Unobserved Heterogeneous Spillover Effects in Instrumental Variable Models

Huan Wu *

This version: September 28, 2025

Abstract

This paper develops a novel framework for identifying and estimating causal effects in settings with spillovers, focusing on cases where spillovers arise within known groups and treatment selection is endogenous. The paper contributes to the literature in two ways. First, I introduce two causal parameters: the marginal controlled spillover effect, capturing how peers' treatments affect an individual's outcome holding own treatment fixed, and the marginal controlled direct effect, capturing how an individual's treatment affects the outcome holding peers' treatments fixed. Conditioning on continuous unobserved group characteristics, these parameters provide the foundation for a broad class of policy-relevant effects studied in the literature, such as local average spillover and direct effects. Second, I point identify these effects under weak assumptions, allowing for flexible spillover structures and unobserved heterogeneity without imposing functional restrictions on spillover patterns or correlations among latent characteristics. To estimate these effects, I propose general semiparametric estimators and derive their asymptotic properties. For practical implementation, I further develop a parametric strategy under intuitive assumptions, with Monte Carlo simulations showing superior finite sample performance. Finally, I apply the framework to study the returns to education within best-friend networks using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). The results reveal substantial heterogeneity in educational spillover effects across best-friend network characteristics.

Keywords: Unobserved heterogeneous spillover/direct effect; violation of SUTVA; causal inference; instrumental variable.

JEL subject classification: C21, C31, C14, C36

*Department of Economics, University of North Carolina at Chapel Hill, the United States. Email address: huan.wu@unc.edu

1 Introduction

In the study of treatment effects, the literature typically relies on the Stable Unit Treatment Value Assumption (SUTVA), which assumes that an individual's potential outcomes depend solely on their own treatment status and are unaffected by the treatment assignments of other units. Although analytically convenient for identification and estimation, this assumption is often unrealistic in empirical settings involving social interactions or network structures. In such settings, an individual's outcome may depend not only on her own treatment assignment but also on the treatment status of her peers, rendering the standard SUTVA framework inadequate.

This paper develops a new framework for identifying and estimating causal effects in settings where spillovers arise within predetermined groups, such as friend pairs and couples, whose composition is formed prior to treatment assignment. Our framework explicitly accounts for two distinct channels through which SUTVA is potentially violated. First, an individual's outcome may directly depend on both her own treatment and the treatments received by her group members. Second, an individual's treatment status may be influenced by both her own and her peers' characteristics that are exogenous to the unobserved determinants of treatment and outcomes. This structure enables the use of such characteristics as instrumental variables. These features imply that outcomes and treatments cannot be modeled as functions of individual-level variables alone, thereby necessitating methods that extend beyond the conventional SUTVA framework. Additionally, our framework allows the unobserved factors that determine treatment assignment and outcomes to be arbitrarily correlated across group members, capturing the dependence that naturally arises within networks. By relaxing SUTVA and incorporating both within-group spillovers and correlated unobservables, our approach provides a more credible and flexible foundation for causal inference in group-based settings.

This paper makes several contributions to the literature. In Section 2.1, we introduce an outcome model within the potential outcomes framework, allowing each unit's potential outcome to depend flexibly on the full treatment vector of her group members, without imposing functional form restrictions. Treatment decisions are modeled using a single-index threshold-crossing structure: an individual receives treatment if her unobserved cost of participation lies below a threshold determined by both her own and her peers' exogenous characteristics. This formulation can be rationalized as the equilibrium behavior of a simultaneous incomplete information game, as in Aradillas-Lopez (2010). Importantly, we refrain from imposing parametric assumptions on either the outcome equation or the threshold function, which accommodates for flexible spillover structures in both treatment assignment and

outcomes. Moreover, our framework places no restrictions on the joint distribution of unobserved factors across group members, allowing for arbitrary dependence between unobserved determinants of treatment and outcomes within groups.

We introduce two novel causal parameters designed to capture treatment effect heterogeneity in the presence of social interactions: the marginal spillover effect (MCSE) and the marginal direct effect (MCDE). On the one hand, the MCSE quantifies the effect of changes in peers' treatment status on an individual's outcome, holding the individual's own treatment fixed and conditioning on the latent characteristics of all group members. On the other hand, the MCDE measures the effect of an individual's own treatment on her outcome, holding peers' treatment assignments fixed and similarly conditioning on the full vector of group-level unobservables. By conditioning on the latent characteristics of all group members, these parameters flexibly account for heterogeneity in both direct and spillover effects. Importantly, these marginal controlled effects serve as fundamental building blocks for a broad class of policy relevant treatment parameters. In particular, they nest objects such as the local average direct and spillover effects studied by Vazquez-Bare (2023), while also providing the foundation for causal analysis under a wide range of counterfactual policy interventions.

In Section 2.3, we provide formal point identification of the MCSE and the MCDE using continuously distributed instrumental variables, without imposing functional form restrictions on the spillover structure. Unlike several approaches that address noncompliance by excluding specific subpopulations, such as one-sided noncompliers in Vazquez-Bare (2023) and DiTraglia et al. (2023), our strategy achieves point identification without such restrictions. Moreover, we nonparametrically identify the joint distribution of latent characteristics across group members, avoiding the parametric or distributional assumptions commonly imposed in earlier work, such as Balat and Han (2023) and Hoshino and Yanagi (2023).

In Section 4, we propose a semiparametric estimation strategy, building on the methods of Carneiro and Lee (2009), to estimate the marginal controlled effects. This approach alleviates the curse of dimensionality while preserving the flexible spillover structure of our framework, as it retains the nonparametric foundation of the model. We further establish the asymptotic properties of the semiparametric estimators. However, when sample sizes are limited or group sizes are large, the semiparametric approach may yield imprecise estimates. To address this, we introduce a complementary parametric framework based on a set of intuitive assumptions. We provide easy-to-implement parameter estimators along with simple nonparametric bootstrap procedures for valid inference. Finally, Monte Carlo simulations demonstrate the superior finite-sample performance of the proposed parametric methods.

In Section 5.2, we apply our framework to a realistic empirical setting with potential

social spillovers. Specifically, we examine the effect of educational attainment on long-term earnings within best-friend networks, using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Our findings reveal different patterns in the marginal controlled direct and spillover effects. On the one hand, the MCDEs of completing 16 years of education are significantly positive when the best friend has also completed 16 years of education, across all selected values of latent characteristics. However, when the best friend has not completed 16 years, the MCDEs become statistically insignificant across all evaluated levels of unobservables. On the other hand, the marginal controlled spillover effects (MCSEs) display a different pattern, particularly among individuals who did not complete 16 years of education. For those who attained 16 years of education, the MCSEs are significantly positive across most values of latent characteristics. In contrast, for individuals who did not complete 16 years, the spillover effects become statistically insignificant, and even significantly negative at certain values of latent characteristics. These findings provide evidence of spillover effects of educational attainment on long-term earnings within best-friend networks and underscore that both the direction and magnitude of these effects vary systematically with the educational status of individuals and their best friends.

Related literature

Recent literature has increasingly focused on identifying and estimating treatment effects in the presence of spillovers. This paper contributes to several key strands within this growing body of work.

A common strategy for addressing interference has been to impose parametric structures on social interactions. For instance, Manski (1993) introduced the linear-in-means model, formulated as a system of linear simultaneous equations to capture endogenous, exogenous, and correlated effects. Building on this result, subsequent work, such as Bramoullé, Djebbari, and Fortin (2009) and Blume et al. (2015), extended the framework to more complex forms of interaction within linear models and derived conditions under which social effects can be identified. However, they fundamentally rely on correct parametric assumptions regarding the structure of social interactions. Such assumptions may lead to model misspecification, particularly in the presence of nonlinear spillovers or heterogeneity across individuals. In contrast, our framework does not rely on parametric restrictions in both the outcome and treatment equations. By avoiding ad hoc functional form assumptions, it allows for flexible and potentially complex patterns of spillovers. This nonparametric foundation provides a more robust framework for analyzing causal effects in settings involving social interactions within groups.

In our main setting, an individual's outcome depends on the full vector of treatments within the group, consistent with the treatment response function framework of Manski (2013). Building on this framework, several studies, such as Aronow and Samii (2017), considered a concept of exposure mapping. Under randomized controlled trials (RCTs), Aronow and Samii (2017) proposed a design-based approach to estimate and conduct inference on both direct and spillover effects. Several studies, including Hudgens and Halloran (2008) and related works, have applied the design-based approach to estimate spillover effects under RCTs. Our contribution departs from this literature by allowing for noncompliance, so that individuals may not adhere to their assigned treatments. This feature is especially important in observational studies, where treatment assignment is not fully controlled by the researcher, and in experimental settings where imperfect compliance is common. We adopt a large-sample approach rather than the design-based framework, to study causal inference in environments where treatment take-up is endogenous.

Vazquez-Bare (2023) applied a potential outcomes framework to identify and estimate spillover effects in the presence of noncompliance using binary instrumental variables. His approach categorized individuals into compliance types and achieved point identification by excluding certain subpopulations in one-sided noncompliance settings, which is a restriction that is also common in related work, such as DiTraglia et al. (2023). The framework developed in this paper employs a continuous instrumental variable, which enables point identification of both direct and spillover effects without the need to exclude any compliance types from the population. A key advantage of our approach is that the point identified direct and spillover effects are defined conditional on continuous values of the latent characteristics within groups. This feature links our analysis to the marginal treatment effect literature and provides a set of building blocks for recovering a wide range of policy-relevant treatment effects. In particular, these marginal effects can be aggregated to obtain local average direct and spillover effects discussed in Vazquez-Bare (2023), as well as a broad class of counterfactual causal effects under alternative policy interventions.

Recent studies, such as Balat and Han (2023) and Hoshino and Yanagi (2023), have employed instrumental variable approaches to identify and estimate spillover effects in environments with direct strategic interactions among agents. Our framework does not model direct strategic interaction in treatment choices. Instead, we allow an individual's treatment decision to depend on the instruments assigned to her group members. This structure can be rationalized as the equilibrium outcome of a simultaneous game with incomplete information, following Aradillas-Lopez (2010), and thus provides a complementary perspective to frameworks that incorporate strategic interactions. Furthermore, we point identify the joint distribution of unobserved characteristics within groups without imposing functional form

restrictions, thereby permitting arbitrary correlation across group members' unobservables. This broadens the scope of existing literature, which has typically relied on parametric or distributional assumptions to model dependence in unobservables.

2 Model and Identification

2.1 Setting

I consider a sample of G independent and identically distributed (i.i.d.) groups, indexed by $g = 1, \cdots, G$. Each group consists of the same number of units, denoted by $n \geq 2$. For example, a group may correspond to a market with several competing firms or to a household with multiple members. Within each group, units are indexed by $i = 0, \cdots, n-1$. Throughout, I assume that spillover effects operate only within groups and do not extend across groups.

Researchers are often interested in how a treatment affects an outcome. Let Y_{ig} denote the outcome of interest for unit i in group g. In some settings, the outcome Y_{ig} may depend not only on unit i's own treatment status but also on the treatment choices of other units within the same group. For example, a firm's market share is influenced both by its own pricing decisions and by those of its competitors. Similarly, in a household, an individual's health status may depend on her own smoking behavior as well as that of other family members. In such contexts, the Stable Unit Treatment Value Assumption (SUTVA) may be violated, which motivates researchers to develop models that explicitly allow for spillover effects in outcomes.

The binary treatment decision of unit i in group g is denoted by $D_{ig} \in \{0, 1\}$, where $D_{ig} = 1$ indicates that unit i adopts the treatment and $D_{ig} = 0$ otherwise. In many applications, treatment is not purely random but instead influenced by unobserved characteristics that also affect outcomes, which gives rise to endogeneity concerns. In group settings, the treatment decision is further complicated by the fact that units may make their decisions simultaneously, taking into account both their own private information and expectations about the choices of others. Each unit's decision depends on its own private information, denoted by V_{ig}, as well as on its expectations about the probability that other members of the group will adopt the treatment. Units form their expectations on the basis of publicly observed signals (Z_{ig}, Z_{-ig}), where Z_{ig} denotes the random assignment received by unit i in group g, and Z_{-ig} denotes the assignments of the remaining group members. This framework corresponds to an incomplete information simultaneous game, as studied by Aradillas-Lopez (2010) and related papers.

To illustrate, consider a group as an oligopoly market with a fixed number of competing firms. Each firm simultaneously decides whether to increase its price, without observing its rivals' choices at the decision stage. The decision of a given firm, D_{ig}, depends on its private demand shock V_{ig}, such as an idiosyncratic change in reputation or advertising effectiveness, that is unobserved by its competitors. This demand shock affects the firm's incentive to raise its price, while at the same time shifting consumer demand and hence its market share Y_{ig}. As a result, the pricing decision is correlated with the unobserved component of the market share equation, giving rise to an endogeneity problem. While a firm does not observe its rivals' pricing decisions, its own decision is shaped by beliefs about rival behavior. These beliefs are formed conditional on publicly observed market signals (Z_{ig}, Z_{-ig}), for example, industry-wide cost shocks such as tariffs.

Building on the setting described above, I construct the following model for unit i in group g. For clarity of exposition, I focus on the case in which each group g consists of two units, indexed by $i = \{0, 1\}$, while noting that the identification and estimation results extend straightforwardly to groups with more than two members. Furthermore, because the analysis is developed at the level of a super-population of groups, I suppress the group subscript g throughout this section to simplify notation.

$$\begin{cases} Y_i = m_i(D_i, D_{-i}, U_i, U_{-i}) \\ D_i = \mathbb{1}\left\{V_i \leq h_i(Z_i, Z_{-i})\right\}. \end{cases} \tag{1}$$

The first line of Equation (1) specifies the outcome equation. In this framework, unit i's outcome Y_i depends on her own treatment D_i and on her group member's treatment, D_{-i}, which explicitly models spillover effects. Importantly, I also allow Y_i to depend on both unit i's own unobservables U_i and the unobservables of her group member, U_{-i}. In the oligopoly market example, this specification captures the possibility that firm i's market share Y_i is influenced not only by its own unobserved product characteristics U_i but also by the unobserved product characteristics of its rival, U_{-i}.

Throughout the paper, I define the potential outcome for unit i when her treatment is set to d and her group member's treatment to d' as $Y_i(d, d') \equiv m_i(d, d', U_i, U_{-i})$. A key benefit of my framework is that it imposes no functional form restrictions on the outcome equation m_i and places no constraints on the dimension of the unobservables (U_i, U_{-i}). As a result, the effect of group member's treatment D_{-i} on unit i's outcome Y_i is left entirely unrestricted. This generality provides a flexible structure that accommodates rich and heterogeneous patterns of spillover effects in outcomes.

The second line of Equation (1) characterizes the treatment take-up mechanism. The treatment decision of unit i, D_i, may be endogenous because it is determined by a continuous

unobserved factor V_i that can also influence the outcome. An important feature of my framework is that D_i depends only on the unit's own unobservable V_i and not directly on her group member's unobservable V_{-i}. This restriction is plausible in many applications. For example, in the oligopoly market discussed above, a firm's pricing decision is driven by its own private demand shock, while the competitor's demand shock is unobserved and therefore cannot directly affect the firm's decision rule.

Crucially, I do not require the unobserved factors V_i and V_{-i} to be independent, and my framework imposes no functional form restrictions on their joint dependence structure. This flexibility accommodates a wide range of empirically relevant correlations. In the oligopoly setting, correlation across firms' idiosyncratic shocks arises naturally. For instance, a market-wide change in consumer tastes or a new advertising regulation may simultaneously affect how all products are perceived by consumers, thereby inducing correlation between the demand shocks V_i and V_{-i}.

I model the treatment take-up mechanism using a single threshold crossing rule: unit i's chooses to take the treatment, $D_i = 1$, if the unobserved factor $V_i \in \mathbb{R}$ does not exceed a threshold $h_i(Z_i, Z_{-i})$, where $h_i : \mathbb{R}^{k_i} \times \mathbb{R}^{k_{-i}} \mapsto \mathbb{R}$ is an unspecified function. Importantly, I do not impose a parametric form on the threshold function h_i, and the subscript i highlights that thresholds may vary across units within the same group. In contrast to complete information games, where unit i's treatment D_i directly depends on the treatment decision of her peer D_{-i}, my framework is consistent with an incomplete information simultaneous game. In this case, each unit i forms a subjective belief about the joint distribution $\mathbb{P}_i(D_i = 1, D_{-i} = 1 \mid Z_i, Z_{-i})$, where (Z_i, Z_{-i}) denotes the vector of public signals observed by all group members, and makes her treatment decision accordingly. As shown by Aradillas-Lopez (2010), the optimal decision rule in this setting can be expressed in a form similar to the treatment specification in Equation (1).

For identification, which I discuss in details later, the public signals (Z_i, Z_{-i}) must be independent of the unobserved heterogeneity $(U_i, U_{-i}, V_i, V_{-i})$ and must not directly affect the outcomes (Y_i, Y_{-i}). These signals thus serve the role of instrumental variables. In studies that focus on complete information setting, such as those analyzed by Balat and Han (2023) and Hoshino and Yanagi (2023), strategic interactions between D_i and D_{-i} are modeled explicitly, but unit i's treatment is not allowed to depend on her group member's instrument Z_{-i}. An important feature of my framework is that it relaxes this restriction: I allow the treatment D_i to depend on both the unit's own instrument Z_i and the peer's instrument Z_{-i}, thereby accommodating potential spillovers from instruments into treatment decisions. Moreover, instruments may be specified as common public signals shared across group members, so that $Z_i = Z_{-i} = Z$, or including unit-specific variables so that $Z_i \neq Z_{-i}$,

providing substantial flexibility in the choice of valid instruments.

Remark 1. (Endogenous effects in the outcome) Some studies (e.g., Bramoullé, Djebbari, and Fortin, 2009) model spillover effects using a system of structural equations in which a unit's outcome may directly depend on the outcomes of their peers. When outcomes among group members influence one another, such interactions are typically referred to as endogenous effects. In contrast, this framework does not explicitly model endogenous effects, since Y_i does not directly depend on the outcomes of unit i's peers, Y_{-i}. However, this should not be interpreted as ruling out the possibility of endogenous effects within the model. As discussed in Manski (2013), the potential outcome $Y_i(d, d')$ can be interpreted as a reduced-form solution of underlying structural models that include endogenous effects.

For example, consider a system of linear structural functions of $Y_i, i \in \{0, 1\}$,

$$\begin{aligned} Y_0 &= \alpha_0 + \alpha_1 D_0 + \alpha_2 D_1 + \alpha_3 Y_1 + U_0 + \gamma_1 U_1, \\ Y_1 &= \beta_0 + \beta_1 D_1 + \beta_2 D_0 + \beta_3 Y_0 + U_1 + \gamma_2 U_0, \alpha_3\beta_3 \neq 1 \end{aligned}$$

where D_i denotes unit i's treatment and U_i captures unobserved factors. One can solve the system to obtain a reduced-form expression for Y_i as a linear function of D_i and D_{-i}, without explicitly involving Y_{-i}:

$$\begin{aligned} Y_0 =& \frac{\alpha_0 + \alpha_3\beta_0 + (\alpha_1 + \alpha_3\beta_2)\, D_0 + (\alpha_2 + \alpha_3\beta_1)\, D_1}{1 - \alpha_3\beta_3} \\ &+ \frac{(1 + \alpha_3\gamma_2)\, U_0 + (\gamma_1 + \alpha_3)\, U_1}{1 - \alpha_3\beta_3}, \\ Y_{1g} =& \frac{\beta_0 + \beta_3\alpha_0 + (\beta_1 + \beta_3\alpha_2)\, D_1 + (\beta_2 + \beta_3\alpha_1)\, D_0}{1 - \alpha_3\beta_3} \\ &+ \frac{(1 + \beta_3\gamma_1)\, U_1 + (\beta_3 + \gamma_2)\, U_0}{1 - \alpha_3\beta_3}. \end{aligned}$$

Therefore, $Y_i(d, d')$ can be interpreted as solutions for Y_i when the treatment assignments are set to $D_i = d$ and $D_{-i} = d'$. This interpretation remains valid when the structural functions are nonlinear. In the nonlinear case, the link between the potential outcome equations and the underlying structural models is less transparent, though it can still be derived by researchers in the context of specific applications. In this paper, however, the focus is on the reduced-form treatment effects of a unit's' own and peers' treatments, rather than on the structural parameters embedded in the structural equations.

The directed acyclic graph in Figure 1 illustrates the causal relationships among the key variables in the model. The red arrows represent spillover channels: unit i's outcome Y_i may depend on her peer's treatment D_{-i}, and her treatment D_i may depend on her peer's

instrument Z_{-i}. Direct interaction between treatments D_i and D_{-i}, however, is ruled out. The unobserved heterogeneity V_i and V_{-i} introduce endogeneity, as they may simultaneously affect both treatments and outcomes. Those are represented by the black dashed arrows. The blue dashed arrow reflects potential dependence between V_i and V_{-i}, for which I do not impose any functional restrictions.

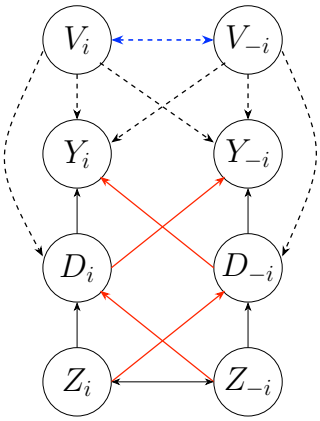

Figure 1: Causal relations under spillover setting

Assumptions 1-3 specify the restrictions required for the key variables.

Assumption 1. (Random assignment) For each group g, the instrumental variables satisfy

$$\big(Z_i, Z_{-i}\big) \perp\!\!\!\perp \big(V_i, V_{-i}, U_i, U_{-i}\big)$$

for $i, -i \in \{0, 1\}$.

Assumption 1 requires that instruments are randomly assigned at the group level, implying that the group-level instrument vector (Z_i, Z_{-i}) is independent of the unobserved heterogeneity of all units in the group. This assumption places no restrictions on the dependence structure between Z_i and Z_{-i} within a group. The instruments may be arbitrarily correlated across units within a group, as long as they remain jointly independent of the unobserved heterogeneity $(V_i, V_{-i}, U_i, U_{-i})$.

Assumption 2. (Exclusion restriction) Given d_0, d_1 and u_0, u_1, the instrumental variables (Z_i, Z_{-i}) do not directly affect the outcome Y_i:

$$m_i\big(d_0, d_1, z_0, z_1, u_0, u_1\big) = m_i\big(d_0, d_1, z_0', z_1', u_0, u_1\big)$$

for any $z_0 \neq z_0'$ and $z_1 \neq z_1'$.

Assumption 2 requires that the instruments affect the outcome only through their influence on treatment take-up, without exerting any direct effect on the outcome. This condition corresponds to the standard exclusion restriction commonly imposed in instrumental variable analyses.

Assumption 3. (Distribution of V) The unobserved variable V_i is continuously distributed.

Assumption 3 requires that the unobserved heterogeneity V_i has a continuous distribution, which is a common condition in the literature. Under this assumption, V_i can be normalized to follow a uniform distribution on the interval $(0, 1)$.

2.2 Conceptual Basis for Identification

Vazquez-Bare (2023) studies a related setting with spillovers affecting both treatment take-up and outcomes, while ruling out direct strategic interactions in treatment decisions. However, without additional restrictions, his framework does not deliver point identification of the parameters of interest. Specifically, he considers a binary instrumental variable $Z_i \in \{0, 1\}$ and classify the population into a finite number of discrete types according to the values of the potential treatment vector $(D_i(0,0), D_i(0,1), D_i(1,0), D_i(1,1))$, which indicates the unit's treatment status under each possible assignment. The variation provided by the instrumental variable is insufficient to point identify the proportions of the different population types. As acknowledged in Vazquez-Bare (2023), "... in general, the simultaneous presence of spillovers on treatment status and outcomes can impede identification of causally interpretable parameters even when the instruments are randomly assigned." To obtain identification, he imposes a one-sided noncompliance restriction, requiring that units can only take treatment when assigned to it. This assumption rules out several population types, enabling point identification for certain causal effects. In contrast, my framework introduces a continuous unobserved characteristic V_i and leverages the variation in instrumental variables to achieve point identification of causal effects without relying on one-sided noncompliance. This extension allows identification of parameters that remain unidentified in Vazquez-Bare (2023), as well as other policy-relevant treatment effects, thereby broadening the range of settings in which causal spillover effects can be recovered.

Before introducing the main parameters of interest and presenting the formal identification results, I provide a heuristic derivation that highlights the differences between this framework and that of Vazquez-Bare (2023), and motivates the definition of parameters of interest and the identification.

I define the propensity score function for unit i as the probability of treatment conditional on the group-level instrument vector, $P_i(Z_i, Z_{-i}) \equiv \mathbb{P}(D_i = 1 \mid Z_i, Z_{-i})$. I denote this

function simply by P_i and define the support of the propensity scores for all group members as $\mathcal{P} \equiv \text{Supp}(P_i, P_{-i})$. Given the independence assumption and the distributional assumption on V_i, the propensity score function identifies the threshold function h_i in the treatment take-up equation, as shown in the following derivation:

$$\begin{aligned} &\mathbb{P}\left(D_i = 1 \mid Z_i = z_0, Z_{-i} = z_1\right) \\ =&\mathbb{P}\left(V_i \leq h_i(z_0, z_1) \mid Z_i = z_0, Z_{-i} = z_1\right) \\ =&\mathbb{P}\left(V_i \leq h_i(z_0, z_1)\right) = h_i(z_0, z_1). \end{aligned} \tag{2}$$

Additionally, the propensity scores (P_i, P_{-i}) are independent of all group members' unobserved heterogeneity, since they are functions of the instruments (Z_i, Z_{-i}).

I next show how to identify the local average spillover and direct effects, defined as

$$\begin{aligned} &\mathbb{E}\big[Y_i(d,1) - Y_i(d,0) \mid (V_i, V_{-i}) \in P\big], \\ &\mathbb{E}\big[Y_i(1,d) - Y_i(0,d) \mid (V_i, V_{-i}) \in P\big], \end{aligned}$$

where the first expression represents the local average spillover effect and the second the local average direct effect, both conditional on (V_i, V_{-i}) lying in a subset P of the propensity score support $\mathcal{P}$.

For any $(p_0, p_1) \in \text{Supp}(P_i, P_{-i})$, consider the average outcome for the subpopulation with $D_i = 1$ and $D_{-i} = 1$, conditional on the group-level propensity scores $P_i = p_0$ and $P_{-i} = p_1$, $\mathbb{E}[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1]$. This conditional expectation identifies

$$\begin{aligned} \mathbb{E}\big[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\big] &= \mathbb{E}\big[Y_i(1,1)\mathbb{1}\big\{V_i \leq p_0\big\}\mathbb{1}\big\{V_{-i} \leq p_1\big\} \mid P_i = p_0, P_{-i} = p_1\big] \\ &= \mathbb{E}\big[Y_i(1,1)\mathbb{1}\big\{V_i \leq p_0\big\}\mathbb{1}\big\{V_{-i} \leq p_1\big\}\big], \end{aligned}$$

where the second equality follows from the independence assumption. The last expression corresponds to the average potential outcome $Y_i(1,1)$ for the subpopulation whose unobserved characteristics lie in the region $\{V_i \leq p_0, V_{-i} \leq p_1\}$. Then, by fixing P_i at p_0 and increasing P_{-i} from p_1 to $p_1' > p_1$, one can identify the average of $Y_i(1,1)$ for the subpopulation with unobserved characteristics satisfying $\{V_i \leq p_0, p_1 < V_{-i} \leq p_1'\}$,

$$\begin{aligned} &\mathbb{E}\big[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1'\big] - \mathbb{E}\big[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\big] \\ =&\mathbb{E}\big[Y_i(1,1)\mathbb{1}\big\{V_i \leq p_0\big\}\mathbb{1}\big\{p_1 < V_{-i} \leq p_1'\big\}\big]. \end{aligned}$$

The above equality likewise holds when P_i is fixed at $p_0' > p_0$. Taking the difference between

the two expressions corresponding to p_0' and p_0 yields

$$\begin{aligned}
&\Big(\mathbb{E}\big[Y_i D_i D_{-i} \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[Y_i D_i D_{-i} \mid P_i = p_0', P_{-i} = p_1\big]\Big) \\
&- \Big(\mathbb{E}[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1'] - \mathbb{E}[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1]\Big) \\
=&\mathbb{E}\big[Y_i(1,1)\mathbb{1}\{p_0 < V_i \leq p_0'\}\mathbb{1}\{p_1 < V_{-i} \leq p_1'\}\big],
\end{aligned} \tag{3}$$

which identifies the average potential outcome $Y_i(1,1)$ for the subpopulation with unobserved characteristics (V_i, V_{-i}) lying in the rectangle $\{p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\}$.

Applying the same strategy to the conditional expectation $\mathbb{E}[D_i(1-D_{-i}) \mid P_i, P_{-i}]$ yields the negative of the average potential outcome $Y_i(1,0)$ for the subpopulation with unobserved characteristics (V_i, V_{-i}) lying in the rectangle $\{p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\}$,

$$\begin{aligned}
&\Big(\mathbb{E}\big[Y_i D_i(1-D_{-i}) \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[Y_i D_i(1-D_{-i}) \mid P_i = p_0', P_{-i} = p_1\big]\Big) \\
&- \Big(\mathbb{E}\big[Y_i D_i(1-D_{-i}) \mid P_i = p_0, P_{-i} = p_1'\big] - \mathbb{E}\big[Y_i D_i(1-D_{-i}) \mid P_i = p_0, P_{-i} = p_1\big]\Big) \\
=&-\mathbb{E}\big[Y_i(1,0)\mathbb{1}\{p_0 < V_i \leq p_0'\}\mathbb{1}\{p_1 < V_{-i} \leq p_1'\}\big].
\end{aligned} \tag{4}$$

Summing Equations (3) and (4) identifies the local average spillover effect with unit i's treatment fixed at 1, conditional on the group-level unobserved heterogeneity lying in the region $\{p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\}$, weighted by the corresponding subpopulation share:

$$\begin{aligned}
&\Big\{\big(\mathbb{E}\big[Y_i D_i \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[Y_i D_i \mid P_i = p_0', P_{-i} = p_1\big]\big) \\
&- \big(\mathbb{E}[Y_i D_i \mid P_i = p_0, P_{-i} = p_1'] - \mathbb{E}[Y_i D_i \mid P_i = p_0, P_{-i} = p_1]\big)\Big\} \\
=&\mathbb{E}\big[(Y_i(1,1) - Y_i(1,0))\mathbb{1}\{p_0 < V_i \leq p_0'\}\mathbb{1}\{p_1 < V_{-i} \leq p_1'\}\big] \\
=&\mathbb{E}\big[Y_i(1,1) - Y_i(1,0) \mid p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\big] \cdot \mathbb{P}\big(p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\big).
\end{aligned} \tag{5}$$

The conditional expectation $E[Y_i(1,1) - Y_i(1,0) \mid p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1']$ represents the local average spillover effect when unit i's own treatment is fixed at 1, conditional on the group-level unobserved heterogeneity lying in the region $\{p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\}$. Since the probability $\mathbb{P}(p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1')$ can be identified by replacing Y_i with $\mathbb{1}\{Y_i \in \mathcal{Y}\}$ in Equation (3), it follows that

$$\begin{aligned}
&\mathbb{E}\big[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1'\big] - \mathbb{E}\big[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\big] \\
=&\mathbb{P}(p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'),
\end{aligned}$$

Thus, the local average spillover effect, with unit i's own treatment fixed at 1 can be identified

as

$$
\begin{aligned}
&\Big\{\big(\mathbb{E}\big[Y_iD_i \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[Y_iD_i \mid P_i = p_0', P_{-i} = p_1\big]\big) \\
&- \big(\mathbb{E}[Y_iD_i \mid P_i = p_0, P_{-i} = p_1'] - \mathbb{E}[Y_iD_i \mid P_i = p_0, P_{-i} = p_1]\big)\Big\}\Big/ \\
&\Big\{\big(\mathbb{E}\big[D_iD_{-i} \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[D_iD_{-i} \mid P_i = p_0', P_{-i} = p_1\big]\big) \\
&- \big(\mathbb{E}\big[D_iD_{-i} \mid P_i = p_0, P_{-i} = p_1'\big] - \mathbb{E}\big[D_iD_{-i} \mid P_i = p_0, P_{-i} = p_1\big]\big)\Big\} \\
=&\mathbb{E}\big[Y_i(1,1) - Y_i(1,0) \mid p_0 < V_i \le p_0', p_1 < V_{-i} \le p_1'\big].
\end{aligned}
$$

This expression illustrates the identification strategy for the local average spillover effect. Intuitively, the idea is to focus on groups in which unit i takes the treatment ($D_i = 1$) and then exploit exogenous shifts in the propensity scores. First, by increasing P_{-i} from p_1 to p_1' while holding P_i fixed, the peer $-i$ transitions from being untreated ($D_{-i} = 0$) to treated ($D_{-i} = 1$). Similarly, by subsequently raising P_i from p_0 to p_0', the analysis isolates the subpopulation with unobserved characteristics in the rectangle $\{p_0 < V_i \le p_0', p_1 < V_{-i} \le p_1'\}$. Within this region, the peer's treatment status changes from 0 to 1 while unit i remains treated, which provides the necessary variation to identify the spillover effect.

Formally, the numerator captures the change in outcomes Y_i induced by this shift in peers' treatment, while the denominator captures the probability mass of units whose treatment status changes due to the variation in (P_i, P_{-i}). The ratio therefore identifies the local average spillover effect, that is, the average causal effect of peers' treatment on unit i's outcome conditional on (V_i, V_{-i}) lying in the specified region.

By repeating the same procedure, the local average spillover effect with unit i's treatment fixed, conditional on the group-level unobserved heterogeneity lying in the region $p_0 < V_i \le p_0', ; p_1 < V_{-i} \le p_1'$, can be identified as

$$
\begin{aligned}
\operatorname{sgn}(2d-1)\cdot&\Big\{\big(\mathbb{E}\big[Y_i\mathbb{1}\{D_i = d\} \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[Y_i\mathbb{1}\{D_i = d\} \mid P_i = p_0', P_{-i} = p_1\big]\big) \\
&- \big(\mathbb{E}[Y_i\mathbb{1}\{D_i = d\} \mid P_i = p_0, P_{-i} = p_1'] - \mathbb{E}[Y_i\mathbb{1}\{D_i = d\} \mid P_i = p_0, P_{-i} = p_1]\big)\Big\}\Big/ \\
&\Big\{\big(\mathbb{E}\big[D_iD_{-i} \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[D_iD_{-i} \mid P_i = p_0', P_{-i} = p_1\big]\big) \\
&- \big(\mathbb{E}\big[D_iD_{-i} \mid P_i = p_0, P_{-i} = p_1'\big] - \mathbb{E}\big[D_iD_{-i} \mid P_i = p_0, P_{-i} = p_1\big]\big)\Big\} \\
=&\mathbb{E}\big[Y_i(d,1) - Y_i(d,0) \mid p_0 < V_i \le p_0', p_1 < V_{-i} \le p_1'\big]
\end{aligned}
\tag{6}
$$

for $d \in \{0, 1\}$, where the function $\operatorname{sgn}(x)$ denotes the sign of x.

Additionally, the local average direct effect is identified by focusing on groups in which the peer is treated ($D_{-i} = 1$) and then exploiting exogenous shifts in the propensity scores.

This identifies the effect of changing unit i's own treatment status while holding the peer's treatment fixed, conditional on (V_i, V_{-i}) lying in the region $\{p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\}$,

$$\begin{aligned}
\operatorname{sgn}(2d-1)\cdot\Big\{&\big(\mathbb{E}\big[Y_i\mathbb{1}\{D_{-i}=d\} \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[Y_i\mathbb{1}\{D_{-i}=d\} \mid P_i = p_0', P_{-i} = p_1\big]\big) \\
&- \big(\mathbb{E}[Y_i\mathbb{1}\{D_{-i}=d\} \mid P_i = p_0, P_{-i} = p_1'] - \mathbb{E}[Y_i\mathbb{1}\{D_{-i}=d\} \mid P_i = p_0, P_{-i} = p_1]\big)\Big\}\Big/ \\
&\Big\{\big(\mathbb{E}\big[D_iD_{-i} \mid P_i = p_0', P_{-i} = p_1'\big] - \mathbb{E}\big[D_iD_{-i} \mid P_i = p_0', P_{-i} = p_1\big]\big) \\
&- \big(\mathbb{E}\big[D_iD_{-i} \mid P_i = p_0, P_{-i} = p_1'\big] - \mathbb{E}\big[D_iD_{-i} \mid P_i = p_0, P_{-i} = p_1\big]\big)\Big\} \\
=&\mathbb{E}\big[Y_i(1,d) - Y_i(0,d) \mid p_0 < V_i \leq p_0', p_1 < V_{-i} \leq p_1'\big]
\end{aligned} \tag{7}$$

for $d \in \{0,1\}$.

Consider the setting in Vazquez-Bare (2023), where the instrumental variable is binary, $Z_i \in \{0,1\}$. In this case, one can determine the ordering of the propensity scores $\mathbb{P}(D_i = 1 \mid Z_i = z, Z_{-i} = z')$ for $z, z' \in 0, 1$, and consequently the ordering of the threshold values $h_i(z,z')$, since the propensity score identifies the threshold function h_i by Equation (2). Without loss of generality, suppose the following order holds: $h_i(0,0) \leq h_i(0,1) \leq h_i(1,0) \leq h_i(1,1)$. Under this ordering, the treatment take-up equation $D_i = \mathbb{1}\{V_i \leq h_i(Z_i, Z_{-i})\}$ directly implies the monotonicity assumption imposed in Vazquez-Bare (2023). Moreover, this framework provides a mapping from unit i's unobserved characteristic V_i to the discrete types defined in Vazquez-Bare (2023). For instance, units with $h_i(0,1) \leq V_i \leq h_i(1,0)$ correspond to the complier type, denoted by $\xi_i = C$ in his framework.

Suppose $p_0 = h_i(0,1)$, $p_0' = h_i(1,0)$, $p_1 = h_{-i}(0,1)$, and $p_1' = h_{-i}(1,0)$. In this case, Equations (6) and (7) are intended to identify the local average spillover and direct effects for units with unobserved characteristics in the region $\{h_i(0,1) \leq V_i \leq h_i(1,0), h_{-i}(0,1) \leq V_{-i} \leq h_{-i}(1,0)\}$, that is, when both group members are compliers. However, with a binary instrument, it is impossible to observe a group with $P_i = h_i(1,0)$ and $P_{-i} = h_{-i}(1,0)$, since this would require each unit to take the treatment when assigned while the peer does not receive it, which is a scenario that cannot be realized with only two instrument values.

In contrast, when the instrumental variable Z_i exhibits richer variation, such as being continuously distributed, the propensity scores (P_i, P_{-i}) also vary continuously over their joint support $\mathcal{P}$. This additional variation makes it possible to take limits as $p_1' \to p_1$ and $p_0' \to p_0$, thereby identifying parameters of the form

$$\mathbb{E}\big[Y_i(d,1) - Y_i(d,0) \mid V_i = p_0, V_{-i} = p_1\big] \text{ and } \mathbb{E}\big[Y_i(1,d) - Y_i(0,d) \mid V_i = p_0, V_{-i} = p_1\big] \tag{8}$$

from Equations (6) and (7). These parameters capture the spillover and direct effects at

specific values of the unobserved characteristics (V_i, V_{-i}) within the support $\mathcal{P}$. By assuming that each point in the complier region $\{h_i(0,1) \leq V_i \leq h_i(1,0), h_{-i}(0,1) \leq V_{-i} \leq h_{-i}(1,0)\}$ can be identified, integrating the parameters in (8) over this region yields the local average spillover and direct effects for groups in which both members are compliers.

The next section provides formal definitions of the parameters in Equation (8) and develops a method for their identification. These parameters serve as building blocks for identifying not only the local average spillover and direct effects discussed above, but also a broader class of policy-relevant treatment effects of interest to researchers.

2.3 Causal Parameters and Identification Results

Definition 1 provides the formal definition of the causal spillover and direct effects in Equation (8).

Definition 1. (Marginal controlled spillover effects (MCSE) and marginal controlled direct effects (MCDE)) Consider the model in Equation (1).

1. Fix the treatment of unit i at $D_i = d$, $d \in \{0,1\}$. The marginal controlled spillover effect (MCSE), given $V_i = p_0$ and $V_{-i} = p_1$, is defined as

$$\text{MCSE}_i^{(d)}(p_0, p_1) \equiv \mathbb{E}[Y_i(d,1) - Y_i(d,0) \mid V_i = p_0, V_{-i} = p_1].$$

2. For unit i, fix the peer's treatment at $D_{-i} = d$, $d \in \{0,1\}$. The marginal controlled direct effect (MCDE), given $V_i = p_0$ and $V_{-i} = p_1$, is defined as

$$\text{MCDE}_i^{(d)}(p_0, p_1) \equiv \mathbb{E}[Y_i(1,d) - Y_i(0,d) \mid V_i = p_0, V_{-i} = p_1].$$

The marginal controlled spillover effect captures the impact of changing the peer's treatment on a unit's potential outcome while controlling the unit's own treatment status fixed, conditional on the unobserved characteristics (V_i, V_{-i}) within the group. Similarly, the marginal controlled direct effect measures the impact of changing a unit's own treatment on her potential outcome while controlling the peer's treatment constant, again conditional on (V_i, V_{-i}). Because both effects are defined relative to the group-level unobserved heterogeneity, they capture treatment effect heterogeneity arising from latent factors within the group.

As noted in the setting, the unobserved heterogeneities V_i and V_{-i} within a group can exhibit dependence. Identification of the parameters of interest therefore requires recovering the joint density of (V_i, V_{-i}). The analysis proceeds by first identifying their joint distribution

and subsequently deriving the joint density. Because the marginal distributions of V_i and V_{-i} are normalized to be uniform on the interval $(0,1)$, their joint distribution is characterized by the copula that captures the dependence structure between them. Formally, the copula is defined as

$$C_{V_i,V_{-i}}(p_0,p_1) \equiv \mathbb{P}\big(V_i \leq p_0, V_{-i} \leq p_1\big), (p_0,p_1) \in (0,1)^2.$$

Lemma 1 provides the identification of this copula on the support of the propensity scores (P_i, P_{-i}), without imposing any functional form assumptions, where P_i denotes unit i's propensity score as defined in Equation (2).

Lemma 1. (Copula of (V_i, V_{-i})) Under Assumptions 1-3, the copula between V_i and V_{-i} is identified as

$$C_{V_i,V_{-i}}(p_0,p_1) = \mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid P_i = p_0, P_{-i} = p_1\big),$$

for $(p_0,p_1) \in \mathcal{P}$, where $\mathcal{P}$ denotes the support of (P_i, P_{-i}).

Proof. See Appendix A.1. □

Identification requires the joint density of (V_i, V_{-i}), which is characterized by their copula density once the marginal distributions are normalized to be uniform. Let $c_{V_i,V_{-i}}(\cdot,\cdot)$ denote the copula density of (V_i, V_{-i}). Since Lemma 1 establishes identification of the copula, the copula density can be obtained provided that the conditional probability $\mathbb{P}(D_i = 1, D_{-i} = 1 \mid P_i, P_{-i})$ is twice differentiable. This differentiability condition requires that P_i and P_{-i} exhibit continuous variation, which in turn implies that at least some components of the instrument vector (Z_i, Z_{-i}) must be continuously distributed. Assumption 4 introduces this continuity requirement.

Assumption 4. (Continuous instruments) At least one component of the instrumental variables (Z_i, Z_{-i}) is continuously distributed.

It then follows that the copula density of (V_i, V_{-i}) is identified, as stated in Corollary 1.

Corollary 1. (Copula density of (V_i, V_{-i})) Suppose that Assumptions 1-4 hold. If $\mathbb{E}[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_0]$ is twice differentiable at $(p_0,p_1) \in \mathcal{P}$, the copula density $c_{V_i,V_{-i}}(p_0,p_1)$ is identified as

$$c_{V_i,V_{-i}}(p_0,p_1) = \frac{\partial^2 \mathbb{E}\big[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\big]}{\partial p_0 \partial p_1}.$$

Proof. See Appendix A.1. □

Following the literature, the conditional expectation of the potential outcome, given the values of the group-level unobserved characteristics (V_i, V_{-i}),

$$m_i^{(d_0,d_1)}(p_0,p_1) \equiv \mathbb{E}\big[Y_i(d_0,d_1) \mid V_i = p_0, V_{-i} = p_1\big], (p_0,p_1) \in (0,1)^2,$$

is defined as the marginal treatment response (MTR) function. The MCSEs and MCDEs introduced in Definition 1 are obtained as differences of the corresponding MTR functions. Hence, identification of the MCSEs and MCDEs requires identifying the underlying MTR functions. Theorem 1 provides the identification of the parameters of interest, MCSEs and MCDEs, while the detailed process for identifying MTR functions is presented in Appendix A.2.

Theorem 1. (Identifying MCSEs and MCDEs) Suppose that Assumptions 1-4 hold. For $d_0, d_1 \in 0, 1$ and $(p_0, p_1) \in \mathcal{P}$, the following additional regularity conditions are imposed: (i) $\mathbb{E}[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1]$ and $\mathbb{E}[Y_i \mathbb{1}\{D_i = d_0\}\mathbb{1}\{D_{-i} = d_1\} \mid P_i = p_0, P_{-i} = p_1]$ are twice differentiable; (ii) the marginal treatment response functions $m_i^{(d_0,d_1)}(p_0, p_1)$ are continuous; and (iii) the copula density $c_{V_i,V_{-i}}(p_0, p_1)$ is bounded from above and away from zero.

Then, the marginal controlled spillover effects (MCSEs) are identified as

$$\operatorname{sgn}(2d-1) \cdot \frac{\partial^2 \mathbb{E}\left[Y_i \mathbb{1}\{D_i = d\} \mid P_i = p_0, P_{-i} = p_1\right]}{\partial p_0 \partial p_1} \Big/ \frac{\partial^2 \mathbb{E}\left[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\right]}{\partial p_0 \partial p_1}$$

for $d \in \{0, 1\}$, and the function $\operatorname{sgn}(x)$ denotes the sign of x.

The marginal controlled direct effects (MCDEs) are identified as

$$\operatorname{sgn}(2d-1) \cdot \frac{\partial^2 \mathbb{E}\left[Y_i \mathbb{1}\{D_{-i} = d\} \mid P_i = p_0, P_{-i} = p_1\right]}{\partial p_0 \partial p_1} \Big/ \frac{\partial^2 \mathbb{E}\left[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\right]}{\partial p_0 \partial p_1}$$

for $d \in \{0, 1\}$.

Proof. See Appendix A.2. □

The identification of MCSEs and MCDEs formalizes the heuristic arguments in Equations (6) and (7) by taking the limits $p_1' \to p_1$ and $p_0' \to p_0$, which is feasible under the assumption of continuous variation in the instrumental variables. Moreover, the identification framework naturally extends to settings with exogenous covariates, with the corresponding results provided in Appendix A.3.

Remark 2. (Groups with multiple individuals) The identification strategy naturally extends to settings with more than two individuals per group. Consider a group of size $n < \infty$, indexed by $i \in \{1, \ldots, n\}$. In this case, the threshold function h_i in Equation (1) depends on the full vector of instrument assignments $(Z_1, \ldots, Z_n)$. The propensity score $\mathbb{P}(D_i = 1 \mid Z_1, \ldots, Z_n)$ identifies the threshold h_i. The joint distribution of the unobserved heterogeneities within the group is then recovered from the conditional probability $\mathbb{P}(D_1 = 1, \ldots, D_n = 1 \mid P_1, \ldots, P_n)$, where P_i denotes the propensity score of individual i.

Once this joint distribution is identified, the marginal treatment response functions can be obtained by differentiating

$$\mathbb{E}\left[Y_i \mathbb{1}\left\{D_1 = d_1\right\} \cdots \mathbb{1}\left\{D_n = d_n\right\} \mid P_1, \cdots, P_n\right]$$

with respect to $(P_1, \ldots, P_n)$, under suitable smoothness conditions. Finally, differences between the resulting MTR functions yield the MCSEs and MCDEs.

Remark 3. (Testing the spillover structure) The identification of marginal treatment response functions makes it possible to test additional structural assumptions about the nature of spillovers. For example, in addition to Assumptions 1-3, suppose that each unit's outcome depends not on the entire treatment vector $\boldsymbol{D} \equiv (D_1, \cdots, D_n)$, but instead on a lower-dimensional function $H(\cdot)$ of this vector. In the literature, $H(\cdot)$ is commonly referred to as the exposure mapping. A standard specification is the average treatment level within the group, $H(\boldsymbol{D}) = \sum_{i=1}^{n} D_i/n$. Under this structure, given unit i's own treatment $D_i = d$, any two treatment vectors $\boldsymbol{d} = (d_1, \cdots, d_n)$ and $\tilde{\boldsymbol{d}} = (\tilde{d}_1, \cdots, \tilde{d}_n)$ that generate the same exposure level, $H(\boldsymbol{d}) = H(\tilde{\boldsymbol{d}})$, should yield identical marginal treatment response functions:

$$\mathbb{E}\big[Y_i(d, \boldsymbol{d}) \mid V_1 = p_1, \cdots, V_n = p_n\big] = \mathbb{E}\big[Y_i(d, \tilde{\boldsymbol{d}}) \mid V_1 = p_1, \cdots, V_n = p_n\big],$$

for all $(p_1, \cdots, p_n)$ in the support of the propensity score functions. This equality provides a testable implication of the assumed spillover structure $H(\cdot)$, thereby linking identification of MTR functions to specification testing of exposure mappings.

2.4 Testable Implications of Identifying Assumptions

In this section, we discuss the testable implications of our model structure and imposed assumptions. Corollary 1 gives identification of the copula density. Since a fundamental property of any copula density is that it must be nonnegative, this leads to a testable inequality:

$$\frac{\partial^2 \mathbb{E}\big[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\big]}{\partial p_0 \partial p_1} \geq 0$$

holds for any $(p_0, p_1) \in \mathcal{P}$ if the cross-derivative correctly identifies the copula density. This constitutes one of the testable implications of our framework.

In addition, the proof of Theorem 1 provides the identification of the conditional distributions of potential outcomes, $\mathbb{P}(Y_i(d, d') \in A \mid V_i = p_0, V_{-i} = p_1)$. By the basic properties of probability, this quantity must lie within the interval $[0, 1]$. Combined with the earlier testable implication derived from the copula density, this leads to a set of inequality restric-

tions that must hold for any Borel set $A \subseteq \mathcal{Y}$ and for all $d \in \{0,1\}$:

$$\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{E}\Big[\mathbb{1}\big\{Y_i(d,d) \in A\big\}\mathbb{1}\big\{D_i = d\big\}\mathbb{1}\big\{D_{-i} = d\big\} \mid P_i = p_0, P_{-i} = p_1\Big] \geq 0,$$
$$-\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{E}\Big[\mathbb{1}\big\{Y_i(d,1-d) \in A\big\}\mathbb{1}\big\{D_i = d\big\}\mathbb{1}\big\{D_{-i} = 1-d\big\} \mid P_i = p_0, P_{-i} = p_1\Big] \geq 0.$$

The above inequalities are commonly referred to as nesting inequalities in the literature.

It is important to note that both the copula function and the marginal treatment response functions depend on the propensity score values rather than directly on the instrument values. This observation gives rise to an additional set of testable implications. Specifically, if there exist distinct instrument values $(z_0, z_1) \neq (\tilde{z}_0, \tilde{z}_1)$ such that the corresponding propensity scores are equal, i.e., $P_i(z_0, z_1) = P_i(\tilde{z}_0, \tilde{z}_1) = p_0$ and $P_{-i}(z_0, z_1) = P_{-i}(\tilde{z}_0, \tilde{z}_1) = p_1$, then the identified objects that depend solely on (p_0, p_1) must also be equal. This leads to the following testable equality

$$\begin{aligned}
&\mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid Z_i = z_0, Z_{-i} = z_1\big) \\
=&\mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid P_i(z_0, z_1) = p_0, P_{-i}(z_1, z_0) = p_1\big) \\
=&\mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid P_i(\tilde{z}_0, \tilde{z}_1) = p_0, P_{-i}(\tilde{z}_1, \tilde{z}_0) = p_1\big) \\
=&\mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid Z_i = \tilde{z}_0, Z_{-i} = \tilde{z}_1\big),
\end{aligned}$$

since both conditional probabilities $\mathbb{P}(D_i = 1, D_{-i} = 1 \mid Z_i = z_0, Z_{-i} = z_1)$ and $\mathbb{P}(D_i = 1, D_{-i} = 1 \mid Z_i = \tilde{z}_0, Z_{-i} = \tilde{z}_1)$ identify the copula function $C_{V_i, V_{-i}}(p_0, p_1)$. Similarly, we should have

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid Z_i = z_0, Z_{-i} = z_1\big] \\
=&\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid P_i(z_0, z_1) = p_0, P_{-i}(z_1, z_0) = p_1\big] \\
=&\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid P_i(\tilde{z}_0, \tilde{z}_1) = p_0, P_{-i}(\tilde{z}_1, \tilde{z}_0) = p_1\big] \\
=&\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid Z_i = \tilde{z}_0, Z_{-i} = \tilde{z}_1\big],
\end{aligned}$$

since both conditional means $\mathbb{E}[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid Z_i = z_0, Z_{-i} = z_1]$ and $\mathbb{E}[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid Z_i = \tilde{z}_0, Z_{-i} = \tilde{z}_1]$ identify the integral of $\mathbb{P}(Y_i(d_0, d_1) \in A \mid V_i = v_0, V_{-i} = v_1)$ weighted by the copula density $c_{V_i, V_{-i}}(v_0, v_1)$, according to the proof of Theorem 1. These types of equalities are commonly referred to in the literature as index sufficiency conditions.

To conclude, these insights lead to the following set of testable implications.

Proposition 1. (Testable implications) The following conditions constitute the testable

implications of Assumptions 1-3, given the model structure specified in Equation (1).

1. (Nesting inequalities) For any Borel set $A \subseteq \mathcal{Y}$, $d \in \{0,1\}$, and $(p_0, p_1) \in \mathcal{P}$, the following inequality should hold:

$$\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{E}\Big[\mathbb{1}\{Y_i(d,d) \in A\}\mathbb{1}\{D_i = d\}\mathbb{1}\{D_{-i} = d\} \mid P_i = p_0, P_{-i} = p_1\Big] \geq 0,$$
$$-\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{E}\Big[\mathbb{1}\{Y_i(d,1-d) \in A\}\mathbb{1}\{D_i = d\}\mathbb{1}\{D_{-i} = 1-d\} \mid P_i = p_0, P_{-i} = p_1\Big] \geq 0.$$

2. (Index sufficiency) For any Borel set $A \subseteq \mathcal{Y}$, $d_0, d_1 \in \{0,1\}$, and different values of instruments, $(z_0, z_1) \neq (\tilde{z}_0, \tilde{z}_1)$, that yield the same propensity score values such that $P_i(z_0, z_1) = P_i(\tilde{z}_0, \tilde{z}_1) = p_0$ and $P_{-i}(z_0, z_1) = P_{-i}(\tilde{z}_0, \tilde{z}_1) = p_1$, the following equalities should hold:

$$\begin{aligned}&\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid Z_i = z_0, Z_{-i} = z_1\big] \\ =&\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}\mathbb{1}\{D_i = d_0\} \cdot \mathbb{1}\{D_{-i} = d_1\} \mid Z_i = \tilde{z}_0, Z_{-i} = \tilde{z}_1\big].\end{aligned}$$

Existing literature, such as Carr and Kitagawa (2021), has developed methods that could be implemented to test the conditions in Proposition 1. We refer interested readers to their paper for details on implementation.

2.5 Other Treatment Effects

MCSEs and MCDEs not only capture heterogeneous spillover and direct effects, but also serve as building blocks for deriving other causal parameters commonly studied in the literature.

Researchers are often interested in identifying average spillover and direct effects among the population. Specifically, we define the average controlled spillover effect (ACSE) as $\text{ACSE}_i(d) \equiv \mathbb{E}[Y_i(d,1) - Y_i(d,0)]$, which captures the expected change in unit i's outcome when her peers' treatment status shifts exogenously from zero to one, holding her own treatment status fixed at $D_i = d$. Similarly, we define the average controlled direct effect (ACDE) as $\text{ACDE}_i(d) \equiv \mathbb{E}[Y_i(1,d) - Y_i(0,d)]$, which reflects the expected change in unit i's outcome when her own treatment status changes exogenously from zero to one, holding her peers' treatment status fixed at $D_{-i} = d$.

Under the assumption that the propensity scores have full support, i.e., $\mathcal{P} = [0,1] \times [0,1]$, we can point identify both the ACSEs and the ACDEs using the MCSEs, the MCDEs, and

the copula density of (V_i, V_{-i}) identified in Section 2.3:

$$\text{ACSE}_i(d) = \int_0^1 \int_0^1 \text{MCSE}_i(d; p_0, p_1) c_{V_i, V_{-i}}(p_0, p_1) dp_0 dp_1,$$
$$\text{ACDE}_i(d) = \int_0^1 \int_0^1 \text{MCDE}_i(d; p_0, p_1) c_{V_i, V_{-i}}(p_0, p_1) dp_0 dp_1.$$

In cases where the propensity scores do not exhibit full support, we are still able to partially identify the average controlled spillover and direct effects under the additional assumption that potential outcomes are almost surely bounded, i.e., $|Y_i(d, d')| \leq B, B < \infty$. An alternative approach is to apply an extrapolation method similar to that proposed by Mogstad, Santos, and Torgovitsky (2018), which can extend identification beyond the identified support of the propensity scores.

In addition, we can identify the policy relevant treatment effect (PRTE) using our identified MCSEs and MCDEs. The PRTE, developed by Heckman and Vytlacil (2001, 2005), measures the expected change in outcomes resulting from a policy-induced change in treatment selection. This parameter is particularly useful in contexts with heterogeneous treatment effects, where policy interventions change the selection mechanism into treatment.

A change from a baseline policy to an alternative policy may modify the incentives for treatment selection within a subpopulation. Therefore, the PRTE can be represented as a weighted average of MCDEs and MCSEs, averaged over region where the policy shift alters the treatment status of at least one group member. The local average effect is one of the commonly studied policy treatment effects. We define the local average controlled spillover effect (LACSE) and the local average controlled direct effect (LACDE) and discuss how they can be constructed from the MCDEs and MCSEs in Section 3.2.2.

We can also characterize other PRTEs arising from exogenous policy changes. Let $\mathcal{A}$ denote a set of potential policies, and define D^a as the treatment status under policy $a \in \mathcal{A}$. For each unit i, the treatment decision under policy a is given by $D_i^a = \mathbb{1}\{V_i^a \leq P_i^a(Z_i^a, Z_{-i}^a)\}$, where $P_i^a(Z_i^a, Z_{-i}^a) = \mathbb{P}(D_i^a = 1 \mid Z_i^a, Z_{-i}^a)$ denotes the policy-specific propensity score. We assume that the implementation of a given policy a does not affect the distribution of potential outcomes or the distribution of unobserved heterogeneities, as decribed in Assumption 5. Rather, the policy only affects the selection mechanism through changes in the propensity score function.

Assumption 5. (Policy invariance) The distribution of

$$\left\{(Y_i^a(d, d'), Y_{-i}^a(d, d'), V_i^a, V_{-i}^a)\right\}_{d,d' \in \{0,1\}}$$

is invariant to any policy $a \in \mathcal{A}$.

The PRTE is defined as

$$\text{PRTE}(a, a') \equiv \frac{\mathbb{E}[Y_i^{a'} - Y_i^a]}{\Delta P},$$

where ΔP represents the proportion of groups in which at least one member changes treatment status due to the policy intervention. Under Assumption 5, we can express the expected outcome under policy a, denoted $\mathbb{E}[Y_i^a]$, as follows:

$$\begin{aligned}\mathbb{E}\big[Y_i^a\big] = \int_0^1 \int_0^1 \Big\{ & m_i^{(1,1)}(p_0, p_1)\mathbb{P}\big(P_i^a \geq p_0, P_{-i}^a \geq p_1\big) \\ & + m_i^{(1,0)}(p_0, p_1)\mathbb{P}\big(P_i^a \geq p_0, P_{-i}^a < p_1\big) \\ & + m_i^{(0,1)}(p_0, p_1)\mathbb{P}\big(P_i^a < p_0, P_{-i}^a \geq p_1\big) \\ & + m_i^{(0,0)}(p_0, p_1)\mathbb{P}\big(P_i^a < p_0, P_{-i}^a < p_1\big)\Big\} c_{V_i, V_{-i}}(p_0, p_1) dp_0 dp_1.\end{aligned}$$

Since $\mathbb{E}[Y_i^a]$ can be expressed as a weighted average of marginal treatment response functions and the PRTE is defined as the difference in $\mathbb{E}[Y_i^a]$ under different policies a, the PRTE can be interpreted as a weighted average of the MCDEs and MCSEs over specific regions of unobserved heterogeneity (V_i, V_{-i}). Therefore, once the MCDEs and MCSEs are identified, they can be used to construct the PRTE for a wide range of policy interventions. In Appendix B, we provide explicit derivations of the PRTE under three common types of policy changes.

3 Comparisons With Relevant Literature

3.1 Comparing With Marginal Treatment Effect

If spillover effects are present but researchers ignore them and apply standard marginal treatment effect (MTE) identification strategies, we can show that the following expressions,

$$\frac{\partial}{\partial p_0}\mathbb{E}\big[Y_i \mid P_i(Z_i) = p_0\big], P_i(z_0) \equiv \mathbb{P}\big(D_i = 1 \mid Z_i = z_0\big),$$

no longer identify the MTE, $\mathbb{E}[Y_i(1) - Y_i(0) \mid V_i = p_0]$.

In the presence of the spillover structure specified in Equation (1), the conventional

propensity score identifies

$$\begin{aligned}
&\mathbb{P}\left(D_i = 1 \mid Z_i = z_0\right) \\
=&\mathbb{P}\left(V_i \leq h_i(z_0, Z_{-i}) \mid Z_i = z_0\right) \\
=&\int_{z_1 \in \mathcal{Z}} \mathbb{P}\left(V_i \leq h_i(z_0, z_1) \mid Z_i = z_0, Z_{-i} = z_1\right) f_{Z_{-i}|Z_i=z_0}\left(z_1\right) dz_1 \\
=&\int_{z_1 \in \mathcal{Z}} h_i\left(z_0, z_1\right) f_{Z_{-i}|Z_i=z_0}\left(z_1\right) dz_1,
\end{aligned}$$

where the second equality follows from the law of iterated expectations, while the final equality holds under the independence assumption 1 and the distributional normalization in assumption 3. This expression reveals that the conventional propensity score is a weighted average of the threshold function $h_i(z_0, z_1)$ over Z_{-i}, conditional on Z_i.

Additionally, we show that when spillover effects are present, the conventional MTE identifier, $\partial \mathbb{E}[Y_i \mid P_i\left(Z_i\right) = p_0]/\partial p_0$ identifies a weighted average of MCDEs, along with residuals capturing the bias induced by the peer treatment spillovers and the correlation between unobserved heterogeneities V_i and V_{-i}.

Proposition 2. (Breakdown of MTE causal interpretation) Consider the model with spillovers in Equation (1). Suppose that the assumptions in Theorem 1 hold. Then, the MTE identifier, $\partial \mathbb{E}[Y_i \mid P_i\left(Z_i\right) = p_0]/\partial p_0$, identifies

$$\begin{aligned}
&\int_0^1 \int_0^{p_1} \text{MCDE}_i(1; p_0, v_1) c_{V_i, V_{-i}}(p_0, v_1) f_{P_{-i}|P_i=p_0}(p_1) dv_1 dp_1 \\
+&\int_0^1 \int_{p_1}^1 \text{MCDE}_i(1; p_0, v_1) c_{V_i, V_{-i}}(p_0, v_1) f_{P_{-i}|P_i=p_0}(p_1) dv_1 dp_1 \\
&+ \mathcal{R}_{11} + \mathcal{R}_{10} + \mathcal{R}_{01} + \mathcal{R}_{00}.
\end{aligned}$$

The residuals $R_{dd'}, d, d' \in \{0, 1\}$, are generally nonzero and depend on both the individual's and the peer's treatment statuses, as well as the copula density between the unobserved heterogeneities V_i and V_{-i}. The explicit forms of $R_{dd'}$ are provided in Appendix C.1.

Proof. See Appendix C.1. □

If spillover effects are absent in our setting, which means that $Y_i(D_i, d) = Y_i(D_{ig}, d') \equiv Y_i(D_i)$, and $h_i(Z_i, z) = h_{ig}(Z_i, z') \equiv h_i(Z_i)$, then our identification results reduce to the

standard MTE framework. In this case, the propensity score we defined identifies

$$\begin{aligned}
&\mathbb{P}\big(D_i = 1 \mid Z_i = z_0, Z_{-i} = z_1\big) \\
=&\mathbb{P}\big(V_i \leq h_i(z_0) \mid Z_i = z_0, Z_{-i} = z_1\big) \\
=&\mathbb{P}\big(V_i \leq h_i(z_0)\big) = h_i(z_0),
\end{aligned}$$

which is the threshold function from the standard MTE setting, where peer instruments do not influence individual treatment decisions.

If we take the cross-partial derivative of $\mathbb{P}(D_i = 1, D_{-i} = 1 \mid P_i = p_0, P_{-i} = p_1)$, we identify the copula density evaluated at the point (p_0, p_1), $c_{V_i,V_{-i}}(p_0, p_1)$, as

$$\begin{aligned}
&\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{P}(D_i = 1, D_{-i} = 1 \mid P_i = p_0, P_{-i} = p_1) \\
=&\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{P}(V_i \leq p_0, V_{-i} \leq p_1 \mid P_i = p_0, P_{-i} = p_1) \\
=&\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{P}\big(V_i \leq p_0, V_{-i} \leq p_1\big) \\
=&\frac{\partial}{\partial p_0}\mathbb{P}\big(V_i \leq p_0\big) \cdot \frac{\partial}{\partial p_1}\mathbb{P}\big(V_{-i} \leq p_1\big) = 1.
\end{aligned}$$

The copula density equals one since V_i and V_{-i} are independent. This aligns with the standard MTE framework, where the individuals' unobserved heterogeneities are independent.

Finally, by taking the cross-partial derivative of $\mathbb{E}[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1]$, we can derive

$$\begin{aligned}
&\frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[Y_i D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\big] \\
=&\frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[Y_i \mathbb{1}\{V_i \leq p_0\}\mathbb{1}\{V_{-i} \leq p_1\} \mid P_i = p_0, P_{-i} = p_1\big] \\
=&\frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[Y_i \mathbb{1}\{V_i \leq p_0\}\mathbb{1}\{V_{-i} \leq p_1\}\big] \\
=&\frac{\partial^2}{\partial p_0 \partial p_1}\int_0^{p_1}\int_0^{p_0}\mathbb{E}\big[Y_i(1) \mid V_i = v_0, V_{-i} = v_1\big]c_{V_i,V_{-i}}(v_0, v_1)dv_0 dv_1 \\
=&\frac{\partial^2}{\partial p_0 \partial p_1}\int_0^{p_1}\int_0^{p_0}\mathbb{E}\big[Y_i(1) \mid V_i = v_0\big]dv_0 dv_1 \\
=&\mathbb{E}\big[Y_i(1) \mid V_i = p_0\big],
\end{aligned}$$

where the third line follows from Assumption 1, the fifth line holds because V_{-i} is independent with $(Y_i(d), V_i)$, and the copula density equals to one. As a result, we identify the marginal treatment response function as in the standard MTE framework, where there are no spillover effects from the peer's treatment.

One may interpret the model in Equation (1) as a standard MTE framework, treating (D_{-i}, Z_{-i}) as covariates. A natural question arises: can we identify the marginal treatment response functions using standard MTE methods by conditioning on the peer's treatment D_{-i} as a covariate? The answer is affirmative, provided that D_{-i} is exogenous with respect to unit i's potential outcomes and unobserved heterogeneity. Specifically, this requires the independence assumption

$$(Z_i, Z_{-i}, D_{-i}) \perp\!\!\!\perp \{(V_i, Y_i(d, d'))\}_{d\in\{0,1\}, d'\in\{0,1\}}. \tag{9}$$

When this condition holds, the marginal treatment response functions conditional on the individual's own unobserved heterogeneity, $\mathbb{E}[Y_i(d, d') \mid V_i]$, can be identified using the standard MTE identification strategy, extended to include D_{-i} as an additional covariate. For instance,

$$\begin{aligned}
&\frac{\partial}{\partial p}\mathbb{E}\left[Y_i \cdot D_i \mid P_i(Z_i, Z_{-i}) = p, D_{-i} = d'\right] \\
=&\frac{\partial}{\partial p}\mathbb{E}\left[Y_i(1, d') \cdot \mathbb{1}\{V_i \le p\} \mid P_i(Z_i, Z_{-i}) = p, D_{-i} = d'\right] \\
=&\frac{\partial}{\partial p}\int_0^p \mathbb{E}\left[Y_i(1, d') \mid V_i = v\right] dv \\
=&\mathbb{E}\left[Y_i(1, d') \mid V_i = p\right].
\end{aligned}$$

It is important to note that the second equality in the above derivation holds only if D_{-i} satisfies the independence condition in Equation (9).

However, the assumption of exogeneity for D_{-i} is not realistic within our spillover framework. As illustrated in Figure 1, the peer's unobserved heterogeneity V_{-i} affects her treatment D_{-i} and may also be correlated with unit i's unobserved confounder V_i, thereby threatening the exogeneity of D_{-i}. In the marginal spillover setting, where V_i and V_{-i} are allowed to be arbitrarily correlated, applying the standard MTE identification strategy by conditioning on D_{-i} would lead to

$$\begin{aligned}
&\frac{\partial}{\partial p}\mathbb{E}\left[Y_i \cdot D_i \mid P_i(Z_i, Z_{-i}) = p, D_{-i} = d'\right] \\
=&\frac{\partial}{\partial p}\mathbb{E}\left[Y_i(1, d') \cdot \mathbb{1}\{V_i \le p\} \mid P_i(Z_i, Z_{-i}) = p, D_{-i} = d'\right] \\
=&\frac{\partial}{\partial p}\int_0^p \mathbb{E}\left[Y_i(1, d') \mid V_i = v, D_{-i} = 1\right] f_{V_i|D_{-i}=d}(v) dv \\
=&\mathbb{E}\left[Y_i(1, d') \mid V_i = p, D_{-i} = d'\right],
\end{aligned}$$

with the distribution of $V_i \mid D_{-i} = d$ being normalized to follow a uniform distribution on

$[0,1]$. However, $\mathbb{E}\left[Y_i(1,d') \mid V_i = p, D_{-i} = d'\right]$ is not equal to $\mathbb{E}\left[Y_i(1,d') \mid V_i = p\right]$, because D_{-i} is determined by V_{-i}, which is dependent on $Y_i(d,d')$ even conditioning on V_i, as illustrated in Figure 1.

One may also interpret the model in Equation (1) as an MTE framework with multivalued treatments, such as the framework in Lee and Salanié (2018). We discuss the connection between our spillover setting and the multivalued treatment literature in Appendix C.2.

3.2 Comparison with Spillover under Discrete Instrument

3.2.1 Underidentification With Discrete Instrument

In this section, we compare our spillover framework to a related setting with discrete population types, as studied in Vazquez-Bare (2023). The main difference is that Vazquez-Bare (2023) assumes each unit i is randomly assigned a binary instrument, $Z_i \in \{0,1\}$, and specifies the treatment equation as

$$\begin{aligned} D_i = &\left[D_i(1,1)Z_{-i} + D_i(1,0)(1-Z_{-i})\right]Z_i \\ &+ \left[D_i(0,1)Z_{-i} + D_i(0,0)(1-Z_{-i})\right](1-Z_i). \end{aligned}$$

$D_i(z,z')$, where $z,z' \in \{0,1\}$, denotes the potential treatment status of unit i, given that her own instrument is fixed at $Z_i = z$ and her peer's instrument at $Z_{-i} = z'$ exogenously. As in the continuous instrument setting, we require the instrument to be randomly assigned and excluded from the outcome equation.

Based on values of potential decisions $D_i(z,z')$, one can divide the population into different types. Vazquez-Bare (2023) introduces a monotonicity assumption to reduce types in the population in this spillover setting .

Assumption 6. (Monotonicity) For all i and g, $D_i(1,1) \geq D_i(1,0) \geq D_i(0,1) \geq D_i(0,0)$.

The order specified in Assumption 6 is without loss of generality and can be rearranged, provided that the potential treatments can be ordered. However, Assumption 6 requires that this ordering holds uniformly across all units in the population. This makes it a stronger assumption than what we impose in the marginal spillover setting, where we allow group members to have different propensity scores $P(Z_i, Z_{-i})$. As a result, the ordering of potential treatments $D_i(z,z')$ may vary across individuals within a group.

By imposing Assumption 6, Vazquez-Bare (2023) divides the population into five types, as listed in Table 1. Let $T_i \in \{at, sc, c, gc, nt\}$ denote the type of unit i. Vazquez-Bare (2023) defines the parameters of interest as $\mathbb{E}[Y_i(d,d') \mid T_i = t, T_{-i} = t']$, referring to them as "local average potential outcomes." These are conceptually equivalent to our marginal treatment

response functions, except they are conditioned on discrete types rather than continuous unobserved heterogeneity.

Table 1: Population types

$D_i(1,1)$	$D_i(1,0)$	$D_i(0,1)$	$D_i(0,0)$	Type (T_i)
1	1	1	1	always-taker(at)
1	1	1	0	social-interaction complier (sc)
1	1	0	0	complier (c)
1	0	0	0	group complier (gc)
0	0	0	0	never-taker (nt)

Proposition 1 in Vazquez-Bare (2023) provides a method for point identifying the marginal distribution of types in the population, $\mathbb{P}(T_i = t)$, using the observed probabilities $\mathbb{P}(D_i = d \mid Z_i = z, Z_{-i} = z')$ and their differences. However, the joint distribution of types, $\mathbb{P}(T_i = t, T_{-i} = t')$ for $t, t' \in \{at, sc, c, gc, nt\}$, cannot be point identified using the joint probabilities $\mathbb{P}(D_i = d, D_{-i} = d' \mid Z_i = z, Z_{-i} = z')$. We can only point identify two probabilities, $\mathbb{P}(T_i = at, T_{-i} = at)$ and $\mathbb{P}(T_i = nt, T_{-i} = nt)$ as

$$\mathbb{P}\big(T_i = at, T_{-i} = at\big) = \mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid Z_i = 0, D_{-i} = 0\big),$$
$$\mathbb{P}\big(T_i = nt, T_{-i} = nt\big) = \mathbb{P}\big(D_i = 0, D_{-i} = 0 \mid Z_i = 1, D_{-i} = 1\big).$$

The remaining probabilities, $\mathbb{P}(D_i = d, D_{-i} = d' \mid Z_i = z, Z_{-i} = z')$, along with their differences, reflect mixtures over the joint distribution of types. Therefore, it is not possible to point identify each proportion of the joint types $(T_i = t, T_{-i} = t')$ separately. For example, the difference between $\mathbb{P}(D_i = 1, D_{-i} = 1 \mid Z_i = 0, Z_{-i} = 0)$ and $\mathbb{P}(D_i = 1, D_{-i} = 1 \mid Z_i = 1, Z_{-i} = 0)$ corresponds to a sum over five different joint type probabilities,

$$\begin{aligned}
&\mathbb{P}\big(D_{ig} = 1, D_{-i,g} = 1 \mid Z_{ig} = 1, Z_{-i,g} = 0\big) - \mathbb{P}\big(D_{ig} = 1, D_{-i,g} = 1 \mid Z_{ig} = 0, Z_{-i,g} = 0\big) \\
=&\mathbb{P}\big(T_{ig} = at, T_{-i,g} = sc\big) + \mathbb{P}\big(T_{ig} = sc, T_{-i,g} = at\big) + \mathbb{P}\big(T_{ig} = sc, T_{-i,g} = sc\big) \\
&+ \mathbb{P}\big(T_{ig} = c, T_{-i,g} = at\big) + \mathbb{P}\big(T_{ig} = c, T_{-i,g} = sc\big).
\end{aligned}$$

Nevertheless, none of these individual joint type probabilities can be point identified.

In fact, the joint probabilities of types are underidentified. Our objective is to identify them by decomposing the following observed treatment probabilities,

$$\mathbb{P}\big(D_i = d, D_{-i} = d' \mid Z_i = 1, Z_{-i} = 1\big),$$

for $(d, d') = (0,0), (1,0), (1,1)$,

$$\mathbb{P}\big(D_i = d, D_{-i} = d' \mid Z_i = 1, Z_{-i} = 0\big),$$

for $(d, d') = (0,0), (1,0), (0,1), (1,1)$, and

$$\mathbb{P}\big(D_i = d, D_{-i} = d' \mid Z_i = 0, Z_{-i} = 0\big),$$

for $(d, d') = (0,0), (1,0), (1,1)$. Thus, we have ten equations available to identify the joint type probabilities $\mathbb{P}(T_i = t, T_{-i} = t')$. However, there are fifteen such joint probabilities, which exceeds the number of available equations. Therefore, the system is underidentified, and we cannot point identify all joint type probabilities. Since the joint distribution of types is not point-identified, the "local average potential outcomes" $\mathbb{E}[Y_i(d, d') \mid T_i = t, T_{-i} = t']$ cannot be point identified either.

3.2.2 Connection to the Continuous Instrument Case

In the setting with continuous instrument, the treatment equation is modeled as $D_i = \mathbb{1}\{V_i \leq h_i(Z_i, Z_{-i})\}$. Under our imposed assumptions, the function $h_i(z, z')$ can be identified using the propensity score $\mathbb{P}(D_i = 1 \mid Z_i = z, Z_{-i} = z')$. If there exist two values $z_0, z_1 \in \mathcal{Z}$ such that $h_i(z, z'), z, z' \in \{z_0, z_1\}$, is ordered consistently across all units, then the potential treatments $D_i(z, z'), z, z' \in \{z_0, z_1\}$, satisfy the monotonicity assumption in Assumption 6. Without loss of generality, we assume the following ordering holds: $h_i(z_0, z_0) \leq h_i(z_0, z_1) \leq h_i(z_1, z_0) \leq h_i(z_1, z_1)$.

According to the definitions of types in Table 1, we can construct a mapping from a unit's unobservable V_i to the corresponding type T_i,

$$\begin{cases} T_i = at, & \text{if } V_i \leq h_i(z_0, z_0) \\ T_i = sc, & \text{if } h_i(z_0, z_0) < V_i \leq h_i(z_0, z_1) \\ T_i = c, & \text{if } h_i(z_0, z_1) < V_i \leq h_i(z_1, z_0) \\ T_i = gc, & \text{if } h_i(z_1, z_0) < V_i \leq h_i(z_1, z_1) \\ T_i = nt, & \text{if } V_i \geq h_i(z_1, z_1) \end{cases} \tag{10}$$

Figure 2 presents the mapping from the space of unobservables (V_0, V_1) to the joint types of corresponding units, (T_0, T_1). As shown in Figure 2, when using a binary instrument, changing the instrument values from (z, z') to $(\tilde{z}, \tilde{z}')$, where $z, z', \tilde{z}, \tilde{z}' \in \{z_0, z_1\}$, does not allow us to point identify the probability of any pair of types (t_0, t_1), $t_0, t_1 \in \{nt, gc, c, sc, at\}$.

However, in the case of a continuous instrument, as the instrument vector $(\tilde{z}, \tilde{z}')$ approaches (z, z'), we can identify the joint density at $(V_0, V_1) = (p_0(z, z'), p_1(z', z))$.

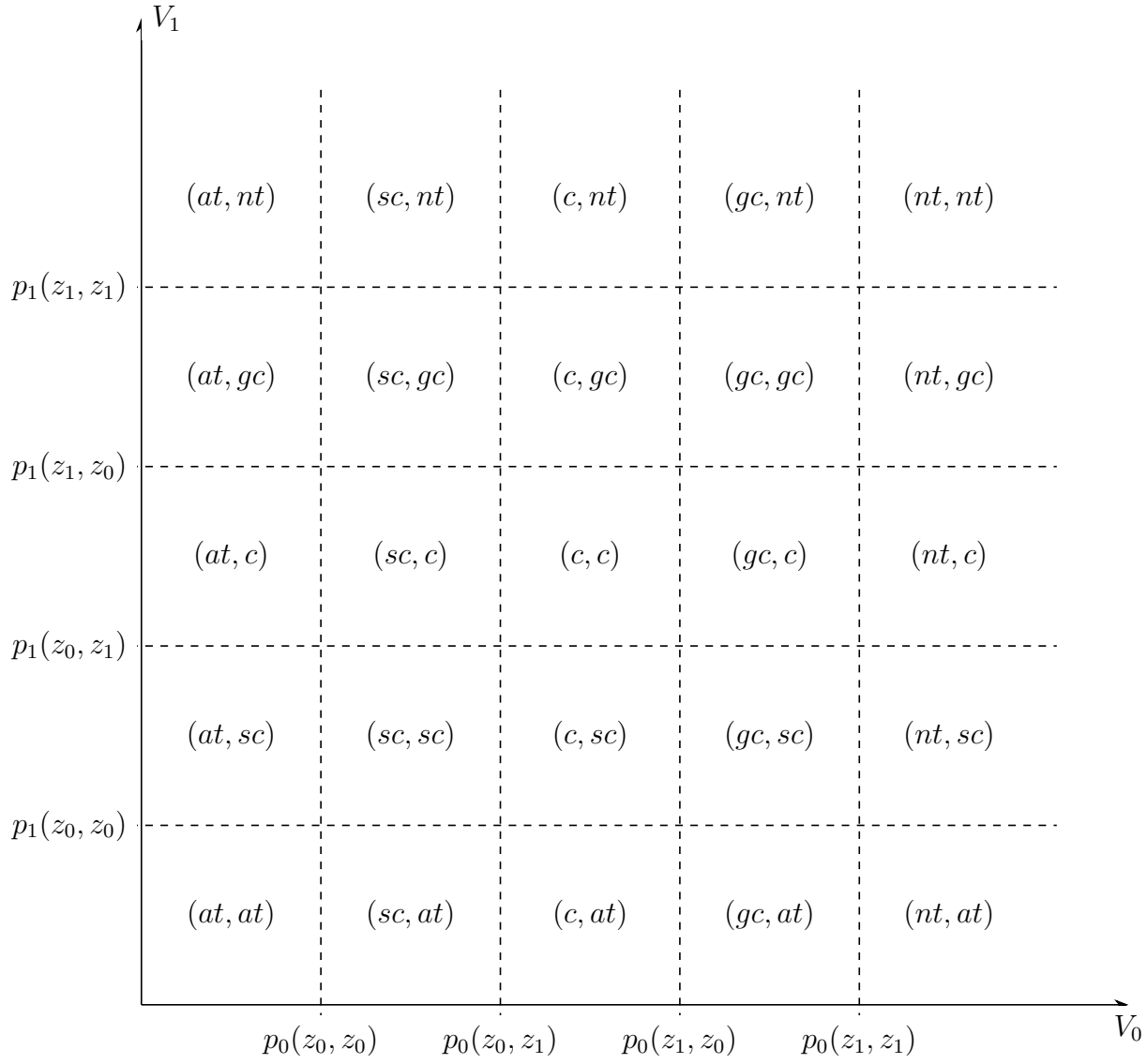

Figure 2: Mapping From Unobservables to Types Under Binary Instrument

In the marginal spillover setting, we are able to identify the joint density of the unobservables (V_i, V_{-i}) on the support of the propensity score function $\mathcal{P}$. Leveraging the relationships in Equation (10), if $\mathcal{P}$ has full support over $(0,1) \times (0,1)$, we can point identify the joint distribution of types $\mathbb{P}(T_i = t, T_{-i} = t')$, the local average controlled spillover effects (LACSE), defined as $\mathbb{E}[Y_i(d,1) - Y_i(d,0) \mid T_i = t, T_{-i} = t']$, and the local average controlled direct effects (LACDE), defined as $\mathbb{E}[Y_i(1,d) - Y_i(0,d) \mid T_i = t, T_{-i} = t']$, where $d \in \{0,1\}$ and $t, t' \in \{at, sc, c, gc, nt\}$. These point identification results are not attainable with a discrete instrument, as discussed in Section 3.2.1.

For example, we can point identify the probability that both units in the group are compliers as

$$\mathbb{P}(T_i = c, T_{-i} = c) = \int_{h_{-i}(z_0,z_1)}^{h_{-i}(z_1,z_0)} \int_{h_i(z_0,z_1)}^{h_i(z_1,z_0)} c_{V_i,V_{-i}}(v_0, v_1) dv_0 dv_1,$$

where the thresholds $h_i(Z_i, Z_{-i})$ are point identified through the propensity score $\mathbb{P}(D_i = 1 \mid Z_i, Z_{-i})$.

In addition, the local average controlled spillover effects, conditional on both units in the group being compliers, can be identified as

$$\begin{aligned}
&\mathbb{E}\left[Y_i(d_0, 1) - Y_i(d_0, 0) \mid T_i = c, T_{-i} = c\right] \\
=&\frac{1}{\mathbb{P}(T_i = c, T_{-i} = c)} \int_{h_{-i}(z_0,z_1)}^{h_{-i}(z_1,z_0)} \int_{h_i(z_0,z_1)}^{h_i(z_1,z_0)} \mathrm{MCSE}_i\left(d_0; v_0, v_1\right) c_{V_i,V_{-i}}(v_0, v_1) dv_0 dv_1,
\end{aligned}$$

where $d_0 \in \{0, 1\}$, and Theorem 1 provides identification of $\mathrm{MCSE}_i\left(d_0; v_0, v_1\right)$.

4 Estimation and Inference

4.1 Semiparametric estimation and inference

4.1.1 Estimation procedure

Consider a sample of groups with size g, where the observations $\{Y_{0g}, Y_{1g}, \ldots, Y_{(n-1)g}, D_{0g}, D_{1g}, \ldots, D_{(n-1)g}, Z_{0g}, Z_{1g}, \ldots, Z_{(n-1)g}, X_{0g}, X_{1g}, \ldots, X_{(n-1)g}\}_{g=1,\ldots,G}$ are independently and identically distributed (i.i.d.) according to the probability measure $\mathbb{P}$. In this section, we adopt a semiparametric estimation approach similar to that proposed in Carneiro and Lee (2009), which allows for the inclusion of exogenous covariates X_{ig} in the model. To illustrate the estimation procedure, we consider the setting with exogenous covariates discussed in Appendix A.3, and focus on a simple case where each group consists of two units, $i \in \{0, 1\}$. The extension to the general case with $n > 1$ follows in an analogous manner.

$$\begin{cases}
Y_{ig} = &\big[Y_{ig}(X_{ig}, X_{(1-i)g}, 1, 1) D_{(1-i)g} + Y_{ig}(X_{ig}, X_{(1-i)g}, 1, 0)(1 - D_{(1-i)g})\big] D_{ig} \\
&+ \big[Y_{ig}(X_{ig}, X_{(1-i)g}, 0, 1) D_{(1-i)g} + Y_{ig}(X_{ig}, X_{(1-i)g}, 0, 0)(1 - D_{(1-i)g})\big](1 - D_{ig}), \\
D_{ig} = &\mathbb{1}\big\{V_{ig} \leq h(W_{ig}, W_{(1-i)g})\big\},
\end{cases}$$

where $W_{ig} \equiv (X_{ig}, Z_{ig})$.

Assumption 7. (Estimation assumptions) Assume that (i) $\mathbb{E}|Y_i(d, d')| < \infty, d, d' \in \{0, 1\}$.

(ii) Propensity scores, $P_i, i \in \{0,1\}$ are nondegenerate continuous random variable. (iii) $\mathbb{E}[Y_{idd'} \mid X = \mathbf{x}, P_0 = p_0, P_1 = p_1]$ and $\mathbb{E}[D_0 D_1 \mid P_0 = p_0, P_1 = p_1]$ are twice differentiable. (iv) $\partial^2 \mathbb{E}[D_0 D_1 \mid P_0 = p_0, P_1 = p_1]/\partial p_0 \partial p_1$ is bounded from above and away from zero.

Based on the identification results with exogenous covariates in Appendix A.3, our goal is to estimate the marginal treatment response functions,

$$m_{ig}^{(\mathbf{x},d,d')}(p_0,p_1) = \frac{\partial^2 \mathbb{E}\left[Y_{idd'g} \mid X_g = \mathbf{x}, P_{0g} = p_0, P_{1g} = p_1\right]}{\partial p_0 \partial p_1} \Big/ \frac{\partial^2 \mathbb{E}\left[D_{0g} D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\right]}{\partial p_0 \partial p_1},$$
$$Y_{idd'g} \equiv Y_{ig} \mathbb{1}\left\{D_{0g} = d, D_{1g} = d'\right\}\left\{2\,\mathrm{sgn}\left(1 - |d - d'|\right) - 1\right\}, i, d, d' \in \{0,1\}, g \in \{1, \cdots, G\}, \tag{11}$$

where $\mathrm{sgn}(x)$ is the sign function indicating the sign of a scalar x. The expression $\{2\,\mathrm{sgn}(1 - |d - d'|) - 1\}$ evaluates to 1 when $d = d'$, and to -1 when $d \neq d'$.

The propensity score functions $P_{0g} \equiv P_0(W_g)$ and $P_{1g} \equiv P_1(W_g)$ with $W_g = (W_{0g}, W_{1g}) \in \mathbb{R}^\ell$, as defined in Equation (25), are unobserved and should be identified from the data. Our first stage of estimation is to estimate the propensity score functions using series regression methods. To mitigate the curse of dimensionality, we adopt the partially linear additive regression by assuming

$$\mathbb{P}\left(D_{ig} = 1 \mid W_g = w\right) = \varphi_1\left(w_1^{cts}\right) + \cdots + \varphi_{\ell_1}\left(w_{\ell_1}^{cts}\right) + w_1^{disc}\vartheta_1 + \cdots + w_{\ell_2}^{disc}\vartheta_{\ell_2}. \tag{12}$$

We allow w to consist of both continuous and discrete components, represented as $w = (w^{cts}, w^{disc})$, where $w^{cts} = (w_1^{cts}, \cdots, w_{\ell_1}^{cts})$ is an ℓ_1-dimensional vector of continuous random variables, and $w^{disc} = (w_1^{disc}, \cdots, w_{\ell_2}^{disc})$ is an ℓ_2-dimensional vector of discrete random variables. To allow for greater model flexibility, no functional form assumptions are imposed on the unknown functions of the continuous parts, $\varphi_1, \cdots, \varphi_{\ell_1}$, and the parameters on discrete variables $\vartheta_1, \cdots, \vartheta_{\ell_2}$ are also unknown.

Series estimation relies on constructing a basis for smooth functions defined on $\mathbb{R}$, denoted as $\{p_k : k = 1, 2, \ldots\}$, such that each continuous function φ_ℓ, for $\ell = 1, \ldots, \ell_1$, can be approximated arbitrarily well by a linear combination of these basis functions as $k \to \infty$. Commonly used basis functions include polynomial basis functions, splines, and wavelets. Given a positive integer κ, we construct a regressor vector

$$P_\kappa(w) = \left[p_1\left(w_1^{cts}\right), \ldots, p_\kappa\left(w_1^{cts}\right), \ldots, p_1\left(w_{\ell_1}^{cts}\right), \ldots, p_\kappa\left(w_{\ell_1}^{cts}\right), w_1^{disc}, \cdots, w_{\ell_2}^{disc}\right]',$$

where first $\kappa \times \ell_1$ elements correspond to basis expansions of the continuous covariates, while the remaining ℓ_2 elements retain the discrete covariates in their original form.

The series estimator of the conditional probability $\mathbb{P}(D_{ig} = 1 \mid W_g)$, for $i \in \{0,1\}$, is

computed as

$$\tilde{P}_i\left(W_g\right) = P_\kappa\left(W_g\right)' \hat{\theta}^i_\kappa,$$

where $\hat{\theta}^i_\kappa$ is obtained as the solution to the following least squares problem

$$\hat{\theta}^i_\kappa = \arg\min_{\theta^i_\kappa \in \mathbb{R}^{\widetilde{\kappa}}} \frac{1}{G}\sum_{g=1}^{G}\left(D_{ig} - P_\kappa(W_g)'\theta^i_\kappa\right)^2,$$

where $\widetilde{\kappa} = 2\kappa + \ell_2$ is the dimension of the coefficient $\hat{\theta}^i_\kappa$.

Remark 4. (Series with Lasso regression) To increase flexibility in selecting relevant basis terms, we can combine nonparametric series estimation with Lasso regression, which enables automatic selection and regularization of basis terms. The estimation errors of such estimators have been studied in Bickel, Ritov, and Tsybakov (2009) and other related works cited therein. We can select the positive integer κ such that $\kappa \gg g$. Then, the series estimator, $\tilde{P}_i\left(W_g\right)$, with l_1-penalization is derived as

$$\begin{aligned}
\hat{\theta}^i_\kappa &= \arg\min_{\theta^i_\kappa \in \mathbb{R}^{\widetilde{\kappa}}} \frac{1}{G}\sum_{g=1}^{G}\left(D_{ig} - P_\kappa(W_g)'\theta^i_\kappa\right)^2 + 2\lambda\frac{1}{\widetilde{\kappa}}\sum_{j=1}^{\widetilde{\kappa}}\|p_j\|_G|\theta^i_j|, \\
\tilde{P}_i\left(W_g\right) &= P_\kappa\left(W_g\right)' \hat{\theta}^i_\kappa,
\end{aligned}$$

where $\lambda > 0$ is the tuning constant, $p_j(\cdot)$ denotes the j-th component of the basis expansion $P_\kappa(\cdot)$, and $\|\cdot\|_G$ stands for the empirical norm, $\|p_j\|_G = \sqrt{1/G\sum_{g=1}^{G} p_j^2(W_g)}$. In the estimation procedure, the tuning parameter λ is selected using cross-validation.

A finite-sample issue is that $\tilde{P}_i\left(W_g\right)$ may lie outside the unit interval $[0,1]$. Following the literature, we address this by applying a trimming procedure. The trimmed estimator is

$$\begin{aligned}
\hat{P}_i\left(W_g\right) =& \tilde{P}_i\left(W_g\right) + \left(1 - \delta - \tilde{P}_i\left(W_g\right)\right)\mathbb{1}\left\{\tilde{P}_i\left(W_g\right) > 1\right\} \\
&+ \left(\delta - \tilde{P}_i\left(W_g\right)\right)\mathbb{1}\left\{\tilde{P}_i\left(W_g\right) < 0\right\},
\end{aligned}$$

for some small $\delta > 0$ chosen by researchers. Then, $\hat{P}_i\left(W_g\right)$ provides an appropriate estimate of the propensity score $\mathbb{P}(D_{ig} = 1 \mid W_g)$, which we denote by $\hat{P}_{ig}$ for simplicity.

We next estimate the cross-partial derivatives that appear in the numerator and denominator of the estimand in Equation (11). We begin with the cross-partial derivative $\partial^2\mathbb{E}[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1]/\partial p_0 \partial p_1$. Local polynomial regression offers a convenient method for estimating derivatives of conditional means. Following the recommendation in Fan and Gijbels (1996), we choose the polynomial order to be $p = d + 1$, where d is the

order of the derivative to be estimated. As we are estimating a second-order derivative, we implement a local polynomial regression of order three:

$$\begin{aligned}\min_{b_0,\cdots,b_9}\sum_{g=1}^{G}\Big[D_{0g}D_{1g}-b_0-b_1\left(\hat{P}_{0g}-p_0\right)-\cdots b_4\left(\hat{P}_{0g}-p_0\right)\left(\hat{P}_{1g}-p_1\right)\\ -\cdots-b_9\left(\hat{P}_{1g}-p_1\right)^3\Big]^2 K_{h_{G1}}\left(\hat{P}_g-p\right)\\ K_{h_{G1}}\left(\hat{P}_g-p\right)=K\left(\frac{\hat{P}_{0g}-p_0}{h_{G1}}\right)\times K\left(\frac{\hat{P}_{1g}-p_1}{h_{G1}}\right),\end{aligned} \tag{13}$$

where $K(\cdot)$ is a kernel function, and h_{G1} is the selected bandwidth. For all kernel regressions, bandwidths are selected using either K-fold cross-validation or nearest-neighbor methods. The estimated coefficient $\hat{b}_4(p_0,p_1)$ serves as an estimator of the cross-partial derivative $\partial^2\mathbb{E}[D_{0g}D_{1g}\mid P_{0g}=p_0,P_{1g}=p_1]/\partial p_0\partial p_1$.

The next step involves estimating the cross-partial derivative appearing in the numerator of the estimand, $\partial^2\mathbb{E}[Y_{idd'g}\mid X_g=\mathbf{x},P_{0g}=p_0,P_{1g}=p_1]/\partial p_0\partial p_1$. Given that the covariates X_g may be multidimensional, we follow the semiparametric approach of Carneiro and Lee (2009) to address the curse of dimensionality. Specifically, we impose the partially linear structure on potential outcomes, $Y_{ig}(\mathbf{x},d,d')=\mathbf{x}'\beta_{idd'}+U_{ig}(d,d')$, where the vector $\beta_{idd'}$ represents unknown coefficients that may vary across individuals i and treatment states (d,d'), while $U_{ig}(d,d')$ denotes an unrestricted, nonparametric function. Under this specification, the marginal treatment response function admits the following representation,

$$\begin{aligned}&\frac{\partial^2\mathbb{E}\left[Y_{idd'g}\mid X_g=\mathbf{x},P_{0g}=p_0,P_{1g}=p_1\right]}{\partial p_0\partial p_1}\Big/\frac{\partial^2\mathbb{E}\left[D_{0g}D_{1g}\mid P_{0g}=p_0,P_{1g}=p_1\right]}{\partial p_0\partial p_1}\\ =&\mathbf{x}'\beta_{idd'}+\frac{\partial^2\mathbb{E}\left[U_{idd'g}\mid P_{0g}=p_0,P_{1g}=p_1\right]}{\partial p_0\partial p_1}\Big/\frac{\partial^2\mathbb{E}\left[D_{0g}D_{1g}\mid P_{0g}=p_0,P_{1g}=p_1\right]}{\partial p_0\partial p_1},\\ &U_{idd'g}\equiv U_{ig}(d,d')\mathbb{1}\left\{D_{0g}=d,D_{1g}=d'\right\}\left\{2\,\mathrm{sgn}\left(1-|d-d'|\right)-1\right\},\end{aligned} \tag{14}$$

and we intend to estimate the coefficients $\beta_{idd'}$ that appear in the partially linear component of the outcome. The conditional mean $\mathbb{E}\left[Y_{ig}\mid D_{0g}=d,D_{1g}=d',P_{0g},P_{1g}\right]$ equals to

$$\begin{aligned}\mathbb{E}[Y_{ig}\mid D_{0g}=d,D_{1g}=d',P_{0g},P_{1g}]=\left(\mathbb{E}[X_g\mid D_{0g}=d,D_{1g}=d',P_{0g},P_{1g}]\right)'\beta_{idd'}\\ +\mathbb{E}[U_{ig}\mid D_{0g}=d,D_{1g}=d',P_{0g},P_{1g}].\end{aligned} \tag{15}$$

Therefore, conditional on the subsample with $\{D_{0g}=d,D_{1g}=d'\}$, we can estimate coeffi-

cients $\beta_{idd'}$ via least squares regression as

$$\begin{aligned}\widehat{\beta}_{idd'} &= \left[\sum_{g=1}^{G}\widetilde{X_g}\widetilde{X_g}'\right]^{-1} \times \left[\sum_{g=1}^{G}\widetilde{X_g}\Big\{Y_{ig} - \hat{E}_h\left[Y_{ig} \mid \hat{P}_0\left(W_g\right), \hat{P}_1\left(W_g\right)\right]\Big\}\right], \\ \widetilde{X_g} &= X_g - \hat{E}_h\left[X_g \mid \hat{P}_0\left(W_g\right), \hat{P}_1\left(W_g\right)\right]\end{aligned} \tag{16}$$

where $\hat{E}_h[\cdot \mid \cdot]$ represents a kernel regression estimator with selected bandwidth h. We can calculate the residual as,

$$\widehat{U}_{idd'g} = Y_{idd'g} - X_g'\widehat{\beta}_{idd'}\left\{2\,\mathrm{sgn}\left(1 - |d - d'|\right) - 1\right\}$$

which provides a consistent estimator of $U_{idd'g}$.

In the last step, we can use data $\left\{\left(\hat{U}_{idd'g}, \hat{P}_0(W_g), \hat{P}_0(W_g)\right) : g = 1, \ldots, G\right\}$ to estimate the cross-partial derivative $\partial^2\mathbb{E}[U_{idd'g} \mid P_{0g} = p_0, P_{1g} = p_1]/\partial p_0 \partial p_1$ through local polynomial regression of order three,

$$\begin{aligned}\min_{c_0,\cdots,c_9}\sum_{g=1}^{G}\Big[\hat{U}_{iddg'} - c_0 - c_1\left(\hat{P}_{0g} - p_0\right) - \cdots c_4\left(\hat{P}_{0g} - p_0\right)\left(\hat{P}_{1g} - p_1\right) \\ - \cdots - c_9\left(\hat{P}_{1g} - p_1\right)^3\Big]^2 K_{h_{G2}}\left(\hat{P}_g - p\right), \\ K_{h_{G2}}\left(\hat{P}_g - p\right) = K\left(\frac{\hat{P}_{0g} - p_0}{h_{G2}}\right) \times K\left(\frac{\hat{P}_{1g} - p_1}{h_{G2}}\right).\end{aligned}$$

The estimated coefficient $\hat{c}_4(d, d'; p_0, p_1)$ is an estimator of $\partial^2\mathbb{E}[U_{idd'g} \mid P_{0g} = p_0, P_{1g} = p_1]/\partial p_0 \partial p_1$.

Finally, the marginal treatment response functions $m_{ig}^{(\mathbf{x},d,d')}(p_0, p_1)$ can be estimated by

$$\widehat{m}_{ig}^{(\mathbf{x},d,d')}(p_0, p_1) = \mathbf{x}'\hat{\beta}_{idd'} + \frac{\hat{c}_4(d, d'; p_0, p_1)}{\hat{b}_4(p_0, p_1)}.$$

The next step is to derive the asymptotic distribution of the marginal response functions, abstracting from the effects of covariates,

$$\frac{\hat{c}_4(d, d'; p_0, p_1)}{\hat{b}_4(p_0, p_1)}. \tag{17}$$

4.1.2 Inference results

Uniform consistent rate of propensity score

We choose the cubic spline basis functions $\{p_k : k = 1, 2, \cdots\}$ to approximate the non-parametric components of the propensity score functions. The following assumptions, as imposed in Belloni et al. (2015), are required to derive the uniform convergence rate of the series estimated propensity score functions.

Assumption 8. (Series estimation) (i) The eigenvalues of $\mathbb{E}[P_\kappa(w_g)P_\kappa(w_g)']$ are bounded above and away from zero uniformly over G. (ii) Each function $\varphi_i \in \mathcal{G}$ in Equation (12), where $\mathcal{G}$ is a set of functions f in Hölder classes with exponent s, $\Sigma_s(\mathcal{W})$, such that $\|f\|_s$ is bounded from above uniformly over $\mathcal{G}$. (iii) The support of W^{cts} is known and is a Cartesian product of compact connected intervals on which W^{cts} has a probability density function that is bounded away from zero.

Lemma 2. (Uniform rate of propensity score) Under Assumptions 7-8, we have

$$\max_{g=1,\ldots,G} \left| \hat{P}_i\left(W_g\right) - P_i\left(W_g\right) \right| = O_p\left[\sqrt{\frac{\kappa \log \kappa}{G}} + \kappa^{-s} \right],$$

where $\kappa \to \infty$ as $G \to \infty$, $\kappa^{m/(m-2)} \log \kappa / G = O(1)$ for any m ¿ 2, and $\kappa^{2-2s}/G = O(1)$.

Asymptotic properties of cross-derivative estimators

As described in Equation (17), our estimator consists of the ratio of two estimated cross-derivatives of conditional mean functions. In this subsection, we derive the asymptotic properties of two cross-derivative estimators under the estimation procedure proposed in Section 4.1.1.

We need to impose the following assumptions for deriving the asymptotic properties of the cross-derivative estimators.

Assumption 9. (Local polynomial regression) (i) $\mathbb{E}[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1]$ and $\mathbb{E}[U_{idd'g} \mid P_{0g} = p_0, P_{1g} = p_1]$ are $(p+1)$-time continuously differentiable, $p \geq 3$. (ii) The conditional distributions of $D_{0g}D_{1g} \mid P_{0g}, P_{1g}$ and $U_{idd'g} \mid P_{0g}, P_{1g}$ are continuous at the point (p_0, p_1). (iii) The kernel $K \in L_1$ is bounded with compact support, and $\|u\|^{4p} K(u) \in L_1$, $\|u\|^{4p+2} K(u) \to 0$ as $\|u\| \to \infty$.

Lemma 3. (Convergence rates of cross-derivative estimators) Under Assumptions 7-9, the convergence rates of estimators $\hat{b}_4(p_0, p_1)$ and $\hat{c}_4(d, d'; p_0, p_1)$, as calculated in Section 4.1.1,

can be derived as

$$\begin{aligned}
&\hat{b}_4(p_0,p_1) - \frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g}=p_0, P_{1g}=p_1\big] \\
=&O_P\Big[(Gh_{G1}^6)^{-1/2} + \max_{g:1\le g\le G}|\hat{P}_{0g}-P_{0g}| + \max_{g:1\le g\le G}|\hat{P}_{1g}-P_{1g}| + h_{G1}^4\Big], \\
&\hat{c}_4(d,d';p_0,p_1) - \frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[U_{idd'g} \mid P_{0g}=p_0, P_{1g}=p_1\big] \\
=&O_P\Big[(Gh_{G2}^6)^{-1/2} + \max_{g:1\le g\le G}|\hat{P}_{0g}-P_{0g}| + \max_{g:1\le g\le G}|\hat{P}_{1g}-P_{1g}| + h_{G2}^4\Big],
\end{aligned}$$

where h_{G1}, h_{G2} are the bandwidths selected to estimate $\hat{b}_4(p_0,p_1)$ and $\hat{c}_4(d,d';p_0,p_1)$.

Proof. See Appendix D.1. □

Asymptotic distribution of the marginal treatment response

In this section, we aim to characterize the asymptotic distribution of the marginal treatment response functions excluding the covariate effects, as outlined in Equation (17). To establish this result, we impose the following assumptions.

Assumption 10. (Asymptotic distribution) (i) $\max_{g=1,\dots,G}\big|\hat{P}_i(Z_{0g},Z_{1g}) - P_i(Z_{0g},Z_{1g})\big| = o_p\big[(Gh_{G1}^6)^{-1/2}\big]$. (ii) $h_{G1}, h_{G2} \to 0, Gh_{G1}^6, Gh_{G2}^6 \to \infty$ as $G\to\infty$, $h_{G2} = o(h_{G1})$, $h_{G1}, h_{G2} = o(G^{-1/10})$.

Theorem 2. (Asymptotic distributions of the ratio estimator) Under Assumptions 7-10, the asymptotic distribution of $\hat{c}_4(d,d';p_0,p_1)/\hat{b}_4(d,d';p_0,p_1)$, $d,d'\in\{0,1\}$, can be characterized as

$$\begin{aligned}
&\big(Gh_{G2}^6\big)^{1/2}\left\{\frac{\hat{c}_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)} - \frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}\right\} \\
\xrightarrow{d} &N\left(0, \frac{\sigma^2(p_0,p_1)}{\big(b_4(p_0,p_1)\big)^2 f(p_0,p_1)}\big(M^{-1}\Gamma M^{-1}\big)_{5,5}\right),
\end{aligned}$$

where $\sigma^2(d,d';p_0,p_1) = \mathrm{Var}(U_{idd'g} \mid P_{0g}=p_0, P_{1g}=p_1)$, $f(p_0,p_1)$ denotes the density of (P_{0g},P_{1g}) evaluated at the point (p_0,p_1), and $A_{5,5}$ denotes the element located in the fifth row and fifth column of a matrix A. The definitions of the matrices M and Γ are presented in the Appendix D.2.

Proof. See Appendix D.2. □

Finally, we intend to derive the asymptotic distributions of MCSEs and MCDEs. The

MCSEs and MCDEs with covariates are identified as

$$\begin{aligned}
\text{MCSE}_i(\mathbf{x}, d; p_0, p_1) &= m_{ig}^{(\mathbf{x},d,1)}(p_0, p_1) - m_{ig}^{(\mathbf{x},d,0)}(p_0, p_1), d \in \{0, 1\}, \\
\text{MCDE}_i(\mathbf{x}, d; p_0, p_1) &= m_{ig}^{(\mathbf{x},1,d)}(p_0, p_1) - m_{ig}^{(\mathbf{x},0,d)}(p_0, p_1), d \in \{0, 1\},
\end{aligned}$$

and we estimate them using the corresponding estimated values of the marginal treatment response functions,

$$\begin{aligned}
\widehat{\text{MCSE}}_i(\mathbf{x}, d; p_0, p_1) &= \widehat{m}_{ig}^{(\mathbf{x},d,1)}(p_0, p_1) - \widehat{m}_{ig}^{(\mathbf{x},d,0)}(p_0, p_1), d \in \{0, 1\}, \\
\widehat{\text{MCDE}}_i(\mathbf{x}, d; p_0, p_1) &= \widehat{m}_{ig}^{(\mathbf{x},1,d)}(p_0, p_1) - \widehat{m}_{ig}^{(\mathbf{x},0,d)}(p_0, p_1), d \in \{0, 1\}.
\end{aligned}$$

If we assume that $\hat{c}_4(d, d'; p_0, p_1)/\hat{b}_4(p_0, p_1) - c_4(d, d'; p_0, p_1)/b_4(p_0, p_1)$ are asymptotically independent across different values of $d, d' \in \{0, 1\}$, the asymptotic distributions of $\widehat{\text{MCSE}}_i(\mathbf{x}, d; p_0, p_1)$ and $\widehat{\text{MCDE}}_i(\mathbf{x}, d; p_0, p_1)$ follow directly from Theorem 2.

Corollary 2. Suppose that $\big(\hat{c}_4(d, d'; p_0, p_1)/\hat{b}_4(p_0, p_1) - c_4(d, d'; p_0, p_1)/b_4(p_0, p_1)\big)$ are asymptotically independent across different values of $d, d' \in \{0, 1\}$, and that Assumptions 7–10 are satisfied. Then, the asymptotic distributions of $\widehat{\text{MCSE}}_i(\mathbf{x}, d; p_0, p_1)$ and $\widehat{\text{MCDE}}_i(\mathbf{x}, d; p_0, p_1)$ can be characterized as

$$\begin{aligned}
&\big(Gh_{G2}^6\big)^{1/2} \left\{ \widehat{\text{MCSE}}(\mathbf{x}, d; p_0, p_1) - \text{MCSE}(\mathbf{x}, d; p_0, p_1) \right\} \\
&\xrightarrow{d} N\left(0, \frac{\sigma^2(1, d; p_0, p_1) + \sigma^2(0, d; p_0, p_1)}{\big(b_4(p_0, p_1)\big)^2 f(p_0, p_1)} \big(M^{-1}\Gamma M^{-1}\big)_{5,5}\right), \\
&\big(Gh_{G2}^6\big)^{1/2} \left\{ \widehat{\text{MCDE}}(\mathbf{x}, d; p_0, p_1) - \text{MCDE}(\mathbf{x}, d; p_0, p_1) \right\} \\
&\xrightarrow{d} N\left(0, \frac{\sigma^2(d, 1; p_0, p_1) + \sigma^2(d, 0; p_0, p_1)}{\big(b_4(p_0, p_1)\big)^2 f(p_0, p_1)} \big(M^{-1}\Gamma M^{-1}\big)_{5,5}\right)
\end{aligned}$$

4.2 Parametric procedure

4.2.1 Parametric estimation

The methods introduced in Section 4.1 still rely on large sample sizes to deliver reliable estimates. When the available sample is limited, it may be preferable to impose parametric assumptions to estimate the marginal treatment response functions of interest. This section develops a parametric approach for estimation and inference.

We continue to impose the estimation assumptions in Assumption 7 and additionally introduce the following parametric assumptions.

Assumption 11. We impose the following specifications in the parametric setting.

1. (Propensity score) For the treatment assignment equation $D_{ig} = \mathbb{1}\{\widetilde{V}_{ig} \leq h_i(W_g)\}$ of individual i in group g, $i \in \{0,1\}$, we assume that $h_i(\cdot)$ is a K_1-th order polynomial function of $W_g = (W_{g,1}, \cdots, W_{g,\ell})' \in \mathbb{R}^\ell$:

$$h_i(W_g) = \sum_{\boldsymbol{k} \in \mathcal{K}_{\ell,K_1}} \theta_{i\boldsymbol{k}} \cdot \prod_{j=1}^{\ell} W_{g,j}^{k_j},$$

   where $\boldsymbol{k} = (k_1, \ldots, k_\ell) \in \mathbb{N}_0^\ell$ is a multi-index, $\mathcal{K}_{\ell,K_1} = \{\boldsymbol{k} \in \mathbb{N}_0^\ell : \sum_{j=1}^{\ell} k_j \leq K_1\}$, and $(\theta_{i\boldsymbol{k}})_{\boldsymbol{k} \in \mathcal{K}_{\ell,K_1}} \equiv \theta_i$ are polynomial coefficients. Additionally, we assume that the unobserved heterogeneity $\widetilde{V}_{ig}$ follows a standard normal distribution: $\widetilde{V}_{ig} \sim N(0,1)$.

2. (Copula) We assume that the joint dependence structure of the unobserved heterogeneities V_{0g} and V_{1g} is characterized by a Gaussian copula with correlation parameter $\rho \in [-\varepsilon, \varepsilon]$, where ε is a constant such that $0 < \varepsilon < 1$. Specifically, let $V_{ig} = \Phi(\widetilde{V}_{ig})$ for $i \in \{0,1\}$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. The copula of (V_{0g}, V_{1g}), denoted by $C_{V_{0g},V_{1g}}(\cdot,\cdot)$, is then given by the Gaussian copula with correlation ρ:

$$C_{V_{0g},V_{1g}}(v_0, v_1) = \Phi_\rho\big(\Phi^{-1}(v_0), \Phi^{-1}(v_1)\big), \forall (v_0, v_1) \in [0,1]^2,$$

   where Φ_ρ is the bivariate normal CDF with zero means, unit variances, and correlation ρ, Φ^{-1} denotes the inverse of the standard normal CDF, and ρ is an unknown parameter.

3. (Marginal treatment response) We assume that the potential outcome is partially linear in the covariates, $Y_{ig}(\mathbf{x}, d, d') = \mathbf{x}'\beta_{idd'} + U_{ig}(d, d')$, and $U_{ig}(d, d')$ satisfies the following condition:

$$\begin{aligned}\mathbb{E}[U_{ig}(d, d') \mid V_{0g} = v_0, V_{1g} = v_1] =& \alpha_{idd',0} + \alpha_{idd',1}\Phi^{-1}(v_0) \\ &+ \alpha_{idd',2}\Phi^{-1}(v_1) + \alpha_{idd',3}\Phi^{-1}(v_0)\Phi^{-1}(v_1),\end{aligned}$$

   for all $(v_0, v_1) \in (0,1)^2$, and $\alpha_{idd'} \equiv (\alpha_{idd',0}, \alpha_{idd',1}, \alpha_{idd',2}, \alpha_{idd',3})'$ denotes the vector of unknown coefficients that may be heterogeneous across individuals i and treatment states (d, d').

Our objective is to estimate the marginal treatment response function, $m_{ig}^{(\mathbf{x},d,d')}(v_0, v_1) = \mathbb{E}[Y_{ig}(\mathbf{x}, d, d') \mid V_{0g} = v_0, V_{1g} = v_1]$, for any $\mathbf{x} \in \mathcal{X}$, $d, d' \in \{0,1\}$, and $(v_0, v_1) \in (0,1)^2$.

As in the semiparametric case, the first step involves estimating the propensity score functions $P_0(W_g), P_1(W_g)$, where $W_{ig} = (Z_{ig}, X_{ig}), W_g = (W_{0g}, W_{1g}) \in \mathbb{R}^\ell$. Under the first specification in Assumption 11, and assuming that the instruments and covariates are independent of the unobserved heterogeneity V_{ig}, we can express the propensity score function as

$$P_{ig} \equiv \mathbb{P}(D_{ig} = 1 \mid W_g) = \Phi\left(\sum_{\boldsymbol{k} \in \mathcal{K}_{\ell,K_1}} \theta_{i\boldsymbol{k}} \cdot \prod_{j=1}^{\ell} W_{g,j}^{k_j} \right).$$

The polynomial coefficients can be estimated using standard maximum likelihood methods:

$$\hat{\theta}_i = \arg \max_{(\theta_{i\boldsymbol{k}})_{\boldsymbol{k} \in \mathcal{K}_{\ell,K_1}}} \sum_{g=1}^{G} \left[D_{ig} \log \Phi\Big(\sum_{\boldsymbol{k} \in \mathcal{K}_{\ell,K_1}} \theta_{i\boldsymbol{k}} \cdot \prod_{j=1}^{\ell} W_{g,j}^{k_j} \Big) + (1 - D_{ig}) \log \left(1 - \Phi\Big(\sum_{\boldsymbol{k} \in \mathcal{K}_{\ell,K_1}} \theta_{i\boldsymbol{k}} \cdot \prod_{j=1}^{\ell} W_{g,j}^{k_j} \Big) \right) \right].$$

Once the polynomial coefficients $\hat{\theta}_i$ are estimated, they can be substituted into P_{ig} to obtain the estimated propensity score as

$$\widehat{P}_{ig} = \Phi\left(\sum_{\boldsymbol{k} \in \mathcal{K}_{\ell,K_1}} \widehat{\theta}_{i\boldsymbol{k}} \cdot \prod_{j=1}^{\ell} W_{g,j}^{k_j} \right).$$

The next step is to estimate the joint dependence structure of V_{0g} and V_{1g}. Under the second specification in Assumption 11, this dependence is modeled by a Gaussian copula with correlation parameter ρ. Consequently, the second step of our procedure focuses on estimating ρ. Based on our identification results, we can derive the following equations,

$$\begin{aligned}
&\mathbb{P}(D_{0g} = 1, D_{1g} = 1 \mid P_{0g} = p_0, P_{1g} = p_1) = \Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big), \\
&\mathbb{P}(D_{0g} = 1, D_{1g} = 0 \mid P_{0g} = p_0, P_{1g} = p_1) = p_0 - \Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big), \\
&\mathbb{P}(D_{0g} = 0, D_{1g} = 1 \mid P_{0g} = p_0, P_{1g} = p_1) = p_1 - \Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big), \\
&\mathbb{P}(D_{0g} = 0, D_{1g} = 0 \mid P_{0g} = p_0, P_{1g} = p_1) = 1 - p_0 - p_1 + \Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big).
\end{aligned}$$

Therefore, we estimate ρ using the maximum likelihood, substituting the first-stage esti-

mates $\widehat{P}_{0g}$ and $\widehat{P}_{1g}$ for the true propensity scores P_{0g} and P_{1g},

$$\begin{aligned}\hat{\rho} = \arg\max_{\rho\in[-\varepsilon,\varepsilon]} \sum_{g=1}^{G} \Bigg\{ & D_{0g}D_{1g}\log\Big(\Phi_\rho\big(\Phi^{-1}(\widehat{P}_{0g}),\Phi^{-1}(\widehat{P}_{1g})\big)\Big) + \\ & D_{0g}(1-D_{1g})\log\Big(\widehat{P}_{0g}-\Phi_\rho\big(\Phi^{-1}(\widehat{P}_{0g}),\Phi^{-1}(\widehat{P}_{1g})\big)\Big) + \\ & (1-D_{0g})D_{1g}\log\Big(\widehat{P}_{1g}-\Phi_\rho\big(\Phi^{-1}(\widehat{P}_{0g}),\Phi^{-1}(\widehat{P}_{1g})\big)\Big) + \\ & (1-D_{0g})(1-D_{1g})\log\Big(1-\widehat{P}_{0g}-\widehat{P}_{1g}+\Phi_\rho\big(\Phi^{-1}(\widehat{P}_{0g}),\Phi^{-1}(\widehat{P}_{1g})\big)\Big)\Bigg\}.\end{aligned}$$

Our final step involves estimating the marginal treatment response $\mathbb{E}[Y_{ig}(\mathbf{x},d,d') \mid V_{0g}=v_0, V_{1g}=v_1]$. Under the third specification in Assumption 11, this function admits the following parametric representation,

$$\begin{aligned} m_{ig}^{(\mathbf{x},d,d')}(v_0,v_1) \equiv & \mathbb{E}\big[Y_{ig}(\mathbf{x},d,d') \mid V_{0g}=v_0, V_{1g}=v_1\big] \\ = & \mathbf{x}'\beta_{idd'} + \alpha_{idd',0} + \alpha_{idd',1}\Phi^{-1}(v_0) \\ & + \alpha_{idd',2}\Phi^{-1}(v_1) + \alpha_{idd',3}\Phi^{-1}(v_0)\Phi^{-1}(v_1).\end{aligned}$$

Hence, the last stage of our procedure focuses on estimating the coefficient vectors $\beta_{idd'}$ and $c_{idd'}$. For illustration, consider the case $d=1$ and $d'=1$.

By combining our identification results with the third specification in Assumption 11, we obtain the following equation,

$$\begin{aligned} & \mathbb{E}[Y_{ig}D_{0g}D_{1g} \mid X_g=\mathbf{x}, P_{0g}=p_0, P_{1g}=p_1] \\ = & \int_0^{p_1}\int_0^{p_0} \mathbb{E}\big[U_{ig}(1,1) \mid V_{0g}=v_0, V_{1g}=v_1\big] c_{V_{0g},V_{1g}}(v_0,v_1)dv_0dv_1 \\ & + \mathbf{x}'\beta_{i11}\mathbb{P}\big(D_{0g}D_{1g} \mid P_{0g}=p_0, P_{1g}=p_1\big) \\ = & \alpha_{i11,0}\int_0^{p_1}\int_0^{p_0} c_{V_{0g},V_{1g}}(v_0,v_1)dv_0v_1 + \alpha_{i11,1}\int_0^{p_1}\int_0^{p_0} \Phi^{-1}(v_0)c_{V_{0g},V_{1g}}(v_0,v_1)dv_0v_1 \\ & + \alpha_{i11,2}\int_0^{p_1}\int_0^{p_0} \Phi^{-1}(v_1)c_{V_{0g},V_{1g}}(v_0,v_1)dv_0v_1 \\ & + \alpha_{i11,3}\int_0^{p_1}\int_0^{p_0} \Phi^{-1}(v_0)\Phi^{-1}(v_1)c_{V_{0g},V_{1g}}(v_0,v_1)dv_0v_1 \\ & + \mathbf{x}'\beta_{i11}\mathbb{P}\big(D_{0g}D_{1g} \mid P_{0g}=p_0, P_{1g}=p_1\big) \\ \equiv & \alpha_{i11,0}I_{11}^0(p_0,p_1,\rho) + \alpha_{i11,1}I_{11}^1(p_0,p_1,\rho) + \alpha_{i11,2}I_{11}^2(p_0,p_1,\rho) + \alpha_{i11,3}I_{11}^3(p_0,p_1,\rho) \\ & + \mathbf{x}'\beta_{i11}\Phi_\rho\big(p_0,p_1\big).\end{aligned}$$

In the last line, $I_{11}^1(p_0, p_1, \rho), I_{11}^2(p_0, p_1, \rho)$, and $I_{11}^3(p_0, p_1, \rho)$ denote integrals that depend on p_0, p_1, and the correlation parameter ρ of the Gaussian copula density $c_{V_{0g}, V_{1g}}(\cdot, \cdot)$ when $d = 1$ and $d' = 1$. Given that the propensity scores and the correlation have been estimated in the previous two steps, we substitute $\widehat{P}_{0g}$, $\widehat{P}_{1g}$, and $\hat{\rho}$ for the true values. Then, we estimate coefficient vectors α_{i11} and β_{i11} using the following least squares regression,

$$\begin{aligned}\left(\hat{\alpha}_{i11}', \hat{\beta}_{i11}'\right)' = \arg \min_{(\alpha_{i11}', \beta_{i11}')'} \sum_{g=1}^{G} \Bigg[& Y_{ig} D_{0g} D_{1g} - \alpha_{i11,0} I_{11}^1(\widehat{P}_{0g}, \widehat{P}_{1g}, \hat{\rho}) - \alpha_{i11,1} I_{11}^1(\widehat{P}_{0g}, \widehat{P}_{1g}, \hat{\rho}) \\ & - \alpha_{i11,2} I_{11}^3(\widehat{P}_{0g}, \widehat{P}_{1g}, \hat{\rho}) - \alpha_{i11,3} I_{11}^4(\widehat{P}_{0g}, \widehat{P}_{1g}, \hat{\rho}) - \widetilde{X}_g' \beta_{i11} \Bigg]^2,\end{aligned}$$

where $\widetilde{X}_g$ equals to $X_g \mathbb{P}(D_{0g} D_{1g} \mid \widehat{P}_{0g}, \widehat{P}_{1g})$. We can apply the similar procedure to estimate the coefficient vectors $\beta_{idd'}$, $\alpha_{idd'}$ for other treatment values (d, d'). The estimated marginal treatment response function, $\widehat{m}_{ig}^{(\mathbf{x},d,d')}(v_0, v_1)$, is obtained by substituting $(\hat{\alpha}_{i11}', \hat{\beta}_{i11}')'$ for $(\alpha_{i11}', \beta_{i11}')'$.

Finally, we can estimate MCDEs and MCSEs by taking differences of the estimated marginal treatment response functions,

$$\begin{aligned}\widehat{\text{MCSE}}_i(\mathbf{x}, d; p_0, p_1) &= \widehat{m}_{ig}^{(\mathbf{x},d,1)}(p_0, p_1) - \widehat{m}_{ig}^{(\mathbf{x},d,0)}(p_0, p_1), d \in \{0, 1\}, \\ \widehat{\text{MCDE}}_i(\mathbf{x}, d; p_0, p_1) &= \widehat{m}_{ig}^{(\mathbf{x},1,d)}(p_0, p_1) - \widehat{m}_{ig}^{(\mathbf{x},0,d)}(p_0, p_1), d \in \{0, 1\}.\end{aligned}$$

4.2.2 Parametric asymptotic results

In this section, we introduce a set of assumptions under which the parametric estimators are consistent.

Assumption 12. (Parametric first stage) In the first stage of propensity score estimation, for each individual $i \in \{0, 1\}$, we assume that

1. $\theta_i \in \Theta_i$, where the parameter space Θ_i is compact.

2. The true parameter θ_{i0} is unique.

3. Let $l(\theta_i; D_{ig}, W_g)$ denote the log-likelihood of individual i's treatment in group g:

$$\begin{aligned}l(\theta_i; D_{ig}, W_g) = & D_{ig} \log \Phi\Bigg(\sum_{k \in \mathcal{K}_{\ell, K_1}} \theta_{ik} \cdot \prod_{j=1}^{\ell} W_{g,j}^{k_j} \Bigg) \\ & + (1 - D_{ig}) \log \left(1 - \Phi\Bigg(\sum_{k \in \mathcal{K}_{\ell, K_1}} \theta_{ik} \cdot \prod_{j=1}^{\ell} W_{g,j}^{k_j} \Bigg) \right).\end{aligned}$$

The log-likelihood function $l(\theta_i; D_{ig}, W_g)$ satisfies the following conditions:

(i) $\mathbb{E}\big[\sup_{\theta_i \in \Theta_i} |l(\theta_i; D_{ig}, W_g)|\big] < \infty$.

(ii) $\mathbb{E}\big[\nabla^2_{\theta_i} l(\theta_i; D_{ig}, W_g)\big]$ exists and is invertible.

(iii) $\mathbb{E}\big[\sup_{\theta_i \in \Theta_i} ||\nabla^2_{\theta_i} l(\theta_i; D_{ig}, W_g)||\big] < \infty$.

Under Assumptions 11 and 12, the estimator $\hat{\theta}_i$ obtained in the first stage is consistent.

Lemma 4. Suppose Assumptions 7, 11 and 12 hold. Then, for each $i \in \{0, 1\}$, the estimator $\hat{\theta}_i \xrightarrow{a.s.} \theta_{i0}$ as $G \to \infty$.

Proof. See Appendix E.1. □

To establish the consistency of the second-stage estimator of the Gaussian copula correlation parameter ρ, we impose the following additional assumption.

Assumption 13. (Parametric second stage) In the second stage, to estimate the correlation parameter ρ of the Gaussian copula, we impose the following conditions.

1. The true parameter ρ_0 is unique.

2. Define the log-likelihood of joint treatments in group g as

$$
\begin{aligned}
& l(\rho, \theta; D_g, W_g) = \tilde{l}(\rho; D_g, P_g) \\
\equiv & D_{0g} D_{1g} \log\Big(\Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big)\Big) + \\
& D_{0g}(1 - D_{1g}) \log\Big(P_{0g} - \Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big)\Big) + \\
& (1 - D_{0g}) D_{1g} \log\Big(P_{1g} - \Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big)\Big) + \\
& (1 - D_{0g})(1 - D_{1g}) \log\Big(1 - P_{0g} - P_{1g} + \Phi_\rho\big(\Phi^{-1}(P_{0g}), \Phi^{-1}(P_{1g})\big)\Big),
\end{aligned}
$$

where $P_g \equiv (P_{0g}, P_{1g})$, P_{ig}, $i \in \{0, 1\}$, is the function of the first stage parameter θ_i and the variable W_g, $\theta \equiv (\theta_0', \theta_1')'$, and $D_g \equiv (D_{0g}, D_{1g})$. The log-likelihood needs to satisfy

(i) $\mathbb{E}\big[\sup_{\rho \in [-\varepsilon, \varepsilon]} |l(\rho, \theta; D_g, W_g)|\big] < \infty$.

(ii) There exists a function $L(\cdot)$ with $|L(D_g)| < \infty$ almost surely such that for all $d \in \{0, 1\}^2$, $(p, p') \in (0, 1)^2$, and $\rho \in [-\varepsilon, \varepsilon]$, $|\tilde{l}(\rho; d, p) - \tilde{l}(\rho; d, p')| \leq L(d)||p - p'||$, where $\tilde{l}(\rho; d, p)$ is defined as the second stage log-likelihood given $(P_{0g}, P_{1g}) = p$.

(iii) $\mathbb{E}\big[\partial^2 l(\rho, \theta; D_g, W_g)/\partial\rho^2\big]$ is bounded away from zero.

(iv) $\mathbb{E}\big[\sup_{\rho\in[-\varepsilon,\varepsilon]} |\partial^2 l(\rho; D_g, P_g)/\partial\rho^2|\big] < \infty$ and $\mathbb{E}\big[\sup_{\theta\in\Theta_0\times\Theta_1} ||\nabla_\theta \frac{\partial l(\rho,\theta;D_g,W_g)}{\partial\rho}||\big] < \infty$.

These conditions allow us to establish the consistency of the second-stage estimator of ρ.

Lemma 5. Suppose Assumptions 7, 11, 12, and 13 hold. Then, $\hat{\rho} \xrightarrow{a.s.} \rho_0$ as $G \to \infty$.

Proof. See Appendix E.2. □

Finally, we establish the consistency of the estimated coefficient vector $(\hat{\alpha}'_{i11}, \hat{\beta}'_{i11})'$, which in turn implies the consistency of the estimated marginal controlled effects.

Theorem 3. (Consistency of parametric MCDEs and MCSEs) Suppose that Assumptions 7, 11, 12, and 13 hold. Also, assume that

1. $X'_{Pdd'}X_{Pdd'}$ and $\mathbb{E}[X'_{Pdd'}X_{Pdd'}]$ are nonsingular, where $X_{Pdd'}$ is defined as a $G \times K$ matrix with the g-th row as

$$X_{Pdd'_g} \equiv \big[\Phi_\rho(P_{0g}, P_{1g}), I^1_{dd'}(P_{0g}, P_{1g}, \rho), I^2_{dd'}(P_{0g}, P_{1g}, \rho), I^3_{dd'}(P_{0g}, P_{1g}, \rho), \widetilde{X}'_g\big],$$

   where P_{ig}, $i \in \{0, 1\}$, is the function of the first stage parameter θ_i and the variable W_g.

2. $\|\widehat{X}_{Pdd'} - X_{Pdd'}\|_F^2/G \xrightarrow{a.s.} 0$, where $\widehat{X}_{Pdd'}$ is obtained by replacing the true values P_{0g}, P_{1g}, and ρ in $X_{Pdd'}$ with their estimates $\widehat{P}_{0g}$, $\widehat{P}_{1g}$, and $\hat{\rho}$, respectively. The notation $\|\cdot\|_F$ denotes the Frobenius norm.

3. Set $\varepsilon_{idd'} = Y_{ig}\mathbb{1}\{D_{0g} = d\}\mathbb{1}\{D_{1g} = d'\} - X_{Pdd'_g}(\alpha'_{idd'}, \beta'_{idd'})'$. Then, $\mathrm{Var}(\varepsilon_{idd'}) = \sigma^2_{idd'} < \infty$.

4. Define $\psi_{idd}(\alpha_{idd'}, \beta_{idd'}, \rho, \theta; Y_{ig}, D_{0g}, D_{1g}, W_g) = (Y_{idd'g} - X_{P_g}(\alpha'_{idd'}, \beta'_{idd'})')X_{P_g}$, where $Y_{idd'g} \equiv Y_{ig}\mathbb{1}\{D_{0g} = d, D_{1g} = d'\}$. It satisfies

   (i) $\mathbb{E}[\sup_{\rho\in[-\varepsilon,\varepsilon]} ||\nabla_\rho \psi_{idd}(\alpha_{idd'}, \beta_{idd'}, \rho, \theta; Y_{ig}, D_{0g}, D_{1g}, W_g)||] < \infty$.

   (ii) $\mathbb{E}[\sup_{\theta\in\Theta_0\times\Theta_1} ||\nabla_\theta \psi_{idd}(\alpha_{idd'}, \beta_{idd'}, \rho, \theta; Y_{ig}, D_{0g}, D_{1g}, W_g)||] < \infty$.

Under the above conditions, $(\hat{\alpha}'_{idd'}, \hat{\beta}'_{idd'})' \xrightarrow{p} (\alpha'_{idd'}, \beta'_{idd'})'$ as $G \to \infty$.

Proof. See Appendix E.3. □

Inference regarding the estimator $(\hat{\alpha}'_{idd'}, \hat{\beta}'_{idd'})'$ is performed using standard nonparametric bootstrap methodologies. Based on Assumptions 12 and 13, the conditions in Theorem 3, and standard regularity conditions, the bootstrap distribution converges uniformly to the sampling distribution of the estimator[1] (Romano and Shaikh, 2012). Therefore, standard nonparametric bootstrap methods, such as resampling the data and recomputing all stages, are expected to yield valid inference.

Once we establish the consistency and the asymptotic distibutions of $(\hat{\alpha}'_{idd'}, \hat{\beta}'_{idd'})'$, the consistency and asymptotic distributions of $\widehat{\text{MCSE}}_i(\mathbf{x}, d; p_0, p_1)$ and $\widehat{\text{MCDE}}_i(\mathbf{x}, d; p_0, p_1)$ follow directly from the continuous mapping theorem. This result obtains by the continuity of $\widehat{\text{MCSE}}_i$ and $\widehat{\text{MCDE}}_i$ as functions of $(\hat{\alpha}'_{idd'}, \hat{\beta}'_{idd'})'$.

5 Simulation and Application

5.1 Parametric Simulation

In this section, we present a Monte Carlo simulation to assess the validity of the proposed parametric estimation methods.

For each Monte Carlo replication, we generate G i.i.d. groups, where each group g consists of two members indexed by $i \in \{0, 1\}$. We draw the group instrument vector, $Z_g = (Z_{0g}, Z_{1g})$, i.i.d. from a bivariate normal distribution $N(0, \Sigma_Z)$ with $\Sigma_Z = (1, 0.1; 0.1, 1)$. The correlation of Z_{0g} and Z_{1g} is not zero, since we allow the instruments of group members to be correlated. We also the group-level unobserved heterogeneity vector, $(\widetilde{V}_{0g}, \widetilde{V}_{1g})$, i.i.d. from a bivariate normal distribution $N(0, \Sigma_V)$ with $\Sigma_V = (1, 0.2; 0.2, 1)$ and independent of the instrument vector Z_g. By construction, $\widetilde{V}_{ig}$, $i \in \{0, 1\}$, follows a standard normal distribution. Additionally, the copula linking the normalized unobserved heterogeneity V_{0g} and V_{ig}, where $\widetilde{V}_{ig} = \Phi(\widetilde{V}_i)$, is a Gaussian copula with correlation $\rho = 0.2$. These specifications are consistent with Assumption 3 and the first two conditions in Assumption 11.

We construct the following model to generate individual's treatment and potential out-

[1]The regularity conditions include stochastic equicontinuity and a quadratic remainder condition. A formal proof is left for future work.

come.

$$\begin{cases} D_{0g} = \mathbb{1}\{\widetilde{V}_{0g} \leq Z_{0g} + 0.5Z_{1g}\} \\ D_{1g} = \mathbb{1}\{\widetilde{V}_{1g} \leq Z_{1g} - 0.5Z_{0g}\} \\ Y_{ig}(1,1) = 1 + 0.5U_g + 2\widetilde{V}_{ig} + \widetilde{V}_{(1-i)g} - \widetilde{V}_{ig}\widetilde{V}_{(1-i)g}, i = 0,1 \\ Y_{ig}(1,0) = 3 + 0.5U_g + 2\widetilde{V}_{ig} + \widetilde{V}_{(1-i)g} - \widetilde{V}_{ig}\widetilde{V}_{(1-i)g}, i = 0,1 \\ Y_{ig}(0,1) = 3 + 0.5U_g + 2\widetilde{V}_{ig} - \widetilde{V}_{ig}\widetilde{V}_{(1-i)g}, i = 0,1 \\ Y_{ig}(0,0) = 2 + 0.5U_g + 2\widetilde{V}_{ig} - \widetilde{V}_{ig}\widetilde{V}_{(1-i)g}, i = 0,1, \end{cases}$$

where the group-level disturbance $U_g \in \mathbb{R}$ is generated i.i.d. from the uniform distribution $\mathcal{U}(0,1)$ and is independent of $(Z_{0g}, Z_{1g}, \tilde{V}_{0g}, \tilde{V}_{1g})$. The observed individual outcome Y_i is derived from

$$\begin{aligned} Y_{ig} =& \big[Y_{ig}(1,1)D_{(1-i)g} + Y_{ig}(1,0)(1 - D_{(1-i)g})\big] D_{ig} \\ &+ \big[Y_{ig}(0,1)D_{(1-i)g} + Y_{ig}(0,0)(1 - D_{(1-i)g})\big](1 - D_{ig}), i = 0,1. \end{aligned}$$

In our data generating process, the instrument vector Z_g is independent of the unobserved heterogeneities and potential outcomes, $(\widetilde{V}_{ig}, Y_{ig}(d,d'))_{i,d,d' \in \{0,1\}}$, satisfying Assumption 1. Moreover, Z_g does not directly affect the outcome $Y_{ig}(d,d')$, in accordance with Assumption 2. The threshold function $h_i(\cdot)$ in the treatment assignment equation is specified as a first-order polynomial in the instrument, satisfying the first condition in Assumption 11. For the potential outcomes, their conditional means given V_{0g} and V_{1g} satisfy the third condition in Assumption 11. Therefore, the data generating process satisfies all identification and parametric assumptions.

We apply the method in Section 4.2 to estimate and construct 95% confidence intervals for the marginal controlled spillover (MCSE) and direct effects (MCDE) at selected evaluated points (p_0, p_1). In the final step of computation, directly evaluating the integrals $I^j_{dd'}$, $j = 0,1,2,3,4$, at each estimated $(\widehat{P}_{0g}, \widehat{P}_{1g})$ is analytically intractable. To address this, we approximate the integrals using numerical integration. Specifically, we employ the Gauss-Hermite quadrature method, which we have verified to be both accurate and computationally efficient.

We arbitrarily select the following evaluation points,

$$(p_0, p_1) = (0.3, 0.7), (0.4, 0.6), (0.5, 0.5), (0.6, 0.4), (0.7, 0.3),$$

for which the true MCSEs and MCDEs can be readily computed. We conduct 500 Monte Carlo replications for each of four sample sizes, $G = 1000, 3000, 5000, 10000$. Table 2 reports the coverage rates for the MCSEs, MCDEs, and the correlation parameter ρ.

Table 2: 95% confidence interval of parametrically estimated parameters

	Coverage rate					
	(0.3,0.7)	(0.4,0.6)	(0.5,0.5)	(0.6,0.4)	(0.7,0.3)	ρ
			Panel A1: MCDE ($G = 1000$)			
$d = 1$	0.95	0.952	0.964	0.972	0.958	0.948
$d = 0$	0.96	0.958	0.962	0.96	0.958	
			Panel A2: MCSE ($G = 1000$)			
$d = 1$	0.958	0.962	0.972	0.966	0.96	0.948
$d = 0$	0.964	0.968	0.956	0.954	0.942	
			Panel B1: MCDE ($G = 3000$)			
$d = 1$	0.952	0.944	0.946	0.946	0.958	0.942
$d = 0$	0.952	0.944	0.946	0.936	0.936	
			Panel B2: MCSE ($G = 3000$)			
$d = 1$	0.952	0.948	0.948	0.94	0.942	0.942
$d = 0$	0.938	0.93	0.932	0.944	0.932	
			Panel C1: MCDE ($G = 5000$)			
$d = 1$	0.942	0.95	0.95	0.954	0.932	0.94
$d = 0$	0.952	0.94	0.928	0.918	0.936	
			Panel C2: MCSE ($G = 5000$)			
$d = 1$	0.944	0.948	0.922	0.924	0.926	0.94
$d = 0$	0.958	0.956	0.938	0.946	0.966	
			Panel D1: MCDE ($G = 10000$)			
$d = 1$	0.932	0.938	0.948	0.954	0.944	0.94
$d = 0$	0.926	0.936	0.948	0.946	0.958	
			Panel D2: MCSE ($G = 10000$)			
$d = 1$	0.938	0.95	0.946	0.94	0.954	0.94
$d = 0$	0.95	0.952	0.95	0.954	0.95	

For the MCSEs and MCDEs, when the sample size is $G = 1000$, the coverage rates are already close to, but slightly above, 95% for most parameters. As the sample size increases to $G = 3000$, the coverage rates decrease slightly yet remain close to 95%, with a few parameters falling just below this threshold. For larger sample sizes, the coverage rates for all parameters stabilize around 95%. The coverage rates for ρ are also close to 95% across all sample sizes. These simulation results support the validity of our identification strategy and parametric estimation methods.

5.2 Application: Returns to education in best-friend relationships

In the empirical analysis, we estimate the direct and spillover effects of returns to education among the best-friend groups. We use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative longitudinal survey that follows a cohort of U.S. adolescents from grades 7-12 (1994-95 school year) into adulthood. The dataset contains rich information on respondents' family background and detailed friendship networks during adolescence, as well as educational attainment and income in adulthood. This unique combination of longitudinal social, demographic, and economic data makes Add Health well suited for studying the long-term effects of adolescent friendships.

The Add Health dataset collects detailed friendship information during adolescence in both the in-home and in-school components of the Wave I survey. In each component, respondents are asked to list up to five male and five female friends, ranked from best to fifth best. We construct best-friend groups, each consisting of two respondents, by matching individuals who mutually nominate each other as their best friend. Following Card and Giuliano (2013), we first identify best-friend pairs from the Wave I in-home interviews. We then match any remaining mutually nominated best-friend pairs from the Wave I in-school interviews. Because respondents can nominate the best friend of each gender, we prioritize opposite-gender pairs: if a respondent appears in two different best-friend groups, we retain the group consisting of opposite-gender best friends.

The relationship between an individual's own education and their income has been extensively studied in the economics literature. In contrast, relatively little attention has been paid to how a best friend's educational attainment influences an individual's earnings. Such an effect may operate through two competing channels.

In this empirical study, we investigate the effect of a best friend's educational attainment on an individual's earnings and assess which channel, information sharing or competition, plays the dominant role within best-friend networks. Importantly, our identification framework assumes that spillover effects occur only within the same network and do not extend

across different networks. In the context of returns to education, this implies that any effect of another person's education is restricted to the identified best friend, with no cross-pair spillovers. We take the total personal yearly pre-tax income from the Wave III in-home survey and apply a natural logarithm transformation to construct the outcome variable Y. The binary treatment variable D is set to 1 if the individual has completed at least 16 years of education and 0 otherwise. We include the age, gender, race, health status, and family income as the controlled covariates X. We assume that, conditional on the observed covariates, the coefficients $(\alpha'_{idd'}, \beta'_{idd'})'$ in the potential outcome equations are identical for individuals $i \in \{0, 1\}$ within the same group.

For the continuous instruments Z, we construct measures based on the average parental education level of the individual's non-best friends, defined as all listed friends who are not ranked as the best friend. The average parental education of non-best friends may influence an individual's educational attainment through channels such as shaping aspirations, fostering self-confidence, or behavioral sharing (Cools, Fernández, and Patacchini, 2019). Furthermore, conditional on covariates capturing demographic and socioeconomic characteristics, the family background of non-best friends is plausibly independent of the individual's unobserved heterogeneity. This is because weaker social ties, such as those with non-best friends, are less likely to exhibit the strong peer spillovers characteristic of best-friend relationships, and any residual correlation in unobservables is unlikely to persist once observed similarities are controlled for. Therefore, Assumption 1 is likely to hold in this context.

The average parental education level of non-best friends during adolescence is unlikely to have a direct effect on an individual's yearly income in adulthood. This is because weaker social ties, such as those with non-best friends, generally lack the sustained and intensive interactions needed to shape long-term labor market outcomes. Unlike best friends, non-best friends are less likely to share close personal networks, exchange detailed career information, or provide direct referrals in the labor market. Moreover, by adulthood, many of these weaker ties from adolescence are no longer active, further limiting the scope for any direct influence on earnings. Therefore, any effect of non-best friends' parental education on the individual's income is likely to operate indirectly through its influence on the individual's own educational attainment, rather than through direct channels. Hence, Assumption 2 is plausibly satisfied in this setting.

We also require that the educational decisions of best friends do not directly affect one another. This assumption is plausible because, while best friends may share aspirations or study habits, the final decision on how many years of education to pursue is typically determined by individual specific factors, such as academic ability, that are not directly changed by the best friend's decision. Therefore, any influence between best friends' educational

outcomes is more likely to operate indirectly through shared environments or information exchange, which aligns with the simultaneous incomplete information framework underlying our setting, rather than through direct strategic interaction in determining each other's years of schooling.

Table 3: MCDEs and MCSEs of returns to education using Add Health data

Eval. Points		(0.3, 0.7)	(0.7, 0.3)	(0.5, 0.5)	(0.4, 0.6)	(0.6, 0.4)
			MCDE			
$d=1$	Est.	16.88	11.08	13.17	14.76	11.96
	90% CI	[2.76, 27.85]	[1.72, 18.39]	[4.83, 19.41]	[4.35, 22.79]	[3.92, 17.51]
	95% CI	[1.39, 30.05]	[0.35, 19.38]	[3.04, 21.35]	[2.44, 24.64]	[2.07, 18.66]
$d=0$	Est.	-0.28	-0.04	-2.44	-1.96	-1.85
	90% CI	[-4.23, 3.45]	[-11.76, 8.20]	[-7.15, 2.22]	[-5.60, 1.55]	[-9.75, 4.43]
	95% CI	[-4.71, 3.89]	[-13.57, 10.40]	[-9.16, 2.97]	[-6.14, 2.16]	[-11.26, 5.66]
			MCSE			
$d=1$	Est.	2.78	13.71	9.12	6.27	11.56
	90% CI	[-2.66, 7.90]	[0.07, 26.57]	[0.95, 16.11]	[0.01, 11.67]	[1.08, 21.08]
	95% CI	[-3.79, 9.13]	[-1.67, 28.64]	[-0.25, 17.53]	[-0.67, 12.78]	[-0.80, 23.09]
$d=0$	Est.	-14.38	2.6	-6.49	-10.45	-2.25
	90% CI	[-24.52, -1.65]	[-2.98, 6.87]	[-11.55, -0.47]	[-17.75, -1.51]	[-5.85, 1.50]
	95% CI	[-27.54. 0.79]	[-3.81, 7.64]	[-12.87, 0.34]	[-19.68, -0.06]	[-6.74, 1.84]

After excluding best-friend pairs in which both members have missing values for the treatment D or the instrument Z, the sample comprises 1,019 best-friend pairs. The estimated correlation between best friends' unobservables V_{0g} and V_{1g} is 0.36, indicating a positive dependence structure among unobservables within best-friend networks. Table 3 reports the point estimates and the 90% and 95% confidence intervals for our parameters of interest, $\text{MCDE}_i(\mathbf{x}, d; p_0, p_1)$ and $\text{MCSE}_i(\mathbf{x}, d; p_0, p_1)$, for $i \in \{0, 1\}$ and $d \in \{0, 1\}$, evaluated at a set of arbitrarily chosen points (p_0, p_1). The covariates $\mathbf{x}$ are fixed at their sample means. The results reveal substantial heterogeneity across these parameters. In particular, the estimates of $\text{MCDE}_i(\mathbf{x}, d; p_0, p_1)$ with $d = 1$, which capture the direct effect of completing at least 16-year education given the best friend has completed at least 16 years, are positive and statistically significant at the 5% level for all evaluation points. However, the estimates of $\text{MCDE}_i(\mathbf{x}, d; p_0, p_1)$ with $d = 0$, which measure the direct effect of completing at least 16 years of education given the best friend has not completed this level, are not statistically significant even at the 10% level at any evaluation point. This discrepancy may reflect complementarities in human capital accumulation within best-friend pairs, consistent with the

first channel discussed earlier: a highly educated best friend can provide valuable labor market information and opportunities that enhance the returns to one's own education. When both friends attain higher education, they may reinforce each other's labor market prospects through stronger professional networks, mutual encouragement in career development, or joint access to high-return opportunities. In contrast, when the best friend has lower educational attainment, such reinforcing mechanisms may be absent, weakening the direct effect of one's own education on earnings.

Table 3 also reports the estimates of $\text{MCSE}_i(\mathbf{x}, d; p_0, p_1)$ for $d \in \{0, 1\}$, where covariates $\mathbf{x}$ are fixed at their mean values. The parameters with $d = 1$, which capture the spillover effect of the best friend completing at least 16 years of education given the individual has completed 16 years, are positive across all evaluation points, statistically significant at the 90% level for most points, and significant at the 95% level for several points. In contrast, the parameters with $d = 0$, which measure the spillover effect of the best friend completing at least 16 years of education given the individual has not completed 16 years, are negative at most evaluation points and significantly negative at the 90% level for some points. These patterns are consistent with the two channels through which a best friend's educational attainment may affect an individual's earnings. The findings suggest that the information and opportunity channel dominates the competition channel when the individual is also highly educated, leading to positive and significant spillover effects. Conversely, when the individual has not completed 16 years of education, the competition channel appears to dominate, yielding negative spillover estimates. This asymmetry highlights the role of complementarities in human capital and opportunity sharing among equally educated peers, and the potential for relative disadvantage when educational attainment differs within a best-friend pair.

Additionally, when best-friend pairs differ in their values of unobservable V_i, the parameters $\text{MCSE}_i(\mathbf{x}, d; p_0, p_1)$ for $d = 0$ capture heterogeneity in spillover effects between the two friends. For example, in groups with $(V_i = 0.3, V_{1-i} = 0.7)$ or with $(V_i = 0.4, V_{1-i} = 0.6)$, the marginal spillover effects conditional on the individual not having completed 16 years of education are significantly negative at the 90% level for the individual with the lower value of unobserved heterogeneity, but statistically insignificant for the individual with the higher value. This heterogeneity provides additional insight into the role of unobserved characteristics in affecting spillover effects. In our framework, V_i represents a latent variable summarizing unobserved factors, such as innate ability, motivation, or persistence, that affect an individual's propensity to complete at least 16 years of education. Individuals with lower values of V_i tend to have a higher propensity for completing 16 years of education, even if they did not ultimately achieve it. For these individuals, having a highly educated

best friend may generate stronger direct competition as they share similar labor markets, thus producing more negative spillover effects on earnings. This is consistent with the competition channel dominating the information-sharing channel in this subgroup. By contrast, individuals with higher values of V_i, and thus a lower unobserved propensity for education, are less likely to compete directly with their highly educated best friends in the same labor markets, reducing the scope for competitive pressure. As a result, the estimated spillover effects for individuals with high values of the unobserved factor are statistically insignificant.

6 Conclusion

In this paper, we examine causal direct and spillover effects in settings where the Stable Unit Treatment Value Assumption (SUTVA) may be violated. We focus on a framework with independently and identically distributed groups of fixed size—such as pairs of friends or household members—where spillovers are restricted to occur within groups. We allow for potential spillovers in outcomes: a unit's outcome may be influenced by the treatment status of her group members. Treatment compliance is imperfect in our setting, and we address the resulting endogeneity using instrument variables. In addition, we permit treatment spillovers, meaning that a unit's treatment decision may be affected by instruments assigned to other members within the same group.

We formulate the outcome model under the potential outcomes framework, allowing the outcome to flexibly depend on the treatment statuses of her peers without imposing functional form restrictions. Treatment is modeled as a binary decision governed by a threshold-crossing rule: a unit selects into treatment if an unobserved confounding variable falls below a threshold determined by both her own and her peers' instruments. We allow these unobserved confounding variables to be arbitrarily dependent across group members, imposing no assumptions on the joint distribution.

We introduce two central causal parameters: the marginal controlled spillover effect (MCSE) and the marginal controlled direct effect (MCDE), defined conditional on the unobserved heterogeneities of all group members. These effects capture individual heterogeneity in both spillover and direct treatment effects. We point identify these parameters using continuous instruments and show that they serve as foundational components for constructing a range of widely studied causal estimands, including (local) average spillover/direct effects and policy-relevant treatment effects (PRTEs). Additionally, we propose semiparametric estimators for these marginal effects and derive their asymptotic properties.

We propose a semi-nonparametric method to estimate the identified marginal controlled spillover and direct effects, and derive their asymptotic distributions. When the sample size

is small or group size is large, we introduce a parametric estimation method and recommend its use. Confidence intervals for the parametrically estimated effects are constructed using the nonparametric bootstrap. We conduct Monte Carlo simulations to assess the validity of the parametric confidence intervals. We apply our parametric estimation and inference procedure to study the impact of educational attainment on long-term earnings within best-friend networks, using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). These results indicate that educational attainment generates spillover effects on long-term earnings within best-friend networks, and that the nature of these effects differs depending on individuals' educational attainment.

Future research will extend the proposed framework to more general settings, such as those in which individual outcomes depend on a correctly specified exposure mapping of the group treatment vector, rather than the full vector itself.

Reference

Aradillas-Lopez, Andres. 2010. "Semiparametric estimation of a simultaneous game with incomplete information." *Journal of Econometrics* 157 (2):409–431.

Aronow, Peter M and Cyrus Samii. 2017. "ESTIMATING AVERAGE CAUSAL EFFECTS UNDER GENERAL INTERFERENCE, WITH APPLICATION TO A SOCIAL NETWORK EXPERIMENT." *THE ANNALS of APPLIED STATISTICS* :1912–1947.

Balat, Jorge F and Sukjin Han. 2023. "Multiple treatments with strategic substitutes." *Journal of Econometrics* 234 (2):732–757.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. 2015. "Some new asymptotic theory for least squares series: Pointwise and uniform results." *Journal of Econometrics* 186 (2):345–366.

Bickel, Peter J, Ya'acov Ritov, and Alexandre B Tsybakov. 2009. "Simultaneous analysis of Lasso and Dantzig selector." *The Annals of Statistics* 37 (4):1705–1732.

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

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of econometrics* 150 (1):41–55.

Card, David and Laura Giuliano. 2013. "Peer effects and multiple equilibria in the risky behavior of friends." *Review of Economics and Statistics* 95 (4):1130–1149.

Carneiro, Pedro and Sokbae Lee. 2009. "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality." *Journal of Econometrics* 149 (2):191–208.

Carr, Thomas and Toru Kitagawa. 2021. "Testing instrument validity with covariates." *arXiv preprint arXiv:2112.08092* .

Cools, Angela, Raquel Fernández, and Eleonora Patacchini. 2019. "Girls, boys, and high achievers." Tech. rep., National Bureau of Economic Research.

DiTraglia, Francis J, Camilo García-Jimeno, Rossa O'Keeffe-O'Donovan, and Alejandro Sánchez-Becerra. 2023. "Identifying causal effects in experiments with spillovers and non-compliance." *Journal of Econometrics* 235 (2):1589–1624.

Fan, J. and I. Gijbels. 1996. *Local polynomial modelling and its applications*. Chapman & Hall/CRC.

Heckman, James J and Edward Vytlacil. 2001. "Policy-relevant treatment effects." *American Economic Review* 91 (2):107–111.

———. 2005. "Structural equations, treatment effects, and econometric policy evaluation 1." *Econometrica* 73 (3):669–738.

Hoshino, Tadao and Takahide Yanagi. 2023. "Treatment effect models with strategic interaction in treatment decisions." *Journal of Econometrics* 236 (2):105495.

Hudgens, Michael G and M Elizabeth Halloran. 2008. "Toward causal inference with interference." *Journal of the American Statistical Association* 103 (482):832–842.

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76 (3):1071–1102.

Lee, Sokbae and Bernard Salanié. 2018. "Identifying effects of multivalued treatments." *Econometrica* 86 (6):1939–1963.

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

———. 2013. "Identification of treatment response with social interactions." *The Econometrics Journal* 16 (1):S1–S23.

Masry, Elias. 1996. "Multivariate regression estimation local polynomial fitting for time series." *Stochastic Processes and their Applications* 65 (1):81–101.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2018. "Using instrumental variables for inference about policy relevant treatment parameters." *Econometrica* 86 (5):1589–1619.

Romano, Joseph P and Azeem M Shaikh. 2012. "ON THE UNIFORM ASYMPTOTIC VALIDITY OF SUBSAMPLING AND THE BOOTSTRAP." *The Annals of Statistics* 40 (6):2798–2822.

Vazquez-Bare, Gonzalo. 2023. "Causal spillover effects using instrumental variables." *Journal of the American Statistical Association* 118 (543):1911–1922.

Appendices

A Proof of identification for MCSEs and MCDEs

A.1 Identifying the copula

Under Assumptions 1-3, the propensity score for unit i can be shown as

$$\begin{aligned}
&\mathbb{P}\left(D_i = 1 \mid Z_i = z_0, Z_{-i} = z_1\right) \\
=&\mathbb{P}\left(V_i \leq h_i(Z_i, Z_{-i}) \mid Z_i = z_0, Z_{-i} = z_1\right) \\
=&\mathbb{P}\left(V_i \leq h_i(z_0, z_1) \mid Z_i = z_0, Z_{-i} = z_1\right) \\
=&\mathbb{P}\left(V_i \leq h_i(z_0, z_1)\right) \\
=&h_i(z_0, z_1),
\end{aligned} \tag{18}$$

given $z_0 \in \text{Supp}(Z_i) = \mathbb{R}^{k_i}, z_1 \in \text{Supp}(Z_{-i}) = \mathbb{R}^{k-i}$. The third equality follows directly from Assumption 1. Furthermore, under Assumption 3, we normalize V_i to follow a uniform distribution $\mathcal{U}(0,1)$, which justifies the final equality. Equation (18) shows that the threshold function $h_i(Z_i, Z_{-i})$ in the treatment mechanism is identified by the propensity score function P_i on its support $\mathcal{P}_i$.

Once the propensity scores $(P_i, P_{-i}) \in \mathcal{P}$ of all group members are identified, the copula $C_{V_i,V_{-i}}(p_0, p_1)$, which characterizes the dependence structure between the unobserved heterogeneities within the group, can be identified as

$$\begin{aligned}
&\mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid P_i\big(Z_i, Z_{-i}\big) = p_0, P_{-i}\big(Z_{-i}, Z_i\big) = p_1\big) \\
=&\mathbb{P}\big(V_i \leq h_i(Z_i, Z_{-i}), V_{-i} \leq h_{-i}(Z_{-i}, Z_i) \mid P_i\big(Z_i, Z_{-i}\big) = p_0, P_{-i}\big(Z_{-i}, Z_i\big) = p_1\big) \\
=&\mathbb{P}\big(V_i \leq p_0, V_{-i} \leq p_1 \mid P_i\big(Z_i, Z_{-i}\big) = p_0, P_{-i}\big(Z_{-i}, Z_i\big) = p_1\big) \\
=&\mathbb{P}\big(V_i \leq p_0, V_{-i} \leq p_1\big) \\
=&C_{V_i,V_{-i}}(p_0, p_1)
\end{aligned} \tag{19}$$

for $(p_0, p_1) \in \mathcal{P}$, where the second equality follows from the identification of the threshold function h_i by the propensity score P_i, and the last equality holds under Assumption 1.

If Assumption 4 holds and $\mathbb{E}[D_i D_{-i} \mid P_i, P_{-i}]$ is twice differentiable at $(p_0, p_1) \in \mathcal{P}$, then

the copula density can be identified by taking second-order derivatives,

$$\begin{aligned}&\frac{\partial^2 \mathbb{E}\big[D_i D_{-i} \mid P_i = p_0, P_{-i} = p_1\big]}{\partial p_0 \partial p_1} \\ =&\frac{\partial^2 \mathbb{P}\big(V_i \le p_0, V_{-i} \le p_1\big)}{\partial p_0 \partial p_1} = c_{V_i, V_{-i}}(p_0, p_1).\end{aligned} \tag{20}$$

A.2 Identifying the marginal treatment response functions

Given the values of propensity scores $P_i(Z_i, Z_{-i}) = p_0$, $P_{-i}(Z_{-i}, Z_i) = p_1$, and any Borel set $A \subset \mathcal{Y}$, we have

$$\begin{aligned}&\mathbb{E}\big[\mathbb{1}\{Y_i \in A\} D_i D_{-i} \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\big] \\ =&\mathbb{E}\big[\mathbb{1}\{Y_i(1,1) \in A\} \cdot \mathbb{1}\{V_i \le h_i(Z_i, Z_{-i})\} \cdot \mathbb{1}\big\{V_{-i} \le h_{-i}(Z_{-i}, Z_i)\big\} \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\big] \\ =&\mathbb{E}\big[\mathbb{1}\{Y_i(1,1) \in A\} \cdot \mathbb{1}\{V_i \le p_0\} \cdot \mathbb{1}\big\{V_{-i} \le p_1\big\} \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\big] \\ =&\mathbb{E}\big[\mathbb{1}\{Y_i(1,1) \in A\} \cdot \mathbb{1}\{V_i \le p_0\} \cdot \mathbb{1}\big\{V_{-i} \le p_1\big\}\big] \\ =&\int_0^{p_1} \int_0^{p_0} \mathbb{P}\big(Y_i(1,1) \in A \mid V_i = v_0, V_{-i} = v_1\big) c_{V_i, V_{-i}}(v_0, v_1) dv_0 dv_1,\end{aligned}$$

where the second equality follows from Equation (18), and the third equality holds under Assumption 1. If the function $\mathbb{E}\big[\mathbb{1}\{Y_i(1,1) \in A\} D_i D_{-i} \mid \cdot, \cdot\big]$ is twice differentiable and $m_i^{(1,1)}(\cdot, \cdot)$ is continunous at (p_0, p_1), by the Leibniz integral rule,

$$\begin{aligned}&\frac{\partial^2}{\partial p_1 \partial p_0} \mathbb{E}\big[\mathbb{1}\{Y_i(1,1) \in A\} D_i D_{-i} \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\big] \\ =&\mathbb{P}\big(Y_i(1,1) \in A \mid V_i = p_0, V_{-i} = p_1\big) \cdot c_{V_i, V_{-i}}(p_0, p_1).\end{aligned} \tag{21}$$

Since the copula density of (V_i, V_{-i}), $c(\cdot, \cdot)$, is identified from Corollary 1, we can identify $\mathbb{P}\big(Y_i(1,1) \in A \mid V_i = p_0, V_{-i} = p_1\big)$ from Equation (21). This, in turn, implies that the marginal treatment response function $m_i^{(1,1)}(p_0, p_1)$ is identified.

We can apply the same procedure and obtain

$$
\begin{aligned}
&-\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{E}\big[\mathbb{1}\{Y_i \in A\} D_i(1-D_{-i}) \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\big] \\
=&\mathbb{P}\big(Y_i(1,0) \in A \mid V_i = p_0, V_{-i} = p_1\big) \cdot c_{V_i, V_{-i}}(p_0, p_1), \\
&-\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}(1-D_i) D_{-i} \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\big] \\
=&\mathbb{P}\big(Y_i(0,1) \in A \mid V_i = p_0, V_{-i} = p_1\big) \cdot c_{V_i, V_{-i}}(p_0, p_1), \\
&\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{E}\big[\mathbb{1}\{Y_i \in A\}(1-D_i)(1-D_{-i}) \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\big] \\
=&\mathbb{P}\big(Y_i(0,0) \in A \mid V_i = p_0, V_{-i} = p_1\big) \cdot c_{V_i, V_{-i}}(p_0, p_1).
\end{aligned}
\tag{22}
$$

We can then identify the remaining marginal treatment response functions from Equation (22).

Based on results in Eqs.(21) - (22), we can identify $\text{MCSE}_i(d; p_0, p_1)$, $\text{MCDE}_i(d; p_0, p_1)$ for any $d \in \{0, 1\}$ and $p_0, p_1 \in \mathcal{P}$.

A.3 Identification With Exogenous Covariates

The identification results can be extended to settings with exogenous covariates. Let $X_i \in \mathbb{R}^{d_i}$ denote a vector of covariates that affect both the outcome and the treatment assignment for unit i. For instance, unit i's earnings and educational choices may depend on the family characteristics of both herself and her best friend in our leading example. Given $(X_i, X_{-i}) = \mathbf{x}$, $D_i = d$, and $D_{-i} = d'$, we model the potential outcome as

$$
Y_i(\mathbf{x}, d, d') = \mu_{dd'}\big(\mathbf{x}, U_i(d, d;)\big), \tag{23}
$$

where the functions $\mu_{dd'}(\cdot, \cdot)$ are known and specified by the researcher, while $U_i(d, d')$ captures unobserved factors affecting unit i's potential outcome under own treatment status $D_i = d$ and the peer $-i$'s treatment $D_{-i} = d'$. A common specification for $\mu_{dd'}(\cdot, \cdot)$ assumes additive separability and linearity in covariates: $\mu_{dd'}\big(\mathbf{x}, U_i(d, d')\big) = \mathbf{x}\beta_{dd'} + U_i(d, d')$.

We next introduce a potential outcome model that incorporates exogenous covariates.

$$
\begin{cases}
Y_i = \big[Y_i(X_i, X_{-i}, 1, 1) D_{-i} + Y_i(X_i, X_{-i}, 1, 0)(1 - D_{-i})\big] D_i \\
\qquad + \big[Y_i(X_i, X_{-i}, 0, 1) D_{-i} + Y_i(X_i, X_{-i}, 0, 0)(1 - D_{-i})\big](1 - D_i), \\
D_i = \mathbb{1}\big\{V_i \leq h(W_i, W_{-i})\big\},
\end{cases}
\tag{24}
$$

where $W_i \equiv (X_i, Z_i) \in \mathbb{R}^{d_i} \times \mathbb{R}^{k_i}$.

Under Equation (24), we replace Assumption 1 with Assumption 14, which imposes random assignment of both covariates and instruments.

Assumption 14. (Exogenous covariates and random assignment) The covariates X_i and the instruments Z_i satisfy

$$\big(X_i, X_{-i}, Z_i, Z_{-i}\big) \perp\!\!\!\perp \Big\{\big(V_i, V_{-i}, U_i(d,d'), U_{-i}(d,d'),\big)\Big\}_{d\in\{0,1\},d'\in\{0,1\}}.$$

Under Assumptions 2, 3, and 14, the propensity score with exogenous covariates, defined as $P_i(W_i, W_{-i}) \equiv \mathbb{P}(D_i = 1 \mid W_i, W_{-i})$, can be expressed as

$$\begin{aligned}
&\mathbb{P}\left(D_i = 1 \mid W_i = w_0, W_{-i} = w_1\right)\\
=&\mathbb{P}\left(V_i \le h_i(W_i, W_{-i}) \mid W_i = w_0, W_{-i} = w_1\right)\\
=&\mathbb{P}\left(V_i \le h_i(w, w') \mid W_i = w_0, W_{-i} = w_1\right)\\
=&\mathbb{P}\left(V_i \le h_i(w_0, w_1)\right)\\
=&h_i(w_0, w_1)
\end{aligned} \tag{25}$$

given $W_i = w_0, W_{-i} = w_1$. Equation (25) demonstrates that, in the presence of exogenous covariates, the propensity score $P_i(W_i, W_{-i})$ continues to identify the threshold function h_i over its support $\mathcal{P}_i$.

Similar to Equation (19), we can identify the copula function $c_{V_i,V_{-i}}(p_0, p_1)$ as

$$\begin{aligned}
&\mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid P_i(W_i, W_{-i}) = p_0, P_{-i}(W_{-i}, W_{ig}) = p_1\big)\\
=&\mathbb{P}\big(V_i \le h(W_i, W_{-i}), V_{-i} \le h(W_{-i}, W_i) \mid P_i(W_i, W_{-i}) = p_0, P_{-i}(W_{-i}, W_i) = p_1\big)\\
=&\mathbb{P}\big(V_i \le p_0, V_{-i} \le p_1 \mid P_i(W_i, W_{-i}) = p_0, P_{-i}(W_{-i}, W_i) = p_1\big)\\
=&\mathbb{P}\left(V_i \le p_0, V_{-i} \le p_1\right),
\end{aligned} \tag{26}$$

where the last equality holds under Assumption 14. Then, the copula density of (V_i, V_{-i}), $c_{V_i,V_{-i}}(\cdot,\cdot)$, is identifiable provided that the copula $C_{V_i,V_{-i}}(\cdot,\cdot)$ is twice differentiable,

$$\begin{aligned}
&\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{P}\big(D_i = 1, D_{-i} = 1 \mid P_i(W_i, W_{-i}) = p_0, P_{-i}(W_{-i}, W_i) = p_1\big)\\
=&\frac{\partial^2}{\partial p_1 \partial p_0}\mathbb{P}\big(V_i \le p_0, V_{-i} \le p_1\big) = c_{V_i,V_{-i}}(p_0, p_1).
\end{aligned} \tag{27}$$

The last step is to identify the marginal treatment response functions, defined as $m_i^{(\mathbf{x},d,d')}(p_0, p_1) \equiv \mathbb{E}\big[Y_i(\mathbf{x}, d, d') \mid V_i = p_0, V_{-i} = p_1\big]$ with covariates. Given the covariates and propensity scores

of both units i and her peer $-i$, we can express the following conditional expectation as

$$
\begin{aligned}
&\mathbb{E}\big[Y_i D_i D_{-i} \mid (X_i, X_{-i}) = \mathbf{x}, P(W_i, W_{-i}) = p_0, P(W_{-i}, W_i) = p_1\big] \\
=&\mathbb{E}\big[\mu_{11}(\mathbf{x}, U_i(1,1)) \cdot \mathbb{1}\{V_i \leq h_i(W_i, W_{-i})\} \cdot \mathbb{1}\{V_{-i} \leq h_{-i}(W_{-i}, W_i)\} \mid \\
&\quad (X_i, X_{-i}) = \mathbf{x}, P_i(W_i, W_{-i}) = p_0, P_{-i}(W_{-i}, W_i) = p_1\big] \\
=&\mathbb{E}\big[\mu_{11}(\mathbf{x}, U_i(1,1)) \cdot \mathbb{1}\{V_i \leq p_0\} \cdot \mathbb{1}\{V_{-i} \leq p_1\} \mid \\
&\quad (X_i, X_{-i}) = \mathbf{x}, P_i(W_i, W_{-i}) = p_0, P_{-i}(W_{-i}, W_i) = p_1\big] \\
=&\mathbb{E}\big[\mu_{11}(\mathbf{x}, U_i(1,1)) \cdot \mathbb{1}\{V_i \leq p_0\} \cdot \mathbb{1}\{V_{-i} \leq p_1\}\big] \\
=&\int_0^{p_1} \int_0^{p_0} \mathbb{E}\big[\mu_{11}(\mathbf{x}, U_i(1,1)) \mid V_i = v_0, V_{-i} = v_1\big] c_{V_i, V_{-i}}(v_0, v_1) dv_0 dv_1,
\end{aligned}
$$

where the second equality follows from Equation (25), and the third equality holds under Assumption 14. If the conditional mean $\mathbb{E}[Y_i D_i D_{-i} \mid (X_i, X_{-i}) = \mathbf{x}, \cdot, \cdot]$ is twice differentiable, and the marginal treatment response functions $\mathbb{E}[\mu_{dd'}(\mathbf{x}, U_i(d, d')) \mid \cdot, \cdot]$ are continuous at (p_0, p_1), then the marginal treatment response function $\mathbb{E}[\mu_{11}(\mathbf{x}, U_i(1,1)) \mid V_i = p_0, V_{-i} = p_1]$ is identified by taking the cross-derivative as shown below:

$$
\begin{aligned}
&\frac{\partial^2}{\partial p_1 \partial p_0} \mathbb{E}\left[Y_i D_i D_{-i} \mid (X_i, X_{-i}) = \mathbf{x}, P_i(W_i, W_{-i}) = p_0, P_{-i}(W_{-i}, W_i) = p_1\right] \\
=&\mathbb{E}\left[\mu_{11}(\mathbf{x}, U_i(1,1)) \mid V_i = p_0, V_{-i} = p_1\right] c_{V_i, V_{-i}}(p_0, p_1).
\end{aligned}
$$

Therefore, the marginal treatment response function $m_i^{(\mathbf{x},1,1)}(p_0, p_1)$ is identified, given that the copula density is identified as in Equation (27). By analogous reasoning, the remaining marginal treatment response functions $m_i^{(\mathbf{x},d,d')}(p_0, p_1)$ are also identified for all $d, d' \in \{0, 1\}$ and $\mathbf{x} \in \mathbb{R}^{d_i}$.

B Deriving Policy Relevant Treatment Effects with MCSEs and MCDEs

In this section, we identify the PRTEs under three types of common policy interventions with identified MCSEs and MCDEs.

Case 1: Absolute increase by an exogenous value. Suppose there exists an alternative policy $a' \in \mathcal{A}$ that exogenously increases the propensity score of all units by a constant $\varepsilon > 0$, such that $P_i^{a'} = P_i^a + \varepsilon$ and $P_i^a, P_i^{a'} \in [0, 1]$, for all i in every group. By taking the difference between the expected outcomes under the two policies, $\mathbb{E}[Y_i^a]$ and $\mathbb{E}[Y_i^{a'}]$, we can

express this difference as weighted average of MCDEs and MCSEs as follows,

$$\begin{aligned}
\mathbb{E}\left[Y_i^{a'} - Y_i^a\right] = \int_0^1 \int_0^1 \Big\{ & \text{MCDE}_i(0; p_0, p_1)\mathbb{P}\left(p_0 - \varepsilon \leq P_i^a \leq p_0, P_{-i}^a \leq p_1 - \varepsilon\right) \\
& + \text{MCSE}_i(0; p_0, p_1)\mathbb{P}\left(P_i^a \leq p_0 - \varepsilon, p_1 - \varepsilon \leq P_{-i}^a < p_1\right) \\
& + \text{MCDE}_i(1; p_0, p_1)\mathbb{P}\left(p_0 - \varepsilon \leq P_i^a \leq p_0, p_1 \leq P_{-i}^a\right) \\
& + \text{MCSE}_i(1; p_0, p_1)\mathbb{P}\left(p_0 \leq P_i^a, p_1 - \varepsilon \leq P_{-i}^a < p_1\right) \\
& + \left(\text{MCDE}_i(1; p_0, p_1) + \text{MCSE}_i(0; p_0, p_1)\right) \\
& \quad \mathbb{P}\left(p_0 - \varepsilon \leq P_i^a < p_0, p_1 - \varepsilon \leq P_{-i}^a < p_1\right) \Big\} c_{V_i, V_{-i}}(p_0, p_1) dp_0 dp_1
\end{aligned}$$

Once we can identify MCDEs and MCSEs, as well as the joint distributions of propensity scores (P_i^a, P_{-i}^a) and unobservables (V_i, V_{-i}) over the full support $[0,1] \times [0,1]$, the policy relevant treatment effect can be point identified as

$$\mathbb{E}\left[Y_i^{a'} - Y_i^a\right] / \Delta P,$$

where ΔP denotes the proportion of groups in which at least one member changes treatment status as a result of the policy shift from a to a'. This proportion is identified as

$$\begin{aligned}
\Delta P = \int_0^1 \int_0^1 \Big\{ & \mathbb{P}\left(p_0 - \varepsilon \leq P_i^a \leq p_0, P_{-i}^a \leq p_1 - \varepsilon\right) + \mathbb{P}\left(P_i^a \leq p_0 - \varepsilon, p_1 - \varepsilon \leq P_{-i}^a < p_1\right) \\
& + \mathbb{P}\left(p_0 - \varepsilon \leq P_i^a \leq p_0, p_1 \leq P_{-i}^a\right) + \mathbb{P}\left(p_0 \leq P_i^a, p_1 - \varepsilon \leq P_{-i}^a < p_1\right) \\
& + \mathbb{P}\left(p_0 - \varepsilon \leq P_i^a < p_0, p_1 - \varepsilon \leq P_{-i}^a < p_1\right) \Big\} c_{V_i, V_{-i}}(p_0, p_1) dp_0 dp_1.
\end{aligned}$$

We can also identify the PRTEs for cases where $\varepsilon < 0$, or where the policy shift affects group members in opposite directions—for instance, $\varepsilon_i > 0$ and $\varepsilon_{-i} < 0$—by applying analogous derivations.

Case 2: Proportional increase by an exogenous value. Consider an alternative policy $a' \in \mathcal{A}$ that exogenously increases the propensity score of all individuals proportionally, such that $P_i^{a'} = P_i^a + \varepsilon(1 - P_i^a)$ for all individuals i, where $0 < \varepsilon < 1$ and $P_i^a, P_i^{a'} \in [0,1]$.

Under this policy shift, we can identify the PRTE as $\mathbb{E}[Y_i^{a'} - Y_i^a]/\Delta P$, where

$$
\begin{aligned}
\mathbb{E}\left[Y_i^{a'} - Y_i^a\right] = \int_0^1 \int_0^1 \Bigg\{ & \text{MCDE}_\text{i}(0; p_0, p_1)\mathbb{P}\left(\frac{p_0-\varepsilon}{1-\varepsilon} \le P_i^a \le p_0, P_{-i}^a \le \frac{p_1-\varepsilon}{1-\varepsilon}\right) \\
& + \text{MCSE}_i(0; p_0, p_1)\mathbb{P}\left(P_i^a \le \frac{p_0-\varepsilon}{1-\varepsilon}, \frac{p_1-\varepsilon}{1-\varepsilon} \le P_{-i}^a < p_1\right) \\
& + \text{MCDE}_i(1; p_0, p_1)\mathbb{P}\left(\frac{p_0-\varepsilon}{1-\varepsilon} \le P_i^a \le p_0, p_1 \le P_{-i}^a\right) \\
& + \text{MCSE}_i(1; p_0, p_1)\mathbb{P}\left(p_0 \le P_i^a, \frac{p_1-\varepsilon}{1-\varepsilon} \le P_{-i}^a < p_1\right) \\
& + (\text{MCDE}_i(1; p_0, p_1) + \text{MCSE}_i(0; p_0, p_1)) \\
& \quad \mathbb{P}\left(\frac{p_0-\varepsilon}{1-\varepsilon} \le P_i^a < p_0, \frac{p_1-\varepsilon}{1-\varepsilon} \le P_{-i}^a < p_1\right) \Bigg\} c_{V_i, V_{-i}}(p_0, p_1) dp_0 dp_1,
\end{aligned}
$$

$$
\begin{aligned}
\Delta P = \int_0^1 \int_0^1 \Bigg\{ & \mathbb{P}\left(\frac{p_0-\varepsilon}{1-\varepsilon} \le P_i^a \le p_0, P_{-i}^a \le \frac{p_1-\varepsilon}{1-\varepsilon}\right) \\
& + \mathbb{P}\left(P_i^a \le \frac{p_0-\varepsilon}{1-\varepsilon}, \frac{p_1-\varepsilon}{1-\varepsilon} \le P_{-i}^a < p_1\right) \\
& + \mathbb{P}\left(\frac{p_0-\varepsilon}{1-\varepsilon} \le P_i^a \le p_0, p_1 \le P_{-i}^a\right) \\
& + \mathbb{P}\left(p_0 \le P_i^a, \frac{p_1-\varepsilon}{1-\varepsilon} \le P_{-i}^a < p_1\right) \\
& + \mathbb{P}\left(\frac{p_0-\varepsilon}{1-\varepsilon} \le P_i^a < p_0, \frac{p_1-\varepsilon}{1-\varepsilon} \le P_{-i}^a < p_1\right) \Bigg\} c_{V_i, V_{-i}}(p_0, p_1) dp_0 dp_1.
\end{aligned}
$$

Case 3: Increase the instrument value. The third type of policy intervention involves shifting the value of certain instruments. For example, consider a policy change where the j-th component of the instrument is increased by ε, such that $P_i^{a'} = P_i^a(Z + \varepsilon e_j)$, where e_j denotes the unit vector in the j-th coordinate. In the previous policy changes, the direction of the shift in propensity scores was known for all individuals, allowing us to determine the corresponding changes in treatment responses across the entire range of unobserved characteristics (V_i, V_{-i}). However, when we change the instruments, the effect on propensity scores is not necessarily uniform—some individuals may experience an increase in their propensity scores, while others may see a decrease. The heterogeneous shifts in propensity scores introduce variation in group members' treatment responses, making the analysis more complicated. To address this problem, we decompose the expected outcome

difference, $\mathbb{E}[Y_i^{a'} - Y_i^a]$, as

$$\begin{aligned}\mathbb{E}\big[Y_i^{a'} - Y_i^a\big] =& \mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} \geq P_i^a, P_{-i}^{a'} \geq P_{-i}^a\}\big] \\ &+ \mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} \geq P_i^a, P_{-i}^{a'} < P_{-i}^a\}\big] \\ &+ \mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} < P_i^a, P_{-i}^{a'} \geq P_{-i}^a\}\big] \\ &+ \mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} < P_i^a, P_{-i}^{a'} < P_{-i}^a\}\big],\end{aligned}$$

with applying the law of total probability. Given that the distribution of $P_i^a(\cdot)$ is identified and $P_i^{a'}(Z) = P_i^a(Z + \varepsilon e_j)$, we can also identify the joint distribution of $(P_i^a, P_{-i}^a, P_i^{a'}, P_{-i}^{a'})$.

We can solve each component in the above equation as

$$
\begin{aligned}
&\mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} \geq P_i^a, P_{-i}^{a'} \geq P_{-i}^a\}\big] \\
=&\int_0^1\int_0^1 \Big\{ \mathrm{MCDE_i}(0;p_0,p_1)\mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 > P_{-i}^{a'} \geq P_{-i}^a\big) \\
&+ \mathrm{MCSE}_i(0;p_0,p_1)\mathbb{P}\big(p_0 > P_i^{a'} \geq P_i^a, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big) \\
&+ \mathrm{MCDE}_i(1;p_0,p_1)\mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 \leq P_{-i}^a \leq P_{-i}^{a'}\big) \\
&+ \mathrm{MCSE}_i(1;p_0,p_1)\mathbb{P}\big(p_0 \leq P_i^a \leq P_i^{a'}, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big) \\
&+ \big(\mathrm{MCDE}_i(1;p_0,p_1) + \mathrm{MCSE}_i(0;p_0,p_1)\big)\mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big)\Big\} c_{V_i,V_{-i}}(p_0,p_1)dp_0dp_1, \\
&\mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} \geq P_i^a, P_{-i}^{a'} < P_{-i}^a\}\big] \\
=&\int_0^1\int_0^1 \Big\{ \mathrm{MCDE_i}(0;p_0,p_1)\mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 > P_{-i}^a > P_{-i}^{a'}\big) \\
&- \mathrm{MCSE}_i(0;p_0,p_1)\mathbb{P}\big(p_0 > P_i^{a'} \geq P_i^a, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big) \\
&+ \mathrm{MCDE}_i(1;p_0,p_1)\mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 \leq P_{-i}^{a'} < P_{-i}^a\big) \\
&- \mathrm{MCSE}_i(1;p_0,p_1)\mathbb{P}\big(p_0 \leq P_i^a \leq P_i^{a'}, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big) \\
&+ \big(\mathrm{MCDE}_i(0;p_0,p_1) - \mathrm{MCSE}_i(0;p_0,p_1)\big)\mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big)\Big\} c_{V_i,V_{-i}}(p_0,p_1)dp_0dp_1, \\
&\mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} < P_i^a, P_{-i}^{a'} \geq P_{-i}^a\}\big] \\
=&\int_0^1\int_0^1 \Big\{ -\mathrm{MCDE_i}(0;p_0,p_1)\mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 > P_{-i}^{a'} \geq P_{-i}^a\big) \\
&+ \mathrm{MCSE}_i(0;p_0,p_1)\mathbb{P}\big(p_0 > P_i^a > P_i^{a'}, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big) \\
&- \mathrm{MCDE}_i(1;p_0,p_1)\mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 \leq P_{-i}^a \leq P_{-i}^{a'}\big) \\
&+ \mathrm{MCSE}_i(1;p_0,p_1)\mathbb{P}\big(p_0 \leq P_i^{a'} < P_i^a, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big) \\
&+ \big(-\mathrm{MCDE}_i(0;p_0,p_1) + \mathrm{MCSE}_i(0;p_0,p_1)\big)\mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big)\Big\} c_{V_i,V_{-i}}(p_0,p_1)dp_0dp_1, \\
&\mathbb{E}\big[(Y_i^{a'} - Y_i^a)\mathbb{1}\{P_i^{a'} < P_i^a, P_{-i}^{a'} < P_{-i}^a\}\big] \\
=&\int_0^1\int_0^1 \Big\{ -\mathrm{MCDE_i}(0;p_0,p_1)\mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 > P_{-i}^a > P_{-i}^{a'}\big) \\
&- \mathrm{MCSE}_i(0;p_0,p_1)\mathbb{P}\big(p_0 > P_i^a > P_i^{a'}, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big) \\
&- \mathrm{MCDE}_i(1;p_0,p_1)\mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 \leq P_{-i}^{a'} < P_{-i}^a\big) \\
&- \mathrm{MCSE}_i(1;p_0,p_1)\mathbb{P}\big(p_0 \leq P_i^{a'} < P_i^a, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big) \\
&- \big(\mathrm{MCDE}_i(1;p_0,p_1) + \mathrm{MCSE}_i(0;p_0,p_1)\big)\mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big)\Big\} c_{V_i,V_{-i}}(p_0,p_1)dp_0dp_1,
\end{aligned}
$$

which can be identified once we identify the MCDEs, MCSEs, and the copula density of (V_i, V_{-i}).

Finally, we can identify the PRTE in this case as $\mathbb{E}[Y_i^{a'} - Y_i^a]/\Delta P$, where

$$
\begin{aligned}
&\Delta P = \Delta P_1 + \Delta P_2 + \Delta P_3 + \Delta P_4,\\
&\Delta P_1 = \int_0^1 \int_0^1 \Big\{ \mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 > P_{-i}^{a'} \geq P_{-i}^a\big) + \mathbb{P}\big(p_0 > P_i^{a'} \geq P_i^a, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big)\\
&+ \mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 \leq P_{-i}^a \leq P_{-i}^{a'}\big) + \mathbb{P}\big(p_0 \leq P_i^a \leq P_i^{a'}, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big)\\
&+ \mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big) \Big\} c_{V_i,V_{-i}}(p_0, p_1) dp_0 dp_1,\\
&\Delta P_2 = \int_0^1 \int_0^1 \Big\{ \mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 > P_{-i}^a > P_{-i}^{a'}\big) + \mathbb{P}\big(p_0 > P_i^{a'} \geq P_i^a, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big)\\
&+ \mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, p_1 \leq P_{-i}^{a'} < P_{-i}^a\big) + \mathbb{P}\big(p_0 \leq P_i^a \leq P_i^{a'}, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big)\\
&+ \mathbb{P}\big(P_i^a < p_0 \leq P_i^{a'}, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big) \Big\} c_{V_i,V_{-i}}(p_0, p_1) dp_0 dp_1,\\
&\Delta P_3 = \int_0^1 \int_0^1 \Big\{ \mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 > P_{-i}^{a'} \geq P_{-i}^a\big) + \mathbb{P}\big(p_0 > P_i^a > P_i^{a'}, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big)\\
&+ \mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 \leq P_{-i}^a \leq P_{-i}^{a'}\big) + \mathbb{P}\big(p_0 \leq P_i^{a'} < P_i^a, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big)\\
&+ \mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, P_{-i}^a < p_1 \leq P_{-i}^{a'}\big) \Big\} c_{V_i,V_{-i}}(p_0, p_1) dp_0 dp_1,\\
&\Delta P_4 = \int_0^1 \int_0^1 \Big\{ \mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 > P_{-i}^a > P_{-i}^{a'}\big) + \mathbb{P}\big(p_0 > P_i^a > P_i^{a'}, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big)\\
&+ \mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, p_1 \leq P_{-i}^{a'} < P_{-i}^a\big) + \mathbb{P}\big(p_0 \leq P_i^{a'} < P_i^a, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big)\\
&+ \mathbb{P}\big(P_i^{a'} < p_0 \leq P_i^a, P_{-i}^{a'} < p_1 \leq P_{-i}^a\big) \Big\} c_{V_i,V_{-i}}(p_0, p_1) dp_0 dp_1.
\end{aligned}
$$

C Comparison With Relevant Literature

C.1 Breakdown of MTE causal validity: Proof

We can express $\mathbb{E}\left[Y_i D_i \mid p_i\left(Z_i\right) = p_0\right]$ as

$$
\begin{aligned}
&\mathbb{E}\left[Y_i D_i \mid P_i(Z_i) = p_0\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[Y_i D_i \mid P_i(Z_i) = p_0, P_{-i}(Z_{-i}) = p_1\right] \mid P_i(Z_i) = p_0\right]
\end{aligned}
$$

by applying the law of iterated expectations. The inner conditional expectation can be further expressed as

$$
\begin{aligned}
&\mathbb{E}\left[Y_i D_i D_{-i} \mid P_i(Z_i) = p_0, P_{-i}(Z_{-i}) = p_1\right] \\
&\quad + \mathbb{E}\left[Y_i D_i (1 - D_{-i}) \mid P_i(Z_i) = p_0, P_{-i}(Z_{-i}) = p_1\right] \\
=&\mathbb{E}\left[Y_i(1,1) \mathbb{1}\{V_i \leq h_i(Z_i, Z_{-i})\} \mathbb{1}\{V_{-i} \leq h_{-i}(Z_{-i}, Z_i)\} \mid h_i(Z_i, Z_{-i}) = p_0, h_{-i}(Z_{-i}, Z_i) = p_1\right] \\
&\quad + \mathbb{E}\left[Y_i(1,0) \mathbb{1}\{V_i \leq h_i(Z_i, Z_{-i})\} \mathbb{1}\{V_{-i} > h_{-i}(Z_{-i}, Z_i)\} \mid P_i(Z_i, Z_{-i}) = p_0, P_{-i}(Z_{-i}, Z_i) = p_1\right] \\
=&\mathbb{E}\left[Y_i(1,1) \mathbb{1}\{V_i \leq p_0\} \mathbb{1}\{V_{-i} \leq p_1\}\right] + \mathbb{E}\left[Y_i(1,0) \mathbb{1}\{V_i \leq p_0\} \mathbb{1}\{V_{-i} > p_1\}\right] \\
=&\int_0^{p_1} \int_0^{p_0} \mathbb{E}\left[Y_i(1,1) \mid V_i = v_0, V_{-i} = v_1\right] c_{V_i, V_{-i}}(v_0, v_1) dv_0 dv_1 \\
&\quad + \int_{p_1}^{1} \int_0^{p_0} \mathbb{E}\left[Y_i(1,0) \mid V_i = v_0, V_{-i} = v_1\right] c_{V_i, V_{-i}}(v_0, v_1) dv_0 dv_1.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[Y_i D_i \mid P_i(Z_i) = p_0\right] \\
=&\int_0^{1} \int_0^{p_1} \int_0^{p_0} \mathbb{E}\left[Y_i(1,1) \mid V_i = v_0, V_{-i} = v_1\right] c_{V_i, V_{-i}}(v_0, v_1) f_{P_{-i} \mid P_i = p_0}(p_1) dv_0 dv_1 dp_1 \\
&\quad + \int_0^{1} \int_{p_1}^{1} \int_0^{p_0} \mathbb{E}\left[Y_i(1,0) \mid V_i = v_0, V_{-i} = v_1\right] c_{V_i, V_{-i}}(v_0, v_1) f_{P_{-i} \mid P_i = p_0}(p_1) dv_0 dv_1 dp_1,
\end{aligned}
$$

where $f_{P_{-i} \mid P_i = p_0}(\cdot)$ denotes the conditional density of propensity score function P_{-i} given $P_i = p_0$. If Y_i is bounded, i.e., $|Y_i| < \infty$, then by Fubini's theorem, we can interchange the order of integration in the expression above, yielding the following result.

$$
\begin{aligned}
&\mathbb{E}\left[Y_i D_i \mid P_i(Z_i) = p_0\right] \\
=&\int_0^{p_0} \int_0^{1} \int_0^{p_1} \mathbb{E}\left[Y_i(1,1) \mid V_i = v_0, V_{-i} = v_1\right] c_{V_i, V_{-i}}(v_0, v_1) f_{P_{-i} \mid P_i = p_0}(p_1) dv_1 dp_1 dv_0 \\
&\quad + \int_0^{p_0} \int_0^{1} \int_{p_1}^{1} \mathbb{E}\left[Y_i(1,0) \mid V_i = v_0, V_{-i} = v_1\right] c_{V_i, V_{-i}}(v_0, v_1) f_{P_{-i} \mid P_i = p_0}(p_1) dv_1 dp_1 dv_0.
\end{aligned}
$$

Suppose the above function is continuously differentiable with respect to p_0. In that case,

we can apply the Leibniz integral rule to differentiate and obtain the following equalities,

$$
\begin{aligned}
&\frac{\partial}{\partial p_0}\int_0^{p_0}\int_0^1\int_0^{p_1}\mathbb{E}\left[Y_i(1,1)\mid V_i=v_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(v_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1dv_0\\
=&\int_0^1\int_0^{p_1}\mathbb{E}\left[Y_i(1,1)\mid V_i=p_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(p_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1\\
&+\int_0^{p_0}\int_0^1\int_0^{p_1}\mathbb{E}\left[Y_i(1,1)\mid V_i=v_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(v_0,v_1)\frac{\partial}{\partial p_0}f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1dv_0\\
\equiv&\int_0^1\int_0^{p_1}\mathbb{E}\left[Y_i(1,1)\mid V_i=p_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(p_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1+\mathcal{R}_{11},\\
&\frac{\partial}{\partial p_0}\int_0^{p_0}\int_0^1\int_{p_1}^1\mathbb{E}\left[Y_i(1,0)\mid V_i=v_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(v_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1dv_0\\
=&\int_0^1\int_{p_1}^1\mathbb{E}\left[Y_i(1,0)\mid V_i=p_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(p_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1\\
&+\int_0^{p_0}\int_0^1\int_{p_1}^1\mathbb{E}\left[Y_i(1,0)\mid V_i=v_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(v_0,v_1)\frac{\partial}{\partial p_0}f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1dv_0\\
\equiv&\int_0^1\int_{p_1}^1\mathbb{E}\left[Y_i(1,0)\mid V_i=p_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(p_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1+\mathcal{R}_{10}.
\end{aligned}
$$

The terms $\mathcal{R}_{11}$ and $\mathcal{R}_{10}$ are nonzero because $c_{V_i,V_{-i}}(v_0,v_1)\neq 0$, and $\partial f_{P_{-i}|P_i=p_0}(p_1)/\partial p_0\neq 0$ given that P_i and P_{-i} are dependent as they are both functions of (Z_i,Z_{-i}).

Similarly, under the assumption that Y_i is bounded and that $\mathbb{E}\left[Y_i(1-D_i)\mid P_i(Z_i)=p_0\right]$ is continuously differentiable with respect to p_0, we obtain the following equalities,

$$
\begin{aligned}
&\frac{\partial}{\partial p_0}\mathbb{E}\left[Y_i(1-D_i)\mid P_i(Z_i)=p_0\right]\\
=&-\int_0^1\int_0^{p_1}\mathbb{E}\left[Y_i(0,1)\mid V_i=p_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(p_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1+\mathcal{R}_{01}\\
&-\int_0^1\int_{p_1}^1\mathbb{E}\left[Y_i(0,0)\mid V_i=p_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(p_0,v_1)f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1+\mathcal{R}_{00},\\
&\mathcal{R}_{01}=\int_{p_0}^1\int_0^1\int_0^{p_1}\mathbb{E}\left[Y_i(0,1)\mid V_i=v_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(v_0,v_1)\frac{\partial}{\partial p_0}f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1dv_0,\\
&\mathcal{R}_{00}=\int_{p_0}^1\int_0^1\int_{p_1}^1\mathbb{E}\left[Y_i(0,0)\mid V_i=v_0,V_{-i}=v_1\right]c_{V_i,V_{-i}}(v_0,v_1)\frac{\partial}{\partial p_0}f_{P_{-i}|P_i=p_0}(p_1)dv_1dp_1dv_0.
\end{aligned}
$$

By taking the difference between $\partial\mathbb{E}[Y_iD_i\mid P_i(Z_i)=p_0]/\partial p_0$ and $-\partial\mathbb{E}[Y_i(1-D_i)\mid$

$P_i(Z_i) = p_0]/\partial p_0$, we would identify

$$\begin{aligned}
&\int_0^1 \int_0^{p_1} \mathbb{E}\left[Y_i(1,1) - Y_i(0,1) \mid V_i = p_0, V_{-i} = v_1\right] c_{V_i,V_{-i}}(p_0, v_1) f_{P_{-i}|P_i=p_0}(p_1) dv_1 dp_1 \\
+ &\int_0^1 \int_{p_1}^1 \mathbb{E}\left[Y_i(1,0) - Y_i(0,0) \mid V_i = p_0, V_{-i} = v_1\right] c_{V_i,V_{-i}}(p_0, v_1) f_{P_{-i}|P_i=p_0}(p_1) dv_1 dp_1 \\
&+ \mathcal{R}_{11} + \mathcal{R}_{10} + \mathcal{R}_{01} + \mathcal{R}_{00}.
\end{aligned}$$

The first two lines identify the marginal controlled direct effects for unit i, $\text{MCDE}_i(1; p_0, v_1)$ and $\text{MCDE}_i(0; p_0, v_1)$, averaged with respect to the copula density between unobserved heterogeneities, weighted by the peer $-i$'s propensity scores given $P_i = p_0$. The remaining terms, $\mathcal{R}_{11} + \mathcal{R}_{10} + \mathcal{R}_{01} + \mathcal{R}_{00}$, are nonzero and introduce bias when one attempts to interpret the conventional MTE identifier as a weighted average of MCDEs in the presence of spillover effects.

C.2 Comparing With Multivalued Treatments Literature

In this section, we compare the methods for identifying marginal spillover effects with the framework discussed in Lee and Salanié (2018). We focus on the group level and consider the treatment vector $\boldsymbol{D}_g \equiv (D_{0g}, D_{1g})$, assigned to each group g. The treatment vector $\boldsymbol{D}_g$ takes values from the set $\{(1,1),(1,0),(0,1),(0,0)\}$, consisting of four elements. Therefore, $\boldsymbol{D}_g$ can be regarded as multivalued treatments assigned at the group level. To align with the notation in Lee and Salanié (2018), we relabel the treatment vectors as follows: $(0,0) \equiv 0$, $(0,1) \equiv 1$, $(1,0) \equiv 2$, and $(1,1) \equiv 3$. Consequently, the treatment $\boldsymbol{D}_g$ takes values from $\{0,1,2,3\} \equiv \mathcal{D}$.

Each group is randomly assigned a continuous instrumental variable, with the instrument vector for group g denoted as $\boldsymbol{Z}_g \equiv (Z_{0g}, Z_{1g})$. Let $V_{ig} \in \mathbb{R}$ represent the unobserved characteristics of individual i in group g, and let $\boldsymbol{V}_g \equiv (V_{0g}, V_{1g})$ denote the vector of unobserved heterogeneity for both individuals in group g. For simplicity, we omit the group subscript g from the notation. The parameters of interest in Lee and Salanié (2018), $E(Y_k \mid \boldsymbol{V} = \boldsymbol{v}) - E(Y_{k'} \mid \boldsymbol{V} = \boldsymbol{v})$, where $k \neq k'$ and $k, k' \in \{0,1,2,3\}$, can be interpreted as marginal controlled spillover effects and marginal controlled direct effects within the spillover framework.

According to the model in Equation (1), we have

1. $\boldsymbol{D} = 0$ if and only if $V_0 > h_0(\boldsymbol{Z})$ and $V_1 > h_1(\boldsymbol{Z})$.
2. $\boldsymbol{D} = 1$ if and only if $V_0 > h_0(\boldsymbol{Z})$ and $V_1 \leq h_1(\boldsymbol{Z})$.

3. $\boldsymbol{D} = 2$ if and only if $V_0 \leq h_0(\boldsymbol{Z})$ and $V_1 > h_1(\boldsymbol{Z})$.

4. $\boldsymbol{D} = 3$ if and only if $V_0 \leq h_0(\boldsymbol{Z})$ and $V_1 \leq h_1(\boldsymbol{Z})$.

It is straightforward to see that the treatment $\boldsymbol{D}$ is measurable with respect to the σ-field generated by the events $\{V_i < Q_i(\boldsymbol{Z})\}$ for $i \in \{0, 1\}$, which aligns with the selection mechanism described in Assumption 2.1 of Lee and Salanié (2018). Furthermore, Theorem 3.1 in Lee (2009) is similar to our approach in identifying the joint density of unobserved heterogeneity $\boldsymbol{V}$ and the marginal treatment response functions, once the threshold functions $h_i(\boldsymbol{Z})$ (denoted as $Q_i(\boldsymbol{Z})$ in Lee and Salanié (2018)) are identified.

After reformulation, our marginal spillover model appears similar to the two-way flow model. However, in our setting, fewer assumptions are needed to point identify the thresholds. Specifically, we only require that the instruments $\boldsymbol{Z}$ are randomly assigned at the group level and do not directly influence the outcomes, without relying on the additional exclusion restrictions on treatments imposed in Assumption 4.1 of Lee and Salanié (2018). This is because we have more information on the observed treatments D, allowing us to identify the marginal distributions of V_0 and V_1 from the proportions of observed treatments:

$$\mathbb{P}(\boldsymbol{D} = 3 \mid \boldsymbol{Z} = z) + \mathbb{P}(\boldsymbol{D} = 2 \mid \boldsymbol{Z} = z) = \mathbb{P}(V_0 \leq h_0(z)),$$
$$\mathbb{P}(\boldsymbol{D} = 3 \mid \boldsymbol{Z} = z) + \mathbb{P}(\boldsymbol{D} = 1 \mid \boldsymbol{Z} = z) = \mathbb{P}(V_1 \leq h_1(z)).$$

Remark 5. (Monotonicity for each unit) When we focus on each individual unit i within a group, the monotonicity condition is satisfied. Specifically, consider any two vectors of instruments, denoted as (z_0, z_1) and $(\tilde{z}_0, \tilde{z}_1)$, where $P_i(z_0, z_1) \leq P_i(\tilde{z}_0, \tilde{z}_1)$. Under the monotonicity assumption, this ordering of the propensity scores implies that the corresponding potential treatments satisfy $D_i(z_0, z_1) \leq D_i(\tilde{z}_0, \tilde{z}_1)$. However, if we treat the entire group as a single decision-making unit and reformulate the setting into a multivalued treatment model, the monotonicity assumption may no longer hold. For instance, in the two-way flow model discussed in Lee and Salanié (2018), when the proportion of $D = 2$ changes, it is unclear whether the shift is driven by changes in h_0 or h_1. In other words, shifts in either h_0 or h_1 can induce changes in the proportion of receiving a given treatment, making it impossible to distinguish between the two effects. As a result, the monotonicity condition is violated, and the marginal distributions of V_0 and V_1 cannot be identified.

In the multivalued treatment setting discussed in Lee and Salanié (2018), if we have enough information on the observed treatment D that allows us to identify the threshold $h_j(\boldsymbol{Z})$ for each $j \in \{1, \cdots, J\}$, then we can point identify the joint density of $\boldsymbol{V}$ and the marginal treatment response functions in Theorem 3.1 of Lee and Salanié (2018). Specifically,

for each dimension $j \in \{1, \cdots, J\}$ of the unobservable $\boldsymbol{V}$, we need a subset of the support of the treatments, $\mathcal{K}_j \subseteq \mathcal{K}$, $\mathcal{K} = \{0, \cdots, K-1\}$, such that

$$\sum_{k \in \mathcal{K}_j} \mathbb{P}\left(D = k \mid \boldsymbol{Z}\right) = \mathbb{P}\left(V_j \leq h_j(\boldsymbol{Z}) \mid \boldsymbol{Z}\right) = \mathbb{P}\left(V_j \leq h_j(\boldsymbol{Z})\right) = h_j(\boldsymbol{Z}).$$

If $\mathcal{K}_j$ that satisfies the above conditions does not exist for some j, we can still partially identify the threshold $h_j(\boldsymbol{Z})$. $h_j(\boldsymbol{Z})$ can be partially identified as

$$\sum_{k \in \underline{\mathcal{K}}_j} \mathbb{P}\left(D = k \mid \boldsymbol{Z}\right) \leq \mathbb{P}\left(V_j \leq h_j(\boldsymbol{Z}) \mid \boldsymbol{Z}\right) = \mathbb{P}\left(V_j \leq h_j(\boldsymbol{Z})\right) = h_j(\boldsymbol{Z}) \leq \sum_{k \in \bar{\mathcal{K}}_j} \mathbb{P}\left(D = k \mid \boldsymbol{Z}\right),$$

where $\underline{\mathcal{K}}_j$ is the largest subset $\widetilde{\underline{\mathcal{K}}}_j$ of $\mathcal{K}$ such that

$$\cup_{k \in \widetilde{\underline{\mathcal{K}}}_j} d_k^{-1}\{D = k\} \subseteq \{V_j \leq h_j(\boldsymbol{Z})\},$$

and $\bar{\mathcal{K}}_j$ is the smallest subset $\widetilde{\bar{\mathcal{K}}_j}$ of $\mathcal{K}$ such that

$$\{V_j \leq h_j(\boldsymbol{Z})\} \subseteq \cup_{k \in \widetilde{\bar{\mathcal{K}}_j}} d_k^{-1}\{D = k\}.$$

For example, in the two-way flow model discussed in Lee and Salanié (2018), we can partially identify $h_1(\boldsymbol{Z})$ and $h_2(\boldsymbol{Z})$ as

$$\begin{aligned}
&\mathbb{P}\left(D = 0 \mid \boldsymbol{Z}\right) \leq h_1(\boldsymbol{Z}) \leq \mathbb{P}\left(D = 0 \mid \boldsymbol{Z}\right) + \mathbb{P}\left(D = 2 \mid \boldsymbol{Z}\right), \\
&\mathbb{P}\left(D = 0 \mid \boldsymbol{Z}\right) \leq h_2(\boldsymbol{Z}) \leq \mathbb{P}\left(D = 0 \mid \boldsymbol{Z}\right) + \mathbb{P}\left(D = 2 \mid \boldsymbol{Z}\right).
\end{aligned}$$

D Proof of asymptotic results in semiparametric estimation

D.1 Convergence rate of cross-derivative estimators: Proof

To simplify the notation, we introduce the following matrix definitions,

$$\begin{aligned}
\widehat{\mathbf{X}}_P &\equiv \begin{bmatrix} 1 & (\hat{P}_{01}-p_0) & \cdots & (\hat{P}_{01}-p_0)(\hat{P}_{11}-p_1) & \cdots & (\hat{P}_{11}-p_1)^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (\hat{P}_{0G}-p_0) & \cdots & (\hat{P}_{0G}-p_0)(\hat{P}_{1G}-p_1) & \cdots & (\hat{P}_{1G}-p_1)^3, \end{bmatrix} \\
\widehat{\mathbf{W}}_h &\equiv \operatorname{diag}\Big(K_h(\hat{P}_1-p), \cdots, K_h(\hat{P}_G-p)\Big) \\
\mathbf{D} &\equiv \big[D_{01}D_{11}, \cdots, D_{0G}D_{1G}\big]', \\
\widehat{\mathbf{U}}_{idd'} &= \big[\hat{U}_{idd'1}, \cdots, \hat{U}_{idd'G}\big]',
\end{aligned}$$

where $\widehat{\mathbf{X}}_P$ is a $G \times 10$ matrix of regressors used in the local polynomial regression, $\widehat{\mathbf{W}}_h$ is a $G \times G$ diagonal matrix consisting of the kernel functions, $\mathbf{D}$ is a $G \times 1$ vector entries $D_{0g}D_{1g}$, and $\widehat{\mathbf{U}}_{idd'}$ is a $G \times 1$ vector residuals $\hat{U}_{idd'g}$ as entries. Using this notation, we can express the estimators $\hat{b}_4(p_0, p_1)$ and $\hat{c}_4(d, d'; p_0, p_1)$ derived in Section 4.1.1 as

$$\begin{aligned}
\hat{b}_4(p_0,p_1) &= e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\mathbf{D}, \\
\hat{c}_4(d,d';p_0,p_1) &= e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{U}}_{idd'},
\end{aligned}$$

where e_5 is the 10×1 standard basis vector with a one in the fifth entry and zeros elsewhere.

We aim to characterize the asymptotic behavior of

$$e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\mathbf{D} - \frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g}=p_0, P_{1g}=p_1\big]. \tag{28}$$

We rewrite Equation (28) as

$$\begin{aligned}
&e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\big[\mathbf{D}-\mathbf{D}^*\big] + e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\mathbf{D}^* \\
&- e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\mathbf{D}_* \\
=&e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\big[\mathbf{D}-\mathbf{D}^*\big] + e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\big[\mathbf{D}^*-\widehat{\mathbf{X}}_P\mathbf{D}_*\big],
\end{aligned}$$

and define $\mathbf{D}^*, \mathbf{D}_*$ as

$$\begin{aligned}
\mathbf{D}^* \equiv & \big[\mathbb{E}[D_{01}D_{11} \mid P_{01}, P_{11}], \cdots, \mathbb{E}[D_{0G}D_{1G} \mid P_{0G}, P_{1G}]\big]', \\
\mathbf{D}_* \equiv & \Big[\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big], \cdots, \frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big], \\
& \cdots, \frac{\partial^3}{\partial p_1^3}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big]\Big]',
\end{aligned}$$

where $\mathbf{D}^*$ is a $G \times 1$ vector consisting of the conditional means $\mathbb{E}[D_{0g}D_{1g} \mid P_{0g}, P_{1g}], g = 1, \cdots, G$, and $\mathbf{D}_*$ is a 10×1 vector comprising the partial derivatives of $\mathbb{E}[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1]$ up to third order.

We apply Taylor series expansion to expand the conditional mean $\mathbb{E}[D_{0g}D_{1g} \mid P_{0g}, P_{1g}]$ around (p_0, p_1) and express it as

$$\begin{aligned}
& \mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big] + \frac{\partial}{\partial p_0}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big](P_{0g} - p_0) \\
& + \frac{\partial}{\partial p_1}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big](P_{1g} - p_1) + \cdots \\
& + \frac{\partial^3}{6\partial p_1^3}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big](P_{1g} - p_1)^3 + R_P(p_0, p_1) \\
= & \mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big] + \frac{\partial}{\partial p_0}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big]\big[(P_{0g} - \hat{P}_{0g}) + (\hat{P}_{0g} - p_0)\big] \\
& + \frac{\partial}{\partial p_1}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big]\big[(P_{1g} - \hat{P}_{1g}) + (\hat{P}_{1g} - p_1)\big] + \cdots \\
& + \frac{\partial^3}{6\partial p_1^3}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big]\big[(P_{1g} - \hat{P}_{1g}) + (\hat{P}_{1g} - p_1)\big]^3 + R_P(p_0, p_1)
\end{aligned}$$

where $R_P(p_0, p_1)$ represents the remainder terms from the Taylor expansion, and the last step is to decompose $(P_{ig} - p_i)$ as $[(P_{ig} - \hat{P}_{ig}) + (\hat{P}_{ig} - p_i)]$, $i \in \{0, 1\}$. Noting that the gth entry of $[\mathbf{D}^* - \widehat{\mathbf{X}}_P\mathbf{D}_*]$ equals to

$$\begin{aligned}
\big[\mathbf{D}^* - \widehat{\mathbf{X}}_P\mathbf{D}_*\big]_g = & \mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g}, P_{1g}\big] - \mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big] \\
& - \frac{\partial}{\partial p_0}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big](\hat{P}_{0g} - p_0) - \cdots \\
& - \frac{\partial^3}{\partial p_1^3}\mathbb{E}\big[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1\big](\hat{P}_{1g} - p_1)^3,
\end{aligned}$$

we can leverage the expansion results to show

$$\big[\mathbf{D}^* - \widehat{\mathbf{X}}_P\mathbf{D}_*\big]_g = O_P\Big[\max_{g:1\le g\le G}|\hat{P}_{0g} - P_{0g}| + \max_{g:1\le g\le G}|\hat{P}_{1g} - P_{1g}| + h_{G1}^4\Big].$$

The results in Section 4.1.2, combined with the boundedness of the kernel assumed in Assumption 9, imply that

$$\left\|\widehat{\mathbf{X}}_P - \mathbf{X}_P\right\| = o_P(1), \left\|\widehat{\mathbf{W}}_{h_{G1}} - \mathbf{W}_{h_{G1}}\right\| = o_P(1),$$

where we define $\mathbf{X}_P$ and $\mathbf{W}_{h_{G1}}$ as

$$\begin{aligned}
\mathbf{X}_P &\equiv \begin{bmatrix} 1 & (P_{01}-p_0) & \cdots & (P_{01}-p_0)(P_{11}-p_1) & \cdots & (P_{11}-p_1)^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (P_{0G}-p_0) & \cdots & (P_{0G}-p_0)(P_{1G}-p_1) & \cdots & (P_{1G}-p_1)^3, \end{bmatrix} \\
\mathbf{W}_h &\equiv \operatorname{diag}\Big(K_h(P_1-p), \cdots, K_h(P_G-p)\Big).
\end{aligned}$$

Since $\left(\mathbf{X}_P'\mathbf{W}_{h_{G1}}\mathbf{X}_P\right)^{-1}\mathbf{X}_P'\mathbf{W}_{h_{G1}} = O_P(1)$ by assumptions, it follows that

$$\begin{aligned}
&\left(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\right)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}} = O_P(1) \\
\Longrightarrow\ & e_5'\left(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\right)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\left[\mathbf{D}^* - \widehat{\mathbf{X}}_P\mathbf{D}_*\right] \\
&= O_P\Big[\max_{g:1\le g\le G}|\hat{P}_{0g}-P_{0g}| + \max_{g:1\le g\le G}|\hat{P}_{1g}-P_{1g}| + h_{G1}^4\Big].
\end{aligned}$$

Additionally,

$$e_5'\left(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\right)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\left[\mathbf{D}-\mathbf{D}^*\right] = e_5'\left(\mathbf{X}_P'\mathbf{W}_{h_{G1}}\mathbf{X}_P\right)^{-1}\mathbf{X}_P'\mathbf{W}_{h_{G1}}\left[\mathbf{D}-\mathbf{D}^*\right]\left[1+o_P(1)\right].$$

By applying the results from Masry (1996), the term $e_5'\left(\mathbf{X}_P'\mathbf{W}_{h_{G1}}\mathbf{X}_P\right)^{-1}\mathbf{X}_P'\mathbf{W}_{h_{G1}}\left[\mathbf{D}-\mathbf{D}^*\right]$ converges at the rate $O_P[(Gh_{G1}^6)^{-1/2}]$. the convergence rate of Equation (28) is given by

$$\begin{aligned}
&e_5'\left(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\right)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\mathbf{D} - \frac{\partial^2}{\partial p_0\partial p_1}\mathbb{E}\left[D_{0g}D_{1g} \mid P_{0g}=p_0, P_{1g}=p_1\right] \\
=&O_P\Big[(Gh_{G1}^6)^{-1/2} + \max_{g:1\le g\le G}|\hat{P}_{0g}-P_{0g}| + \max_{g:1\le g\le G}|\hat{P}_{1g}-P_{1g}| + h_{G1}^4\Big]
\end{aligned}$$

Furthermore, we can show that

$$\begin{aligned}
&e_5'\left(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\right)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{U}}_{idd'} - \frac{\partial^2}{\partial p_0\partial p_1}\mathbb{E}\left[U_{idd'g} \mid P_{0g}=p_0, P_{1g}=p_1\right] \\
=&e_5'\left(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\right)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\left[\mathbf{U}_{idd'} - X_i(\hat{\beta}_{dd'}-\beta_{dd'})\right] \\
&- \frac{\partial^2}{\partial p_0\partial p_1}\mathbb{E}\left[U_{idd'g} \mid P_{0g}=p_0, P_{1g}=p_1\right] \\
=&e_5'\left(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\right)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\mathbf{U}_{idd'} - \frac{\partial^2}{\partial p_0\partial p_1}\mathbb{E}\left[U_{idd'g} \mid P_{0g}=p_0, P_{1g}=p_1\right] + O_P(G^{-1/2}),
\end{aligned}$$

where the last equation holds by $\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}} = O_P(1)$, $\hat{\beta}_{dd'} - \beta_{dd'} = O_P(G^{-1/2})$ according to Theorem 3 in Carneiro and Lee (2009), and $\widehat{\mathbf{U}}_{idd'}, \mathbf{U}_{idd'}$ are defined as

$$\widehat{\mathbf{U}}_{idd'} = \big[\hat{U}_{idd'1}, \cdots, \hat{U}_{idd'G}\big]', \mathbf{U}_{idd'} = \big[U_{idd'1}, \cdots, U_{idd'G}\big]'.$$

Then, the convergence rate of $e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{U}}_{idd'}$ can be proven as

$$\begin{aligned}
&e_5'\big(\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{X}}_P\big)^{-1}\widehat{\mathbf{X}}_P'\widehat{\mathbf{W}}_{h_{G1}}\widehat{\mathbf{U}}_{idd'} - \frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[U_{idd'g} \mid P_{0g} = p_0, P_{1g} = p_1\big] \\
=&O_P\Big[(Gh_{G2}^6)^{-1/2} + \max_{g:1\le g\le G}|\hat{P}_{0g} - P_{0g}| + \max_{g:1\le g\le G}|\hat{P}_{1g} - P_{1g}| + h_{G2}^4\Big]
\end{aligned}$$

using an argument analogous to that used in the preceding analysis.

D.2 Asymptotic distribution of the marginal treatment response: Proof

Under Assumption 10, combined with conclusions in Masry (1996), we have

$$\begin{aligned}
&\big(Gh_{G2}^6\big)^{1/2}\Big\{\hat{c}_4(d,d';p_0,p_1) - c_4(d,d';p_0,p_1)\Big\} \xrightarrow{d} N\Big(0, \frac{\sigma^2(d,d';p_0,p_1)}{f(p_0,p_1)}\big(M^{-1}\Gamma M^{-1}\big)_{5,5}\Big), \\
&c_4(d,d';p_0,p_1) \equiv \frac{\partial^2}{\partial p_0 \partial p_1}\mathbb{E}\big[U_{idd'g} \mid P_{0g} = p_0, P_{1g} = p_1\big],
\end{aligned}$$

where $\sigma^2(d,d';p_0,p_1) = \mathrm{Var}(U_{idd'g} \mid P_{0g} = p_0, P_{1g} = p_1)$, and $f(p_0,p_1)$ denotes the density of (P_{0g}, P_{1g}) evaluated at the point (p_0,p_1). The matrices M and Γ are 10×10-matrices composed of multivariate moments of the kernel functions K and K^2, and are defined as

$$M = \begin{bmatrix}
\int u_0^0u_1^0K(u)d(u) & \int u_0^1u_1^0K(u)d(u) & \cdots & \int u_0^1u_1^2K(u)d(u) & \int u_0^0u_1^3K(u)d(u) \\
\int u_0^1u_1^0K(u)d(u) & \int u_0^2u_1^0K(u)d(u) & \cdots & \int u_0^2u_1^2K(u)d(u) & \int u_0^1u_1^3K(u)d(u) \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
\int u_0^0u_1^3K(u)d(u) & \int u_0^1u_1^2K(u)d(u) & \cdots & \int u_0^1u_1^5K(u)d(u) & \int u_0^0u_1^6K(u)d(u)
\end{bmatrix},$$

$$\Gamma = \begin{bmatrix}
\int u_0^0u_1^0K^2(u)d(u) & \int u_0^1u_1^0K^2(u)d(u) & \cdots & \int u_0^1u_1^2K^2(u)d(u) & \int u_0^0u_1^3K^2(u)d(u) \\
\int u_0^1u_1^0K^2(u)d(u) & \int u_0^2u_1^0K^2(u)d(u) & \cdots & \int u_0^2u_1^2K^2(u)d(u) & \int u_0^1u_1^3K^2(u)d(u) \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
\int u_0^0u_1^3K^2(u)d(u) & \int u_0^1u_1^2K^2(u)d(u) & \cdots & \int u_0^1u_1^5K^2(u)d(u) & \int u_0^0u_1^6K^2(u)d(u)
\end{bmatrix}.$$

Since $\partial^2\mathbb{E}[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1]/\partial p_0\partial p_1$ is bounded from above and away from zero by Assumption 9 and $\hat{b}_4(p_0,p_1) \xrightarrow{p} \partial^2\mathbb{E}[D_{0g}D_{1g} \mid P_{0g} = p_0, P_{1g} = p_1]/\partial p_0\partial p_1 \equiv b_4(p_0,p_1)$, it

follows that

$$\begin{aligned}&\left(Gh_{G2}^{6}\right)^{1/2}\left\{\hat{c}_4(d,d';p_0,p_1)-c_4(d,d';p_0,p_1)\right\}\frac{1}{\hat{b}_4(p_0,p_1)}\\ \xrightarrow{d}&N\left(0,\frac{\sigma^2(d,d';p_0,p_1)}{\left(b_4(p_0,p_1)\right)^2 f(p_0,p_1)}\left(M^{-1}\Gamma M^{-1}\right)_{5,5}\right)\end{aligned}$$

We can rewrite $\left(Gh_{G2}^{6}\right)^{1/2}\left\{\hat{c}_4(d,d';p_0,p_1)-c_4(d,d';p_0,p_1)\right\}/\hat{b}_4(p_0,p_1)$ as

$$\begin{aligned}&\left(Gh_{G2}^{6}\right)^{1/2}\left\{\hat{c}_4(d,d';p_0,p_1)-c_4(d,d';p_0,p_1)\right\}\frac{1}{\hat{b}_4(p_0,p_1)}\\ =&\left(Gh_{G2}^{6}\right)^{1/2}\left\{\frac{\hat{c}_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}+\frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}-\frac{c_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)}\right\}\\ =&\left(Gh_{G2}^{6}\right)^{1/2}\left\{\frac{\hat{c}_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}\right\}\\ &+\left(Gh_{G2}^{6}\right)^{1/2}\left\{\frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}-\frac{c_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)}\right\}.\end{aligned}$$

By applying Assumption 10 along with the results in Section 4.1.2, we obtain

$$\frac{1}{\hat{b}_4(p_0,p_1)}-\frac{1}{b_4(p_0,p_1)}=O_P\left[(Gh_{G1}^{6})^{-1/2}\right],$$

which implies that

$$\left(Gh_{G2}^{6}\right)^{1/2}\left\{\frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}-\frac{c_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)}\right\}=o_P(1)$$

under the condition $h_{G2}=o(h_{G1})$. Therefore,

$$\begin{aligned}&\left(Gh_{G2}^{6}\right)^{1/2}\left\{\frac{\hat{c}_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}\right\}\\ =&\left(Gh_{G2}^{6}\right)^{1/2}\left\{\hat{c}_4(d,d';p_0,p_1)-c_4(d,d';p_0,p_1)\right\}\frac{1}{\hat{b}_4(p_0,p_1)}+o_P(1),\end{aligned}$$

and the asymptotic distribution of estimated marginal treatment response function without the covariate effect can be characterized as

$$\begin{aligned}&\left(Gh_{G2}^6\right)^{1/2}\left\{\frac{\hat{c}_4(d,d';p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,d';p_0,p_1)}{b_4(p_0,p_1)}\right\}\\ \xrightarrow{d}&N\left(0,\frac{\sigma^2(d,d';p_0,p_1)}{\left(b_4(p_0,p_1)\right)^2 f(p_0,p_1)}\left(M^{-1}\Gamma M^{-1}\right)_{5,5}\right).\end{aligned}$$

Finally, under the assumptions that $\hat{c}_4(d,d';p_0,p_1)/\hat{b}_4(p_0,p_1)-c_4(d,d';p_0,p_1)/b_4(p_0,p_1)$ are asymptotically independent across different values of $d,d'\in\{0,1\}$, we can derive the asymptotic distributions of $\widehat{\text{MCSE}}(\mathbf{x},d;p_0,p_1)$ as

$$\begin{aligned}&\left(Gh_{G2}^6\right)^{1/2}\left\{\widehat{\text{MCSE}}(\mathbf{x},d;p_0,p_1)-\text{MCSE}(\mathbf{x},d;p_0,p_1)\right\}\\ =&\left(Gh_{G2}^6\right)^{1/2}\left[\mathbf{x}'\left(\hat{\beta}_{d1}-\beta_{d1}\right)+\mathbf{x}'\left(\hat{\beta}_{d0}-\beta_{d0}\right)\right]\\ &+\left(Gh_{G2}^6\right)^{1/2}\left[\left(\frac{\hat{c}_4(d,1;p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,1;p_0,p_1)}{b_4(p_0,p_1)}\right)+\left(\frac{\hat{c}_4(d,0;p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,0;p_0,p_1)}{b_4(p_0,p_1)}\right)\right]\\ =&o_P(1)+\left(Gh_{G2}^6\right)^{1/2}\left[\left(\frac{\hat{c}_4(d,1;p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,1;p_0,p_1)}{b_4(p_0,p_1)}\right)+\left(\frac{\hat{c}_4(d,0;p_0,p_1)}{\hat{b}_4(p_0,p_1)}-\frac{c_4(d,0;p_0,p_1)}{b_4(p_0,p_1)}\right)\right]\\ &\xrightarrow{d}N\left(0,\frac{\sigma^2(d,1;p_0,p_1)+\sigma^2(d,0;p_0,p_1)}{\left(b_4(p_0,p_1)\right)^2 f(p_0,p_1)}\left(M^{-1}\Gamma M^{-1}\right)_{5,5}\right),\end{aligned}$$

where the second equality holds because $\hat{\beta}_{dd'}-\beta_{dd'}=O_P(G^{-1/2})$ applying Theorem 3 in Carneiro and Lee (2009). Similarly, the asymptotic distribution of $\widehat{\text{MCDE}}(\mathbf{x},d;p_0,p_1)$ can be derived as

$$\begin{aligned}&\left(Gh_{G2}^6\right)^{1/2}\left\{\widehat{\text{MCSE}}(\mathbf{x},d;p_0,p_1)-\text{MCSE}(\mathbf{x},d;p_0,p_1)\right\}\\ \xrightarrow{d}&N\left(0,\frac{\sigma^2(1,d;p_0,p_1)+\sigma^2(0,d;p_0,p_1)}{\left(b_4(p_0,p_1)\right)^2 f(p_0,p_1)}\left(M^{-1}\Gamma M^{-1}\right)_{5,5}\right).\end{aligned}$$

E Proof of asymptotic results in parametric estimation

E.1 Consistency of the first-stage parameter

For each $i\in\{0,1\}$, the function $l(\theta_i;d,\mathbf{w})$ is continuous in θ_i for all $d\in\{0,1\}$ and $\mathbf{w}\in\mathcal{W}$. Additionally, the parameter space Θ_i is compact and $\mathbb{E}[\sup_{\theta_i\in\Theta_i}|l(\theta_i;D_{ig},W_g)|]<\infty$ given

Assumption 12, applying the uniform law of large numbers, we have

$$\sup_{\theta_i \in \Theta_i} \left| \frac{1}{G} \sum_{g=1}^{G} l(\theta_i; D_{ig}, W_g) - \mathbb{E}\big[l(\theta_i; D_{ig}, W_g)\big] \right| \xrightarrow{a.s.} 0,$$

as $G \to \infty$. Define $Q(\theta_i) = \mathbb{E}\big[l(\theta_i; D_{ig}, W_g)\big]$ and $Q_G(\theta_i) = \sum_{g=1}^{G} l(\theta_i; D_{ig}, W_g)/G$. We can derive

$$\begin{aligned} 0 \leq Q(\theta_{i0}) - Q(\hat{\theta}_i) &= Q_G(\hat{\theta}_i) - Q(\hat{\theta}_i) + Q(\theta_{i0}) - Q_G(\hat{\theta}_i) \\ &\leq \sup_{\theta_i \in \Theta_i} \big|Q_G(\theta_i) - Q(\theta_i)\big| + Q(\theta_{i0}) - Q_G(\theta_{i0}) \\ &\leq 2 \sup_{\theta_i \in \Theta_i} |Q_G(\theta_i) - Q(\theta_i)| \\ &\xrightarrow{a.s.} 0, \end{aligned}$$

as $G \to \infty$, where the second line holds because $\hat{\theta}_i$ maximizes the function $Q_G(\theta_i)$. Since θ_i is the unique maximizer of $Q(\theta_i)$ and Θ_i is compact based on Assumption 12, $Q_G(\hat{\theta}_i) \xrightarrow{a.s.} Q(\theta_{i0})$ implies that $\hat{\theta}_i \xrightarrow{a.s.} \theta_{i0}$ as $G \to \infty$.

E.2 Consistency of the second-stage correlation

We use $Q_G(\rho)$ and $\widehat{Q}_G(\rho)$ to define

$$\begin{aligned} Q_G(\rho) &= \frac{1}{G} \sum_{g=1}^{G} \tilde{l}\big(\rho; D_g, P_g\big), \\ \widehat{Q}_G(\rho) &= \frac{1}{G} \sum_{g=1}^{G} \tilde{l}\big(\rho; D_g, \widehat{P}_g\big), \end{aligned}$$

where $\widehat{P}_g = (\widehat{P}_{0g}, \widehat{P}_{1g})$ is the vector of propensity scores estimated in the first stage. Then, we can write

$$\begin{aligned} \left| \widehat{Q}_G(\rho) - \mathbb{E}\big[\tilde{l}(\rho; D_g, P_g)\big] \right| \leq & \left| \widehat{Q}_G(\rho) - Q_G(\rho) \right| \\ & + \left| Q_G(\rho) - \mathbb{E}\big[\tilde{l}(\rho; D_g, P_g)\big] \right|. \end{aligned}$$

Since $\tilde{l}(\rho; d, p)$ is continuous in ρ for all $d \in \{0,1\}^2$ and $p \in (0,1)^2$, ρ lies in a compact interval, and $\mathbb{E}\big[\sup_{\rho \in [-\varepsilon, \varepsilon]} |l(\rho, \theta; D_g, W_g)|\big] < \infty$ under Assumption 13, the law of large numbers implies that

$$\sup_{\rho \in [-\varepsilon, \varepsilon]} \left| Q_G(\rho) - \mathbb{E}\big[\tilde{l}(\rho; D_g, P_g)\big] \right| \xrightarrow{a.s.} 0.$$

Assuption 13 also assumes that there exists a function $L(\cdot)$ such that for all $\rho \in [-\varepsilon, \varepsilon]$,

$$\left|\tilde{l}(\rho; D_g, \widehat{P}_g) - \tilde{l}(\rho; D_g, P_g)\right| \leq L(D_g)\left|\left|\widehat{P}_g - P_g\right|\right|.$$

Since $|L(D_g)| < \infty$ almost surely and $\left|\left|\widehat{P}_g - P_g\right|\right| \xrightarrow{a.s.} 0$ by Lemma 4, it follows that $\sup_{\rho\in[\varepsilon,\varepsilon]} \left|\tilde{l}(\rho; D_g, \widehat{P}_g) - \tilde{l}(\rho; D_g, P_g)\right| \xrightarrow{a.s.} 0$, which further implies

$$\sup_{\rho\in[\varepsilon,\varepsilon]} \left|\widehat{Q}_G(\rho) - Q_G(\rho)\right| \xrightarrow{a.s.} 0.$$

Therefore, we have

$$\sup_{\rho\in[\varepsilon,\varepsilon]} \left|\widehat{Q}_G(\rho) - \mathbb{E}\big[\tilde{l}(\rho; D_g, P_g)\big]\right| \xrightarrow{a.s.} 0.$$

Since ρ_0 is the unique maximizer of $\mathbb{E}\big[\tilde{l}(\rho; D_g, P_g)\big]$ and lies within a compact interval, by the similar arguments in the proof of Lemma 4, we can derive $\hat{\rho} \xrightarrow{a.s.} \rho_0$ as $G \to \infty$.

E.3 Consistency of the marginal treatment response coefficients

Based on the identification results the third specification in Assumption 11, for each $i \in \{0, 1\}$ and $g \in \{1, \cdots, G\}$,

$$Y_{ig}\mathbb{1}\big\{D_{0g} = d\big\}\mathbb{1}\big\{D_{1g} = d'\big\} = X_{Pdd'_g}\big(\alpha'_{idd'}, \beta'_{idd'}\big)' + \varepsilon_{idd'g},$$

where the error term $\varepsilon_{idd'g}$ satisfies $\mathbb{E}[\varepsilon_{idd'g} \mid X_{Pdd'_g}] = 0$. The vector of coefficients $(\alpha'_{idd'}, \beta'_{idd'})'$ is estimated by

$$\big(\hat{\alpha}'_{idd'}, \hat{\beta}'_{idd'}\big)' = \big(\widehat{X}'_{Pdd'}\widehat{X}_{Pdd'}\big)^{-1}\widehat{X}'_{Pdd'}\widetilde{Y}_{idd'},$$

where $\widetilde{Y}_{idd'}$ is defined as a $G \times 1$ vector with the g-th element as $Y_{ig}\mathbb{1}\big\{D_{0g} = d\big\}\mathbb{1}\big\{D_{1g} = d'\big\}$, and $\widehat{X}_{Pdd'}$ is obtained by substituting $\widehat{P}_{0g}$, $\widehat{P}_{1g}$, and $\hat{\rho}$ for the true values into $X_{Pdd'}$. Then, we can write the estimated coefficients as

$$\begin{aligned}\big(\hat{\alpha}'_{idd'}, \hat{\beta}'_{idd'}\big)' =& \big(\widehat{X}'_{Pdd'}\widehat{X}_{Pdd'}\big)^{-1}\widehat{X}'_{Pdd'}\big(X_{Pdd'}\big(\alpha'_{idd'}, \beta'_{idd'}\big)' + \varepsilon_{idd'}\big) \\ =& \big(\widehat{X}'_{Pdd'}\widehat{X}_{Pdd'}\big)^{-1}\widehat{X}'_{Pdd'}X_{Pdd'}\big(\alpha'_{idd'}, \beta'_{idd'}\big)' + \big(\widehat{X}'_{Pdd'}\widehat{X}_{Pdd'}\big)^{-1}\widehat{X}'_{Pdd'}\varepsilon_{idd'}.\end{aligned}$$

Let $\widehat{X}_{Pdd'} = X_{Pdd'} + \Delta_G$, where Δ_G is defined as a $G \times K$ matrix such that $\Delta_G =$

$\widehat{X}_{Pdd'} - X_{Pdd'}$. Then, we have

$$\begin{aligned}
\frac{1}{G}\widehat{X}'_{Pdd'}\widehat{X}_{Pdd'} =& \frac{1}{G}\big(X_{Pdd'} + \Delta_G\big)'\big(X_{Pdd'} + \Delta_G\big) \\
=& \frac{1}{G}X'_{Pdd'}X_{Pdd'} + \frac{1}{G}X'_{Pdd'}\Delta_G + \frac{1}{G}\Delta'_G X_{Pdd'} + \frac{1}{G}\Delta'_G\Delta_G, \\
\frac{1}{G}\widehat{X}'_{Pdd'}X_{Pdd'} =& \frac{1}{G}\big(X_{Pdd'} + \Delta_G\big)'X_{Pdd'} \\
=& \frac{1}{G}X'_{Pdd'}X_{Pdd'} + \frac{1}{G}\Delta'_G X_{Pdd'}
\end{aligned}$$

Applying the Cauchy-Schwarz inequalities, we obtain

$$\frac{1}{G}\big\|X_{Pdd'}\Delta_G\big\|_F \leq \sqrt{\frac{\|X_{Pdd'}\|_F^2}{G}} \cdot \sqrt{\frac{\|\Delta_G\|_F^2}{G}} \overset{a.s.}{\to} 0,$$

since $\|\Delta_G\|_F^2/G \overset{a.s.}{\to} 0$, and $\|X_{Pdd'}\|_F^2/G$ is bounded almost surely by $\mathbb{E}[X'_{Pdd'}, X_{Pdd'}]$ is non-singular. Therefore, we should have

$$\begin{aligned}
\frac{1}{G}\widehat{X}'_{Pdd'}\widehat{X}_{Pdd'} &= \frac{1}{G}X'_{Pdd'}X_{Pdd'} + o_{a.s.}(1) \overset{a.s.}{\to} \mathbb{E}\big[X'_{Pdd'g}X_{Pdd'g}\big], \\
\frac{1}{G}\widehat{X}'_{Pdd'}X_{Pdd'} &= \frac{1}{G}X'_{Pdd'}X_{Pdd'} + o_{a.s.}(1) \overset{a.s.}{\to} \mathbb{E}\big[X'_{Pdd'g}X_{Pdd'g}\big],
\end{aligned}$$

and then

$$\left(\frac{1}{G}\widehat{X}'_{Pdd'}\widehat{X}_{Pdd'}\right)^{-1} \overset{a.s.}{\to} \left(\mathbb{E}\big[X'_{Pdd'g}X_{Pdd'g}\big]\right)^{-1}$$

by the continuous mapping theorem and the nonsingularity condition.

We can also express the term $\widehat{X}'_{Pdd'}\varepsilon_{idd'}/G$ as

$$\frac{\widehat{X}'_{Pdd'}\varepsilon_{idd'}}{G} = \frac{1}{G}\big(X_{Pdd'} + \Delta_G\big)'\varepsilon_{idd'} = \frac{1}{G}X'_{Pdd'}\varepsilon_{idd'} + \frac{1}{G}\Delta'_G\varepsilon_{idd'}.$$

The first term $X'_{Pdd'}\varepsilon_{idd'}/G \overset{a.s.}{\to} \mathbb{E}[X'_{Pdd'_g}\varepsilon_{idd'g}] = 0$ as $\mathbb{E}[\varepsilon_{idd'g} \mid X_{Pdd'_g}] = 0$. Applying the Cauchy-Schwarz, the second term becomes

$$\left\|\frac{1}{G}\Delta'_G\varepsilon_{idd'}\right\| \leq \sqrt{\frac{1}{G}\big\|\Delta_G\big\|_F^2} \cdot \sqrt{\frac{1}{G}\|\varepsilon_{idd'}\|^2} \overset{a.s.}{\to} 0,$$

since $\|\Delta_G\|_F^2/G \overset{a.s.}{\to} 0$, and $\|\varepsilon_{idd'}\|^2/G$ is bounded almost surely by $\mathrm{Var}(\varepsilon_{idd'g}) = \sigma_{idd'g} < \infty$. Thus,

$$\frac{\widehat{X}'_{Pdd'}\varepsilon_{idd'}}{G} \overset{a.s.}{\to} 0.$$

Combining the above results, we have

$$\left(\hat{\alpha}'_{idd'}, \hat{\beta}'_{idd'}\right)' \xrightarrow{a.s.} \left(\mathbb{E}\left[X'_{Pdd'g} X_{Pdd'g}\right]\right)^{-1} \mathbb{E}\left[X'_{Pdd'g} X_{Pdd'g}\right] \left(\alpha'_{idd'}, \beta'_{idd'}\right)' = \left(\alpha'_{idd'}, \beta'_{idd'}\right)'.$$