\eta_1) && \text{(since } \mathbf{c}_{0K} = 0\text{)} \\
&= \mathbf{c}_{kK}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{kK} \\
&= \eta_{2k} + \eta_{3sk} && (87)
\end{aligned}
$$

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

$$\begin{aligned}
E(y_i|\mathbf{X},\mathbf{G}) &= E\left(o_i + \sum_{j\in\mathbf{p}_i} p_{ij}\,\middle|\,\mathbf{X},\mathbf{G}\right) && \text{(by (23) in Proposition 1)}\\
&= E\left(o_i + \sum_{j=1}^{N} p_{ij}\mathbb{I}\left(j\in\mathbf{p}_i\right)\,\middle|\,\mathbf{X},\mathbf{G}\right) && \text{(where } \mathbb{I}\left(\cdot\right) \text{ is the indicator function)}\\
&= E(o_i|\mathbf{X},\mathbf{G}) + \sum_{j=1}^{N} E(p_{ij}|\mathbf{X},\mathbf{G})\mathbb{I}\left(j\in\mathbf{p}_i\right)\\
& && \text{(since } \mathbb{I}\left(j\in\mathbf{p}_i\right) \text{ is a function of } \mathbf{G})\\
&= E(o_i|\mathbf{X},\mathbf{G}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X},\mathbf{G})\\
&= E(o_i|\mathbf{X}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X}) && \text{(since RA} \implies (o_i, p_{ij}, \mathbf{X}) \perp\!\!\!\perp (\mathbf{G},\mathbf{p}_i))\\
&= E(o_i|\mathbf{x}_i) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j) && \text{(since (1)} \implies (\tau_i,\tau_j) \perp\!\!\!\perp \tau_{j'})\\
&= \lambda_0 + \mathbf{x}_i\lambda_1 + \sum_{j\in\mathbf{p}_i}(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}_j') && \text{(by (86))}\\
&= \lambda_0 + \mathbf{x}_i\lambda_1 + \eta_0(n-1) + \mathbf{x}_i\eta_1(n-1) + \bar{\mathbf{x}}_i\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\\
&= (\lambda_0 + \eta_0(n-1)) + \mathbf{x}_i(\lambda_1 + \eta_1(n-1)) + \bar{\mathbf{x}}_i\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i' && (88)
\end{aligned}$$

Applying the law of iterated expectations to this result:

$$\begin{aligned}
E(y_i|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) &= E(E(y_i|\mathbf{X},\mathbf{G})|\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}) && \text{(law of iterated expectations)}\\
&= E\left(\begin{matrix}(\lambda_0+\eta_0(n-1)) + \mathbf{x}_i(\lambda_1+\eta_1(n-1))\\ +\bar{\mathbf{x}}_i\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\end{matrix}\,\middle|\,\mathbf{x}_i=\mathbf{x},\bar{\mathbf{x}}_i=\bar{\mathbf{x}}\right)\\
& && \text{(by (88))}\\
&= (\lambda_0+\eta_0(n-1)) + \mathbf{x}(\lambda_1+\eta_1(n-1)) + \bar{\mathbf{x}}\eta_2(n-1) + \mathbf{x}\eta_3(n-1)\bar{\mathbf{x}}'\\
& && (89)
\end{aligned}$$

Applying the law of iterated projections to this result:

$$\begin{aligned}
L(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i\bar{\mathbf{x}}_i') &= L(E(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i)|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i\bar{\mathbf{x}}_i') && \text{(law of iterated projections)}\\
&= L\left(\begin{matrix}(\lambda_0+\eta_0(n-1))\\ +\mathbf{x}_i(\lambda_1+\eta_1(n-1))\\ +\bar{\mathbf{x}}_i\eta_2(n-1)\\ +\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\end{matrix}\,\middle|\,\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i\bar{\mathbf{x}}_i'\right) && \text{(by (89))}\\
&= \underbrace{(\lambda_0+\eta_0(n-1))}_{\beta_0} + \mathbf{x}_i\underbrace{(\lambda_1+\eta_1(n-1))}_{\beta_1} + \bar{\mathbf{x}}_i\underbrace{\eta_2(n-1)}_{\beta_2} + \mathbf{x}_i\underbrace{\eta_3(n-1)}_{\beta_3}\bar{\mathbf{x}}_i'\\
& && (90)
\end{aligned}$$

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

$$\begin{aligned}
HPE_{s,k} &= \eta_{2k} + \eta_{3sk} && \text{(by (87))}\\
&= \frac{\beta_{2k}+\beta_{3sk}}{n-1} && \text{(by (90))}
\end{aligned}$$

which is the result in (36). To get result (37), first note that:

$$\begin{aligned}
E(p_{ij}|\mathbf{x}_j=\mathbf{x}) &= E(E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j)|\mathbf{x}_j=\mathbf{x}) && \text{(law of iterated expectations)}\\
&= E(\eta_0+\mathbf{x}_i\eta_1+\mathbf{x}_j\eta_2+\mathbf{x}_i\eta_3\mathbf{x}_j'|\mathbf{x}_j=\mathbf{x}) && \text{(by (86))}\\
&= \eta_0+E(\mathbf{x}_i|\mathbf{x}_j=\mathbf{x})\eta_1+\mathbf{x}\eta_2+E(\mathbf{x}_i|\mathbf{x}_j=\mathbf{x})\eta_3\mathbf{x}' && \text{(conditioning rule)}\\
&= \eta_0+E(\mathbf{x}_i)\eta_1+\mathbf{x}\eta_2+E(\mathbf{x}_i)\eta_3\mathbf{x}' && \text{(since (1) } \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j)\\
&= (\eta_0+E(\mathbf{x}_i))\,\eta_1+\mathbf{x}\,(\eta_2+\eta_3'E(\mathbf{x}_i')) && (91)
\end{aligned}$$

Equation (25) from Proposition 1 implies:

$$\begin{aligned}
CAPE_k &= E(p_{ij}|\mathbf{x}_j=\mathbf{c}_{kK})-E(p_{ij}|\mathbf{x}_j=\mathbf{c}_{0K}) && \text{(by (25) in Proposition 1)}\\
&= \ ((\eta_0+E(\mathbf{x}_i)\eta_1)+\mathbf{c}_{kK}\,(\eta_2+\eta_3'E(\mathbf{x}_i'))) && \text{(by (91))}\\
&\quad\ -((\eta_0+E(\mathbf{x}_i)\eta_1)+\mathbf{c}_{0K}\,(\eta_2+\eta_3'E(\mathbf{x}_i'))) \\
&= \ ((\eta_0+E(\mathbf{x}_i)\eta_1)+\mathbf{c}_{kK}\,(\eta_2+\eta_3'E(\mathbf{x}_i'))) && \text{(since } \mathbf{c}_{0K}=0)\\
&\quad\ -((\eta_0+E(\mathbf{x}_i)\eta_1)) \\
&= \mathbf{c}_{kK}\,(\eta_2+\eta_3'E(\mathbf{x}_i')) && (92)
\end{aligned}$$

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

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

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

$$\begin{aligned}
L(y_i|\bar{\mathbf{x}}_i) &= L(L(y_i|\mathbf{x}_i,\bar{\mathbf{x}}_i,\mathbf{x}_i'\bar{\mathbf{x}}_i)|\bar{\mathbf{x}}_i) && \text{(law of iterated projections)}\\
&= L\left(\begin{matrix}(\lambda_0+\eta_0(n-1))+\mathbf{x}_i(\lambda_1+\eta_1(n-1))\\ +\bar{\mathbf{x}}_i\eta_2(n-1)+\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'\end{matrix}\,\middle|\,\bar{\mathbf{x}}_i\right) && \text{(by (90))}\\
&= (\lambda_0+\eta_0(n-1))+L(\mathbf{x}_i|\bar{\mathbf{x}}_i)(\lambda_1+\eta_1(n-1)) && \text{(property of linear projection)}\\
&\quad +\bar{\mathbf{x}}_i\eta_2(n-1)+L(\mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}_i'|\bar{\mathbf{x}}_i) \\
&= (\lambda_0+\eta_0(n-1))+E(\mathbf{x}_i)(\lambda_1+\eta_1(n-1)) && (\text{RA} \implies \mathbf{x}_i \perp\!\!\!\perp \bar{\mathbf{x}}_i)\\
&\quad +\bar{\mathbf{x}}_i\eta_2(n-1)+E(\mathbf{x}_i)\eta_3(n-1)\bar{\mathbf{x}}_i' \\
&= \underbrace{(\lambda_0+\eta_0(n-1)+E(\mathbf{x}_i)(\lambda_1+\eta_1(n-1)))}_{\alpha_0+E(\mathbf{x}_i)\alpha_1}+\bar{\mathbf{x}}_i\underbrace{(\eta_2(n-1)+\eta_3'E(\mathbf{x}_i')(n-1))}_{\alpha_2} && (93)
\end{aligned}$$

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

$$\begin{aligned}
CAPE_k &= \mathbf{c}_{kK}\,(\eta_2+\eta_3'E(\mathbf{x}_i')) && \text{(by (92))}\\
&= \mathbf{c}_{kK}\frac{\alpha_2}{n-1} && \text{(by (93))}\\
&= \frac{\alpha_{2k}}{n-1}
\end{aligned}$$

which is the result in (37).

Lemma 1

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

$$\begin{aligned}
\Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{G} = \mathbf{G_A}, \mathbf{X} = \mathbf{X_A}) &= \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} = \mathbf{X_A}) && \text{(by } (\textbf{CRA}))\\
&= \prod_{i=1}^{N} \Pr(\tau_i = \tau_i(\mathbf{T_A}) | \mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A})) && \text{(by (1))}\\
&= \prod_{i=1}^{N} \frac{\Pr(\tau_i = \tau_i(\mathbf{T_A}) \cap \mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A}))}{\Pr(\mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A}))}\\
&= \prod_{i=1}^{N} \mathbb{I}\left(\mathbf{x}_i(\mathbf{X_A}) = \mathbf{x}(\tau_i(\mathbf{T_A}))\right) \frac{\Pr(\tau_i = \tau_i(\mathbf{T_A}))}{\Pr(\mathbf{x}_i = \mathbf{x}_i(\mathbf{X_A}))} && (94)
\end{aligned}$$

Therefore:

$$\begin{aligned}
E(y_i | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E\left(y_i(\mathbf{p}_i) | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}_i) = \bar{\mathbf{x}}\right) && \text{(by (4) and (16))}\\
&= E\left(y\left(\tau_i, \{\tau_j\}_{j \in \mathbf{p}_i}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}_i) = \bar{\mathbf{x}}\right)\\
&= \sum_{\mathbf{T_A} \in \mathcal{T}^N} \sum_{\mathbf{G_A} \in \mathcal{G}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j \in \mathbf{p}_i(\mathbf{G_A})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} \cap \mathbf{G} = \mathbf{G_A} | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}_i(\mathbf{G_A})) = \bar{\mathbf{x}}) \end{pmatrix}
\end{aligned}$$

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

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

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

$$\begin{aligned}
E(y_i | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= \sum_{\mathbf{G_A} \in \mathcal{G}^N} \Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \left(\sum_{\mathbf{T_A} \in \mathcal{T}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j \in \mathbf{p}_i(\mathbf{G_B})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}})) \end{pmatrix} \right)\\
&= \sum_{\mathbf{T_A} \in \mathcal{T}^N} \begin{pmatrix} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j \in \mathbf{p}_i(\mathbf{G_B})}\right) \\ * \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}})) \end{pmatrix} \overbrace{\left(\sum_{\mathbf{G_A} \in \mathcal{G}^N} \Pr(\mathbf{G} = \mathbf{G_A} | \mathbf{X} \in \chi_i(\mathbf{G_A}, \mathbf{x}, \bar{\mathbf{x}})) \right)}^{=1}\\
&= \sum_{\mathbf{T_A} \in \mathcal{T}^N} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j \in \mathbf{p}_i(\mathbf{G_B})}\right) \Pr(\mathbf{T} = \mathbf{T_A} | \mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}}))
\end{aligned} \tag{95}$$

Since equation (95) applies for all $\mathbf{G}$ that satisfy (CRA) it also applies for purely random $\mathbf{G}$. Let $\tilde{\mathbf{p}}$ be a purely random draw from $\mathcal{P}_i^{n-1}$. Then

$$E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}(\tilde{\mathbf{p}})) = \sum_{\mathbf{T_A} \in \mathcal{T}^N} y\left(\tau_i(\mathbf{T_A}), \{\tau_j(\mathbf{T_A})\}_{j \in \mathbf{p}_i(\mathbf{G_B})}\right) \Pr(\mathbf{T} = \mathbf{T_A}|\mathbf{X} \in \chi_i(\mathbf{G_B}, \mathbf{x}, \bar{\mathbf{x}})) \quad \text{(by (95))}$$

$$= E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) \quad \text{(also by (95))}$$

which is the result in (38).

Proposition 4

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

$$E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) = E(y_i(\tilde{\mathbf{p}}_i)|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i) = \bar{\mathbf{x}}) \quad \text{(by (38) in Lemma 1)}$$

Then:

1. Let $\tilde{\delta^S}$ be the coefficients from estimating equation (14) with counterfactual data on outcomes $y_i(\tilde{\mathbf{p}}_i)$ and peer group composition $\mathbf{z}_i^S(\tilde{\mathbf{p}}_i)$. Since $\mathbf{z}_i^S$ is saturated, the events $\mathbf{z}_i^S = \mathbf{c}_{mM^S}$ and $\{\bar{\mathbf{x}}_i\} = \bar{\mathbb{X}}^m$ are identical, implying that:

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

   which implies that:

$$\delta^S = \tilde{\delta^S} \tag{96}$$

   Since $\tilde{\mathbf{G}}$ satisfies (RA), part one of Proposition 3 applies to the counterfactiual $y_i(\tilde{\mathbf{p}}_i)$ and $\mathbf{z}_i^S(\tilde{\mathbf{p}}_i)$. Therefore:

$$HGE_{s,m}^S = \delta_{2m}^{\tilde{S}} + \delta_{3sm}^{\tilde{S}} \quad \text{(by (34) in Proposition 3)}$$
$$= \delta_{2m}^S + \delta_{3sm}^S \quad \text{(by (96))}$$

   Result (39) then follows from substitution of this result into result (33). Result (40) follows from substitution of result (39) into result (31).

2. Let $\tilde{\beta}$ be the coefficients from estimating equation (12) from with counterfactual outcomes $y_i(\tilde{\mathbf{p}}_i)$ and peer group composition $\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)$. Result (38) in Lemma 1 implies that:

$$\beta = \tilde{\beta} \tag{97}$$

   Since peer effects satistfy (PSE) and $\tilde{\mathbf{G}}$ satisfies (RA), part two of Proposition 3 applies to the counterfactiual $y_i(\tilde{\mathbf{p}}_i)$ and $\bar{\mathbf{x}}_i(\tilde{\mathbf{p}}_i)$. Therefore:

$$HPE_{s,k} = \frac{\tilde{\beta}_{2k} + \tilde{\beta}_{3sk}}{n-1} \quad \text{(by (36) in Proposition 3)}$$
$$= \frac{\beta_{2k} + \beta_{3sk}}{n-1} \quad \text{(by (97))}$$

   which is result (41). Result (42) follows from substitution of result (41) into result (30).

Proposition 5

1. Let $\tilde{\mathbf{p}}$ be a purely random draw from $\mathcal{P}_i^{n-1}$. By (45), both $\mathbf{G}_0$ and $\mathbf{G}_1$ satisfy (CRA) and Lemma 1 applies. Therefore:

$$\begin{aligned} E(y_{i0}|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_{i0} = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) && \text{(by (38) in Lemma 1)} \\ &= E(y_{i1}|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_{i1} = \bar{\mathbf{x}}) && (98) \end{aligned}$$

For any $m > 0$, $\bar{\mathbb{X}}^m$ is a singleton $\{\bar{\mathbf{x}}^m\}$ so the events $\bar{\mathbf{x}}_{i1} = \bar{\mathbf{x}}^m$ and $\mathbf{z}_{i1} = \mathbf{c}_{mM}$ are identical. Therefore, for any $m > 0$:

$$\begin{aligned} E(y_{i1}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_{i1} = \mathbf{c}_{mM}) &= E(y_{i1}|\mathbf{x}_i = \mathbf{c}_{sK}, \bar{\mathbf{x}}_{i1} = \bar{\mathbf{x}}^m) && \text{(identical events)} \\ &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}^m) && \text{(by (98))} \\ &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) && \text{(identical events)} \\ &= HGE_{s,m} + E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) && (99) \end{aligned}$$

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

$$\begin{aligned} E(y_{i1}|\mathbf{x}_i = \mathbf{c}_{sK}) &= \sum_{m=0}^{M} E(y_{i1}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_{i1} = \mathbf{c}_{mM}) \Pr(\mathbf{z}_{i1} = \mathbf{c}_{mM}|\mathbf{x}_i = \mathbf{c}_{sK}) \\ &= \sum_{m=1}^{M} E(y_{i1}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_{i1} = \mathbf{c}_{mM}) \Pr(\mathbf{z}_{i1} = \mathbf{c}_{mM}|\mathbf{x}_i = \mathbf{c}_{sK}) \\ &\qquad \text{(since } \Pr(\bar{\mathbf{x}}_{i1} \in \bar{\mathbb{X}}^0) = 0\text{)} \\ &= \sum_{m=1}^{M} (HGE_{s,m} + E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M})) \Pr(\mathbf{z}_{i1} = \mathbf{c}_{mM}|\mathbf{x}_i = \mathbf{c}_{sK}) \\ &\qquad \text{(by (99))} \\ &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \left(\underbrace{\sum_{m=1}^{M} \Pr(\mathbf{z}_{i1} = \mathbf{c}_{mM}|\mathbf{x}_i = \mathbf{c}_{sK})}_{1} \right) \\ &\quad + \sum_{m=1}^{M} HGE_{s,m} \Pr(\mathbf{z}_{i1} = \mathbf{c}_{mM}|\mathbf{x}_i = \mathbf{c}_{sK}) \\ &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) + \sum_{m=1}^{M} HGE_{s,m} \Pr(\mathbf{z}_{i1} = \mathbf{c}_{mM}|\mathbf{x}_i = \mathbf{c}_{sK}) \\ &\qquad (100) \end{aligned}$$

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

$$\begin{aligned}
CARE_s(\mathbf{G}_0,\mathbf{G}_1) &= E(y_{i1}-y_{i0}|\mathbf{x}_i=\mathbf{c}_{sK}) \\
&= E(y_{i1}|\mathbf{x}_i=\mathbf{c}_{sK}) - E(y_{i0}|\mathbf{x}_i=\mathbf{c}_{sK}) \\
&= \left(E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{c}_{0M}) + \sum_{m=1}^{M} HGE_{s,m}\Pr(\mathbf{z}_{i1}=\mathbf{c}_{mM}|\mathbf{x}_i=\mathbf{c}_{sK})\right) \\
&\quad - \left(E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{z}_i(\tilde{\mathbf{p}})=\mathbf{c}_{0M}) + \sum_{m=1}^{M} HGE_{s,m}\Pr(\mathbf{z}_{i0}=\mathbf{c}_{mM}|\mathbf{x}_i=\mathbf{c}_{sK})\right) \\
&\qquad \text{(by (100))} \\
&= \sum_{m=1}^{M} HGE_{s,m}\left(\Pr(\mathbf{z}_{i1}=\mathbf{c}_{mM}|\mathbf{x}_i=\mathbf{c}_{sK}) - \Pr(\mathbf{z}_{i0}=\mathbf{c}_{mM}|\mathbf{x}_i=\mathbf{c}_{sK})\right) \\
&= \mathbf{HGE}_s E(\mathbf{z}'_{i1}-\mathbf{z}'_{i0}|\mathbf{x}_i=\mathbf{c}_{sK})
\end{aligned}$$

which is the result in (53).

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

$$\begin{aligned}
E(y_{i1}|\mathbf{X},\mathbf{G}_1) &= E(y_i(\mathbf{p}_{i1})|\mathbf{X},\mathbf{G}_1) && \text{(definition)} \\
&= E\left(o_i + \sum_{j\in\mathbf{p}_{i1}} p_{ij} \,\middle|\, \mathbf{X},\mathbf{G}_1\right) && \text{(by Proposition 1)} \\
&= E\left(o_i|\,\mathbf{X},\mathbf{G}_1\right) + \sum_{j\in\mathbf{p}_{i1}} E\left(p_{ij}|\,\mathbf{X},\mathbf{G}_1\right) \\
&= E\left(E(o_i|\mathbf{X},\mathbf{G}_1,\sigma)|\,\mathbf{X},\mathbf{G}_1\right) + \sum_{j\in\mathbf{p}_{i1}} E\left(E(p_{ij}|\mathbf{X},\mathbf{G}_1,\sigma)|\,\mathbf{X},\mathbf{G}_1\right) \\
& && \text{(law of iterated expectations)} \\
&= E\left(E(o_i|\mathbf{X},\sigma)|\,\mathbf{X},\mathbf{G}_1\right) + \sum_{j\in\mathbf{p}_{i1}} E\left(E(p_{ij}|\mathbf{X},\sigma)|\,\mathbf{X},\mathbf{G}_1\right) \\
& && \text{(since } \mathbf{G}_1 \text{ is a function of } (\mathbf{X},\sigma)) \\
&= E\left(E(o_i|\mathbf{x}_i)|\,\mathbf{X},\mathbf{G}_1\right) + \sum_{j\in\mathbf{p}_{i1}} E\left(E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j)|\,\mathbf{X},\mathbf{G}_1\right) \\
&= E(o_i|\mathbf{x}_i) + \sum_{j\in\mathbf{p}_{i1}} E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j) \\
&= \lambda_0 + \mathbf{x}_i\lambda_1 + \sum_{j\in\mathbf{p}_{i1}} \eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}'_j \\
& && \text{(where } (\lambda,\eta) \text{ are defined as in (86))} \\
&= (\lambda_0+\eta_0(n-1)) + \mathbf{x}_i(\lambda_1+\eta_1(n-1)) + \bar{\mathbf{x}}_{i1}\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}'_{i1} && (101)
\end{aligned}$$

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

$$\begin{aligned}
E(y_{i1}|\mathbf{x}_i=\mathbf{x}) &= E(E(y_{i1}|\mathbf{X},\mathbf{G}_1)|\mathbf{x}_i=\mathbf{x}) && \text{(Law of iterated expectations)} \\
&= E\left(\begin{array}{l} (\lambda_0+\eta_0(n-1)) + \mathbf{x}_i(\lambda_1+\eta_1(n-1)) \\ +\bar{\mathbf{x}}_{i1}\eta_2(n-1) + \mathbf{x}_i\eta_3(n-1)\bar{\mathbf{x}}'_{i1} \end{array} \middle|\, \mathbf{x}_i=\mathbf{x}\right) && \text{(by (101))} \\
&= (\lambda_0+\eta_0(n-1)) + \mathbf{x}(\lambda_1+\eta_1(n-1)) && (102) \\
&\quad + E(\bar{\mathbf{x}}_{i1}|\mathbf{x}_i=\mathbf{x})\eta_2(n-1) + \mathbf{x}\eta_3(n-1)E(\bar{\mathbf{x}}'_{i1}|\mathbf{x}_i=\mathbf{x})
\end{aligned}$$

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

$$
\begin{aligned}
CARE_s(\mathbf{G}_0, \mathbf{G}_1) &= E(y_{i1} - y_{i0} | \mathbf{x}_i = \mathbf{c}_{sK}) \\
&= \begin{pmatrix} (\lambda_0 + \eta_0(n-1)) + \mathbf{c}_{sK}(\lambda_1 + \eta_1(n-1)) \\ + E(\bar{\mathbf{x}}_{i1}|\mathbf{x}_i = \mathbf{c}_{sK})\eta_2(n-1) + \mathbf{c}_{sK}\eta_3(n-1)E(\bar{\mathbf{x}}'_{i1}|\mathbf{x}_i = \mathbf{c}_{sK}) \end{pmatrix} \\
&\quad - \begin{pmatrix} (\lambda_0 + \eta_0(n-1)) + \mathbf{c}_{sK}(\lambda_1 + \eta_1(n-1)) \\ + E(\bar{\mathbf{x}}_{i0}|\mathbf{x}_i = \mathbf{c}_{sK})\eta_2(n-1) + \mathbf{c}_{sK}\eta_3(n-1)E(\bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK}) \end{pmatrix} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(by (102))} \\
&= (\eta'_2(n-1) + \mathbf{c}_{sK}\eta_3(n-1))\left(E(\bar{\mathbf{x}}'_{i1}|\mathbf{x}_i = \mathbf{c}_{sK}) - E(\bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK})\right) \\
&= \mathbf{HPE}_s E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK})(n-1)
\end{aligned}
$$

which is the result in (54).

3. Given (PSE, OSE), part two of Proposition 1 applies. By equation (27) in Proposition 1, $\mathbf{HPE}_s = \mathbf{CAPE}$ and so the the result in (55) follows from (54) by substitution. The second result follows from:

$$
\begin{aligned}
E(y_{i1} - y_{i0}) &= \sum_{s=0}^{K} E(y_{i1} - y_{i0}|\mathbf{x}_i = \mathbf{c}_{sK}) \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) && \text{(law of total probability)} \\
&= \sum_{s=0}^{K} CARE_s(\mathbf{G}_0, \mathbf{G}_1) \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) && \text{(definition of } CARE_s\text{)} \\
&= \sum_{s=0}^{K} \mathbf{CAPE}\, E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK})(n-1) \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) && \text{(by result (55))} \\
&= \mathbf{CAPE}(n-1) \sum_{s=0}^{K} E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0}|\mathbf{x}_i = \mathbf{c}_{sK}) \Pr(\mathbf{x}_i = \mathbf{c}_{sK}) \\
&= \mathbf{CAPE}(n-1) \underbrace{E(\bar{\mathbf{x}}'_{i1} - \bar{\mathbf{x}}'_{i0})}_{0} \\
&= 0
\end{aligned}
$$

Proposition 6

1. The proof here is essentially the same as the proof for part two of Proposition 3, but conditioning on ℓ_i. Given (PSE), Proposition 1 implies the potential outcome function can be written as in equation (23) and within-location conditional average peer effects can be written in terms of p_{ij}:

$$
\begin{aligned}
HPE^{\ell}_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell) \\
&\quad - E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell) \\
CAPE^{\ell}_{k} &= E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell) - E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell)
\end{aligned} \tag{103}
$$

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

$$
\begin{aligned}
E(o_i|\mathbf{x}_i, \ell_i = \ell) &= \lambda^{\ell}_0 + \mathbf{x}_i\lambda^{\ell}_1 \\
E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell, \ell_j = \ell') &= \eta^{\ell\ell'}_0 + \mathbf{x}_i\eta^{\ell\ell'}_1 + \mathbf{x}_j\eta^{\ell\ell'}_2 + \mathbf{x}_i\eta^{\ell\ell'}_3\mathbf{x}'_j
\end{aligned} \tag{104}
$$

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

$$\begin{aligned}
HPE^\ell_{s,k} &= E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell) && \text{(by (103))} \\
&\quad - E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell) \\
&= \left(\eta_0^{\ell\ell} + \mathbf{c}_{sK}\eta_1^{\ell\ell} + \mathbf{c}_{kK}\eta_2^{\ell\ell} + \mathbf{c}_{sK}\eta_3^{\ell\ell}\mathbf{c}'_{kK}\right) && \text{(by (104))} \\
&\quad - \left(\eta_0^{\ell\ell} + \mathbf{c}_{sK}\eta_1^{\ell\ell} + \mathbf{c}_{0K}\eta_2^{\ell\ell} + \mathbf{c}_{sK}\eta_3^{\ell\ell}\mathbf{c}'_{0K}\right) \\
&= \left(\eta_0^{\ell\ell} + \mathbf{c}_{sK}\eta_1^{\ell\ell} + \mathbf{c}_{kK}\eta_2^{\ell\ell} + \mathbf{c}_{sK}\eta_3^{\ell\ell}\mathbf{c}'_{kK}\right) - \left(\eta_0^{\ell\ell} + \mathbf{c}_{sK}\eta_1^{\ell\ell}\right) && \text{(since } \mathbf{c}_{0K} = 0) \\
&= \mathbf{c}_{kK}\eta_2^{\ell\ell} + \mathbf{c}_{sK}\eta_3^{\ell\ell}\mathbf{c}'_{kK} \\
&= \eta_{2k}^{\ell\ell} + \eta_{3sk}^{\ell\ell} && (105)
\end{aligned}$$

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

$$\begin{aligned}
E(y_i|\mathbf{X}, \mathbf{G}, \mathbf{L}) &= E\left(o_i + \sum_{j\in\mathbf{p}_i} p_{ij} \middle| \mathbf{X}, \mathbf{G}, \mathbf{L}\right) && \text{(by Proposition 1)} \\
&= E\left(o_i + \sum_{j=1}^{N} p_{ij}\mathbb{I}\left(j \in \mathbf{p}_i\right) \middle| \mathbf{X}, \mathbf{G}, \mathbf{L}\right) \\
& && \text{(where } \mathbb{I}\left(\right) \text{ is the indicator function)} \\
&= E(o_i|\mathbf{X}, \mathbf{G}, \mathbf{L}) + \sum_{j=1}^{N} E(p_{ij}\mathbb{I}\left(j \in \mathbf{p}_i\right)|\mathbf{X}, \mathbf{G}, \mathbf{L}) \\
&= E(o_i|\mathbf{X}, \mathbf{G}, \mathbf{L}) + \sum_{j=1}^{N} E(p_{ij}|\mathbf{X}, \mathbf{G}, \mathbf{L})\mathbb{I}\left(j \in \mathbf{p}_i\right) \\
& && \text{(since } \mathbb{I}\left(j \in \mathbf{p}_i\right) \text{ is a function of } \mathbf{G}) \\
&= E(o_i|\mathbf{X}, \mathbf{G}, \mathbf{L}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X}, \mathbf{G}, \mathbf{L}) \\
&= E(o_i|\mathbf{X}, \mathbf{L}) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{X}, \mathbf{L}) && (\text{RAL} \implies \mathbf{T} \perp\!\!\!\perp \mathbf{G}|\mathbf{L}) \\
&= E(o_i|\mathbf{x}_i, \ell_i) + \sum_{j\in\mathbf{p}_i} E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i, \ell_j = \ell_i) && \text{(by (57))} \\
&= \lambda_0^{\ell_i} + \mathbf{x}_i\lambda_1^{\ell_i} + \sum_{j\in\mathbf{p}_i}\left(\eta_0^{\ell_i\ell_i} + \mathbf{x}_i\eta_1^{\ell_i\ell_i} + \mathbf{x}_j\eta_2^{\ell_i\ell_i} + \mathbf{x}_i\eta_3^{\ell_i\ell_i}\mathbf{x}'_j\right) && \text{(by (104))} \\
&= \left(\lambda_0^{\ell_i} + \eta_0^{\ell_i\ell_i}(n-1)\right) + \mathbf{x}_i\left(\lambda_1^{\ell_i\ell_i} + \eta_1^{\ell_i\ell_i}(n-1)\right) && (106) \\
&\quad + \bar{\mathbf{x}}_i\eta_2^{\ell_i\ell_i}(n-1) + \mathbf{x}_i\eta_3^{\ell_i\ell_i}(n-1)\bar{\mathbf{x}}'_i
\end{aligned}$$

Applying the law of iterated projections:

$$
\begin{aligned}
L^{\ell}\left(y_i \middle| \begin{matrix}\mathbf{x}_i, \bar{\mathbf{x}}_i, \\ \mathbf{x}_i'\bar{\mathbf{x}}_i\end{matrix}\right) &= L^{\ell}\left(E\left(y_i \middle| \mathbf{X}, \mathbf{G}, \mathbf{L}\right) \middle| \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)} \\
&= L^{\ell}\left(\begin{matrix}(\lambda_0^{\ell_i} + \eta_0^{\ell_i}(n-1)) \\ + \mathbf{x}_i(\lambda_1^{\ell_i\ell_i} + \eta_1^{\ell_i\ell_i}(n-1)) \\ + \bar{\mathbf{x}}_i\eta_2^{\ell_i\ell_i}(n-1) \\ + \mathbf{x}_i\eta_3^{\ell_i\ell_i}(n-1)\bar{\mathbf{x}}_i'\end{matrix} \middle| \mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}_i'\bar{\mathbf{x}}_i \right) && \text{(by (106))} \\
&= \underbrace{(\lambda_0^{\ell} + \eta_0^{\ell\ell}(n-1))}_{\beta_0^{\ell}} + \mathbf{x}_i \underbrace{(\lambda_1^{\ell} + \eta_1^{\ell\ell}(n-1))}_{\beta_1^{\ell}} && (107) \\
&\quad + \bar{\mathbf{x}}_i \underbrace{\eta_2^{\ell\ell}(n-1)}_{\beta_2^{\ell}} + \mathbf{x}_i \underbrace{\eta_3^{\ell\ell}(n-1)}_{\beta_3^{\ell}} \bar{\mathbf{x}}_i'
\end{aligned}
$$

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

$$
\begin{aligned}
HPE_{s,k}^{\ell} &= \eta_{2k}^{\ell\ell} + \eta_{3sk}^{\ell\ell} && \text{(by (105))} \\
&= \frac{\beta_{2k}^{\ell} + \beta_{3sk}^{\ell}}{n-1} && \text{(by (107))}
\end{aligned}
$$

which is the result in (63). The same procedure can be used to derive the result in (64). First, express $\mathbf{CAPE}^{\ell}$ in terms of $\eta^{\ell\ell}$:

$$
\begin{aligned}
CAPE_k^{\ell} &= E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell) - E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell) && \text{(by (103))} \\
&= E\left(E\left(p_{ij} \middle| \mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell_j = \ell\right) \middle| \mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell\right) \\
&\quad - E\left(E\left(p_{ij} \middle| \mathbf{x}_i, \mathbf{x}_j, \ell_i = \ell_j = \ell\right) \middle| \mathbf{x}_j = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell\right) \\
& && \text{(law of iterated expectations)} \\
&= E\left(\eta_0^{\ell\ell} + \mathbf{x}_i\eta_1^{\ell\ell} + \mathbf{x}_j\eta_2^{\ell\ell} + \mathbf{x}_i\eta_3^{\ell\ell}\mathbf{x}_j' \middle| \mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell\right) && \text{(by 104))} \\
&\quad - E\left(\eta_0^{\ell\ell} + \mathbf{x}_i\eta_1^{\ell\ell} + \mathbf{x}_j\eta_2^{\ell\ell} + \mathbf{x}_i\eta_3^{\ell\ell}\mathbf{x}_j' \middle| \mathbf{x}_j = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell\right) \\
&= \eta_0^{\ell\ell} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell)\eta_1^{\ell\ell} && \text{(since } \mathbf{c}_{0K} = 0\text{)} \\
&\quad + \mathbf{c}_{kK}\eta_2^{\ell\ell} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{c}_{kK}, \ell_i = \ell_j = \ell)\eta_3^{\ell\ell}\mathbf{c}_{kK}' \\
&\quad - \left(\eta_0^{\ell\ell} + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{c}_{0K}, \ell_i = \ell_j = \ell)\eta_1^{\ell\ell}\right) \\
&= \eta_0^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell_j = \ell)\eta_1^{\ell\ell} && \text{((57)} \implies \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j|\mathbf{L}\text{)} \\
&\quad + \mathbf{c}_{kK}\eta_2^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell_j = \ell)\eta_3^{\ell\ell}\mathbf{c}_{kK}' \\
&\quad - \left(\eta_0^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell_j = \ell)\eta_1^{\ell\ell}\right) \\
&= \mathbf{c}_{kK}\eta_2^{\ell\ell} + E(\mathbf{x}_i|\ell_i = \ell)\eta_3^{\ell\ell}\mathbf{c}_{kK}' \\
&= \mathbf{c}_{kK}\left(\eta_2^{\ell\ell} + (\eta_3^{\ell\ell})'E(\mathbf{x}_i'|\ell_i = \ell)\right) && (108)
\end{aligned}
$$

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

$$\begin{aligned}
L^\ell(y_i|\bar{\mathbf{x}}_i) &= L^\ell\left(L^\ell\left(y_i|\,\mathbf{x}_i,\bar{\mathbf{x}}_i\right)\middle|\,\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)}\\
&= L^\ell\left(\alpha_0^\ell+\mathbf{x}_i\alpha_1^\ell+\bar{\mathbf{x}}_i\alpha_2^\ell\middle|\,\bar{\mathbf{x}}_i\right) && \text{(by (58))}\\
&= \alpha_0^\ell+L^\ell(\mathbf{x}_i|\bar{\mathbf{x}}_i)\alpha_1^\ell+\bar{\mathbf{x}}_i\alpha_2^\ell\\
&= \alpha_0^\ell+E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell+\bar{\mathbf{x}}_i\alpha_2^\ell && \text{(by (57))}\\
&= \left(\alpha_0^\ell+E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell\right)+\bar{\mathbf{x}}_i\alpha_2^\ell && (109)\\
L^\ell(y_i|\bar{\mathbf{x}}_i) &= L^\ell\left(E\left(y_i|\,\mathbf{X},\mathbf{G},\mathbf{L}\right)|\,\bar{\mathbf{x}}_i\right) && \text{(law of iterated projections)}\\
&= L^\ell\left(\begin{array}{l|}(\lambda_0^{\ell_i}+\eta_0^{\ell_i}(n-1))\\ +\mathbf{x}_i(\lambda_1^{\ell_i\ell_i}+\eta_1^{\ell_i\ell_i}(n-1))\\ +\bar{\mathbf{x}}_i\eta_2^{\ell_i\ell_i}(n-1)\\ +\mathbf{x}_i\eta_3^{\ell_i\ell_i}(n-1)\bar{\mathbf{x}}_i'\end{array}\;\bar{\mathbf{x}}_i\right) && \text{(by (107))}\\
&= (\lambda_0^\ell+\eta_0^\ell(n-1))+L^\ell(\mathbf{x}_i|\bar{\mathbf{x}}_i)(\lambda_1^{\ell\ell}+\eta_1^{\ell\ell}(n-1))\\
&\quad+\bar{\mathbf{x}}_i\eta_2^{\ell\ell}(n-1)+L^\ell(\mathbf{x}_i|\bar{\mathbf{x}}_i)\eta_3^{\ell\ell}(n-1)\bar{\mathbf{x}}_i'\\
&= (\lambda_0^\ell+\eta_0^\ell(n-1))+E(\mathbf{x}_i|\ell_i=\ell)(\lambda_1^{\ell\ell}+\eta_1^{\ell\ell}(n-1)) && \text{(by (57))}\\
&\quad+\bar{\mathbf{x}}_i\eta_2^{\ell\ell}(n-1)+E(\mathbf{x}_i|\ell_i=\ell)\eta_3^{\ell\ell}(n-1)\bar{\mathbf{x}}_i'\\
&= \underbrace{(\lambda_0^\ell+\eta_0^\ell(n-1))+E(\mathbf{x}_i|\ell_i=\ell)(\lambda_1^{\ell\ell}+\eta_1^{\ell\ell}(n-1))}_{\alpha_0^\ell+E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell} && (110)\\
&\quad+\bar{\mathbf{x}}_i\underbrace{\left(\eta_2^{\ell\ell}(n-1)+(\eta_3^{\ell\ell})'E(\mathbf{x}_i'|\ell_i=\ell)(n-1)\right)}_{\alpha_2^\ell}
\end{aligned}$$

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

$$\begin{aligned}
CAPE_k^\ell &= \mathbf{c}_{kK}(\eta_2^{\ell\ell}+(\eta_3^{\ell\ell})'E(\mathbf{x}_i'|\ell_i=\ell)) && \text{(by (108))}\\
&= \mathbf{c}_{kK}\frac{\alpha_2^\ell}{n-1} && \text{(by (109) and (110))}\\
&= \frac{\alpha_{2k}^\ell}{n-1}
\end{aligned}$$

which is the result in (64).

2. First note that:

$$\begin{aligned}
HPE_{s,k} &= E(p_{ij}|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK})-E(p_{ij}|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{0K}) && \text{(by Proposition 1)}\\
&= E(E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j,\ell_i,\ell_j)|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK}) && \text{(law of iterated expectations)}\\
&\quad-E(E(p_{ij}|\mathbf{x}_i,\mathbf{x}_j,\ell_i,\ell_j)|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{0K})\\
&= E(\eta_0^{\ell_i\ell_j}+\mathbf{x}_i\eta_1^{\ell_i\ell_j}+\mathbf{x}_j\eta_2^{\ell_i\ell_j}+\mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK}) && \text{(by (104))}\\
&\quad-E(\eta_0^{\ell_i\ell_j}+\mathbf{x}_i\eta_1^{\ell_i\ell_j}+\mathbf{x}_j\eta_2^{\ell_i\ell_i}+\mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{0K})\\
&= E(\eta_0^{\ell_i\ell_j}+\mathbf{x}_i\eta_1^{\ell_i\ell_j}+\mathbf{x}_j\eta_2^{\ell_i\ell_j}+\mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{kK}) && (111)\\
&\quad-E(\eta_0^{\ell_i\ell_j}+\mathbf{x}_i\eta_1^{\ell_i\ell_j}|\mathbf{x}_i=\mathbf{c}_{sK},\mathbf{x}_j=\mathbf{c}_{0K})
\end{aligned}$$

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

$$\eta^{\ell\ell'}=\eta \qquad \text{for all } \ell,\ell' \tag{112}$$

Applying (112) to (111) allows $HPE_{s,k}$ to be expressed in terms of η:

$$\begin{aligned}
HPE_{s,k} &= E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}'_j|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) && \text{(by (112))}\\
&\quad - E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}'_j|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{0K})\\
&= (\eta_0 + \mathbf{c}_{sK}\eta_1 + \mathbf{c}_{kK}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{kK}) - (\eta_0 + \mathbf{c}_{sK}\eta_1 + \mathbf{c}_{0K}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{0K})\\
&= \mathbf{c}_{kK}\eta_2 + \mathbf{c}_{sK}\eta_3\mathbf{c}'_{kK}\\
&= \eta_{2k} + \eta_{3sk} && (113)
\end{aligned}$$

Applying (112) to (107) allows $L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}'_i\bar{\mathbf{x}}_i)$ to be expressed in terms of η:

$$L^\ell(y_i|\mathbf{x}_i, \bar{\mathbf{x}}_i, \mathbf{x}'_i\bar{\mathbf{x}}_i) = \underbrace{(\lambda_0^\ell + \eta_0(n-1))}_{\beta_0^\ell} + \mathbf{x}_i\underbrace{(\lambda_1^\ell + \eta_1(n-1))}_{\beta_1^\ell} + \bar{\mathbf{x}}_i\underbrace{\eta_2(n-1)}_{\beta_2^\ell} + \mathbf{x}_i\underbrace{\eta_3(n-1)}_{\beta_3^\ell}\bar{\mathbf{x}}'_i \tag{114}$$

which implies that:

$$\begin{aligned}
HPE_{s,k} &= \eta_{2k} + \eta_{3sk} && \text{(by (113))}\\
&= \frac{E(\beta_{2k}^{\ell_i}) + E(\beta_{3sk}^{\ell_i})}{n-1} && \text{(by (114))}
\end{aligned}$$

which is the result in (65).[5] To prove the result in (66), first note that:

$$\begin{aligned}
CAPE_k &= E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{kK}) - E(p_{ij}|\mathbf{x}_j = \mathbf{c}_{0K}) && \text{(by Proposition 1)}\\
&= \quad E(E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i, \ell_j)|\mathbf{x}_j = \mathbf{c}_{kK}) && \text{(law of iterated expectations)}\\
&\quad -E(E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i, \ell_j)|\mathbf{x}_j = \mathbf{c}_{0K})\\
&= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}'_j|\mathbf{x}_j = \mathbf{c}_{kK}) && \text{(by (104))}\\
&\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_i} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}'_j|\mathbf{x}_j = \mathbf{c}_{0K})\\
&= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}'_j|\mathbf{x}_j = \mathbf{c}_{kK}) && (115)\\
&\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j}|\mathbf{x}_j = \mathbf{c}_{0K})
\end{aligned}$$

Applying (112) to (115) allows $CAPE_k$ to be expressed in terms of η:

$$\begin{aligned}
CAPE_k &= E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}'_j|\mathbf{x}_j = \mathbf{c}_{kK}) && \text{(by (112))}\\
&\quad - E(\eta_0 + \mathbf{x}_i\eta_1 + \mathbf{x}_j\eta_2 + \mathbf{x}_i\eta_3\mathbf{x}'_j|\mathbf{x}_j = \mathbf{c}_{0K})\\
&= \ (\eta_0 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{c}_{kK})\eta_1 + \mathbf{c}_{kK}\eta_2 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{c}_{kK})\eta_3\mathbf{c}'_{kK})\\
&\quad - (\eta_0 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{c}_{0K})\eta_1 + \mathbf{c}_{0K}\eta_2 + E(\mathbf{x}_i|\mathbf{x}_j = \mathbf{c}_{0K})\eta_3\mathbf{c}'_{0K})\\
&= \ (\eta_0 + E(\mathbf{x}_i)\eta_1 + \mathbf{c}_{kK}\eta_2 + E(\mathbf{x}_i)\eta_3\mathbf{c}'_{kK}) && \text{(since } \mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j\text{)}\\
&\quad - (\eta_0 + E(\mathbf{x}_i)\eta_1 + \mathbf{c}_{0K}\eta_2 + E(\mathbf{x}_i)\eta_3\mathbf{c}'_{0K})\\
&= \mathbf{c}_{kK}\eta_2 + E(\mathbf{x}_i)\eta_3\mathbf{c}'_{kK} && \text{(since } \mathbf{c}_{0K} = 0\text{)}\\
&= \mathbf{c}_{kK}(\eta_2 + \eta'_3E(\mathbf{x}'_i)) && (116)
\end{aligned}$$

[5]Note that β_2^ℓ, β_3^ℓ, and α_2^ℓ do not vary by ℓ, so it is not strictly necessary to average across locations in (65) and (66) rather than simply choosing an arbitrary location. In a finite sample, an average of noisy estimators would typically outperform any one of those estimators.

Applying (112) to (110) allows $L^\ell(y_i|\bar{\mathbf{x}}_i)$ to be expressed in terms of η:

$$\begin{aligned}L^\ell(y_i|\bar{\mathbf{x}}_i) = &\underbrace{(\lambda_0^\ell + \eta_0(n-1)) + E(\mathbf{x}_i|\ell_i=\ell)(\lambda_1^{\ell\ell} + \eta_1(n-1))}_{\alpha_0^\ell + E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell} \\ &+ \bar{\mathbf{x}}_i \underbrace{(\eta_2(n-1) + \eta_3' E(\mathbf{x}_i'|\ell_i=\ell)(n-1))}_{\alpha_2^\ell}\end{aligned} \tag{117}$$

which implies that:

$$\begin{aligned}E(\alpha_2^{\ell_i}) &= E\left(\eta_2(n-1) + \eta_3' E(\mathbf{x}_i'|\ell_i)(n-1)\right) && \text{(by (117))} \\ &= \left(\eta_2 + \eta_3' E(E(\mathbf{x}_i'|\ell_i))\right)(n-1) \\ &= \left(\eta_2 + \eta_3' E(\mathbf{x}_i')\right)(n-1) && (118)\end{aligned}$$

and therefore:

$$\begin{aligned}CAPE_k &= \mathbf{c}_{kK}(\eta_2 + \eta_3' E(\mathbf{x}_i')) && \text{(by (116))} \\ &= \frac{\mathbf{c}_{kK} E(\alpha_2^{\ell_i})}{n-1} && \text{(by (118))} \\ &= \frac{E(\alpha_{2k}^{\ell_i})}{n-1}\end{aligned}$$

which is the result in (66).

3. Assumption (OSE) implies that:

$$\begin{aligned}E(p_{ij}|\mathbf{x}_i, \mathbf{x}_j, \ell_i=\ell, \ell_j=\ell') &= E(p_j|\mathbf{x}_i, \mathbf{x}_j, \ell_i=\ell, \ell_j=\ell') && \text{(by OSE)} \\ &= E(p_j|\mathbf{x}_j, \ell_j=\ell') && \text{(by (57))} \\ \implies (\eta_0^{\ell\ell'}, \eta_1^{\ell\ell'}, \eta_2^{\ell\ell'}, \eta_3^{\ell\ell'}) &= (\eta_0^{\ell'}, 0, \eta_2^{\ell'}, 0) && (119)\end{aligned}$$

Applying (119) to (110) produces:

$$L^\ell(y_i|\bar{\mathbf{x}}_i) = \underbrace{(\lambda_0^\ell + \eta_0^\ell(n-1)) + E(\mathbf{x}_i|\ell_i=\ell)\lambda_1^\ell}_{\alpha_0^\ell + E(\mathbf{x}_i|\ell_i=\ell)\alpha_1^\ell} + \bar{\mathbf{x}}_i \underbrace{\eta_2^\ell(n-1)}_{\alpha_2^\ell} \tag{120}$$

Applying (119) to (115) produces:

$$\begin{aligned}CAPE_k &= E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j} + \mathbf{x}_j\eta_2^{\ell_i\ell_j} + \mathbf{x}_i\eta_3^{\ell_i\ell_j}\mathbf{x}_j'|\mathbf{x}_j = \mathbf{c}_{kK}) && \text{(by (115))} \\ &\quad - E(\eta_0^{\ell_i\ell_j} + \mathbf{x}_i\eta_1^{\ell_i\ell_j}|\mathbf{x}_j = \mathbf{c}_{0K}) \\ &= E(\eta_0^{\ell_j} + \mathbf{x}_j\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{kK}) && \text{(by (119))} \\ &\quad - E(\eta_0^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{0K}) \\ &= E(\eta_0^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{kK}) - E(\eta_0^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{0K}) + \mathbf{c}_{kK}E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{kK}) && (121)\end{aligned}$$

The third term in this expression is identified, but the first two terms are not, because η_0^ℓ

cannot be distinguished from λ_0^ℓ. However, assumption (PLI) implies that $\eta_0^{\ell'} = \eta_0$, so:

$$\begin{aligned}
CAPE_k &= E(\eta_0|\mathbf{x}_j = \mathbf{c}_{kK}) - E(\eta_0|\mathbf{x}_j = \mathbf{c}_{0K}) + \mathbf{c}_{kK}E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{kK}) && \text{(by PLI)} \\
&= \eta_0 - \eta_0 + \mathbf{c}_{kK}E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{kK}) \\
&= \mathbf{c}_{kK}E(\eta_2^{\ell_j}|\mathbf{x}_j = \mathbf{c}_{kK}) \\
&= \mathbf{c}_{kK}E(\eta_2^{\ell_i}|\mathbf{x}_i = \mathbf{c}_{kK}) && \text{(by exchangeability)} \\
&= \frac{\mathbf{c}_{kK}E(\alpha_2^{\ell_i}|\mathbf{x}_i = \mathbf{c}_{kK})}{n-1} && \text{(by (120))} \\
&= \frac{E(\alpha_{2k}^{\ell_i}|\mathbf{x}_i = \mathbf{c}_{kK})}{n-1}
\end{aligned}$$

which is the result in (67).

Proposition 7

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

$$\begin{aligned}
y_i(j \cup \mathbf{q}) - y_i(j' \cup \mathbf{q}) &= \Big(h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) + \epsilon_i\Big) - \Big(h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) + \epsilon_i\Big) \\
&= h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) - h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}}) && (122)
\end{aligned}$$

and (PSE) implies:

$$\begin{aligned}
y_i(j \cup \mathbf{q}) - y_i(j' \cup \mathbf{q}) &= y_i(j \cup \mathbf{q}_0) - y_i(j' \cup \mathbf{q}_0) && \text{(by (PSE))} \\
&= h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}_0}) - h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{\mathbf{x}_k^*\}_{k\in\mathbf{q}_0}) && \text{(by (122))} \\
&= h(\mathbf{x}_i^*, \{\mathbf{x}_j^*\} \cup \{0, 0, \ldots, 0\}) - h(\mathbf{x}_i^*, \{\mathbf{x}_{j'}^*\} \cup \{0, 0, \ldots, 0\}) && (123)
\end{aligned}$$

Let:

$$\begin{aligned}
h_1(\mathbf{x}_i^*) &\equiv h\left(\mathbf{x}_i^*, \{0, \ldots, 0\}\right) && (124) \\
h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) &\equiv h\left(\mathbf{x}_i^*, \{\mathbf{x}_j^*, \ldots, 0\}\right) - h\left(\mathbf{x}_i^*, \{0, \ldots, 0\}\right)
\end{aligned}$$

Then:

$$
\begin{aligned}
y_i(\mathbf{p}) &= h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \dots, \mathbf{x}_{\mathbf{p}((n-1))}^*\right\}\right) + \epsilon_i && \text{(by DCE)} \\
&= \begin{pmatrix}
h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right) \\
+h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, 0, \dots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right) \\
+h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \mathbf{x}_{\mathbf{p}(2)}^*, 0, \dots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, 0, \dots, 0\right\}\right) \\
+ \quad \vdots \\
+h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \dots, \mathbf{x}_{\mathbf{p}((n-1))}^*\right\}\right) - h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, \dots, \mathbf{x}_{\mathbf{p}((n-1)-1)}^*, 0\right\}\right)
\end{pmatrix} + \epsilon_i \\
&= \begin{pmatrix}
h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right) \\
+h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(1)}^*, 0, \dots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right) \\
+h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}(2)}^*, 0, \dots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right) \\
+ \quad \vdots \\
+h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_{\mathbf{p}((n-1))}^*, 0, \dots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right)
\end{pmatrix} + \epsilon_i && \text{(by (122) and (123))} \\
&= h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right) + \sum_{j \in \mathbf{p}} h\left(\mathbf{x}_i^*, \left\{\mathbf{x}_j^*, 0, \dots, 0\right\}\right) - h\left(\mathbf{x}_i^*, \{0, \dots, 0\}\right) + \epsilon_i \\
&= h_1(\mathbf{x}_i^*) + \sum_{j \in \mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i
\end{aligned}
$$

which is the first result in equation (71). If $\mathbf{x}^*$ is categorical, then these functions can be written in the form:

$$
\begin{aligned}
h_1(\mathbf{x}_i^*) &= \theta_0 + \mathbf{x}_i^* \theta_1 \\
h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) &= \phi_0 + \mathbf{x}_i^* \phi_1 + \mathbf{x}_j^* \phi_2 + \mathbf{x}_i^* \phi_3 (\mathbf{x}_j^*)'
\end{aligned} \tag{125}
$$

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

$$
\begin{aligned}
y_i(\mathbf{p}) &= h_1(\mathbf{x}_i^*) + \sum_{j \in \mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i && \text{(by (71))} \\
&= \theta_0 + \mathbf{x}_i^* \theta_1 + \sum_{j \in \mathbf{p}} \left(\mathbf{x}_j^* \phi_2 + \mathbf{x}_i^* \phi_3 (\mathbf{x}_j^*)'\right) + \epsilon_i && \text{(by (125))} \\
&= \theta_0 + \mathbf{x}_i^* \theta_1 + \left(\sum_{j \in \mathbf{p}} \mathbf{x}_j^*\right) \phi_2 + \mathbf{x}_i^* \phi_3 \left(\sum_{j \in \mathbf{p}} \mathbf{x}_j^*\right)' + \epsilon_i \\
&= \theta_0 + \mathbf{x}_i^* \theta_1 + \bar{\mathbf{x}}_i^*(\mathbf{p}) \underbrace{\phi_2 (n-1)}_{\theta_2} + \mathbf{x}_i^* \underbrace{\phi_3 (n-1)}_{\theta_2} \bar{\mathbf{x}}_i^*(\mathbf{p})' + \epsilon_i
\end{aligned}
$$

which is the last result in equation (71).

2. First note that:

$$
\begin{aligned}
y_i(\mathbf{p}) - y_i(\mathbf{p}') &= \left(h_1(\mathbf{x}_i^*) + \sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i \right) - \left(h_1(\mathbf{x}_i^*) + \sum_{j\in\mathbf{p}'} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) + \epsilon_i \right) && \text{(by (71))} \\
&= \left(\sum_{j\in\mathbf{p}} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) \right) - \left(\sum_{j\in\mathbf{p}'} h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) \right) && (126)
\end{aligned}
$$

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

$$
\begin{aligned}
y_i(\mathbf{p}) - y_i(\mathbf{p}') &= y_{i'}(\mathbf{p}) - y_{i'}(\mathbf{p}) && \text{(by (OSE))} \\
&= \left(\sum_{j\in\mathbf{p}} h_2(0, \mathbf{x}_j^*) \right) - \left(\sum_{j\in\mathbf{p}'} h_2(0, \mathbf{x}_j^*) \right) && \text{(by (126))} \\
&= \left(\sum_{j\in\mathbf{p}} h_3(\mathbf{x}_j^*) \right) - \left(\sum_{j\in\mathbf{p}'} h_3(\mathbf{x}_j^*) \right)
\end{aligned}
$$

The results in equation (72) follow by substitution into the results in equation (71) .

Proposition 8

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

$$
E\left(h(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\mathbf{p}}) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\mathbf{p}) = \bar{\mathbf{x}} \right) = h(\mathbf{x}, \{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\}) \tag{127}
$$

for all $\mathbf{p} \in \mathcal{P}^{n-1}$. Let $\tilde{\mathbf{p}}$ be a purely random draw from $\mathcal{P}^{n-1}$. Then:

$$
(\mathbf{x}_i^*, \epsilon_i) \perp\!\!\!\perp \{\mathbf{x}_j^*\}_{j\in\tilde{\mathbf{p}}} \tag{128}
$$

which implies:

$$
\begin{aligned}
E(\epsilon_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) &= E(\epsilon_i|\mathbf{x}_i = \mathbf{x}) && \text{(by (128))} \\
&= 0 && (129)
\end{aligned}
$$

By (CRA), Lemma 1 holds and therefore:

$$
\begin{aligned}
E(y_i|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i = \bar{\mathbf{x}}) &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}) && \text{(by (38) in Lemma 1)} \\
&= E\left(h\left(\mathbf{x}_i^*, \{\mathbf{x}_j^*\}_{j\in\tilde{\mathbf{p}}}\right) + \epsilon_i \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) && \text{(by DCE)} \\
&= E\left(h\left(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\tilde{\mathbf{p}}}\right) + \epsilon_i \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) && \text{(by NOV)} \\
&= E\left(h\left(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\tilde{\mathbf{p}}}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) \\
&\quad + E\left(\epsilon_i | \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}}\right) \\
&= E\left(h\left(\mathbf{x}_i, \{\mathbf{x}_j\}_{j\in\tilde{\mathbf{p}}}\right) \middle| \mathbf{x}_i = \mathbf{x}, \bar{\mathbf{x}}_i(\tilde{\mathbf{p}}) = \bar{\mathbf{x}} \right) && \text{(by (129))} \\
&= h\left(\mathbf{x}, \{\mathbf{x}^1, \ldots, \mathbf{x}^{n-1}\}\right) && \text{(by (127))}
\end{aligned}
$$

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

2. (PSE,DCE) implies that the first result in Proposition 7 applies:

$$\begin{aligned}
p_{ij} &= h_2(\mathbf{x}_i^*, \mathbf{x}_j^*) && \text{(by Proposition 7)} \\
&= h_2(\mathbf{x}_i, \mathbf{x}_j) && \text{(by (NOV))}
\end{aligned}$$

It follows by substitution that:

$$\begin{aligned}
E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) &= E(h_2(\mathbf{x}_i, \mathbf{x}_j)|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) && \text{(result above)} \\
&= h_2(\mathbf{c}_{sK}, \mathbf{c}_{kK}) && \text{(conditioning rule)} \\
E\left(p_{ij} \middle| \begin{array}{l} \mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}, \\ \ell_i = \ell, \ell_j = \ell' \end{array}\right) &= E\left(h_2(\mathbf{x}_i, \mathbf{x}_j) \middle| \begin{array}{l} \mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}, \\ \ell_i = \ell, \ell_j = \ell' \end{array}\right) && \text{(result above)} \\
&= h_2(\mathbf{c}_{sK}, \mathbf{c}_{kK}) && \text{(conditioning rule)} \\
&= E(p_{ij}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{x}_j = \mathbf{c}_{kK}) && (130)
\end{aligned}$$

which is condition (LI). Therefore the second result in Proposition 6 applies.

Equation (33)

$$
\begin{aligned}
HGE_{s,m} &= E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \qquad \text{(by (22))} \\
&= \sum_{r=0}^{M^S} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}) \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\
&\quad - \sum_{r=0}^{M^S} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}) \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \\
&\qquad \text{(law of total probability)} \\
&= \sum_{r=0}^{M^S} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}) \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix}. \\
&\qquad \text{(independence of } \mathbf{x}_i \text{ and } (\mathbf{z}_i^S(\tilde{\mathbf{p}}), \mathbf{z}_i(\tilde{\mathbf{p}}))\text{)} \\
&= \sum_{r=0}^{M^S} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}) \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \\
&\quad - \sum_{r=0}^{M^S} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{0M^S}) \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \\
&\quad + \sum_{r=0}^{M^S} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{0M^S}) \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \\
&= \sum_{r=0}^{M^S} \overbrace{\begin{pmatrix} E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}) \\ - E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{0M^S}) \end{pmatrix}}^{= 0 \text{ for } r=0,\ HGE_{s,r} \text{ for } r>0} \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \\
&\quad + E(y_i(\tilde{\mathbf{p}})|\mathbf{x}_i = \mathbf{c}_{sK}, \mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{0M^S}) \underbrace{\begin{pmatrix} \sum_{r=0}^{M^S} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \sum_{r=0}^{M^S} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix}}_{=(1=1)=0} \\
&= \sum_{r=1}^{M^S} HGE_{s,r}^S \begin{pmatrix} \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{mM}) \\ - \Pr(\mathbf{z}_i^S(\tilde{\mathbf{p}}) = \mathbf{c}_{rM^S}|\mathbf{z}_i(\tilde{\mathbf{p}}) = \mathbf{c}_{0M}) \end{pmatrix} \qquad \text{(see (33))}
\end{aligned}
$$