Meng Hsuan "Rex" Hsieh
701 Tappan Ave, R6349
Ann Arbor, Michigan, 48109
rexhsieh@umich.edu
(646) 705-4791
September 15, 2026

Dear Workshop Organizers,

I am writing to apply for consideration for the workshop on network and development economics. The next few pages of this packet contains, in order,

(1) my most up-to-date curriculum vitae with courses completed;
(2) statement of purpose/motivation; and
(3) my job market paper (/working paper), "Mediated Interference".

Below, please find the name and contact information of faculty member who has agreed to serve as my reference, should I be shortlisted for participation:

Name: Kaspar Wüthrich
Email: kasparwu@umich.edu

Thank you in advance for considering my application.

Sincerely,

Meng Hsuan "Rex" Hsieh

MENG HSUAN "REX" HSIEH

701 Tappan Ave, R6349
Ann Arbor, MI 48109

PHONE (646) 705-4791
EMAIL rexhsieh@umich.edu

EDUCATION

Ph.D. Candidate in Business Economics and Public Policy 2021—Present
Ross School of Business, University of Michigan, Ann Arbor, MI

B.A. in Economics and Mathematics, *magna cum laude* 2015—2018
New York University, New York, NY

RESEARCH

Publications

1. How Taxes Affect Growth: Evidence from Cross-Country Panel Data (w/ K. Kawano, J. Olson, J. Slemrod; 2026). ***International Tax and Public Finance***, 33(2), 375-444. [doi]
2. Tangential Wasserstein Projections (w/ F. Gunsilius & M. Lee; 2024). ***Journal of Machine Learning Research***, 25(69), 1-41. [doi]
3. Revisiting the Analysis of Matched-Pair and Stratified Experiments in the Presence of Attrition (w/ Y. Bai, J. Liu & M. Tabord-Meehan; 2024). ***Journal of Applied Econometrics***, 39(2), 256-268. [doi]

Working Papers

4. Mediated Interference (2026). [JMP]
5. Have the Growth Effects of Tax Policy Changed Over Time? (w/ W. Gale, L. Kawano, A. Saini, S. Thorpe, S. Wyse; 2026). *Submitted*. [WP]
6. External Validity in Empirical Public Finance: With an Application to Taxable Income Elasticities (w/ G. Sanz-Maldonado, J. Slemrod; 2026). *Revision at* ***Journal of Public Economics***. [WP]
7. Pricing AI Services with Data-Use Rights (w/ F. Zhou, Z. Hu; 2026). *Submitted*. [WP]
8. Stacked Triple Differences (2026). *Submitted*. [WP]
9. The Rise of Within-Region Inequality: Analysis Using a Shift-Share-Like Instrument (2024). [WP]

TEACHING EXPERIENCE

Instructor of Record, University of Michigan, Ross School of Business
- → BBA, "Applied Microeconomics" (BE 300). [Teaching Evaluation] Fall 2023
- → Ross Summer Connection, "Introduction to Microeconomics". July 2025, 2026

Co-Instruction Specialist, University of Chicago, Harris School of Public Policy
- → "Economic Statistics, Modeling, and Forecasting". November 2025
- → "Applied Econometrics". October–November 2024

Teaching Assistant and Grader, University of Michigan, Ross School of Business
- → BBA, "Macroeconomic Data and Analysis" (BE 470). Winter 2026
- → BBA, "Senior Thesis" (BA 480). AY2024/2025, AY2026/2027
- → Executive MBA, "Economics of Business" (EMBA 604). Fall 2021, 2022, 2023, 2024, 2025, 2026
- → Executive MBA, "Global Finance" (EMBA 604). Spring 2026
- → Executive MBA, "Macroeconomics" (EMBA 637). Summer 2022, 2024
- → Full-Time MBA, "Growth and Stabilization in the Macro Economy" (BE 562). Winter 2026
- → Full-Time MBA, "Competitive Tactics and Policies" (BE 580). Fall 2024, 2025

→ Full-Time MBA, "Economic Decisions and Games" (BE 582). Fall 2022, 2025, 2026
→ MBAN, "Decision Strategies" (MBAN 503). Fall 2024, 2025, 2026
→ Online MBA, "Competitive Tactics and Policies" (BE 780). Summer 2024, 2025, 2026
→ Weekend MBA, "Applied Microeconomics" (WMBA 502). Summer 2022, 2023, 2024, 2025

Academic Awards

Best Ph.D. Student Paper Award, IAAE Conference 2026
Thomas William Leabo Fellowship for Academic and Teaching Excellence (Top 1-2 Students Per Academic Year), UM Ross School of Business 2024
Katherine Terrell Prize for Best Student Paper (Runner Up), UM Ross School of Business 2024
Emeritus Award for Best Student Paper as Second-Year Ph.D. Student, UM Ross School of Business 2024
Spivey Hall Fellow, UM Ross School of Business 2023
Katherine Terrell Prize for Best Student Paper (Winner), UM Ross School of Business 2023

Presentations

2026: IAAE, MEG, NTA
2024: Michigan Student Symposium for Interdisciplinary Statistical Sciences

Reviewer

International Tax and Public Finance, National Tax Journal, Journal of Economic Growth, Journal of Macroeconomics, Conference on Learning Theory

Skills

Programming Languages: (Advanced) R; (Intermediate) Stata (Mata), Python, SAS, SQL, C.
Other Technical Skills: (Advanced) LaTeX, ArcGIS, (Intermediate) Bash
Statistical Packages Developed: stc (Stata package)

Languages

Chinese/Mandarin (Native), **English** (Near-Native)

Community Involvement

Microeconomic Theory Reading Group 2025—Present
Econometrics Methods Reading Group 2022—Present
Ross Ph.D. Forum 2021—2024
First-Year Representative (2021-2022); Facilities and Communications Chair (2022-2023); President (2023-2024)

Other Professional Experience

Analyst, NERA Economic Consulting 2018—2021

Courses Taken

Ph.D. field sequences: Econometrics; Public Finance
Statistics: Computation and Optimization Methods in Statistics
Mathematics: Differential geometry (field)

Last updated: September 15, 2026

Meng Hsuan "Rex" Hsieh
701 Tappan Ave, R6349
Ann Arbor, Michigan, 48109
rexhsieh@umich.edu
(646) 705-4791
September 15, 2026

Statement of Motivation

My research interests lie in network econometrics, currently with a focus on identifying treatment effects in populations connected by a social network and when treatment reshapes social relationships. Agricultural training and other social programs can change both participants' behaviour and social connections through which information and resources flow. I am interested in distinguishing these channels because they have different implications for program design and expansion.

My current paper, "Mediated Interference", attached below, develops a framework for decomposing treatment effects in randomized saturation experiments into direct, network-mediated (treatment effect through network changes due to treatment intervention), treatment–network interaction, and saturation-spillover (which is reduced-form) components. The central challenge is that randomizing treatment does not randomize the relationships that subsequently form: individuals who acquire new connections may differ in ways that also affect their outcomes. I pinpoint the additional assumptions needed to attribute treatment effects to network changes, draw connections to the causal mediation literature, and develop estimation and inference methods that account for dependence across connected individuals. My broader goal of this paper is to make analyses of network mechanisms more useful to empirical researchers by clarifying what their experimental designs and network measurements can establish under reasonable assumptions and arguments.

In my paper, I apply this framework to an agronomy-training experiment among coffee farmers in Rwanda. I find suggestive evidence of negative network-mediated effect and a positive interaction effect with own training. These findings suggest that social connections can shape access to scarce resources as well as transmit information, with different effects for the trained and untrained farmers.

The workshop's combination of lectures and work-in-progress presentations would help me refine my paper, and develop further research questions grounded in the practical challenges of empirical network research. I would particularly value feedback on the plausibility of my identifying assumptions, as well as pinpointing possible ways to extend this analysis to quasi-experimental and observational settings. As such, I would greatly appreciate being considered for participating in this workshop. Thank you in advance for your consideration.

Sincerely,

Meng Hsuan "Rex" Hsieh

MEDIATED INTERFERENCE

MENG HSUAN HSIEH

ABSTRACT. Randomized saturation experiments are a workhorse design for identifying spillovers, but randomization does not reveal whether these spillovers operate through social connections—especially relevant when treatment changes the network structure. In the context of randomized saturation designs, I decompose a reduced-form causal contrast into direct, network-mediated, treatment–network interaction, and saturation-spillover components without imposing functional form assumptions on the potential outcomes. To identify and give causal interpretation to the network components, I introduce an identifying restriction called baseline network sufficiency: conditional on pretreatment network information and baseline covariates, the unobserved determinants of post-treatment network connections are unrelated to outcome shocks at fixed treatment and exposure. Together with experimental design and regularity conditions, this restriction identifies natural and interventional effects through a common mediation formula without needing to specify and estimate a network formation model. I estimate these effects with doubly-debiased machine learning estimators, allowing for dependent data and sidestepping the need for sample splitting provided that a stability condition holds. I develop valid large-sample inference and consistent network covariance estimation when a fixed number of villages grows. Reanalyzing an agronomy-training experiment in Rwanda, I find evidence for a negative network-mediated treatment effect for input use among untreated farmers and a positive interaction between own training and network exposure among them. These results complement the original study's evidence of limited knowledge diffusion and adverse saturation spillovers.

Keywords: Network interference, causal mediation, endogenous network formation, randomized saturation designs, double machine learning, interventional effects, natural effects

Preliminary. Please do not circulate.

Date: September 15, 2026.

All errors are mine. This work came into fruition with enormous support from my advisors, Andreas Hagemann, Florian Gunsilius, Eva Janssens, Joel Slemrod, and Kaspar Wüthrich. I thank Alessandro Casini, Ben Golub, Alejandro Sanchez-Beccera, and participants at IAAE 2026 in Lisbon, Portugal for insightful discussions and feedback.

1. INTRODUCTION

Empirical researchers use randomized saturation experiments to measure how the effects of a program depend on the proportion of a population assigned to treatment. The design first assigns treatment intensities to villages, schools, or other groups, and then assigns treatment to individuals within each group. Comparing untreated individuals across intensities identifies spillover effects; comparing treated and untreated individuals provides additional contrasts that describe the consequences of expanding a program (Hudgens and Halloran, 2008, Baird et al., 2018). These comparisons remain informative when the program changes the social connections through which individuals interact. They do not, however, reveal how much of an effect operates through those changes.

This distinction matters for programs delivered through social groups. An agricultural training program may improve a farmer's knowledge, create new relationships with other participants, and change access to shared labor or inputs. A farmer who receives training may also benefit less from an additional trained contact than a farmer who does not. These responses can reinforce or offset one another. Network measurements help describe them, but a relationship between outcomes and the number of trained friends need not identify a network-mediated effect. Farmers who form productive connections may differ in unobserved ways from those who do not, and assignment can itself affect which connections form. The post-treatment network is then an endogenous mediator, even though treatment was randomized.

In this paper, I develop identification and estimation theory for a decomposition of randomized saturation-policy effects into direct, network-mediated, treatment–network interaction, and spillover components, allowing nonlinear outcome responses and a fixed number of villages with growing village sizes. I propose *mediated interference* (MInt), a framework for tracing how treatment changes outcomes through own assignment, the relationships it creates or dissolves, and the treatment intensity among peers. The analysis separates the information supplied by the experiment from the additional restrictions needed to attribute its effects to changes in the network. It also distinguishes identification under repeated assignment from estimation based on a single realized allocation of village saturations. Both distinctions are necessary for interpreting a decomposition of an experimental treatment effect.

The empirical setting makes these issues concrete. Duflo et al. (2026) study an agronomy-training program for coffee farmers in Rwanda. The experiment assigned 27 villages to treatment concentrations of 25, 50, or 75 percent and randomized training offers among 1,594 applicant households. Friendship networks were measured before and after the program. The authors find gains in knowledge and self-reported adoption, new ties primarily among farmers trained together, and little evidence of knowledge diffusion to untreated farmers through baseline friendships. They also report suggestive evidence of adverse spillovers in villages with higher training concentrations, consistent with competition for shared inputs. Thus, an individual treatment contrast and a comparison across treatment intensities can have different economic interpretations. The application motivates studying the network response alongside the outcome response.

To describe the decomposition, let $R_s(p)$ denote the mean outcome under own assignment $p \in \{0, 1\}$ and village saturation s, averaging over the remaining assignments under the specified experimental policy and over a common target population. For high and low saturations s_H and s_L, the contrast between a treated individual under the high-saturation regime and an untreated individual under the low-saturation regime satisfies

$$\mathrm{TE} := R_{s_H}(1) - R_{s_L}(0) = D + N + I + S\,.$$

Here $S = R_{s_H}(0) - R_{s_L}(0)$ is the spillover contrast among untreated individuals. The direct effect D measures the effect of changing own treatment while fixing network exposure at the value it would take if the individual were assigned to control. The network-mediated effect N measures the effect of changing an untreated individual's network exposure from observed to that as-if they are treated. This comparison holds peers' assignments fixed, so that changes in exposure reflect changes in links induced by counterfactual own-treatment change. The interaction I measures the difference between this network-mediated effect for a treated individual and the corresponding effect for an untreated individual. The decomposition follows by taking differences of potential outcomes and remains valid without functional-form restrictions.

The first part of the analysis provides identification assumptions for the parts of this decomposition. To do so, I summarize the network environment by a specified exposure mapping, such as the pair of treated- and untreated-friend counts. Under an intervention of interest, changing the peers environment while holding exposure fixed must leave an individual's potential outcome unchanged. This exposure-sufficiency assumption is a common restriction in interference analysis that reduces the assignment environment to the arguments needed to define interpretable potential outcomes (Hudgens and Halloran, 2008, Aronow and Samii, 2017, Sävje, 2024). The causal comparisons of interest change own treatment while holding exposure fixed, or change the network-generated exposure while holding own treatment fixed. These are the direct and network-mediated comparisons. Defining an exposure contrast does not by itself establish causal interpretations; Leung (2024) distinguishes the role of exposure mappings in defining estimands from their role in restricting interference, and Leung (2025a) shows that exposure contrasts can even reverse the signs of the relevant individual effects under randomized assignment unless additional conditions hold. Here the interpretation rests on explicit exposure sufficiency and the mediation restrictions that I state later. Each comparison can be defined using an individual's own counterfactual exposure or an independent draw from a specified exposure distribution.

Following the existing mediation analysis literature, I distinguish two interpretations of these means. *Natural effects* compare outcomes under the network exposure that the same individual would receive under another own assignment. In the Rwanda setting, a natural direct effect compares a farmer's knowledge with and without training while retaining the network exposure that this farmer would have without training. A natural network-mediated effect instead changes that farmer's network exposure to the value as if they received training, holding their treatment status at no-training. These comparisons use the same peer offers and formation primitives in both worlds. Their identification requires a cross-world restriction. On the other hand, *interventional effects* draw exposure from an assignment-specific distribution conditional on baseline information. This draw is independent of the selected individual's latent outcome response given baseline information. For the Rwanda application, the interventional network-mediated effect compares an untreated farmer's outcomes under network exposure drawn from the trained versus untrained distributions among farmers with the same baseline characteristics. It therefore averages over an exposure distribution rather than requiring the selected farmer's own counterfactual exposure. Both natural and interventional effects can compare treatments while holding exposure fixed, or exposures while holding treatment fixed; the distinction lies in how exposure is specified in these comparisons (VanderWeele et al., 2014). Interventional effects avoid the cross-world restriction, but still require identification of the potential-outcome response at fixed treatment and exposure.

Identifying these responses requires adjusting for the common causes of exposure and outcomes. To achieve this, I propose an identifying restriction called baseline network sufficiency (BNS). The economic content of BNS is that baseline network information captures the common causes of exposure and outcomes, so conditioning on

this information removes their confounding. In practice, this is a sufficient condition for the cross-world independence assumption required for identifying the natural effects. Together with other identifying assumptions, BNS enables identifying the estimands of interest using conditional outcome regressions and observed exposure distributions, without needing to specify and estimate how each link forms. In the Rwanda application, BNS is plausible if baseline production, farmers' resources, and their relationships capture the characteristics that jointly determine new social connections and returns to a fixed exposure. It can fail if unmeasured reliability, liquidity, or productivity still affects both partner choice and outcomes. Under the identifying assumptions, the formula uses a conditional outcome regression and exposure distributions without specifying or estimating the network formation law. Avoiding a parametric formation model therefore does not eliminate the need for exposure sufficiency, BNS or the relevant mean restrictions, overlap, and the sampling conditions used for estimation.

The second part of my analysis develops estimation and inference procedures for the identified causal estimands. I adapt the augmented mediation score of Tchetgen Tchetgen and Shpitser (2012) and its double machine learning implementation in Chernozhukov et al. (2018), Farbmacher et al. (2022) to dependent data. In particular, I establish two properties of my estimation procedure: Neyman orthogonality, which is a statement about the first-order sensitivity of estimators to nuisance estimation errors; and stochastic equicontinuity, which controls the empirical fluctuations introduced by estimating nuisance parmaeters. My estimation procedure sidesteps the need for sample splitting, and control the dependence between fitted scores and their training observations through a neighborhood stability condition (Chen et al., 2022, Cao and Leung, 2025). This condition bounds the change in a fitted score when a dependent neighborhood is replaced by an independent copy. Together with the regularity conditions I impose, these results justify estimation of the identified estimands while using the full sample for estimating nuisance parameters.

Valid inference presents a challenge because it must account for the fact that each village is observed at only one saturation. Increasing the number of sampled households cannot recover individual outcomes at a different saturation level than that which it is assigned, nor could it average out a shock shared by the entire village. I therefore develop inference conditional on the realized saturation allocation, imposing a mild homogeneity assumption on the villages to do so. Under these restrictions and suitable conditions on dependence and nuisance estimation rates, I establish asymptotic normality of my proposed estimators as village sizes grow, with the number of villages fixed. The result permits growing dependence neighborhoods and dependence across village boundaries. The normal approximation is centered at the causal means for each growing population, hence I also give conditions under which the asymptotic normality result is centered at a fixed population target. I then construct a feasible and consistent network heteroskedasticity and autocorrelation consistent (HAC) variance estimator, with explicit bounds on the errors from covariance weighting and score estimation. These results yield confidence intervals and joint tests for the components of the decomposition.[1] The appendices extend the results to weak dependence and unbounded outcomes under additional rate and moment conditions.

The Rwanda application illustrates the usefulness of the decomposition at the heart of my proposed framework. In my preferred specification, which measures network exposure by the endline number of treated friends, I find positive direct effects on knowledge and self-reported adoption, with no statistically detectable network-mediated contribution or treatment–network interaction for either outcome. For inputs, I find a negative network-mediated

[1]Following the variance-combination principle in Leung (2025b), I also discuss using the larger of the network and village-cluster variance estimates. Network weights can capture cross-village dependence, while village clustering retains within-village covariance terms that a narrow network bandwidth omits. For intervals with the same center and critical value, taking the larger variance preserves coverage whenever either procedure is valid.

effect for untreated farmers and a positive interaction with own training; this suggests that having more friends who received training reduces input use among untreated farmers, while own-training shifts response to an expanding network in the positive direction. These findings are consistent with the mechanism of farmers competing for shared inputs, as proposed in Duflo et al. (2026), and suggests that own training changes how farmers respond to exposure to trained friends.

1.1. **Related literature.** The paper builds on the analysis of randomized experiments with interference. Duflo and Saez (2003) use randomized encouragement within randomly selected departments to study information and social interactions in retirement-plan decisions. Sobel (2006) shows why spillovers complicate the interpretation of randomized housing-mobility studies. Hudgens and Halloran (2008) define direct, indirect, total, and overall effects under two-stage randomization and partial interference; Baird et al. (2018) study identification and design tradeoffs in saturation experiments. Vazquez-Bare (2023) studies nonparametric identification, estimation, and inference for direct and spillover effects under within-group interference. He characterizes how differences in means and reduced-form linear-in-means coefficients combine these effects, including spillover combinations with negative weights. Aronow and Samii (2017) develop design-based estimation with general exposure mappings, and Leung (2022a) allow interference to extend beyond a fixed neighborhood but vanishes with distance. Gao and Ding (2025) develop design-based guarantees for regression estimators, covariate adjustment, and network-robust inference in network experiments. Sävje (2024) distinguish exposure-effect definitions from assumptions about correct exposure specification. Tchetgen Tchetgen and VanderWeele (2012) develop inverse-probability methods for causal effects under interference, and Athey et al. (2018) construct exact tests by conditioning on an artificial experiment that sharpens the null. These approaches clarify what treatment-assignment contrasts measure and how the experimental design supports inference.

Decomposing those contrasts through a treatment-responsive network additionally requires identifying interventions on the mediator. Other individuals' assignments affect a focal outcome through a specified network exposure, and own assignment can change the relationships generating that exposure. The saturation-spillover component can also contain channels outside the within-saturation network contrast. Related mediation analyses separate contagion from other spillover pathways (VanderWeele et al., 2012), whereas the network itself is treatment-responsive here. In related work on saturation designs, Lu et al. (2026) distinguish within- and between-cluster spillovers using specified exposure neighborhoods. Their network describes interference across units; the present analysis also treats changes in connections as a mediating response to treatment. Borusyak et al. (2026) optimize experimental assignment and analysis for spillover effects under a known exposure mapping. Here the exposure is additionally generated by an endogenous network response, requiring restrictions on its relation to potential outcomes.

Several papers study interventions that alter social networks. Heß et al. (2021) use a randomized development program in rural Gambia to document reductions in informal economic transfers, with evidence consistent with changes in wealth, market participation, and the distribution of program benefits. Feigenberg et al. (2013) show that experimentally increasing microfinance-group meeting frequency changes subsequent social interaction and risk sharing. Cai and Szeidl (2018) organize randomized business meetings that expand partnerships and improve firm performance. These experiments show that network responses can concern access to resources and transactions as well as information. Comola and Prina (2021) develop a treatment-response model accounting for network changes and an instrumental-variable identification strategy. Banerjee et al. (2024) document changes in social

structure following exposure to formal credit markets, and Derksen and Souza (forthcoming) show that experimentally changing access to information induces students to form and maintain new connections. The closest methodological comparison is Gao (2025), who defines direct and network-mediated effects for a treatment-responsive network and studies identification of peer effects under a linear outcome model. To address the endogeneity of post-treatment network, she proposes an instrumental variable approach using the baseline network and randomized peer assignments, in the spirit of a shift-share intrumental variable. I study a different identifying restriction, and develop estimation of nonlinear mediation functionals within a randomized saturation experiment. Finally, the economic distinction also appears in Espinosa and Stanton (2026). Employee training changes requests for managerial assistance and reallocates managers' time. Such allocation responses motivate studying network channels even when the mechanism is not transmission of knowledge between peers.

The causal estimands draw on the distinction between natural and controlled effects developed by Robins and Greenland (1992) and Pearl (2001), and on nonparametric mediation identification in Imai et al. (2010). Mediation and interaction are explicitly separated in VanderWeele (2014). The decomposition here combines a within-saturation direct–mediation–interaction identity with a randomized saturation spillover contrast. Its four terms therefore answer a different question from the four terms in that paper. Vansteelandt and Daniel (2017) develop interventional effects and explain the conditions under which they decompose a total effect. I use their distinction between natural and interventional targets to state the additional assumptions behind the network interpretation. VanderWeele et al. (2014) explain how randomized mediator interventions can define effects when exposure-induced mediator–outcome confounding obstructs natural-effect identification. The distinction remains relevant when the mediator is a network exposure. A separate issue is whether comparisons indexed by an exposure mapping represent the intended individual-level interventions (Leung, 2024, 2025a). In particular, the sign-preserving representations in Leung (2025a) depend on restrictions on interference and the assignment mechanism; they do not identify interventions on an endogenous post-treatment network. I therefore specify both the exposure intervention and the adjustment conditions that support its causal interpretation. In recent work, Wu and Xiao (2026) study nonlinear mediation with treatment and mediator spillovers conditional on an observed adjacency matrix. The distinguishing feature here is that the network generating exposure can itself respond to the randomized treatment.

The problem is also related to identifying the channels that carry spillovers. Ritzwoller (2025) shows how correlated measures of economic proximity can confound interpretations of spillover regressions and develops regression adjustments for the causal effect of proximity on spillover intensity. I share the concern that an association between connections and treatment responses need not identify a mechanism. The target here is a decomposition of an experimental regime contrast through the network response to own assignment. Models of latent network heterogeneity, including Graham (2017), and network-based adjustment methods such as Auerbach (2022), provide relevant foundations for using network information to learn about confounding. Their identification results rely on their respective formation and outcome restrictions and do not make a baseline degree a generally sufficient mediation control.

For estimation, Tchetgen Tchetgen and Shpitser (2012) derive efficient influence functions and multiply robust estimators for mediation functionals. Farbmacher et al. (2022) combine these scores with double machine learning and nested nuisance estimation, building on the general orthogonal-score framework of Chernozhukov et al. (2018). The ordinary mediation score and its independent-data orthogonality are therefore existing ingredients. The work required here is to align that score with the target population and conditional saturation profile, and to control its remainder under network dependence. Chen et al. (2022) give stability conditions for debiased estimation without sample splitting under independent sampling; Cao and Leung (2025) extend this approach to dependent data in

general metric spaces. I specialize the neighborhood-replacement argument to the complete data entering the mediation score and state the conditions needed for its application to the estimator using all villages. Emmenegger et al. (2025) develop cross-fitting for treatment-effect estimation on an observed network using its dependency graph. Balkus et al. (2026) examine ordinary cross-fitting with correlated units and argue that removing all dependence between folds can be unnecessary. Huber and Oberhänsli (2026) develop mediation estimators under conditional parallel trends, providing a related alternative to adjustment in outcome levels.

Finally, the inference analysis draws on limit theory and variance estimation for network-dependent observations. Ogburn et al. (2024) develop semiparametric causal inference from a single social network, allowing dependence through both transmission and latent similarities. Chen and Shao (2004) provide normal-approximation bounds under local dependence using Stein's method. The cumulant approach used here has a complementary development in Janson (1988), Döring et al. (2022), Féray (2018); Appendix C relates its connected-tuple calculation to applications in statistical mechanics and stochastic geometry. Kojevnikov et al. (2021) establish laws of large numbers, central limit theorems, and network HAC results under restrictions linking dependence to network growth. Leung (2023) studies inference with a fixed number of growing network clusters and characterizes the role of low conductance in obtaining asymptotic independence. The additional issue in a saturation experiment is that conditioning on the village assignments changes which populations supply each treatment cell. The analysis here links a correctly centered mediation estimator to a variance benchmark under that conditioning, making explicit the comparability restrictions needed before local network growth can justify shrinking confidence intervals. Leung (2025b) studies cluster-randomized trials with interference across cluster boundaries and develops conservative inference. The related concern here is that village-based and network-based covariance estimates can account for different pairs of dependent scores.

The remainder of the paper follows the identification and estimation argument. Section 2 introduces the design, exposure mapping, and counterfactual means. Section 3 discusses identification and includes the role of baseline network information. Sections 4 and 5 develop estimation and inference procedures, respectively. Sections 6 and 7 present the empirical application and simulation exercises, and Section 8 concludes. Proofs and additional analyses are collected in the appendices.

2. Design, networks, and causal parameters

A saturation experiment changes both the assignment of individuals and the environments in which they interact. To attribute an outcome change to a network response, the intervention on links must be distinguished from the assignment labels carried by those links. I first define these objects without imposing a particular exposure coding or a statistical model for estimating them. Section 4 then states the additional restrictions under which a fixed number of villages supplies increasing information about the resulting parameters.

2.1. Population and assignment.

For each n, the population $[n] = \{1, \dots, n\}$ is partitioned into J villages. Village j contains n_j individuals, $j(i)$ is individual i's village, and $w_{j,n} = n_j/n$. In the asymptotic analysis J is fixed, $n_j \to \infty$, and $w_{j,n} \to w_j > 0$. An individual is the sampling unit. Repeated measurements can be collected in the individual's observation.

Let $T = (T_1, \dots, T_J)$ denote village saturations, taking values in a finite set $\mathfrak{T} \subseteq \mathcal{S}^J$, where $\mathcal{S} \subseteq [0, 1]$ is finite. As a shorthand, I write $T_i = T_{j(i)}$. The binary variable P_i denotes *assignment or offer of treatment*. It denotes receipt only under perfect compliance, which I will assume throughout. No identification result below interprets

randomized offers as randomized take-up under noncompliance. Write $Z_i = (T_i, P_i)$ for the observed assignment cell and $z = (s, a)$ for a generic cell, with saturation s and own offer a.

Assumption 2.1 (Assignment design)**.** The first-stage law P_n^T on $\mathfrak{T}$ is known and independent of all pretreatment and outcome primitives. Every target village has positive first-stage probability of every saturation used to define its target counterfactuals. Conditional on $T = t$, the assignments are mutually independent, with

$$\mathbb{P}(P = p \mid T = t) = \prod_{1 \le i \le n} t_{j(i)}^{p_i} (1 - t_{j(i)})^{1 - p_i} . \tag{2.1}$$

The independent random variables used to generate these assignments are independent of the primitives. For fixed-profile asymptotics, $\mathfrak{T}$ is fixed, $\inf_{n, t \in \mathfrak{T}} P_n^T(t) > 0$, and each supported profile contains every saturation needed by the parameter.

Thus the design probability of (t, p) is $P_n^T(t)$ times the product in (2.1). The first stage may fix arm counts and need not assign villages independently. Let Z contain the required supported cells.

2.2. **A network that responds to treatment.** Let $\mathcal{W}_n$ collect the pretreatment network and covariate primitives, and let ε_i be the outcome shock, valued in a standard Borel space. The observed baseline graph A^{pre} and covariates X^n are measurable in $\mathcal{W}_n$. Let $\mathscr{W}_n$ be the standard Borel state space of $\mathcal{W}_n$, and let $\mathscr{A}_n = \{\ell \in \{0,1\}^{n \times n} : \ell = \ell^\top,\ \ell_{ii} = 0 \text{ for every } i \in [n]\}$, equipped with its discrete σ-field. A post-treatment undirected graph is generated by the measurable map

$$\begin{aligned} \mathcal{A}_n : \mathscr{W}_n \times \mathfrak{T} \times \{0,1\}^n &\longrightarrow \mathscr{A}_n , \\ (w, t, p) &\longmapsto \big(\mathcal{A}_{ik,n}(w, t, p)\big)_{i,k=1}^n , \\ A(t, p) &= \mathcal{A}_n(\mathcal{W}_n, t, p) . \end{aligned} \tag{2.2}$$

The definition includes all latent formation noise in $\mathcal{W}_n$. Using the same primitives across interventions specifies a joint counterfactual model; the observational distribution alone need not determine that model.

One useful construction fixes a pretreatment eligibility graph K_n and sets

$$\begin{aligned} A_{ik}(t, p) &= K_{ik,n} \mathbb{1}\{\eta_{ik}^1 \le \zeta_{ik,n}(U_i, U_k, X_i, X_k, t_{j(i)}, p_i, t_{j(k)}, p_k)\} , \\ A_{ik}^{\mathrm{pre}} &= K_{ik,n} \mathbb{1}\{\eta_{ik}^0 \le \zeta_{ik,n}^0(U_i, U_k, X_i, X_k)\} . \end{aligned} \tag{2.3}$$

Acceptance functions are symmetric in endpoints and take values in $[0, 1]$. Correlation between $(\eta_{ik}^0, \eta_{ik}^1)$ allows persistent pair-specific affinity while training changes the probability of a tie. In Duflo et al. (2026), repeated meetings within training groups created opportunities for new coffee-friend relationships; ties outside those groups could also disappear. The model allows both responses. Independence across pairs is needed only for particular statistical constructions, such as that in Appendix D.

Remark 2.1 (Eligibility and dynamic formation)**.** Separating contact opportunities from link acceptance has precedents in dynamic formation models. Jackson and Rogers (2007) let arriving individuals meet others both at random and through existing friends, and then form links with compatible individuals among those met. Their opportunity sets are endogenous and evolve with the network size. In the dynamic model discussed by Leung and Moon (2026a, Section 3.1), a special hard-cutoff specification forbids links when latent social distance exceeds r_n. Conditional on positions $x_1, \ldots, x_n$ (e.g. geographic locations), its admissible pairs are

$$K_{ik,n} = \mathbb{1}\{\|x_i - x_k\| \le r_n\} \mathbb{1}\{i \ne k\} .$$

The same paper's more general homophily condition makes distant links unlikely without restricting them from forming ex-ante. That decay restriction is distinct from a deterministic bound on which links can occur. A related hard-distance exclusion appears in Leung and Moon (2026b, Section 2.1) as a random-geometric variant of their strategic formation model. Their general analysis uses stabilization and restrictions on strategic feedback and equilibrium selection, allowing the determining neighborhood to be random. This supplies a formation-model precedent for admissible pairs, while also showing why deterministic finite dependence is a stronger specialization.

Here eligibility specifies the possible links across the interventions being compared. Identification uses the exposure and conditional-response restrictions, without estimating a formation likelihood. For the exact-dependence results, a known deterministic eligibility graph can also help enumerate the primitive inputs shared by observations. Random latent positions or endogenous meeting sets do not automatically provide such a graph. Conditioning on them changes the probability law, and their observability and the resulting conditional independence must be justified. Even a hard cutoff does not suffice when acceptance depends on distant links or shared shocks. In Lemma D.1, I explore additional sufficient conditions on the separation of primitive inputs that justify exact local dependence. Appendix K treats nonzero dependence beyond a chosen neighborhood. ♢

Let $\mathcal{B}$ be the standard Borel space of baseline adjustment variables. The baseline adjustment variable is

$$B_i := \widetilde{X}_i = (X_i, N_i^{\mathrm{pre}}), \qquad N_i^{\mathrm{pre}} = v_{i,n}(A^{\mathrm{pre}}, X^n)\,. \tag{2.4}$$

The map $v_{i,n}$ is specified before treatment; it may read neighbor covariates, in which case those covariates belong to $\mathcal{W}_n$. A baseline graph weighted by *realized* treatment labels is not pretreatment information. The number of graph steps read by $v_{i,n}$ matters for dependence, even though measurability alone suffices to define the causal parameter.

2.3. **Exposure and potential outcomes.** Let G be the exposure space. For a network adjacency matrix $\ell \in \mathscr{A}_n$, saturation profile $t \in \mathfrak{T}$, and peer-assignment vector $p_{-i} \in \{0,1\}^{n-1}$, define the exposure mapping by

$$e_{i,n}(\ell, t, p_{-i}, X^n) \in \mathsf{G}, \qquad G_i = e_{i,n}(A(T,P), T, P_{-i}, X^n)\,. \tag{2.5}$$

The mapping summarizes the network and peers' assignments in the features relevant for outcomes. I allow counts and shares as well as exposures that retain peer identities, covariates, link weights, or paths; in other words, anonymity is not essential for identification. Own assignment is excluded from the *transmission map*. At fixed saturation and peer assignments, changing own assignment therefore changes exposure only through the network. This gives the mediated contrast its interpretation as an effect of link changes. Including own assignment directly in the mapping would also introduce a mechanical change in exposure.

Example 2.1 (Counts as an exposure specification)**.** An analyst may choose

$$G_i = (M_i^P, M_i^C), \qquad M_i^P = \sum_{k \in [n]\setminus\{i\}} A_{ik}P_k, \qquad M_i^C = \sum_{k \in [n]\setminus\{i\}} A_{ik}(1-P_k)\,. \tag{2.6}$$

These are the actual counts, so their support may grow with n.

Let $\mathcal{Y}_i(t,p,\ell)$ denote the potential outcome when assignments and the network are set to (t,p,ℓ).

Assumption 2.2 (Intervention-level exposure sufficiency)**.** Almost surely, simultaneously for every intervention configuration used by the parameter,

$$\mathcal{Y}_i(t,p,\ell) = f_{i,n}(t_{j(i)}, p_i, e_{i,n}(\ell, t, p_{-i}, X^n), \varepsilon_i)\,. \tag{2.7}$$

The structural maps are jointly measurable and the relevant potential outcomes are integrable.

For fixed saturation s, own assignment a, and exposure g, write $Y_i^s(a,g) = f_{i,n}(s,a,g,\varepsilon_i)$ for individual i's potential outcome when these three arguments are set to the specified values, with the same outcome shock ε_i across interventions. This is sometimes referred to as the *controlled response function* (VanderWeele, 2014). In particular, exposure is held at g rather than generated by the individual's treatment-responsive network. The exposure interventions below specify which fixed values or random draws are substituted for g.

It is worth noting that (2.7) restricts potential outcomes at fixed treatment and exposure, including unobserved interventions. An observed regression representation $Y_i = f(Z_i, G_i, \varepsilon_i)$ alone does not supply it. For the count example, identity, link intensity, omitted paths, and other peer characteristics must have no additional effect once the specified arguments are fixed. This distinction follows the separation between exposure definitions and specification assumptions emphasized by Sävje (2024); randomization does not verify it.

2.4. **Network-change worlds and target means.** Select a target individual uniformly from $[n]$, independently of the primitives. Unless expressly designated finite-population, expectations below average over this selection, the primitive sampling law, and the specified assignment intervention. The finite-n target is defined before taking any limit.

An exposure intervention specifies both the assignments that generate relationships and the exposure used to evaluate the controlled response. I distinguish the exposure produced by the focal individual's own counterfactual network from a draw from a conditional exposure distribution. This distinction applies whether the comparison changes the response assignment a or the assignment b that generates exposure.

Definition 2.1 (Exposure interventions)**.** Fix a focal individual i and saturation s.

(i) *Natural exposure.* Let $Q_{j(i),s}$ be the conditional law of the saturations assigned to the other villages, i.e.

$$Q_{j(i),s} = \mathcal{L}_D(T_{-j(i)} \mid T_{j(i)} = s)\,.$$

Draw these saturations and then all peers' Bernoulli assignments under the experimental design. Denote the complete draw by H_i^s, its peer-assignment coordinates by $H_{i,P}^s$, and its full saturation profile by T^s. Make this draw independently of the primitives and use it in both focal-assignment worlds. Let $A_i^s(b)$ be the entire graph generated when the focal assignment is b, with this same draw and the same formation primitives. Define

$$G_i^s(b) = e_{i,n}(A_i^s(b), T^s, H_{i,P}^s, X^n), \qquad Y_i^s(a,g) = f_{i,n}(s,a,g,\varepsilon_i)\,. \tag{2.8}$$

(ii) *Interventional exposure.* Let I be an independent uniform target index and define the probability kernel

$$q_b^\star(\cdot \mid x) = \mathcal{L}(G_I^s(b) \mid B_I = x)\,.$$

Draw $\widetilde{G}_I^s(b)$ from $q_b^\star(\cdot \mid B_I)$ independently of (I, ε_I) conditional on B_I. Thus every individual uses the same target-population kernel, evaluated at that individual's baseline variable.

In Duflo et al. (2026), natural exposure describes the connections a farmer would have with and without training. Relationships adjust to the farmer's offer status, while other farmers' offers and the underlying network-formation factors remain fixed. Exposure could, for example, count connected farmers with and without training offers. A natural cross-world outcome asks how a farmer would perform with training but with the netwokr exposure that same farmer would have developed without training. Interventional exposure assigns the farmer a random draw

from the exposure distribution that would arise without an offer among farmers with the same baseline characteristics. The draw need not match the farmer's own counterfactual exposure. This intervention specifies the farmer's exposure but does not, by itself, specify how to rewire the entire network.

Under partial interference, an individual's outcome and relevant network response depend only on assignments within the individual's group (Hudgens and Halloran, 2008). The external draw then drops out of the intervention. Villages need not be such groups. Duflo et al. (2026) assigned some farmers to training groups in neighboring villages and measured relationships beyond village boundaries. The external assignment law $Q_{j,s}$ accommodates these connections. In the contrast $Y_i^s(a, G_i^s(1)) - Y_i^s(a, G_i^s(0))$, own outcome assignment, peer offers and formation primitives are held fixed while the network responds to the focal offer. This compares the relationships a farmer would have with and without an offer of training, at the same peer assignments. The entire graph is regenerated, so the model can include responses beyond the focal row. Under constrained first-stage randomization, $Q_{j,s}$ retains the corresponding dependence between village assignments.

Define the natural and interventional arrays by

$$\theta_{s,n}^{\mathrm{nat}}(a,b) = \frac{1}{n}\sum_{1\le i\le n} \mathbb{E}[Y_i^s(a, G_i^s(b))] \,, \tag{2.9}$$

$$\theta_{s,n}^{\mathrm{int}}(a,b) = \frac{1}{n}\sum_{1\le i\le n} \mathbb{E}[Y_i^s(a, \widetilde{G}_i^s(b))] \,. \tag{2.10}$$

The natural mean retains the dependence between an individual's counterfactual exposure and response primitives. The interventional mean instead averages the response to a draw that is independent of those primitives conditional on baseline information. Holding a fixed and changing b measures an exposure-mediated response; holding b fixed and changing a measures an own-treatment response. Either comparison can be defined using the natural or the interventional array. Natural versus interventional therefore describes the exposure construction, rather than which treatment argument changes (VanderWeele et al., 2014, Vansteelandt and Daniel, 2017).

Drawing exposure from a separate conditional exposure distribution for each individual would define a different interventional target. These distributions specify which exposure values can be drawn and their probabilities; they are not estimation weights. Under the convention in Definition 2.1(ii), the structural response may still differ across individuals, but exposure is drawn from their common target-population conditional distribution. Separate draws for different individuals need not correspond to a compatible undirected graph. The mean describes an intervention on a focal individual's exposure; a simultaneous network policy additionally requires a joint graph law with the proposed exposure marginals.

The randomized regime mean is

$$R_{s,n}(a) = \frac{1}{n}\sum_{1\le i\le n} \mathbb{E}[Y_i^s(a, G_i^s(a))] \,. \tag{2.11}$$

Natural means with matching assignments equal $R_{s,n}(a)$ by definition and consistency. Interventional means with matching assignments require the within-world identification conditions of the next section. The natural and interventional distinction follows Pearl (2001), Imai et al. (2010), Vansteelandt and Daniel (2017). Coupling formation worlds does not remove their different identifying assumptions.

Fix reference saturations $s_H > s_L$ for which $(s_H, 0), (s_H, 1), (s_L, 0)$ are supported. Suppressing n, I define

$$\mathrm{TE} = R_{s_H}(1) - R_{s_L}(0), \qquad S = R_{s_H}(0) - R_{s_L}(0) \,,$$

$$D = \theta_{s_H}(1,0) - \theta_{s_H}(0,0), \qquad N = \theta_{s_H}(0,1) - \theta_{s_H}(0,0) \,,$$

$$I = \theta_{s_H}(1,1) - \theta_{s_H}(1,0) - \big(\theta_{s_H}(0,1) - \theta_{s_H}(0,0)\big) \ . \tag{2.12}$$

For either the natural or interventional array with $\theta_s(a,a) = R_s(a)$, the components satisfy

$$\mathrm{TE} = S + D + N + I \ . \tag{2.13}$$

The decomposition depends on the reference saturation and own assignment. Following the direct, indirect, and interaction decomposition of VanderWeele (2014), it adds the saturation spillover S among untreated individuals. On the interpretation of the reduced-form total effect, it is worth noting that one can also look at the population policy mean, $V_n(s) = sR_{s,n}(1) + (1-s)R_{s,n}(0)$, which averages over treated and untreated recipients at a saturation s. Thus $V_n(s_H) - V_n(s_L)$ compares two saturation policies, which is also identified through randomization, whereas TE compares offered recipients at high saturation with untreated recipients at low saturation. These reduced-form estimates take on different interpretations. Furthermore, the within-saturation contrast between treated and untreated individuals isolates exposure changes induced by own treatment while holding peers' offers fixed. Across saturations, exposure can also change because more peers received treatment, even if all relationships remain unchanged. The fixed-network example in Appendix I illustrates why a cross-saturation exposure effect does not isolate relationship formation. Finally, I show, in Appendix M, the interpretations of the main analysis specifications in Duflo et al. (2026, Section 4.3). Under regularity conditions, I show that these regressions identify reduced-form total effects, but cannot separately identify D, N, and I.

3. Identification

Randomization identifies assignment-regime contrasts even when treatment changes the network. Attributing those contrasts to a network-mediated channel requires additional restrictions. I state the two identification arguments separately and distinguish the law of a randomized experiment from the law conditional on one realized village allocation. This distinction will determine the weights used in estimation.

The central distinction is between identifying a total effect and identifying its components. With supported assignment cells, randomization identifies TE, the untreated saturation contrast S, the within-saturation own-offer total $D + N + I$, and population saturation-policy contrasts. It does not, by itself, identify D, N, and I separately: these components require the response to an exposure generated under a different own assignment. The mediation assumptions below identify those alternative-exposure means. A separate issue is what an identified exposure contrast measures. Across saturations, exposure can change because peers' offers change even on a fixed network, so identifying that contrast does not isolate the effect of relationship formation.

3.1. What the design identifies.

Fix the primitive population and let $\mathbb{E}_D$ denote expectation over assignment only. For known positive inclusion probabilities $p_i^D(s,a) = \mathbb{P}_D(T_i = s, P_i = a)$, define

$$R^D_{s,a,n} = \frac{1}{n}\sum_{1\le i\le n} \mathbb{E}_D[Y_i(T,P) \mid T_i = s, P_i = a], \qquad \widehat{R}^{HT}_{s,a} = \frac{1}{n}\sum_{1\le i\le n} \frac{\mathbb{1}\{T_i = s, P_i = a\}Y_i}{p_i^D(s,a)} \ . \tag{3.1}$$

Proposition 3.1 (Design identification)**.** *Under Assumptions 2.1 and 2.2, $\mathbb{E}_D\widehat{R}^{HT}_{s,a} = R^D_{s,a,n}$. Thus the contrasts* TE *and S are design identified without mediator exchangeability. Averaging over the primitives gives the regime means in (2.11), when the external law is $Q_{j,s}$.*

This is the Horvitz–Thompson argument (Horvitz and Thompson, 1952). Related inverse-probability arguments under interference appear in Aronow and Samii (2017), who weight by exposure probabilities, and Liu et al.

(2016), who establish unbiasedness under known assignment probabilities. Each weighted observed outcome has expectation equal to the corresponding design-conditional potential-outcome mean. The probability $p_i^D(s,a)$ is a known assignment propensity for the joint saturation–offer cell: it equals $\mathbb{P}_D(T_{j(i)} = s)$ times s when $a = 1$ and times $1 - s$ when $a = 0$. It is determined by the two-stage randomization, not by a fitted model for treatment or post-treatment exposure. A combined unweighted conditional mean has the uniformly weighted target only with equal inclusion probabilities or appropriate weighting. Design unbiasedness does not imply consistency as n_j grows with fixed J; Section 4 supplies a counterexample.

3.2. **Mediation identification.** For this subsection fix s, n and the target population. Let B denote its baseline variable and let $Y^s(a,g)$ and $G^s(b)$ denote the controlled response and common-peer exposure for a randomly selected target individual. Recall from Section 2.3 that the controlled response is the potential outcome with saturation s, own assignment a, and exposure g held fixed.

Each assignment cell must represent the same equally weighted target population. I therefore define its distribution by averaging individuals' cell-conditional distributions with equal weights, equivalently weighting observed contributions by their inverse design inclusion probabilities. This follows the common-target weighting approach of Li et al. (2018, Section 2). For each supported cell (s,a), let

(3.2)
$$P_{s,a}^{\mathrm{tar}} = \frac{1}{n}\sum_{1\le i\le n} \mathcal{L}(B_i, G_i, Y_i \mid T_i = s, P_i = a),$$

$$\int h \, \mathrm{d}P_{s,a}^{\mathrm{tar}} = \frac{1}{n}\sum_{1\le i\le n} \mathbb{E}[h(B_i,G_i,Y_i) \mid T_i = s, P_i = a] = \frac{1}{n}\sum_{1\le i\le n} \mathbb{E}\left[\frac{\mathbb{1}\{T_i = s, P_i = a\}}{p_i^D(s,a)} h(B_i,G_i,Y_i)\right],$$

for bounded measurable h, where $p_i^D(s,a)$ is the design inclusion probability in (3.1). Under Assumption 2.1, these cell laws share the target baseline distribution F_B; with equal inclusion probabilities, they coincide with ordinary cell conditioning for a uniformly selected individual.

Following the observed-law formulation of mediation analysis in Tchetgen Tchetgen and Shpitser (2012, Section 2.1), define the outcome regression and exposure distribution under these cell laws:

$$m_a(g,x) = \mathbb{E}_{P_{s,a}^{\mathrm{tar}}}[Y \mid G = g, B = x], \qquad q_b(\mathrm{d}g \mid x) = P_{s,b}^{\mathrm{tar}}(G \in \mathrm{d}g \mid B = x)\,. \tag{3.3}$$

Throughout this subsection, conditioning on an assignment cell refers to its target law; expressions involving potential outcomes use the same design weighting on the underlying population and assignment space. Appendix A gives the joint target-law construction and the standardization argument. Section 4 states the additional common-law conditions used after conditioning on the realized village allocation.

The assumptions enter in two stages. First, mediator alignment matches the intervention exposure law to an observed assignment cell, while controlled-response exchangeability identifies the mean response to a fixed exposure after adjustment for B. Together with overlap and target standardization, these conditions identify interventional effects. Second, a conditional mean restriction linking the response to the exposure in the other assignment world identifies natural effects. Baseline network sufficiency provides a primitive condition for both exchangeability requirements. I state these requirements separately because the interventional interpretation needs less than the natural interpretation.

Assumption 3.1 (Mediator alignment and controlled-response exchangeability)**.** For every required pair a, b and target-almost every baseline value x.

(i) For every bounded measurable function u on G,

$$\mathbb{E}[u(G^s(b)) \mid B = x] = \mathbb{E}[u(G) \mid T_i = s, P_i = b, B = x] = \int u(g) q_b(\mathrm{d}g \mid x) \, .$$

(ii) Let $h_a : \mathsf{G} \times \mathcal{B} \to \mathbb{R}$ be a jointly measurable version of $h_a(g, x) = \mathbb{E}[Y^s(a, g) \mid B = x]$ on the intervention support. The controlled-response conditional mean satisfies

$$\mathbb{E}[Y^s(a, g) \mid T_i = s, P_i = a, G, B] = h_a(g, B) \, , \qquad m_a(g, B) = h_a(g, B) \, . \tag{3.4}$$

Part (i) states mediator alignment through conditional expectations of test functions, equivalently equality of conditional distributions. It cannot be replaced by equality of exposure means. A nonlinear response can depend on the rest of the exposure distribution. Bernoulli assignment and the design-aligned external law give the individual alignment in Appendix A; the chosen population weights must also agree. Part (ii) is the substantive exposure–outcome adjustment condition. The first equality in (3.4) imposes conditional mean independence on a jointly measurable conditional-response version whose evaluation at the observed G represents the conditional mean of $Y^s(a, G)$. The second follows from consistency on the outcome-arm support. Randomizing own assignment alone does not randomize post-treatment exposure. Under the design-aligned target law (3.2), standardization is a consequence of randomization and consistency.

$$\int \mathbb{E}[Y \mid T_i = s, P_i = a, B = x] F_B(\mathrm{d}x) = R_{s,n}(a) \, . \tag{3.5}$$

Lemma A.2 derives this equality without an additional identification assumption. Using an unweighted conditional law with unequal inclusion probabilities changes the population and need not give the same equality.

Assumption 3.2 (Overlap)**.** All state spaces are standard Borel. For every required a, b,

$$\begin{gathered} q_b(\cdot \mid x) \ll q_a(\cdot \mid x) \quad (F_B\text{-a.e. } x) \, , \\ r_{ab}(g, x) \coloneqq \frac{\mathrm{d}q_b(\cdot \mid x)}{\mathrm{d}q_a(\cdot \mid x)}(g) \quad \text{has a jointly measurable version} \, , \\ \int \int \big\{ |m_a(g, x)| + \mathbb{E}[|Y^s(a, g)| \mid B = x] \big\} q_b(\mathrm{d}g \mid x) F_B(\mathrm{d}x) < \infty \, . \end{gathered} \tag{3.6}$$

The natural responses used below also satisfy $\mathbb{E}|Y^s(a, G^s(b))| < \infty$.

Additional uniform bounds for estimation are specified in Assumption 4.3. For a continuous exposure, r_{ab} is a conditional density ratio when densities exist. For a discrete exposure it equals $q_b(g \mid x)/q_a(g \mid x)$ on the required support. Estimating $\theta_s(a, b)$ requires outcome observations at exposures arising under b; the reverse support inclusion is needed only if the reverse cell is estimated. Infinite support therefore does not require a uniform positive lower bound on every atom.

Theorem 3.2 (Interventional identification)**.** *Under Assumptions 2.1, 2.2, 3.1, and 3.2,*

$$\theta^{\mathrm{int}}_{s,n}(a, b) = \mathbb{E}_B \left[\int m_a(g, B) q_b(\mathrm{d}g \mid B) \right] =: \Psi_{s,ab,n} \, . \tag{3.7}$$

Moreover $\Psi_{s,aa,n} = R_{s,n}(a)$, so the interventional components satisfy (2.13).

By Definition 2.1(ii), the exposure draw from the target-population conditional distribution $q^\star_b(\cdot \mid B)$ is independent of the controlled-response process given B. Integrating that process against $q^\star_b$ and then using Assumption 3.1 to identify the response mean and exposure law gives (3.7). The complete disintegration and regime-mean

arguments are in Appendix A. The result follows the mediation-functional construction in Tchetgen Tchetgen and Shpitser (2012), Vansteelandt and Daniel (2017). It neither randomizes the observed network nor obtains controlled-response exchangeability from the experiment.

Assumption 3.3 (Natural-response mean exchangeability)**.** For each required a, b, let $h_a : \mathsf{G} \times \mathcal{B} \to \mathbb{R}$ be the conditional-response map in Assumption 3.1(ii). There is a jointly measurable conditional mean map $H_{a,b} : \mathsf{G} \times \mathcal{B} \times \mathsf{G} \to \mathbb{R}$ representing $\mathbb{E}[Y^s(a,g) \mid B = x, G^s(b) = v]$, such that on the relevant support $H_{a,b}(g,x,v) = h_a(g,x)$ and

$$\mathbb{E}[Y^s(a,g) \mid B, G^s(b)] = h_a(g,B), \qquad \mathbb{E}[Y^s(a,G^s(b)) \mid B, G^s(b)] = h_a(G^s(b),B)\ . \tag{3.8}$$

The first equality in (3.8) concerns the controlled response at a fixed exposure on the intervention support. The second requires that the conditional mean can also be evaluated at the random exposure. This evaluation requirement avoids an ambiguity that can arise from choosing unrelated almost-sure versions for different fixed values of a continuous exposure. For $b \neq a$, Assumption 3.3 links an outcome under one assignment to an exposure under another. It excludes unmeasured common determinants of the alternative exposure and the response to a fixed exposure. For example, if $Y^s(a,g) = \alpha_a + \kappa g + u$, it requires $g\{\mathbb{E}[\kappa \mid B, G^s(b)] - \mathbb{E}[\kappa \mid B]\} + \mathbb{E}[u \mid B, G^s(b)] - \mathbb{E}[u \mid B] = 0$ for the required g. This can fail even in a randomized experiment. It is a mean restriction, so it is weaker than conditional independence of the full controlled-response distribution. Its credibility depends on the outcome and the adjustment set. Interventional identification requires Assumption 3.1(ii), but does not require this additional relation between the two assignment worlds.

Theorem 3.3 (Natural identification)**.** *Under Assumptions 2.1, 2.2, 3.1, 3.2, and 3.3,* $\theta^{\mathrm{nat}}_{s,n}(a,b) = \Psi_{s,ab,n} = \theta^{\mathrm{int}}_{s,n}(a,b)$.

Theorem 3.3 shows that the natural and interventional means coincide under these stronger assumptions. Conditional on baseline information, the alternative natural exposure does not select individuals with different controlled-response means. Replacing that exposure by an independent draw from the same conditional law therefore leaves its mean response unchanged. The resulting equality is an equality of causal parameters and of their identified functionals; it does not equate the two counterfactual constructions. This distinction is central to mediation analysis. Natural effects require restrictions linking different treatment worlds, whereas interventional effects permit dependence between an individual's natural exposure and the response in the alternative assignment world (Robins and Greenland, 1992, Pearl, 2001, VanderWeele et al., 2014, Vansteelandt and Daniel, 2017). The stronger natural interpretation is available when baseline information removes the relevant common causes, as formalized next.

The equality itself has precedents in statistical mediation analysis. VanderWeele and Tchetgen Tchetgen (2017, p. 921) explain that natural and randomized interventional effects agree under the assumptions identifying the natural effects. Yu et al. (2025, Proposition 2) express their difference through conditional covariances between the mediator response and controlled outcomes. Those covariances vanish under cross-world exchangeability, although cancellation can also yield equality without it. The mediation functional and orthogonal estimation in Farbmacher et al. (2022) provide the corresponding natural-effect estimation framework. Theorem 3.3 applies these ideas to a network-generated exposure, specifying the peer-assignment law, target-population weights, and conditional-response versions needed for the equality to hold.

A single conditional independence assumption, baseline network sufficiency (BNS), supplies both exchangeability restrictions under a common conditional-response mean. BNS is an unconfoundedness assumption. The chosen baseline information captures the dependence between formation primitives and the shock determining responses

to a fixed exposure. Assignment alignment, overlap, and integrability then make the common mediation functional observable.

Assumption 3.4 (Baseline-network sufficiency, BNS)**.** For each target individual, with B_i measurable in the complete pretreatment field $\mathcal{W}_n$,

$$\varepsilon_i \perp\!\!\!\perp \mathcal{W}_n \mid B_i \,. \tag{3.9}$$

The assignment randomization is jointly independent of $(\mathcal{W}_n, \varepsilon^n)$, and the controlled response reads the same shock across interventions.

Assumption 3.4 says that, among individuals with the same baseline information, the formation primitives contain no further information about the shock governing outcomes at fixed treatment and exposure. Because B_i is itself part of the pretreatment information, this restriction permits formation and outcomes to share observed determinants. It excludes residual common causes, such as unmeasured productivity that affects both partner choice and the return to a given exposure. The network may still respond to treatment and to unobserved pair affinity, provided those formation inputs have no remaining association with the response shock after adjustment. Thus randomizing treatment does not establish BNS, and including a baseline degree does not automatically make it sufficient.

The same formation primitives generate exposure in both own-assignment worlds. Their conditional independence from the response shock therefore supports both within-world adjustment and the cross-world comparison. To carry this individual argument over to the target population, the proposition also requires a common conditional response mean: selecting individuals with a particular exposure must not change their mean response at the same baseline value. Section 3.3 gives economic examples of when these restrictions hold and fail.

Proposition 3.4. *Suppose Assumptions 2.1, 2.2, and 3.4 hold. Let* $Q_i(\mathrm{d}e \mid x) \coloneqq \mathcal{L}(\varepsilon_i \mid B_i = x)$ *and* $h_{i,a}(g,x) \coloneqq \int f_{i,n}(s,a,g,e) Q_i(\mathrm{d}e \mid x)$*. For each target individual and every required* a, b*,*

$$\mathbb{E}[Y_i^s(a, G_i^s(b)) \mid B_i, G_i^s(b)] = h_{i,a}(G_i^s(b), B_i) \,, \tag{3.10}$$

$$\mathbb{E}[Y_i \mid T_i = s, P_i = a, G_i, B_i] = h_{i,a}(G_i, B_i) \,. \tag{3.11}$$

If $h_{i,a}(g,x) = h_a(g,x)$ *for all target individuals on the relevant supports, the corresponding conditional-response versions for a uniformly selected target individual satisfy Assumptions 3.1(ii) and 3.3 under the aligned observed and intervention laws. Under Assumptions 3.1(i) and 3.2, both causal means are identified by the observable integral*

$$\theta_{s,n}^{\mathrm{nat}}(a,b) = \theta_{s,n}^{\mathrm{int}}(a,b) = \int\!\!\int m_a(g,x) q_b(\mathrm{d}g \mid x) F_B(\mathrm{d}x) \,.$$

Proposition 3.4 shows that BNS is a sufficient condition for the two mean exchangeability restrictions to hold. BNS removes dependence between the response shock and all formation inputs after conditioning on baseline information; independent assignment draws preserve that separation. For the natural mean, it also permits evaluation of the response at the alternative exposure generated from those inputs. The common-mean condition ensures that selecting individuals into an observed assignment–exposure cell does not select a different response function at the same baseline value. Alignment and overlap remain separate conditions for identifying the integral. The discussion in Section 3.3 gives examples of the required adjustment.

The functions m_a, q_b, and F_B describe the target-standardized observed law; the causal restrictions above determine whether their mediation functional identifies an interventional or natural mean. Weighting alone does not remove exposure–outcome confounding: Appendix I gives two observationally identical randomized models with

different interventional means. The usual mediation weighting identity (Tchetgen Tchetgen and Shpitser, 2012, Section 2.2) gives, whenever the weighted outcome is absolutely integrable,

$$\Psi_{s,ab,n} = \frac{1}{n} \sum_{1 \le i \le n} \mathbb{E}\left[\frac{\mathbb{1}\{T_i = s, P_i = a\}}{p_i^D(s,a)} r_{ab}(G_i, B_i) Y_i \right] . \tag{3.12}$$

The expectation is under the actual experiment with its design inclusion probabilities; Appendix A gives the calculation. This representation is useful for identification but is first-order sensitive to estimating r_{ab}. The augmented score in Section 4 replaces that first-order bias by products of nuisance errors.

3.3. **What baseline network information can establish.** BNS is an unconfoundedness assumption for network-generated exposure. It asks that baseline information account for the characteristics that jointly determine relationship formation and outcomes at a fixed exposure. For instance, farmers who attract more connections may also be more productive. Comparing their output with that of less connected farmers could then confound the consequences of connections with differences in productivity. Baseline adjustment supports a causal interpretation if it accounts for the productivity differences responsible for this selection. Randomizing treatment alone does not resolve the problem because it does not randomize the resulting network exposure.

Although BNS is strong, it provides a sufficient condition for the two exchangeability restrictions used in the mediation analysis. Under the common conditional response means required by Proposition 3.4, it ensures that individuals with different observed exposures have the same mean response when treatment, exposure, and baseline information are held fixed. This restriction supports identification of interventional effects. BNS also ensures that an individual's exposure under an alternative assignment carries no additional information about the mean response at a fixed exposure under the assignment of interest. This cross-world restriction supports identification of natural effects. Both implications follow because the same formation primitives generate exposure in the two assignment worlds. BNS is stronger than the two conditional mean restrictions themselves.

These restrictions allow homophily, which describes individuals with similar characteristics are more likely to form relationships; in other words, it is a form of selection. BNS concerns whether the information governing that selection remains associated with the response shock after conditioning on baseline information. Farmers may therefore select friends with similar observed productivity, and productivity may affect their returns to a given exposure, without violating BNS. The adjustment must account for the productivity differences through which this selection confounds the exposure–outcome comparison.

To see the distinction, consider two farmers with the same recorded baseline characteristics. One has more trained contacts than the other. If the difference reflects matching on an unobserved learning ability that also raises the return to a given number of trained contacts, their outcome difference combines exposure and ability. Such selection generally violates the identifying mean restrictions. By contrast, matching on an unobserved characteristic can be compatible with BNS if that characteristic, together with the other formation inputs, is conditionally independent of the response shock. Unobserved determinants of friendship need not be unobserved determinants of outcomes at a fixed exposure.

Homophily is also only one possible source of confounding. If all farmers prefer highly productive partners, productive farmers may attract more connections even without a preference for similar partners. Unmeasured productivity can then confound the exposure comparison through popularity. A common market condition affecting both relationship formation and production can create a similar problem. BNS excludes residual dependence of

these kinds as well. Consequently, the presence of homophily does not by itself refute BNS, and its absence does not establish the assumption.

An input-allocation model makes the adjustment requirement concrete. Motivated by the labor-allocation mechanism in Duflo et al. (2026), suppose that g measures labor or inputs obtained through the specified relationships and that

$$Y_i^s(a,g) = \alpha(s,a,B_i) + \ell(s,a,g,B_i) + u_i\,, \qquad u_i \perp\!\!\!\perp \mathcal{W}_n \mid B_i\,. \tag{3.13}$$

Baseline characteristics may affect both partner choice and production. Unobserved pair affinity may also affect relationships and the allocation of labor, provided the formation inputs contain no further information about u_i conditional on B_i. Writing $\varepsilon_i = (B_i, u_i)$ gives BNS. If $\mathbb{E}[u_i \mid B_i = x]$ is common across target individuals, the model also gives the common conditional response mean required by Proposition 3.4. Training relationships can redirect labor in this model even without a productivity gain. Interpreting coffee-friend counts as the exposure additionally requires that those counts capture the relevant allocation consequences.

Heterogeneous productivity fits this model when baseline information reveals the productivity differences that affect selection. For example, suppose

$$Y_i^s(a,g) = \alpha(s,a,B_i) + \kappa(B_i)g + u_i\,, \qquad u_i \perp\!\!\!\perp \mathcal{W}_n \mid B_i\,.$$

Farmers may obtain different returns from the same labor input. They may also prefer partners with similar returns, or farmers with high returns may attract more partners. Within a group sharing the same B_i, however, the coefficient $\kappa(B_i)$ is fixed. Differences in exposure within that group therefore do not select farmers with different productivity coefficients. The assumption that productivity is a function of B_i is substantive: a noisy baseline production measure need not reveal the type well enough to justify it.

If productivity instead contains an unobserved component, write

$$Y_i^s(a,g) = \alpha(s,a) + \kappa_i g + u_i\,.$$

When κ_i predicts exposure after conditioning on B_i, farmers with different exposures may have different mean returns to the same allocation. This generally violates the identifying mean restrictions, whether the selection arises through homophily or through a common preference for productive partners. Unmeasured liquidity or reliability presents the same difficulty when it affects both partner choice and production. Baseline degree is sufficient for adjustment only if it removes the dependence that matters for these response comparisons.

The argument also applies to diffusion. Suppose B_i records baseline knowledge and learning ability, g counts trained contacts, and

$$Y_i^s(a,g) = \alpha(s,a,B_i) + \kappa(B_i)\{1-\exp(-g)\} + u_i\,, \qquad \kappa(B_i) \geq 0\,, \qquad u_i \perp\!\!\!\perp \mathcal{W}_n \mid B_i\,.$$

The gain from an additional trained contact diminishes with the number already present. Farmers may form ties with others who have similar knowledge or learning ability. If baseline information captures those characteristics and the remaining shock is conditionally independent of formation, BNS holds. As in the allocation example, a common conditional mean for u_i gives a common conditional response mean across individuals. Unobserved learning ability that predicts both connections and learning gains can invalidate this argument. Choosing an allocation or diffusion interpretation therefore leaves an empirical question: whether the available baseline information accounts for the selection relevant to outcomes under that mechanism.

4. ESTIMATION

Throughout this section, Section 5, and their proofs, all probabilities and expectations are conditional on the entire supported village-saturation profile $T = t$. I suppress this conditioning in notation. The number of villages J remains fixed, $n_j \to \infty$ for every j, and $n_j/n \to w_j > 0$. For a vector $v \in \mathbb{R}^{d_\theta}$, $\|v\| = (\sum_{1\le c\le d_\theta} v_c^2)^{1/2}$ denotes its Euclidean norm; matrix norms are operator norms unless specified otherwise.

Let $O_i = (B_i, Z_i, G_i, Y_i)$ and let $O = O_I$ for an index I selected uniformly from $[n]$, independently of the observations. For an integrable function h, write $\mathsf{P}_i h = \int h(o)\mathcal{L}(O_i)(\mathrm{d}o)$ and $\mathsf{P}h = n^{-1}\sum_{1\le i\le n}\mathsf{P}_i h$. Thus $\mathsf{P}h = \mathbb{E}[h(O)]$. For a fitted function $\widehat{h}$, $\mathsf{P}_i\widehat{h}$ is a random integral.

Assumption 4.1 (Common conditional laws)**.** For each n, the following equalities hold almost surely on every required assignment-cell support. There exist a probability measure $F_{B,n}$, measurable probability kernels $q_{z,n}$ and measurable functions $m_{z,n}$, common across individuals and supported profiles, such that, whenever $t_{j(i)} = s$ and $z = (s, a)$ is required,

$$\mathcal{L}(B_i) = F_{B,n}, \qquad \mathbb{P}(P_i = a \mid B_i) = s^a(1-s)^{1-a},$$
$$\mathcal{L}(G_i \mid B_i, Z_i = z) = q_{z,n}(\cdot \mid B_i), \qquad \mathbb{E}[Y_i \mid B_i, G_i, Z_i = z] = m_{z,n}(G_i, B_i).$$

For every required saturation,

$$\inf_{n,t\in\mathfrak{T}} \sum_{1\le j\le J} w_{j,n}\mathbb{1}\{t_j = s\} > 0.$$

Assumption 4.1 excludes village differences that remain confounded with saturation after conditioning on the allocation. To see why we need this restriction, consider two equally sized villages, one at each saturation, with $Y_{ji} = \alpha_j + \tau\mathbb{1}\{T_j = s_H\} + u_{ji}$ and $(\alpha_1, \alpha_2) = (1, 0)$. Let the errors be mutually independent, independent of assignment, and centered, with uniformly bounded variances. As village sizes grow, the high-minus-low difference in sample means converges in probability to $\tau + 1$ or $\tau - 1$, depending on the allocation. More individuals average out the errors but leave the village difference. With unrestricted village intercepts, the conditional comparison cannot separate that difference from τ. Design unbiasedness under repeated randomization, as in Proposition 3.1, does not resolve this conditional identification problem.

Assumption 4.1 requires common baseline distributions, conditional exposure laws, and conditional outcome means across individuals and supported saturation profiles. It allows several forms of heterogeneity. Treatment and network-mediated effects can vary with observed characteristics: training can have different effects for farmers with different productivity, landholdings, or network positions recorded in B_i. The return to an additional treated contact can depend on these characteristics, contacts summarized in G_i, own assignment, and saturation; the response can be nonlinear. Exposure distributions can also differ across baseline types and assignment cells through $q_{z,n}(\cdot \mid B_i)$. Unobserved individual heterogeneity is permitted insofar as the formation and response functions in Section 2 induce the required common exposure laws and conditional means. Conditional outcome variances and higher moments may differ across individuals, subject to the stated moment and dependence conditions. Villages may differ in size, realized covariate composition, and realized networks even though the specified population laws agree.

For $z = (s, a)$, define the realized saturation share and the assignment-cell probability of the selected observation by[2]

$$\lambda_{s,n} = \sum_{1\le j\le J} w_{j,n}\mathbb{1}\{t_j = s\}, \qquad e_z = \lambda_{s,n}s^a(1-s)^{1-a}. \tag{4.1}$$

Let $\mathcal{H}_n$ be a known deterministic graph on $[n]$, specified before the stochastic data are observed. Its closed neighborhood and maximum neighborhood size are

$$N[i] = \{i\} \cup \{k : \{i,k\} \in E(\mathcal{H}_n)\}, \qquad \Delta_n = \max_{i\in[n]} |N[i]| .$$

The graph describes possible dependence between complete observations.

Assumption 4.2 (Local dependence)**.** For disjoint $A, C \subseteq [n]$ with no edge between them,

$$\sigma(O_i : i \in A) \perp\!\!\!\perp \sigma(O_i : i \in C) .$$

The setwise condition requires independence of $\sigma(O_i : i \in A)$ and $\sigma(O_k : k \in C)$ whenever disjoint sets $A, C \subseteq [n]$ have no connecting edge. Pairwise independence is insufficient. If independent fair signs X_1, X_2 generate $X_3 = X_1X_2$, each pair is independent but X_3 is determined by the other two. For a cumulant, disconnection must split the entire tuple into independent groups; for neighborhood replacement, the evaluation observation must be independent of the complete retained training array. These are different uses of the same setwise condition. Dependency graphs built from shared primitive inputs appear in probability (Janson, 1988, Chen and Shao, 2004), while network causal-inference constructions track the observations' shared treatment and covariate inputs (Ogburn et al., 2024, Emmenegger et al., 2025). General network dependence need not be exact. Kojevnikov et al. (2021) instead bound dependence between separated sets, and Cao and Leung (2025) permit absolute regularity beyond the selected neighborhoods.

Local dependence is discussed by Cao and Leung (2025, Example 5); Assumption 4.2 imposes its setwise dependency-graph form, as in Chen and Shao (2004, Section 2.4.1). Defining a graph does not itself establish independence. The assumption requires that two collections with no connecting edge are independent under the sampling law. It applies to baseline summaries and exposures as well as outcomes because each enters the score. Appendix D derives this property when observations depend on specified sets of independent node and pair inputs. Dependence may cross village boundaries, and its extent is measured by Δ_n. A shared village shock can connect all n_j observations and violate the required growth rates. Appendix K allows dependence beyond these neighborhoods under explicit absolute-regularity rates, relating the argument to Kojevnikov et al. (2021), Cao and Leung (2025).

Suppress the subscript n on nuisance parameters. For a pair (z, z') set

$$r_{zz'}(g,x) = \frac{\mathrm{d}q_{z'}(\cdot \mid x)}{\mathrm{d}q_z(\cdot \mid x)}(g), \qquad \nu_{zz'}(x) = \int m_z(g,x) q_{z'}(\mathrm{d}g \mid x), \qquad \theta_{zz',n} = \int \nu_{zz'}(x) F_B(\mathrm{d}x) . \tag{4.2}$$

The statistical target θ_n collects a fixed number d_θ of these cells. It equals the corresponding causal target under the identification conditions in Section 3, applied to these common laws. If identification is approximate, write $b_{\mathrm{ID},n} = \|\theta_n - \theta_n^{\mathrm{causal}}\|$; consistency for the causal target requires $b_{\mathrm{ID},n} \to 0$, and its centered normal approximation requires $\sqrt{n} b_{\mathrm{ID},n} \to 0$. The results below concern the exact statistical functional.

With $\eta = (e_z, e_{z'}, m_z, r_{zz'}, \nu_{zz'})$, define

$$\psi_{zz'}(O;\eta) = \nu(B) + \frac{\mathbb{1}\{Z=z\}}{e_z(B)} r(G,B)\{Y - m(G,B)\} + \frac{\mathbb{1}\{Z=z'\}}{e_{z'}(B)}\{m(G,B) - \nu(B)\} . \tag{4.3}$$

[2]Under Assumption 4.1, $\mathbb{P}(Z = z \mid B) = e_z$ for the uniformly selected observation O. For a fixed individual the village saturation is already conditioned on. Equation (4.1) averages over individuals in villages assigned to arm s; the unconditional design inclusion probability in (3.1) additionally averages the first-stage lottery.

The score is the augmented mediation functional of Tchetgen Tchetgen and Shpitser (2012, Theorem 1); its nested-regression form is used for double machine learning by Farbmacher et al. (2022). The expectations and rates below account for dependence between individuals and for fitting on those same individuals.

Assumption 4.3 (Uniform boundedness)**.** There are constants $M \geq 1$, $R \geq 1$ and $e_* > 0$, independent of n and the supported profile, such that

$$|Y_i| \leq M, \quad |m_z|, |\nu_{zz'}| \leq M, \quad e_* \leq e_z \leq 1, \quad 0 \leq r_{zz'} \leq R\,.$$

All fitted functions, including those from the replacement samples below, are measurable and satisfy the same bounds.

Assumption 4.3 keeps outcomes and score weights uniformly bounded. It rules out assignment cells with vanishing probabilities and exposure distributions whose density ratios become arbitrarily large. The exposure itself may be continuous or have growing support. The fitted bounds can be enforced by clipping predictions to the corresponding intervals, with the same clipping rule applied to replacement fits. Under the assumed population bounds, this does not increase prediction errors or changes in individual nuisance predictions across replacement samples.[3] Clipping enforces restrictions on the estimates; the population boundedness and overlap conditions remain assumptions. Restricting the population to improve overlap would require corresponding changes to the target and assignment probabilities. Appendix L relaxes outcome boundedness under additional moment conditions. For my use below, these restrictions give

$$|\psi| \leq B_\psi := M + 2M(R+1)/e_*, \qquad |\psi - \theta| \leq B_\phi := B_\psi + M\,.$$

To measure estimation errors with nuisance parameters, define

$$\|h\|_z^2 = \int\int h(g,x)^2 q_z(\mathrm{d}g \mid x) F_B(\mathrm{d}x), \qquad \|v\|_B^2 = \int v(x)^2 F_B(\mathrm{d}x)\,.$$

For a collection $(h_c)_{c\in\mathcal{C}}$ indexed by the finite set $\mathcal{C}$ of required score cells, its nuisance norm is $\max_{c\in\mathcal{C}} \|h_c\|_c$, where each coordinate uses the baseline or exposure norm just defined. These norms integrate evaluation covariates and exposures; for estimated functions they are random variables. The density-ratio bound also gives $\|h\|_{z'} \leq \sqrt{R}\|h\|_z$, allowing prediction errors to be compared across exposure distributions.

Proposition 4.1 (Neyman orthogonality)**.** *Under Assumptions 4.1 and 4.3, $\mathbb{E}\psi(O;\eta_0) = \theta$ and $\mathbb{E}\psi(O_i;\eta_0) = \theta$ for every i. Expanding the expected score at any admissible candidate $\widetilde{\eta}$ gives the following identity: for $\delta m = \widetilde{m} - m$, $\delta r = \widetilde{r} - r$, $\delta\nu = \widetilde{\nu} - \nu$ and $\kappa_c = e_c/\widetilde{e}_c$, we have*

$$\mathbb{E}\psi(O;\widetilde{\eta}) - \theta = \mathbb{E}_B\big[(1-\kappa_{z'})\delta\nu + (\kappa_{z'} - \kappa_z)\mathbb{E}_{q_{z'}}[\delta m \mid B] - \kappa_z \mathbb{E}_{q_z}[\delta r \delta m \mid B]\big]\,. \tag{4.4}$$

The derivative of this expectation at η_0 is zero along every bounded differentiable admissible nuisance path. With the known probabilities (4.1), the remainder is $-\mathbb{E}_B \int \delta r \delta m \, \mathrm{d}q_z$, whose absolute value is at most $\|\delta r\|_z \|\delta m\|_z$.

With known assignment probabilities, the score is doubly robust. Its mean identifies θ if either the outcome regression or the exposure ratio is correct, even when the other is misspecified (Tchetgen Tchetgen and Shpitser, 2012, Farbmacher et al., 2022). Neyman orthogonality is the related local property that the first derivative of the mean with respect to the nuisances vanishes at their joint truth (Belloni et al., 2014, Chernozhukov et al.,

[3]For projection Π_I onto an interval I, $|\Pi_I(u) - v| \leq |u - v|$ for $v \in I$, and $|\Pi_I(u) - \Pi_I(u')| \leq |u - u'|$. The intervals are $[-M, M]$ for conditional means, $[0, R]$ for density ratios, and $[e_*, 1]$ for estimated assignment probabilities, with endpoints fixed before fitting. Known assignment probabilities require no clipping.

2022, 2018). The displayed product remainder establishes both properties. The mean cancellation allows any bounded outer regression; approximating the particular oracle score used for inference also requires learning its outer regression. If probabilities are fitted, the remainder is bounded by

$$\mathcal{B}(\widetilde{\eta}) \coloneqq e_*^{-1}\left(\|\delta e\|_B\|\delta\nu\|_B + 2\sqrt{R}\|\delta e\|_B\|\delta m\|_z + \|\delta r\|_z\|\delta m\|_z\right) ,$$

where $\|\delta e\|_B = \max_{c=z,z'} \|\widetilde{e}_c - e_c\|_B$. Appendix B shows the derivation of each term in detail. Upon estimating the nuisance parameters without sample splitting, I set

$$\widehat{\theta}_{zz'} = \frac{1}{n}\sum_{1\le i\le n} \psi_{zz'}(O_i;\widehat{\eta}), \qquad \widehat{\beta} = C\widehat{\theta}\,. \tag{4.5}$$

Here C is the fixed 4×5 matrix in (4.7), which converts the cell means into $(D, N, I, S)^\top$. Algorithm 1 fits pooled regressions using all n observations and assignment-cell indicators as predictors. Write $m(g,x,u) = \mathbb{E}[Y \mid G=g, B=x, Z=u]$, so $m_z(g,x) = m(g,x,z)$. A single fitted response surface $\widehat{m}$ supplies every required $\widehat{m}_z$ by setting its assignment argument to z at prediction. The learner allows assignment to interact with the other predictors; pooling imposes no common-slope restriction across cells.

For each required outcome arm z, generate $\widehat{m}(G_i, B_i, z)$ for every individual and regress these predictions on (B_i, Z_i) using all n observations. Denote the resulting bounded surface by $\widehat{v}_z(x,u)$ and set $\widehat{\nu}_{zz'}(x) = \widehat{v}_z(x,z')$ for $z \neq z'$. The population identity underlying this step is

$$\mathbb{E}[m(G,B,z) \mid B=x, Z=z'] = \int m_z(g,x)q_{z'}(\mathrm{d}g \mid x) = \nu_{zz'}(x)\,.$$

The assignment argument is fixed at z when constructing the generated response, whereas the observed assignment Z_i enters the outer regression as a predictor. Each z indexes a different generated response on the same full sample.

For $z = z'$, set $\widehat{r}_{zz} = 1$ and fit a single pooled regression $\widehat{h}(x,u)$ of Y_i on (B_i, Z_i) using all n observations. Set $\widehat{\nu}_{zz}(x) = \widehat{h}(x,z)$ after projection onto $[-M, M]$. This direct regression excludes the exposure and supplies all required regime means. The score becomes

$$\widehat{\psi}_{zz}(O_i) = \widehat{\nu}_{zz}(B_i) + \frac{\mathbb{1}\{Z_i = z\}}{e_z}(Y_i - \widehat{\nu}_{zz}(B_i))\,. \tag{4.6}$$

Use each estimated regime mean in every component that contains it. Order the cells as

$$\theta = (\theta_{H,11}, \theta_{H,10}, \theta_{H,01}, \theta_{H,00}, R_L(0))^\top\,.$$

Then

$$\begin{pmatrix}\widehat{D}\\ \widehat{N}\\ \widehat{I}\\ \widehat{S}\end{pmatrix} = \begin{pmatrix}0 & 1 & 0 & -1 & 0\\ 0 & 0 & 1 & -1 & 0\\ 1 & -1 & -1 & 1 & 0\\ 0 & 0 & 0 & 1 & -1\end{pmatrix}\widehat{\theta}\,. \tag{4.7}$$

The column sums are $(1,0,0,0,-1)$, so the estimated components sum to the same estimated total in every sample.

Having stated the estimation procedure, I turn to the challenge of establishing its validity when the same observations are used to fit the nuisance functions and evaluate the score. Neyman orthogonality controls the effect of nuisance estimation on the population mean of the score, but full-sample estimation also requires controlling the effect of using the same observations to fit and evaluate the nuisance functions. These are distinct issues in double/debiased machine learning: orthogonal scores reduce sensitivity to nuisance estimation errors, while cross-fitting controls the additional error arising from sample reuse (Chernozhukov et al., 2018). Without sample

Algorithm 1: Estimating the cell means and mediated-interference components

Input: One record $O_i = (B_i, Z_i, G_i, Y_i)$ per individual, $i \in [n]$; village membership and realized saturations; required directed cell pairs $\mathcal{C}$; contrast matrix C; regression learners $\mathcal{L}_m, \mathcal{L}_\nu$; a multiclass probability learner $\mathcal{L}_\pi$; fixed prediction bounds M, R; posterior denominator floor $\epsilon_{\pi,n} > 0$; fixed tuning rules and random seeds.

Output: Cell estimates $\widehat{\theta}$, component estimates $C\widehat{\theta}$, total estimate, and the $n \times |\mathcal{C}|$ matrix of individual cell scores.

Construct the prespecified individual outcomes, pretreatment variables and exposure map. Encode $Z_i = (T_i, P_i)$ by assignment-cell indicators, allowing interactions with other regression predictors. Every fit below uses all n observations. Retain their order and record missingness and the analysis population;

For every required $z = (s, a)$ compute $\lambda_{s,n} = n^{-1} \sum_{1 \le i \le n} \mathbb{1}\{T_i = s\}$ and $e_z = \lambda_{s,n} s^a (1-s)^{1-a}$. Stop and report an unsupported assignment cell if $e_z = 0$ or no training observations occur in a required arm;

Fit $\mathcal{L}_\pi$ to classify Z_i from (G_i, B_i) in the full analysis sample. At every prediction argument define

$$\widehat{\pi}_z(g, x) = \frac{\max(\widehat{\pi}_z^{\mathrm{raw}}(g, x), \epsilon_{\pi,n})}{\sum_{u \in \mathsf{Z}} \max(\widehat{\pi}_u^{\mathrm{raw}}(g, x), \epsilon_{\pi,n})} .$$

Fit $\mathcal{L}_m$ once to Y_i on (G_i, B_i, Z_i) using all $i \in [n]$. Project predictions onto $[-M, M]$ to obtain $\widehat{m}(g, x, u)$; set $\widehat{m}_z(g, x) = \widehat{m}(g, x, z)$;

Fit $\mathcal{L}_\nu$ once to Y_i on (B_i, Z_i) using all $i \in [n]$. Project predictions onto $[-M, M]$ to obtain $\widehat{h}(x, u)$ for the regime means;

foreach *distinct* z *appearing in a pair with* $z \neq z'$ **do**

  Generate $\widehat{m}(G_i, B_i, z)$ for every $i \in [n]$. Fit $\mathcal{L}_\nu$ to these responses on (B_i, Z_i) using all $i \in [n]$; project predictions onto $[-M, M]$ to obtain $\widehat{v}_z(x, u)$;

end

foreach *distinct pair* $c = (z, z') \in \mathcal{C}$ **do**

  if $z = z'$ **then**

    Set $\widehat{\nu}_{zz}(x) = \widehat{h}(x, z)$;

    Set $\widehat{r}_{zz} = 1$ exactly and, for every $i \in [n]$, set $\widehat{\psi}_{i,c} = \widehat{\nu}_{zz}(B_i) + \mathbb{1}\{Z_i = z\}\{Y_i - \widehat{\nu}_{zz}(B_i)\}/e_z$;

  else

    Set $\widehat{r}_{zz'}(g, x) = \mathrm{clip}_{[0,R]}(\widehat{\pi}_{z'}(g, x) e_z / [\widehat{\pi}_z(g, x) e_{z'}])$;

    Set $\widehat{\nu}_{zz'}(x) = \widehat{v}_z(x, z')$;

    For every $i \in [n]$, evaluate all three terms of (4.3) using $(e_z, e_{z'}, \widehat{m}_z, \widehat{r}_{zz'}, \widehat{\nu}_{zz'})$, and store their sum as $\widehat{\psi}_{i,c}$;

  end

  Set $\widehat{\theta}_c = n^{-1} \sum_{1 \le i \le n} \widehat{\psi}_{i,c}$. Store a shared regime mean only once;

end

Compute $(\widehat{D}, \widehat{N}, \widehat{I}, \widehat{S})^\top = C\widehat{\theta}$ and $\widehat{\mathrm{TE}} = \widehat{\theta}_{H,11} - \widehat{\theta}_{L,00}$;

return $\widehat{\theta}$, $C\widehat{\theta}$, $\widehat{\mathrm{TE}}$, *and* $(\widehat{\psi}_{i,c})_{i \in [n], c \in \mathcal{C}}$.

splitting, an observation can influence its own fitted score, and, under local dependence, information about that observation can also enter the fit through its dependence neighbors. Consequently, small prediction errors at an independent observation do not by themselves ensure that nuisance estimation has a negligible effect on the empirical score average. Moreover, even independent observations can have dependent fitted scores because they share the estimated nuisance functions.

I address this issue through neighborhood stability, following Cao and Leung (2025), who extend the stability approach of Chen et al. (2022) from i.i.d. to dependent data. The condition requires the complete fitting procedure to be insensitive, at the level of score predictions, to replacing dependence neighborhoods. Under the setwise local-dependence condition in Assumption 4.2, replacing an observation's entire closed neighborhood with observations from an independent copy of the array produces a nuisance estimate independent of that original observation. Stability then allows the empirical score to be compared with this independent-fit benchmark; replacing two neighborhoods provides the corresponding comparison needed to control covariance terms in the score

average. The nuisance accuracy and neighborhood stability requirements in Assumption 4.4, together with the product-remainder bound in Proposition 4.1 and the rates in Theorem 4.2, justify the asymptotic linear expansion while retaining all observations for estimation.

To state neighborhood stability, let $\widetilde{\boldsymbol{O}}$ be an independent copy of the entire array $\boldsymbol{O} = (O_i)_{i=1}^n$. For i, k, replace observations in $N[i] \cup N[k]$ by their copied observations and refit the complete procedure, obtaining $\widehat{\eta}^{(-i,-k)}$; put $\widehat{\eta}^{(-i)} = \widehat{\eta}^{(-i,-i)}$. This follows the replacement notation of Cao and Leung (2025). If the algorithm is randomized, write it as a measurable map of its data and an auxiliary random element U, independent of both arrays, and use the same U for all nuisance estimates. Let O_i^* have the marginal law of O_i and be independent of both arrays and U. Nuisance estimates computed from replacement observations are used only in the proof.

Assumption 4.4 (Nuisance accuracy and neighborhood stability)**.** There are deterministic sequences $\varpi_{\psi,n}, s_n, \varpi_m, \varpi_r, \varpi_\nu, \varpi_e \geq 0$ such that, for each required scalar score,

$$\max_i \max_{\eta^\dagger \in \{\widehat{\eta}, \widehat{\eta}^{(-j,-k)} : j,k \in [n]\}} \|\psi(O_i^*; \eta^\dagger) - \psi(O_i^*; \eta_0)\|_2 \leq \varpi_{\psi,n} \,,$$
$$\max_{i,k} \|\psi(O_i; \widehat{\eta}) - \psi(O_i; \widehat{\eta}^{(-i,-k)})\|_2 \;\vee\; \max_{i,k} \|\psi(O_i^*; \widehat{\eta}) - \psi(O_i^*; \widehat{\eta}^{(-i,-k)})\|_2 \leq s_n \,.$$

Here $\|X\|_2 = (\mathbb{E}|X|^2)^{1/2}$ integrates with respect to the joint distribution of the original array $\boldsymbol{O}$, its independent copy $\widetilde{\boldsymbol{O}}$, the independent evaluation observation O_i^* when present, and the auxiliary random element U. An in-sample O_i is a coordinate of $\boldsymbol{O}$ and is not integrated as an independent draw. In addition,

$$\|\widehat{m} - m\|_z = \mathrm{O}_{\mathbb{P}}(\varpi_m), \quad \|\widehat{r} - r\|_z = \mathrm{O}_{\mathbb{P}}(\varpi_r), \quad \|\widehat{\nu} - \nu\|_B = \mathrm{O}_{\mathbb{P}}(\varpi_\nu), \quad \|\widehat{e} - e\|_B = \mathrm{O}_{\mathbb{P}}(\varpi_e) \,.$$

All bounds are uniform over the finite profile support and required cells. With known probabilities take $\varpi_e = 0$.

The neighborhood-stability requirement is the score-specific counterpart of Cao and Leung (2025, Assumption 2). Their Assumption 3 separately controls learner accuracy relative to average neighborhood size. Here the first displayed bound controls mean-square prediction error at O_i^* for nuisance parameters estimated from either the original data or data with neighborhood replacement observations. The condition $\Delta_n \varpi_{\psi,n}^2 \to 0$ is sufficient for their accuracy requirement when the maximum neighborhood size bounds the average. The second displayed bound controls how score predictions change after replacing $N[i] \cup N[k]$. Its first term evaluates both nuisance estimates on the original observation O_i; its second evaluates them on O_i^*, independent of both data arrays. Requiring $\sqrt{n} s_n \to 0$ gives their neighborhood-stability rate. Their theorem also allows nonzero absolute regularity beyond the chosen neighborhoods. Exact setwise independence here sets those additional dependence errors to zero, at the cost of a stronger dependence restriction.

The replacement data need not have the same joint law as the original array. The assumption therefore applies to the complete nuisance estimation procedure under each replacement law, including generated responses and tuning choices. Consistency established only for the original sample does not verify these mean-square bounds. Lemma B.3 shows exactly where independent evaluation and evaluation on a training observation enter the proof.

Theorem 4.2 (Consistency and asymptotic linearity)**.** *Maintain Assumptions 4.1, 4.2, 4.3, and 4.4, and put* $b_n = \varpi_e \varpi_\nu + \varpi_e \varpi_m + \varpi_r \varpi_m$.

(i) If $\Delta_n / n \to 0$, $s_n \to 0$, *and* $b_n \to 0$, *then* $\|\widehat{\theta}_n - \theta_n\| \xrightarrow{\mathbb{P}} 0$.

(ii) If

$$\Delta_n \varpi_{\psi,n}^2 \to 0, \qquad \sqrt{n}\, s_n \to 0, \qquad \sqrt{n}\, b_n \to 0 \,, \tag{4.8}$$

then, with $\phi_{i,n} = \psi(O_i; \eta_{0,n}) - \theta_n$,

$$\sqrt{n}(\widehat{\theta}_n - \theta_n) = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \phi_{i,n} + \mathrm{o}_{\mathbb{P}}(1) \,, \qquad \mathbb{E}\phi_{i,n} = 0 \,. \tag{4.9}$$

Part (i) proves consistency in the Euclidean norm of the fixed-dimensional cell vector. Consistency of the decomposition follows from $\|C(\widehat{\theta}_n - \theta_n)\| \le \|C\|_{\mathrm{op}} \|\widehat{\theta}_n - \theta_n\|$. The cells are the causal means under Section 3's assumptions. If those means converge to a fixed causal parameter θ, the triangle inequality also gives $\|\widehat{\theta}_n - \theta\| \xrightarrow{\mathbb{P}} 0$; inference centered at θ requires the stronger approximation rate in Section 5. With known assignment probabilities, bounded nuisance errors and either $\varpi_m \to 0$ or $\varpi_r \to 0$ give $b_n \to 0$; both need not be consistently estimated for the result to hold. For consistency, bounded component rates and vanishing errors in at least one factor of every nonzero product suffice. For example, with known assignment probabilities and $\varpi_m = \mathrm{O}(n^{-a})$, $\varpi_r = \mathrm{O}(n^{-b})$, $a, b > 0$, the condition is automatic; asymptotic linearity instead requires $a + b > 1/2$. This is the same product-rate condition as in Chernozhukov et al. (2018, Assumption 3.2 and Section 5). Higher-order orthogonalization can reduce sensitivity to slower nuisance estimation in other models (Bonhomme et al., 2026a,b); using those constructions here would require a new score and proof. The present theorem retains its explicit second-order product restriction. The proof uses neighborhood replacement and double centering as in Cao and Leung (2025), Chen et al. (2022), with the setwise independence in Assumption 4.2 justifying each conditional expectation.

Remark 4.1 (A prespecified analysis population)**.** For an analysis restricted to a prespecified set $\mathcal{I}_n \subseteq [n]$, write $n^\circ = |\mathcal{I}_n|$ and $n_j^\circ = |\mathcal{I}_n \cap \{i : j(i) = j\}|$. Replace the analysis size by n°, require $n_j^\circ / n^\circ \to w_j^\circ > 0$, and form cell shares from these retained counts. Neighborhoods $\mathcal{N}_n(i, r)$ and their unions are defined on the original dependence structure before restriction, following the replacement construction in Cao and Leung (2025). Appendix D specifies the conditions for this restriction. Trimming based on observed exposures or estimated overlap changes the conditional laws and can change the causal population; it requires a corresponding selection analysis. ♢

Remark 4.2 (Cross-fitting with separated training sets)**.** Cross-fitting with separated training sets partitions the evaluation observations and, for each part, removes that part and all its dependence neighbors from nuisance training. Under Assumption 4.2, the retained training observations are independent of the entire evaluation part. This is a network adaptation of cross-fitting, related to the construction of Emmenegger et al. (2025); ordinary random folds do not in general provide this exact separation. It is useful when the influence of observations on the complete learner is difficult to bound and the retained training regions have enough observations in every required arm. Appendix J proves consistency and asymptotic linearity for that procedure. On an expanding spatial village, interior regions can remain large after a boundary buffer is removed. On a densely connected network, the same buffer may remove almost the entire training sample. In that case the proposed split does not meet the theorem's assumptions. Full-sample fitting under Assumption 4.4 avoids discarding nearby training observations and removes randomness due to fold selection (Cao and Leung, 2025, Ritzwoller and Romano, forthcoming). Separation is sufficient for the conditional-independence proof in Appendix J, but is not necessary for every analysis of cross-fitting with dependent data. Balkus et al. (2026) examine ordinary folds with correlated observations and argue that complete separation can be unnecessary. ♢

Remark 4.3 (g-estimation and structural mean models)**.** Structural nested mean models specify a treatment "blip", the change in a conditional outcome mean induced by removing a treatment, potentially as a function of the observed treatment and covariate history. In g-estimation, the blip is subtracted from the observed outcome and its

parameters are chosen to make this adjusted outcome orthogonal to treatment residuals under sequential exchangeability (Robins, 1994, Vansteelandt and Joffe, 2014). The model is semiparametric when the blip has finitely many parameters while baseline response and assignment laws remain unrestricted. This approach has been used for controlled direct effects (Vansteelandt, 2009) and time-varying effect moderation (Almirall et al., 2010, Jaman et al., 2025). The identification strategy here instead uses exposure sufficiency, baseline adjustment, and conditional exposure laws to identify a nonparametric mediation functional. It leaves the treatment–exposure response unrestricted and estimates its contrasts with an augmented score. Although the conditional-mean cancellations in (4.3) resemble g-estimation moments, they do not impose a finite-dimensional blip model. Appendix H gives such a model as an additional specification and derives the moment and rank conditions needed for its coefficients to have the proposed causal interpretation. $\diamondsuit$

5. Inference

Individuals connected through assignments, relationships or baseline information can have correlated scores, so their covariance contributions must be included. I first give a normal approximation under the growing-neighborhood conditions, then construct a feasible, consistent variance estimator. As in Section 4, all probability statements condition on the complete village-saturation profile, with that conditioning suppressed.

5.1. Normal approximation.

Let $\phi_{i,n}$ be the oracle score vector in (4.9). Define

$$V_n = \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{1\le i\le n}\phi_{i,n}\right) = \frac{1}{n}\sum_{1\le i,k\le n}\mathbb{E}[\phi_{i,n}\phi_{k,n}^\top]\,, \tag{5.1}$$

where $\phi_{i,n} = \psi(O_i;\eta_{0,n}) - \theta_n$. The sum includes within-village and between-village covariances permitted by the dependence graph. The number of villages stays fixed. A finite variance limit requires the aggregate of these covariances to remain of order n even when individual neighborhoods grow.

Assumption 5.1 (Variance limit and degree growth)**.** For each supported profile there is a deterministic symmetric $d_\theta \times d_\theta$ matrix V such that

$$V \succeq 0, \qquad \|V\|_{\mathrm{op}} < \infty, \qquad \|V_n - V\|_{\mathrm{op}} \to 0, \qquad \frac{\Delta_n^4}{n} \to 0\,,$$

and $V = \lim_{n\to\infty} V_n$ exists.

The limit in Assumption 5.1 is positive semidefinite and may be singular. For a sufficient condition for studentization, I impose the additional restriction

$$\liminf_{n\to\infty} c^\top V_n c > 0\,, \qquad \text{or, for a joint test,} \qquad \liminf_{n\to\infty} \lambda_{\min}(R_0 V_n R_0^\top) > 0\,, \tag{5.2}$$

where $c \in \mathbb{R}^{d_\theta}$ is fixed and $R_0 \in \mathbb{R}^{r\times d_\theta}$ is fixed with full row rank. Under Assumption 5.1, these restrictions are equivalent to $c^\top V c > 0$ and $R_0 V R_0^\top \succ 0$, respectively. A lower bound on the smallest eigenvalue of the entire V_n would exclude exact identities among the reported cells or components. For a joint test, the rows of R_0 specify the r combinations whose sampling variances must remain positive. The displayed eigenvalue condition provides a constant $\underline{v} > 0$ such that $u^\top R_0 V_n R_0^\top u \ge \underline{v}\|u\|^2$ for every $u \in \mathbb{R}^r$ and all sufficiently large n. It bounds the norm of the inverse covariance used in the Wald statistic by $\underline{v}^{-1}$. In Emmenegger et al. (2025, Assumption 2), the scalar limiting variance lies between fixed positive constants L and U. Their Lemma E.3 uses that lower bound to control the inverse powers σ_N^{-2} and σ_N^{-3} in its Stein-method error bound, equation (A4). Equation (5.2) plays

the corresponding role for scalar studentization or the chosen joint test here. Their result is uniform over a class of laws, whereas my result is pointwise.

Let θ denote the vector of natural or interventional means under a fixed target causal law. Assumption D.1 in Appendix D gives sufficient conditions on the response functions, exposure kernels, and baseline distributions for the finite-n causal means to approach θ faster than $n^{-1/2}$. This approximation is separate from nuisance estimation.

Theorem 5.1 (Joint normal approximation)**.** *Suppose Assumptions 4.1, 4.2, 4.3, 4.4, and 5.1 hold, together with the learning rates in (4.8). Then,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow{d} \mathcal{N}(0, V) \,,$$

where V is the long-run variance stated in Assumption 5.1. Suppose, further, that for every required within-saturation cell $z = (s, a)$, $z' = (s, b)$, Assumptions 2.1, 2.2, 3.1, 3.2, and 4.1 hold. Then,

$$\theta_{zz',n} = \theta^{\mathrm{int}}_{s,n}(a, b) \,.$$

If, on top, Assumption 3.3 holds as well, then $\theta_{zz',n} = \theta^{\mathrm{nat}}_{s,n}(a, b)$. Finally, if Assumption D.1 additionally holds, then the causal vector θ defined in Appendix D.2 satisfies

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V) \,. \tag{5.3}$$

The first conclusion combines an asymptotic-linear expansion with a normal approximation for the oracle score sum. Assumption 4.1 enables an appropriate centering argument as detailed in (B.1), and Assumption 4.3 bounds the scores. Assumption 4.2 supplies the setwise independence used in both neighborhood replacement and the connected-cumulant argument. Under Assumption 4.4, the rates $\Delta_n \varpi^2_{\psi,n} \to 0$ and $\sqrt{n}\, s_n \to 0$ make the centered score error negligible by Lemma B.3, while $\sqrt{n}\, b_n \to 0$ controls the mean remainder in Proposition 4.1. Theorem 4.2(ii) then gives the asymptotic-linear expansion. For the oracle sum, Appendix C bounds each cumulant of fixed order $k \geq 3$ by a constant times $n^{1-k/2}\Delta_n^{k-1}$. The degree restriction in Assumption 5.1 makes these cumulants vanish, and its covariance restriction gives the limit V. The cumulant method follows the dependency graph approach of Janson (1988); see Döring et al. (Section 3.4, 2022) for an exposition. Extensions to weighted dependency graphs (Féray, 2018), specified Ising-model regimes (Dousse and Féray, 2019), and stabilizing geometric functionals (Eichelsbacher et al., 2015) require their own dependence and moment conditions. The causal interpretations follow from the identification results and subsequent discussions in Section 3.

The final conclusion additionally requires Assumption D.1. Inference centered at θ_n does not require $\theta_n - \theta = o(n^{-1/2})$. Emmenegger et al. (2025, Theorem 2.2) center their result at the sample-size-dependent expected average treatment effect θ^0_N. Parts 1 and 3 of their Assumption A4 control nuisance-error products and their relation to dependency degree, respectively; these correspond to the learning conditions here and do not provide a fixed-target recentering rate. In a different setup, Leung (2025b, Section 4.2) require the interference-approximation error to vanish faster than the standard error. Here the corresponding requirement concerns the difference between the finite-n causal mean and the fixed population causal mean. Appendix D.2 gives sufficient conditions using invariant component laws and coupled exposure laws, and explains how cancellation can weaken the requirement for a particular contrast.

5.2. **Estimating the covariance.** I now estimate the covariance of the oracle score sum. The empirical products $\widehat{\phi}_i \widehat{\phi}_k^\top$ estimate contributions to this covariance. A feasible estimator must retain enough of those cross-products to make its truncation or weighting bias vanish, while averaging enough distinct contributions to control its sampling error. This is the covariance-estimation problem studied in spatial and network inference (Conley, 1999, Kelejian

and Prucha, 2007, Kojevnikov et al., 2021, Leung, 2022a). For limit theory under spatial mixing and heterogeneous random fields, see also Jenish and Prucha (2009). I use deterministic positive-semidefinite weights and impose a separate rate on the error from replacing oracle scores by fitted scores.

Choose a deterministic symmetric matrix $W_n = (w_{ik,n})$, which may depend on the conditioned profile and the specified dependence geometry. Write $L_n = \max_i |\{k : w_{ik,n} \neq 0\}|$ for the maximum number of its nonzero entries in a row. Let $\widehat{\phi}_i = \psi(O_i; \widehat{\eta}) - \widehat{\theta}_n$ and $\|u\|_{2,n}^2 = n^{-1} \sum_{1 \le i \le n} \|u_i\|^2$.

Assumption 5.2 (Covariance weights and score error)**.** The weights and fitted scores satisfy

$$W_n \succeq 0, \qquad w_{ii,n} = 1, \qquad |w_{ik,n}| \le 1 \,,$$

$$\left\| \frac{1}{n} \sum_{1 \le i,k \le n} (w_{ik,n} - 1) \mathbb{E}[\phi_i \phi_k^\top] \right\|_{\mathrm{op}} \to 0 \,, \qquad \frac{\Delta_n L_n^2}{n} \to 0 \,, \tag{5.4}$$

$$L_n \|\widehat{\phi} - \phi\|_{2,n} \xrightarrow{\mathbb{P}} 0 \,. \tag{5.5}$$

For cross-fitting with separated training sets, the same conditions apply to the fold-specific scores.[4]

The three restrictions address the issues that also arise in spatial and network HAC estimation. The covariance bias corresponds to the approximation controlled by bandwidth and dependence-tail conditions in Conley (1999), Kelejian and Prucha (2007), Kojevnikov et al. (2021). Under exact dependence, only pairs joined by the dependency graph contribute to that bias. The rate $\Delta_n L_n^2/n \to 0$ controls sampling variation by counting pairs of score products that can be dependent. It is a sufficient bound and need not be sharp for a particular geometry. The last restriction controls estimated scores. Each score error enters at most L_n weighted products, explaining why prediction consistency alone is insufficient when the bandwidth grows. Related nuisance-error restrictions appear in Emmenegger et al. (2025); here L_n records the cost of estimating a covariance with expanding support. Positive semidefiniteness ensures nonnegative estimated variances, while the other restrictions control approximation and estimation error.

Proposition 5.2 (Feasible network covariance)**.** *Under Assumptions 4.1, 4.2, 4.3, 5.1, and 5.2,*

$$\widehat{V}_n^{\mathrm{net}} = \frac{1}{n} \sum_{1 \le i,k \le n} w_{ik,n} \widehat{\phi}_i \widehat{\phi}_k^\top \,, \qquad \|\widehat{V}_n^{\mathrm{net}} - V\|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0 \,. \tag{5.6}$$

Under Theorem 4.2(ii), a sufficient fitted-score rate for (5.5) is $L_n(\varpi_{\psi,n} + s_n + n^{-1/2}) \to 0$. The proof first bounds the variance of each oracle quadratic form by $8B_\phi^4 \Delta_n L_n^2/n$, then expands the feasible-minus-oracle covariance into two cross-products and an error product. The replacement argument bounds the prediction error on the observations used in training.

Remark 5.1 (Constructing covariance weights)**.** A covariance kernel built from the observed friendship graph is random and can depend on treatment and outcome primitives. The construction in Appendix D instead uses the specified dependence geometry to form a positive-semidefinite matrix of neighborhood overlaps. It ensures nonnegative variance estimates and gives explicit bounds on the number of observations receiving weight and the resulting covariance bias. These properties make the conditions in Assumption 5.2 verifiable in expanding lattice examples. Using an estimated or treatment-responsive geometry additionally requires controlling its effect on these bounds. ♢

[4] For i in evaluation part k, use $\widehat{\eta}_k$, its evaluation-cell probabilities, and the corresponding oracle score from Appendix J. This preserves the conditional centering used in the covariance proof.

Corollary 5.1 (Studentized contrasts)**.** *Under the finite-n conclusion of Theorem 5.1 and Proposition 5.2, a contrast $c^\top \theta_n$ with $c^\top V c > 0$ has an asymptotic $1-\alpha$ confidence interval*

$$c^\top \widehat{\theta}_n \pm z_{1-\alpha/2}\sqrt{c^\top \widehat{V}_n^{\mathrm{net}} c/n}\,, \tag{5.7}$$

where $z_{1-\alpha/2}$ is the standard-normal quantile. Under $R_0\theta_n = b_0$, with R_0 as specified after Assumption 5.1,

$$n(R_0\widehat{\theta}_n - b_0)^\top (R_0 \widehat{V}_n^{\mathrm{net}} R_0^\top)^{-1}(R_0\widehat{\theta}_n - b_0) \xrightarrow{d} \chi^2_r\,.$$

If Assumption D.1 also holds, these conclusions apply to the fixed contrast $c^\top\theta$ and the fixed null $R_0\theta = b_0$.

Finally, in Appendix C.3, I discuss practical guidance for using the larger of the network and village-cluster variance estimates in order to account for cross-village dependence while retaining the wider interval when village clustering yields a larger standard error. This follows Leung (2025b), who emphasizes that cross-cluster covariance terms can matter in finite samples even when they vanish asymptotically. I show that the resulting interval preserves asymptotic coverage whenever at least one candidate provides valid inference for the correctly centered contrast, and extend the argument to joint tests. Appendix G.6 reports how these variance choices affect the conclusions from my empirical illustration.

6. Empirical Illustration to Duflo et al. (2026)

Duflo et al. (2026) study TechnoServe's agronomy-training program for coffee farmers in Southern Rwanda. Training improved knowledge and self-reported adoption and created friendships among participants, but the authors find little evidence that knowledge diffused to untreated farmers through baseline friendships. They also find suggestive evidence of adverse spillovers at higher treatment concentrations. I use their data to examine whether the outcome response to training operates through changes in exposure to trained friends, and whether the response to that exposure differs between trained and untreated farmers.

6.1. Setting and data.

The experiment assigned 27 villages to treatment concentrations of 25%, 50%, or 75%, then allocated 855 of 1,594 applicant households to training and 739 to control. Training groups met monthly; some included farmers from neighboring villages. Throughout, treatment means an offer of training, regardless of attendance. The data include ten survey rounds and "coffee" friend networks measured before and after the program.

A coffee friend is "farmers to whom the head or spouse in each household talked about growing coffee" (Duflo et al., 2026). The mediator that I use in my main reanalysis is the number of treated "coffee friends" in the endline network. It averages 1.77 among households with a network record; 22.9% report none. I retain the count in levels. I examine knowledge, self-reported adoption, audited adoption, leaf health, inputs, yields, and net revenue per tree. The first five retain the study's index scales; yields are kilograms of coffee cherries per tree and revenue is Rwandan francs per tree. Knowledge, adoption and leaf health use rounds 6, 8, and 9; the other outcomes use rounds 6–9. Tree audits are averaged within household and round before estimation. Each available household–round observation receives equal weight, so households with more observed rounds contribute more. The outcome samples contain 1,566–1,577 households.

To account for baseline differences that may jointly shape farmers' social connections and outcomes, I condition on 16 pretreatment variables. Baseline harvest per tree and total coffee trees measure initial productivity and the scale of production, while ten indicators of agronomic practices record how farmers managed their coffee before

Algorithm 2: Covariance estimation, confidence intervals, and joint tests

Input: Cell estimates $\widehat{\theta}$ and individual scores $\widehat{\psi}_i = \psi(O_i; \widehat{\eta})$ from the pooled full-sample fits in Algorithm 1; deterministic covariance weights W_n satisfying Assumption 5.2; scalar contrasts c and null values b_c; optional joint restriction (R_0, b_0); error probability $\alpha \in (0,1)$; optional nonnegative scalar variance candidates or symmetric matrix candidates on the scale of V_n.

Output: The cell covariance estimate, contrast standard errors and intervals, scalar p-values, and any prespecified joint test.

Use the saved scores from Algorithm 1 for all n individuals, with their original individual and cell ordering. Check $\widehat{\theta} = n^{-1}\sum_{i\in[n]}\widehat{\psi}_i$;

Check $W_n = W_n^\top$, $w_{ii,n} = 1$, $|w_{ik,n}| \le 1$, and positive semidefiniteness, up to a recorded numerical tolerance. Record the dependence graph, bandwidth and maximum row support L_n;

Set $\widehat{\phi}_i = \widehat{\psi}_i - \widehat{\theta}$ for every $i \in [n]$ and form the matrix $\widehat{\Phi}$ whose ith row is $\widehat{\phi}_i^\top$;

Compute $\widehat{V}^{\mathrm{net}} = \widehat{\Phi}^\top W_n \widehat{\Phi}/n$. Symmetrize only for floating-point roundoff. The covariance estimate for $\widehat{\theta}$ is $\widehat{V}^{\mathrm{net}}/n$, and that for the components is $C\widehat{V}^{\mathrm{net}}C^\top/n$;

foreach *prespecified scalar contrast* c **do**

  Set $\widehat{b}_c = c^\top\widehat{\theta}$ and $\widehat{v}_c = c^\top\widehat{V}^{\mathrm{net}}c$;

  if *a second variance candidate* $\widehat{v}_{2,c}$ *is reported* **then**

    Record both candidates and set $\widehat{v}_c = \max(\widehat{v}_c, \widehat{v}_{2,c})$. Record which candidate is the justified benchmark;

  end

  if $\widehat{v}_c > 0$ *and finite* **then**

    Set $\widehat{\mathrm{se}}_c = \sqrt{\widehat{v}_c/n}$ and $T_c = (\widehat{b}_c - b_c)/\widehat{\mathrm{se}}_c$;

  end

end

if *a joint restriction is requested* **then**

  Check that the rows of the fixed matrix R_0 are linearly independent after the exact component identities, and set $r = \mathrm{rank}(R_0)$;

  Set $A = \widehat{V}^{\mathrm{net}}$. If a second symmetric matrix B is used, diagonalize $B - A = U\,\mathrm{diag}(d_\ell)U^\top$ and set $A \leftarrow A + U\,\mathrm{diag}(\max(d_\ell, 0))U^\top$;

  Set $\widehat{\Sigma}_0 = R_0 A R_0^\top$. If it is positive definite, compute $Q = n(R_0\widehat{\theta} - b_0)^\top\widehat{\Sigma}_0^{-1}(R_0\widehat{\theta} - b_0)$ and report $1 - F_{\chi^2_r}(Q)$; reject at level α when $Q > \chi^2_{r,1-\alpha}$;

end

return *Estimates, intervals,* p*-values.*

training.[5] I also include the household head's sex and years of schooling, household size, and the number of baseline coffee friends outside the randomized roster. These variables account for differences in household composition, education, and existing contacts beyond the experimental sample. Such differences are relevant in this setting because social connections may affect both access to farming information and access to labor and other inputs, as discussed in Duflo et al. (2026, Section 3 and Online Appendix A).

The main identifying assumption, BNS, is that, conditional on these baseline characteristics, the factors governing friendship formation contain no further information about a farmer's outcome at fixed treatment and exposure. The case for this adjustment set is that farmers' production and management choices before training may reveal persistent differences that also shape whom they discuss coffee farming with. Baseline harvest and practices provide information about farming ability and prior adoption, while tree holdings and household size help account for differences in production scale and potential family labor availability. Conditioning on these measures therefore makes BNS more plausible to the extent that they capture the sources of selection into coffee friendships that

[5] The indicators cover rejuvenation, removal of dead branches and suckers, weeding, lime and NPK use, pesticide application, mulching, composting, and record keeping.

also matter for subsequent outcomes. The assumption remains substantive: these measures only indirectly capture labor and financing constraints, and the number of non-roster friends provides limited information about existing network position. This matters especially given the original study's proposed link between social ties and input allocation. I therefore interpret the decomposition under the assumption that these remaining differences do not jointly determine friendship formation and outcomes after adjustment.[6]

6.2. **Estimation and interpretation.** Following the original analysis, my main reanalysis is between a farmer offered training in a village with 75% treatment saturation and a farmer not offered training in a village with 25% saturation. Setting $s_H = 0.75$ and $s_L = 0.25$ in (6.1) gives

$$\text{TE} = R_{0.75}(1) - R_{0.25}(0) = S + D + N + I \,. \tag{6.1}$$

The saturation component S measures the effect of moving between these regimes for untreated farmers. The remaining components decompose the treatment contrast within the 75% regime: D changes own treatment while holding exposure at its control distribution, N changes the exposure distribution while holding the response at its control level, and I captures how the effect of that exposure change differs with own treatment. Thus, $N + I$ measures the exposure effect for a treated recipient. These components concern changes in the distribution of treated-friend exposure rather than the effect of one additional friend. Their interventional interpretation uses independent exposure draws conditional on baseline information; a natural-effects interpretation additionally requires the cross-world restriction.

I estimate the decomposition using the pooled regressions and augmented scores in Algorithm 1, with random forests used for estimating the conditional means and assignment probabilities. Specifically, I estimate the outcome response and the probabilities of assignment cells conditional on exposure and baseline information, then regress the generated outcomes on baseline information and assignment cells. A separate outcome regression excluding exposure supplies the regime means. All forests resample households, keeping their observed rounds together, and form out-of-bag predictions by averaging over trees whose fitting samples exclude the household.[7]

To relate these estimates to the original study, Appendix M characterizes the probability limits of its linear regression coefficients. Under the conditions stated there, equation (2) of Duflo et al. (2026) estimates a weighted within-village total $D_s + N_s + I_s$, while the high-saturation coefficient in equation (5) estimates S. Equation (3) instead reports slopes on assigned baseline friends separately for treated and control farmers, and equation (4) estimates a population saturation-policy contrast. Comparisons with the original results therefore depend on which of these quantities is being estimated.

6.3. **Results.** Table 1 and Figure 1 displays the estimates and 95% intervals. Overall, results show that training improves knowledge and self-reported adoption, with both gains concentrated in the direct component. For each outcome, the estimated direct component exceeds the total effect, while the network component and interaction are smaller and statistically indistinguishable from zero. Thus, the decomposition provides little evidence that changes in exposure to trained friends account for these improvements. The knowledge result is qualitatively consistent with the limited evidence of diffusion in Duflo et al. (2026), although their analysis concerns baseline friendships and a different estimand.

[6]Using treated-friend count as the exposure measure additionally requires that friends' identities and other characteristics have no further effect on outcomes conditional on own treatment, the count, and baseline information.

[7]Each forest has 300 trees and a minimum node size of five. I floor posterior assignment probabilities at 0.01 and cap exposure-density ratios at 100. The nuisance regressions include assignment-cell indicators where appropriate and omit round fixed effects. Appendix G.1 provides implementation details.

TABLE 1. Training, network exposure, and saturation spillovers

Outcome	Total (TE)	Spillover (S)	Direct (D)	Network (N)	Interaction (I)
Knowledge	1.298***	-0.122**	1.574***	0.029	-0.183
	(0.107)	(0.059)	(0.194)	(0.055)	(0.151)
Self-reported adoption	0.324***	-0.061	0.345***	-0.054	0.094
	(0.093)	(0.106)	(0.071)	(0.052)	(0.070)
Audited adoption	-0.069	-0.152**	0.036	-0.055	0.102**
	(0.053)	(0.069)	(0.060)	(0.045)	(0.049)
Leaf health	-0.001	-0.038	0.005	0.025	0.007
	(0.032)	(0.040)	(0.047)	(0.038)	(0.046)
Inputs	0.075	-0.110	0.135*	-0.124**	0.174**
	(0.069)	(0.072)	(0.075)	(0.052)	(0.068)
Yields (kg/tree)	-0.028	-0.113*	0.064	0.024	-0.003
	(0.054)	(0.063)	(0.050)	(0.043)	(0.065)
Net revenue (RWF/tree)	-12.036	-26.181**	11.019	3.929	-0.802
	(11.441)	(13.210)	(12.091)	(8.246)	(13.541)

Notes. The total compares 75%-village treatment with 25%-village control. Components are defined in (2.12) and sum to the total before rounding. The mediator is the endline treated-friend count. Network standard errors with household score aggregation and radius four are in parentheses. Units are index points, kg/tree for yields and RWF/tree for net revenue. Inference is pointwise, without adjustment for multiple testing, under the qualifications in Appendix G.6. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The input results illustrate what the decomposition adds to the analysis. A small total effect masks a negative network component and a positive interaction, both statistically significant and comparable in magnitude to the direct component. The negative network component implies that shifting exposure from its distribution without training to its distribution with training reduces inputs when the recipient's own treatment is held at control. The positive interaction shows that own training changes the response to this same exposure shift. Its magnitude more than offsets the negative network component, leaving a small positive, but statistically insignificant, exposure effect for trained recipients. Thus, the evidence is strongest for a difference in how trained and untreated recipients respond to the exposure change. This difference would be obscured by considering only the total effect or the exposure effect for trained recipients.

This distinction goes beyond the original study's comparison of exposure responses along baseline friendships. Those regressions allow responses to differ by own treatment, but do not isolate the contribution of the change in endline exposure induced by own training; see Appendix M.3. Under the identifying assumptions of the decomposition, the estimates separate this contribution from the extent to which own training changes its effect on inputs. Audited adoption exhibits a similar positive and statistically significant interaction, although its direct and network components are imprecisely estimated. Together, these findings suggest that training can alter the consequences of changes in social exposure, in addition to changing exposure itself. This interpretation is compatible with the input-allocation mechanism in Duflo et al. (2026, Section 3 and Online Appendix A), without identifying labor exchange as the channel measured by the coffee friend count.[8]

[8]The broader comparisons are in Appendix G.3. They show that the estimated within-village effects on knowledge and self-reported adoption are close in magnitude to the original estimates, as is the population-policy effect on knowledge. These similarities hold despite differences in weighting and estimation. Statistical significance is less directly comparable because the original study uses randomization inference, whereas the tests here use a normal approximation.

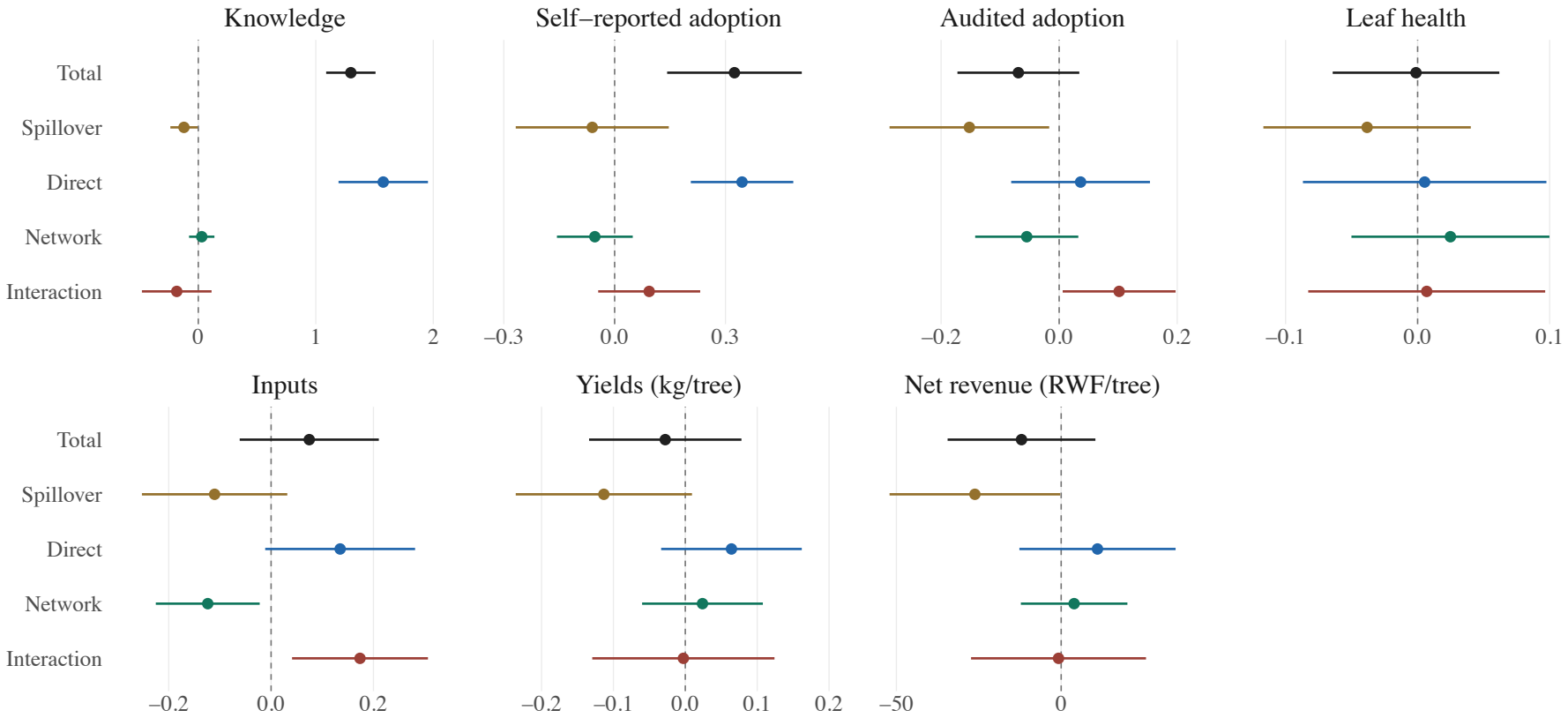


FIGURE 1. The four-term decomposition with endline treated-friend exposure. Dots reproduce the estimates presented in Table 1; bars indicate 95% confidence intervals using the reported network HAC standard errors from Algorithm 2. Each panel has its own horizontal scale.

7. Monte Carlo evidence

The simulations examine the finite-sample properties of the proposed estimation and inference procedures under the following DGP.

There are $J = 18$ villages, with nine assigned to each saturation in $\{0.25, 0.75\}$. Only these two saturations enter the five target cells, so a third arm would add individuals without adding information about any required conditional law; the design therefore uses two arms. The primary size sequence is

$$v_j = 0.70 + 0.60(j-1)/17, \qquad n_j = \lfloor v_j \overline{n} + 1/2 \rfloor, \qquad \overline{n} \in \{450, 900, 1800, 3600\} .$$

Thus total size is $n = 18\overline{n} \in \{8100, 16200, 32400, 64800\}$, and village sizes range from 315 to 4680. The primary exercise conditions on one profile assigning villages 1 through 9 to saturation 0.25 and villages 10 through 18 to saturation 0.75. A separate exercise redraws the first-stage allocation by uniformly permuting its entries. Conditional on that allocation, $P_i = \mathbb{1}\{\xi_i \le T_i\}$ for independent uniform ξ_i. The realized arm shares $\lambda_{s,n}$ in (4.1), rather than $1/2$, enter the known score probabilities; under the fixed profile the high-saturation share is 0.5794 because the larger villages are assigned to it.

The eligibility graph within each village is a deterministic cycle. An individual can link to the nearest neighbor on either side. The vertices are positions in the experimental construction, and the baseline and post-treatment graphs are random subgraphs. Let X_i be independent Bernoulli$(1/2)$, U_i independent uniform on $\{-1, 1\}$, and all node primitives be independent across individuals. On each eligible unordered pair draw independent $V_{ik}, V'_{ik} \sim \mathrm{Unif}[0,1]$ and $H_{ik} \sim \mathrm{Bernoulli}(1/2)$, independently across pairs and of node primitives. Set $\eta^0_{ik} = V_{ik}$ and $\eta^1_{ik} = H_{ik}V_{ik} + (1 - H_{ik})V'_{ik}$. These variables are shared across all assignment interventions. The link probabilities are

$$\begin{aligned} p^0(u,v) &= \Lambda(-0.40 + 0.35(u+v)) , \\ p^1(u,v,a,b) &= \Lambda(-2.30 + 0.35(u+v) + 0.70ab + 0.35(a+b)) , \end{aligned} \tag{7.1}$$

where $\Lambda(x) = (1 + \exp(-x))^{-1}$. Use $A^{\mathrm{pre}}_{ik} = K_{ik}\mathbb{1}\{\eta^0_{ik} \le p^0(U_i, U_k)\}$ and $A_{ik} = K_{ik}\mathbb{1}\{\eta^1_{ik} \le p^1(U_i, U_k, P_i, P_k)\}$. Partial persistence permits both link formation and dissolution. The baseline adjustment is $B_i = (X_i, d^0_i)$, where $d^0_i = \sum_{1 \le k \le n} A^{\mathrm{pre}}_{ik}$. The exposure is the uncapped count pair $G_i = (M^P_i, M^C_i)$ in (2.6); its six possible values satisfy $M^P_i + M^C_i \le 2$.

Use independent $\nu_i \sim \mathrm{Unif}[-0.6, 0.6]$, independent of formation and assignment, and set

$$Y_i = 1 + 0.8P_i + 0.3T_i + 0.25T_iP_i + (1.2 + 0.7P_i)h(G_i) - 0.3k(G_i) + 0.5X_i + 0.15d^0_i + \nu_i . \tag{7.2}$$

The linear design uses $h(m_p, m_c) = m_p$ and $k(m_p, m_c) = m_c$. The nonlinear design uses

$$h(m_p, m_c) = \log(1 + m_p) + 0.25 m_p m_c , \qquad k(m_p, m_c) = m_c^2/2 .$$

The latter contains curvature and a treated-by-untreated exposure interaction on its actual support. Both designs use $\varepsilon_i = (B_i, \nu_i)$, so BNS holds by construction: the noise remaining after conditioning on B_i is independent of every formation input.

The targets are the five means in (4.7) and their $(D, N, I, S, \mathrm{TE})$ contrasts, with $s_H = 0.75$ and $s_L = 0.25$. Counterfactual interventions share peer-assignment uniforms and pair primitives. Exact enumeration of the finite

local distribution gives

$$(D, N, I, S, \mathrm{TE})_{\mathrm{linear}} = (1.1287,\ 0.2910,\ 0.1728,\ 0.3412,\ 1.9337),$$
$$(D, N, I, S, \mathrm{TE})_{\mathrm{nonlinear}} = (1.0841,\ 0.1913,\ 0.1132,\ 0.2750,\ 1.6635),$$

where the displayed values are rounded. These population-law targets do not change with village size. Each configuration has $5{,}000$ completed replications.

Table 2 reports results for the proposed estimator with true nuisance functions and with estimated nuisance parameters using a machine learner under both linear and nonlinear outcome models. With true nuisance functions, bias is negligible, RMSE falls approximately at the square-root rate, and coverage is near or at the nominal level for all five effects. The results support the asymptotic framework of the proposed procedure and the validity of Algorithm 2 for constructing confidence intervals. With estimated nuisances, D, S, and TE retain small bias and near-nominal coverage, whereas N and I exhibit opposing biases and substantial undercoverage in smaller samples, consistent with sparse exposure cells affecting the cross-world mean needed to define these parameters. The shared mean cancels from $N + I$, so accurate estimation of the total can mask errors in its decomposition. As sample size increases, bias and RMSE decline; and, at $\overline{n} = 3{,}600$, coverage of the proposed estimators are close to the nominal levels, further supporting the asymptotic framework of the proposed procedure despite the sparse exposure cells by design.

TABLE 2. Bias, RMSE, and coverage of all five effects

		TE			D			N			S			I		
Nuisances	$\overline{n}$	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.
Panel A: Linear model																
Oracle	450	-0.0002	0.0285	94.1	0.0001	0.0147	95.4	0.0000	0.0241	94.1	-0.0002	0.0210	95.0	-0.0001	0.0146	94.5
Oracle	900	0.0001	0.0200	94.7	0.0002	0.0104	95.2	-0.0002	0.0171	94.2	0.0001	0.0149	94.7	-0.0001	0.0102	94.9
Oracle	1800	0.0002	0.0141	94.5	0.0001	0.0075	94.4	0.0003	0.0120	94.4	-0.0003	0.0107	95.0	0.0002	0.0073	94.2
Oracle	3600	0.0001	0.0100	94.5	-0.0001	0.0051	95.5	0.0002	0.0085	94.9	-0.0001	0.0075	95.1	0.0001	0.0052	94.6
Categorical	450	-0.0004	0.0285	94.0	-0.0005	0.0149	95.3	-0.0326	0.0453	68.4	-0.0001	0.0211	94.7	0.0329	0.0449	54.0
Categorical	900	0.0000	0.0200	94.8	0.0000	0.0105	95.2	-0.0119	0.0223	86.5	0.0001	0.0149	94.5	0.0118	0.0187	76.9
Categorical	1800	0.0002	0.0141	94.6	0.0000	0.0076	94.4	-0.0043	0.0134	91.4	-0.0003	0.0107	95.0	0.0047	0.0100	85.7
Categorical	3600	0.0001	0.0100	94.5	-0.0001	0.0051	95.5	-0.0010	0.0087	94.0	-0.0001	0.0075	95.1	0.0014	0.0057	92.0
Panel B: Nonlinear model																
Oracle	450	-0.0003	0.0193	94.2	-0.0001	0.0135	94.8	-0.0003	0.0161	94.3	0.0001	0.0169	94.8	0.0001	0.0107	95.1
Oracle	900	-0.0002	0.0137	94.2	-0.0001	0.0094	95.3	-0.0002	0.0113	94.7	0.0001	0.0118	95.1	0.0000	0.0077	94.3
Oracle	1800	0.0003	0.0096	94.9	0.0001	0.0067	94.8	0.0001	0.0080	94.9	0.0000	0.0082	95.7	0.0000	0.0055	94.4
Oracle	3600	0.0000	0.0070	94.3	-0.0001	0.0047	95.0	0.0000	0.0056	94.8	0.0001	0.0059	94.9	0.0000	0.0038	94.9
Categorical	450	-0.0003	0.0194	94.2	-0.0009	0.0137	94.4	-0.0275	0.0362	59.8	0.0001	0.0169	94.7	0.0279	0.0360	45.3
Categorical	900	-0.0003	0.0137	94.2	-0.0004	0.0095	95.2	-0.0101	0.0165	81.6	0.0001	0.0119	95.1	0.0101	0.0151	71.8
Categorical	1800	0.0003	0.0096	94.8	0.0001	0.0067	94.9	-0.0036	0.0093	90.2	0.0000	0.0082	95.5	0.0038	0.0076	85.0
Categorical	3600	0.0000	0.0070	94.3	-0.0001	0.0047	95.0	-0.0011	0.0059	93.8	0.0001	0.0059	95.0	0.0011	0.0043	91.8

TE is the total effect; DE is the direct component D; IE is the network-mediated (indirect) component N; SE is the saturation spillover S; and Interaction is I. Oracle uses the true nuisance functions; Categorical uses estimated nuisance parameters. Each configuration has 5,000 completed replications, with no failed fits or studentizations in these rows. Bias is signed; RMSE is root mean squared error; Cov. is coverage in percent for nominal 95% intervals using Algorithm 2. Total sample size is $n = 18\overline{n}$.

8. CONCLUSION

Randomized saturation experiments can identify treatment and spillover contrasts even when the program changes the social network. The additional task of attributing those contrasts to network changes requires assumptions about both exposure and confounding. I developed a framework for decomposing a treatment–saturation regime contrast into direct, network-mediated, treatment–network interaction, and spillover components. The decomposition accommodates nonlinear outcome responses and separates an own-treatment-induced change in connections from a change in the treatment composition of a fixed set of peers.

The main identifying assumption is baseline network sufficiency (BNS): conditional on baseline information, network-formation inputs contain no further information about the shock governing outcomes at fixed own-treatment and exposure. BNS permits treatment to reshape relationships and allows network formation and outcomes to share observed determinants. It also permits unobserved pair affinity to influence connections, but rules out residual common causes of connections and outcomes, such as unmeasured productivity that affects both partner choice and returns to exposure. The input-allocation mechanism in Duflo et al. (2026) motivates this assumption when baseline production, resources, and relationships capture these common determinants and the measured exposure captures the relevant allocation consequences. Under the stated exposure, support, and assignment conditions and a common conditional response mean, BNS identifies natural and interventional components without specifying or estimating a network-formation law. An orthogonal mediation score permits flexible estimation of the required outcome and exposure relationships. The inference theory accommodates network dependence with a fixed number of villages under the stated learning, dependence, and variance conditions.

Applied to the agronomy-training experiment in Rwanda, the decomposition adds to the findings of Duflo et al. (2026) by separating gains from own training from those due to training-induced changes in exposure to trained friends. Knowledge and self-reported adoption gains are concentrated in the direct component, with no statistically significant network component for either outcome. For inputs, the primary specification yields a negative network component and a positive treatment–network interaction, and the interpretation is that the same exposure-distribution change reduces input use for an untreated recipient but has a more favorable effect for a trained recipient. Adverse saturation spillovers accompany imprecise headline effects on yields and revenue. These results support the original findings of limited knowledge diffusion and adverse spillovers while adding evidence that changes in network exposure have different consequences for trained and untreated farmers. The input results are consistent with the authors' account of social connections reallocating scarce resources: training can change access to inputs even without detectable knowledge diffusion.

REFERENCES

Abadie, A. (2005): "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72, 1–19. (Cited on page 30.)

Almirall, D., T. Ten Have, and S. A. Murphy (2010): "Structural Nested Mean Models for Assessing Time-Varying Effect Moderation," *Biometrics*, 66(1), 131–139. https://doi.org/10.1111/j.1541-0420.2009.01238.x. (Cited on page 26.)

Armstrong, T. B., and M. Kolesár (2020): "Simple and Honest Confidence Intervals in Nonparametric Regression," *Quantitative Economics*, 11(1), 1–39. https://doi.org/10.3982/QE1199. (Cited on page 27.)

Aronow, P. M., and C. Samii (2017): "Estimating Average Causal Effects under General Interference, with Application to a Social Network Experiment," *Annals of Applied Statistics*, 11, 1912–1947. (Cited on pages 3, 5, and 12.)

Athey, S., D. Eckles, and G. W. Imbens (2018): "Exact p-Values for Network Interference," *Journal of the American Statistical Association*, 113(521), 230–240. https://doi.org/10.1080/01621459.2016.1241178. (Cited on pages 5 and 22.)

Auerbach, E. (2022): "Identification and Estimation of a Partially Linear Regression Model Using Network Data," *Econometrica*, 90, 347–365. (Cited on page 6.)

Baird, S., J. A. Bohren, C. McIntosh, and B. Özler (2018): "Optimal Design of Experiments in the Presence of Interference," *Review of Economics and Statistics*, 100, 844–860. (Cited on pages 2, 5, and 21.)

Balkus, S. V., H. Laith, and N. S. Hejazi (2026): "On the Use of Cross-Fitting in Causal Machine Learning with Correlated Units," working paper, arXiv:2601.10899v3, July 29. https://arxiv.org/abs/2601.10899v3. (Cited on pages 7, 25, and 53.)

Banerjee, A., E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan (2024): "Changes in Social Network Structure in Response to Exposure to Formal Credit Markets," *Review of Economic Studies*, 91(3), 1331–1372. (Cited on page 5.)

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *Review of Economic Studies*, 81, 608–650. (Cited on page 21.)

Bhatia, R. (2007): *Positive Definite Matrices*, Princeton University Press. (Cited on page 21.)

Bonhomme, S., K. Jochmans, and M. Weidner (2026a): "A Neyman-Orthogonalization Approach to the Incidental Parameter Problem in Likelihood Models," conditionally accepted at *Econometrica*; manuscript arXiv:2412.10304v4, August 17. https://arxiv.org/abs/2412.10304v4. (Cited on page 25.)

Björklund, M., and A. Gorodnik (2017): "Central Limit Theorems for Group Actions Which Are Exponentially Mixing of All Orders," working paper, arXiv:1706.09167v1, June 28. https://arxiv.org/abs/1706.09167v1. (Cited on page 15.)

Bonhomme, S., K. Jochmans, W. K. Newey, and M. Weidner (2026b): "Higher-Order Neyman Orthogonality in Moment-Condition Models," working paper, arXiv:2605.10842v1, May 11. https://arxiv.org/abs/2605.10842v1. (Cited on page 25.)

Borusyak, K., P. Hull, and E. Munro (2026): "Robust Signal Maximization in Spillover Experiments," working paper, arXiv:2607.18601v2, July 22; NBER Working Paper 35570. https://arxiv.org/abs/2607.18601v2. (Cited on page 5.)

Bousquet, O., and A. Elisseeff (2002): "Stability and Generalization," *Journal of Machine Learning Research*, 2, 499–526. https://www.jmlr.org/papers/v2/bousquet02a.html.

Brooks, R. L. (1941): "On Colouring the Nodes of a Network," *Mathematical Proceedings of the Cambridge Philosophical Society*, 37(2), 194–197. https://doi.org/10.1017/S030500410002168X. (Cited on page 36.)

Cai, J., and A. Szeidl (2018): "Interfirm Relationships and Business Performance," *Quarterly Journal of Economics*, 133(3), 1229–1282. https://doi.org/10.1093/qje/qjx049. (Cited on page 5.)

Cao, J., and M. P. Leung (2025): "Neighborhood Stability in Double/Debiased Machine Learning with Dependent Data," working paper, arXiv:2511.10995. (Cited on pages 4, 6, 20, 23, 24, 25, 12, 19, 34, 45, 53, 56, and 57.)

Chandrasekhar, A. G., and M. O. Jackson (2014): "Tractable and Consistent Random Graph Models," NBER Working Paper 20276.

Chatterjee, S., and P. Diaconis (2013): "Estimating and Understanding Exponential Random Graph Models," *Annals of Statistics*, 41, 2428–2461.

Chatterjee, S., P. Diaconis, and A. Sly (2011): "Random Graphs with a Given Degree Sequence," *Annals of Applied Probability*, 21, 1400–1435.

Chen, J., and D. M. Ritzwoller (2023): "Semiparametric Estimation of Long-Term Treatment Effects," *Journal of Econometrics*, 237(2), 105545. https://doi.org/10.1016/j.jeconom.2023.105545. (Cited on pages 9, 11, and 30.)

Chen, L. H. Y., and Q.-M. Shao (2004): "Normal Approximation under Local Dependence," *Annals of Probability*, 32(3A), 1985–2028. https://doi.org/10.1214/009117904000000450. (Cited on pages 7, 20, and 14.)

Chen, Q., V. Syrgkanis, and M. Austern (2022): "Debiased Machine Learning without Sample-Splitting for Stable Estimators," *Advances in Neural Information Processing Systems*, 35, 3096–3109. (Cited on pages 4, 6, 23, 25, and 34.)

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/Debiased Machine Learning for Treatment and Structural Parameters," *Econometrics Journal*, 21, C1–C68. (Cited on pages 4, 6, 22, 25, 52, and 55.)

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2022): "Locally Robust Semiparametric Estimation," *Econometrica*, 90(4), 1501–1535. https://doi.org/10.3982/ECTA16294. (Cited on pages 21 and 6.)

Comola, M., and S. Prina (2021): "Treatment Effect Accounting for Network Changes," *Review of Economics and Statistics*, 103, 597–604. (Cited on page 5.)

Conley, T. G. (1999): "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92(1), 1–45. https://doi.org/10.1016/S0304-4076(98)00084-0. (Cited on pages 27 and 28.)

Döring, H., S. Jansen, and K. Schubert (2022): "The Method of Cumulants for the Normal Approximation," *Probability Surveys*, 19, 185–270. https://doi.org/10.1214/22-PS7. (Cited on pages 7, 27, 14, and 15.)

Derksen, L., and P. C. L. Souza (forthcoming): "Who Knows? Information Access and Endogenous Network Formation," *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.20230126. (Cited on page 6.)

Dousse, J., and V. Féray (2019): "Weighted Dependency Graphs and the Ising Model," *Annales de l'Institut Henri Poincaré D*, 6(4), 533–571. https://doi.org/10.4171/AIHPD/78. (Cited on pages 27, 14, and 15.)

Duflo, E., D. Keniston, T. Suri, and C. Zipfel (2026): "Chat over Coffee? Diffusion of Agronomic Practices and Market Spillovers in Rwanda," *Journal of Development Economics*, 182, 103733. (Cited on pages 2, 5, 8, 10,

11, 12, 18, 29, 30, 31, 32, 37, 41, and 63.)

Duflo, E., and E. Saez (2003): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *Quarterly Journal of Economics*, 118(3), 815–842. https://doi.org/10.1162/00335530360698432. (Cited on page 5.)

Eichelsbacher, P., M. Raič, and T. Schreiber (2015): "Moderate Deviations for Stabilizing Functionals in Geometric Probability," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 51(1), 89–128. https://doi.org/10.1214/13-AIHP576. (Cited on pages 27 and 14.)

Emmenegger, C., M.-L. Spohn, T. Elmer, and P. Bühlmann (2025): "Treatment Effect Estimation with Observational Network Data Using Machine Learning," *Journal of Causal Inference*, 13(1), 2023-0082. https://doi.org/10.1515/jci-2023-0082. (Cited on pages 7, 20, 25, 26, 27, 28, 52, and 54.)

Espinosa, M., and C. T. Stanton (2026): "Training, Communications Patterns, and Spillovers inside Organizations," *Journal of Political Economy*, 134(7), 1933–1966. https://www.nber.org/papers/w30224. (Cited on page 6.)

Feigenberg, B., E. Field, and R. Pande (2013): "The Economic Returns to Social Interaction: Experimental Evidence from Microfinance," *Review of Economic Studies*, 80(4), 1459–1483. https://doi.org/10.1093/restud/rdt016. (Cited on page 5.)

Féray, V. (2018): "Weighted Dependency Graphs," *Electronic Journal of Probability*, 23, paper 93, 1–65. https://doi.org/10.1214/18-EJP222. (Cited on pages 7, 27, 14, and 15.)

Farbmacher, H., M. Huber, L. Lafférs, H. Langen, and M. Spindler (2022): "Causal Mediation Analysis with Double Machine Learning," *Econometrics Journal*, 25, 277–300. (Cited on pages 4, 6, 15, 21, 52, 53, and 55.)

Féray, V. (2019): "Central Limit Theorems for Patterns in Multiset Permutations and Set Partitions," working paper, arXiv:1811.03923v2, May 23. https://arxiv.org/abs/1811.03923v2. (Cited on page 15.)

Frisch, R., and F. V. Waugh (1933): "Partial Time Regressions as Compared with Individual Trends," *Econometrica*, 1(4), 387–401. https://doi.org/10.2307/1907330. (Cited on page 68.)

Gao, M. (2025): "Endogenous Interference in Randomized Experiments," working paper, arXiv:2412.02183. (Cited on page 6.)

Gao, M., and P. Ding (2025): "Causal Inference in Network Experiments: Regression-Based Analysis and Design-Based Properties," *Journal of Econometrics*, 252, 106119. https://doi.org/10.1016/j.jeconom.2025.106119. (Cited on page 5.)

Graham, B. S. (2017): "An Econometric Model of Network Formation with Degree Heterogeneity," *Econometrica*, 85, 1033–1063. (Cited on page 6.)

Hoeffding, W. (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13–30. (Cited on page 36.)

Heß, S., D. Jaimovich, and M. Schündeln (2021): "Development Projects and Economic Networks: Lessons from Rural Gambia," *Review of Economic Studies*, 88(3), 1347–1384. https://doi.org/10.1093/restud/rdaa033. (Cited on page 5.)

Horvitz, D. G., and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685. (Cited on page 12.)

Huber, M., and S. J. Oberhänsli (2026): "Difference-in-Differences for Mediation Analysis Using Double Machine Learning," working paper, arXiv:2602.23877, February 27. https://arxiv.org/abs/2602.23877. (Cited on pages 7 and 30.)

Hudgens, M. G., and M. E. Halloran (2008): "Toward Causal Inference with Interference," *JASA*, 103, 832–842. (Cited on pages 2, 3, 5, and 11.)

Imai, K., L. Keele, and T. Yamamoto (2010): "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25, 51–71. (Cited on pages 6 and 11.)

Jackson, M. O., and B. W. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review*, 97(3), 890–915. https://doi.org/10.1257/aer.97.3.890. (Cited on page 8.)

Jaman, A., G. Wang, A. Ertefaie, M. Bally, R. Lévesque, R. W. Platt, and M. E. Schnitzer (2025): "Penalized G-Estimation for Effect Modifier Selection in a Structural Nested Mean Model for Repeated Outcomes," *Biometrics*, 81(1), ujae165. https://doi.org/10.1093/biomtc/ujae165. (Cited on page 26.)

Janson, S. (1988): "Normal Convergence by Higher Semiinvariants with Applications to Sums of Dependent Random Variables and Random Graphs," *Annals of Probability*, 16(1), 305–312. https://doi.org/10.1214/aop/1176991903. (Cited on pages 7, 20, 27, and 14.)

Janson, S. (2004): "Large Deviations for Sums of Partly Dependent Random Variables," *Random Structures & Algorithms*, 24(3), 234–248. https://doi.org/10.1002/rsa.20008. (Cited on page 36.)

Janson, S. (2011): "Graphons, Cut Norm and Distance, Couplings and Rearrangements," arXiv:1009.2376v3, revised June 3, 2011. https://arxiv.org/abs/1009.2376v3.

Jenish, N., and I. R. Prucha (2009): "Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields," *Journal of Econometrics*, 150(1), 86–98. https://doi.org/10.1016/j.jeconom.2009.02.009. (Cited on page 28.)

Kallenberg, O. (2002): *Foundations of Modern Probability*, 2nd ed., Springer. (Cited on pages 1 and 22.)

Kelejian, H. H., and I. R. Prucha (2007): "HAC Estimation in a Spatial Framework," *Journal of Econometrics*, 140(1), 131–154. https://doi.org/10.1016/j.jeconom.2006.09.005. (Cited on pages 27 and 28.)

Kojevnikov, D., V. Marmer, and K. Song (2021): "Limit Theorems for Network Dependent Random Variables," *Journal of Econometrics*, 222, 882–908. (Cited on pages 7, 20, and 28.)

Leung, M. P., and H. R. Moon (2026a): "Nonparametric Inference in Dynamic Network Formation: Comments on 'Homophily and Transitivity in Dynamic Network Formation'," *Journal of Business & Economic Statistics*, accepted author version published online August 10. https://doi.org/10.1080/07350015.2026.2715055. (Cited on page 8.)

Leung, M. P., and H. R. Moon (2026b): "Normal Approximation in Large Network Models," *Review of Economic Studies*, rdag022, published online March 19. https://doi.org/10.1093/restud/rdag022. (Cited on page 9.)

Leung, M. P. (2022): "Causal Inference Under Approximate Neighborhood Interference," *Econometrica*, 90, 267–293. (Cited on pages 5 and 28.)

Leung, M. P. (2023): "Network Cluster-Robust Inference," *Econometrica*, 91(2), 641–667. (Cited on page 7.)

Leung, M. P. (2024): "Discussion of 'Causal Inference with Misspecified Exposure Mappings: Separating Definitions and Assumptions'," *Biometrika*, 111(1), 17–20. https://doi.org/10.1093/biomet/asad040. (Cited on pages 3 and 6.)

Leung, M. P. (2025a): "Identifying Treatment and Spillover Effects Using Exposure Contrasts," working paper, arXiv:2403.08183v4, December 8 revision; first circulated March 2024. https://arxiv.org/abs/2403.08183v4. (Cited on pages 3 and 6.)

Leung, M. P. (2025b): "Cluster-Randomized Trials with Cross-Cluster Interference," *Journal of the American Statistical Association*, published online December 10. https://doi.org/10.1080/01621459.2025.2566416. (Cited on pages 4, 7, 27, 29, 20, 21, 22, and 26.)

Li, F., K. L. Morgan, and A. M. Zaslavsky (2018): "Balancing Covariates via Propensity Score Weighting," *Journal of the American Statistical Association*, 113(521), 390–400. https://doi.org/10.1080/01621459.2016.1260466. (Cited on page 13.)

Liu, L., M. G. Hudgens, and S. Becker-Dreps (2016): "On inverse probability-weighted estimators in the presence of interference," *Biometrika*, 103, 829–842. (Cited on page 12.)

Lu, S., L. Shi, and P. Ding (2026): "Estimating Within-Cluster and Between-Cluster Spillover Effects in Randomized Saturation Designs," working paper, arXiv:2603.19573. (Cited on page 5.)

Lovell, M. C. (1963): "Seasonal Adjustment of Economic Time Series and Multiple Regression," Cowles Foundation Discussion Paper 151. https://elischolar.library.yale.edu/cowles-discussion-paper-series/380/. (Cited on page 68.)

Miao, W., Z. Geng, and E. J. Tchetgen Tchetgen (2018): "Identifying Causal Effects with Proxy Variables of an Unmeasured Confounder," *Biometrika*, 105, 987–993.

Nourdin, I., and G. Peccati (2010): "Cumulants on the Wiener Space," *Journal of Functional Analysis*, 258(11), 3775–3791. https://doi.org/10.1016/j.jfa.2009.10.024. (Cited on page 15.)

Ogburn, E. L., O. Sofrygin, I. Díaz, and M. J. van der Laan (2024): "Causal Inference for Social Network Data," *Journal of the American Statistical Association*, 119(545), 597–611. https://doi.org/10.1080/01621459.2022.2131557. (Cited on pages 7 and 20.)

Pearl, J. (2001): "Direct and Indirect Effects," *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, 411–420. (Cited on pages 6, 11, and 15.)

Penrose, M. D., and J. E. Yukich (2001): "Central Limit Theorems for Some Graphs in Computational Geometry," *Annals of Applied Probability*, 11, 1005–1041.

Ritzwoller, D. M., and J. P. Romano (forthcoming): "Reproducible Aggregation of Sample-Split Statistics," *American Economic Review*. https://doi.org/10.1257/aer.20231651. (Cited on pages 25 and 55.)

Ritzwoller, D. M. (2025): "Regression Adjustments for Disentangling Spillover Effects," working paper, November 15. (Cited on page 6.)

Robins, J. M., and S. Greenland (1992): "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143–155. (Cited on pages 6 and 15.)

Robins, J. M., and T. S. Richardson (2011): "Alternative Graphical Causal Models and the Identification of Direct Effects," in *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, ed. by P. E. Shrout, K. M. Keyes, and K. Ornstein, New York: Oxford University Press, 103–158.

Robins, J. M. (1994): "Correcting for Non-Compliance in Randomized Trials Using Structural Nested Mean Models," *Communications in Statistics—Theory and Methods*, 23(8), 2379–2412. https://doi.org/10.1080/03610929408831393. (Cited on page 26.)

Sävje, F. (2024): "Causal Inference with Misspecified Exposure Mappings: Separating Definitions and Assumptions," *Biometrika*, 111, 1–15. (Cited on pages 3, 5, and 10.)

Sobel, M. E. (2006): "What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference," *Journal of the American Statistical Association*, 101(476), 1398–1407. https://doi.org/10.1198/016214506000000636. (Cited on page 5.)

Soshnikov, A. (2002): "Gaussian Limit for Determinantal Random Point Fields," *Annals of Probability*, 30(1), 171–187. https://doi.org/10.1214/aop/1020107764. (Cited on page 15.)

Tchetgen Tchetgen, E. J., and I. Shpitser (2012): "Semiparametric Theory for Causal Mediation Analysis: Efficiency Bounds, Multiple Robustness, and Sensitivity Analysis," *Annals of Statistics*, 40, 1816–1845. (Cited on pages 4, 6, 13, 15, 17, 21, 9, 11, and 55.)

Tchetgen Tchetgen, E. J., and T. J. VanderWeele (2012): "On Causal Inference in the Presence of Interference," *Statistical Methods in Medical Research*, 21(1), 55–75. https://doi.org/10.1177/0962280210386779. (Cited on page 5.)

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge: Cambridge University Press. (Cited on page 14.)

VanderWeele, T. J., E. J. Tchetgen Tchetgen, and M. E. Halloran (2012): "Components of the Indirect Effect in Vaccine Trials: Identification of Contagion and Infectiousness Effects," *Epidemiology*, 23(5), 751–761. https://doi.org/10.1097/EDE.0b013e31825fb7a0. (Cited on page 5.)

VanderWeele, T. J., S. Vansteelandt, and J. M. Robins (2014): "Effect Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder," *Epidemiology*, 25(2), 300–306. https://doi.org/10.1097/EDE.0000000000000034. (Cited on pages 3, 6, 11, and 15.)

VanderWeele, T. J. (2014): "A Unification of Mediation and Interaction: A 4-Way Decomposition," *Epidemiology*, 25(5), 749–761, doi:10.1097/EDE.0000000000000121. (Cited on pages 6, 10, and 12.)

VanderWeele, T. J., and E. J. Tchetgen Tchetgen (2017): "Mediation Analysis with Time Varying Exposures and Mediators," *Journal of the Royal Statistical Society, Series B*, 79(3), 917–938. https://doi.org/10.1111/rssb.12194. (Cited on page 15.)

Vansteelandt, S., and R. M. Daniel (2017): "Interventional Effects for Mediation Analysis with Multiple Mediators," *Epidemiology*, 28(2), 258–265, doi:10.1097/EDE.0000000000000596. (Cited on pages 6, 11, and 15.)

Vansteelandt, S., and M. Joffe (2014): "Structural Nested Models and G-Estimation: The Partially Realized Promise," *Statistical Science*, 29, 707–731. (Cited on pages 26 and 47.)

Vansteelandt, S. (2009): "Estimating Direct Effects in Cohort and Case-Control Studies," *Epidemiology*, 20(6), 851–860. https://doi.org/10.1097/EDE.0b013e3181b6f4c9. Corrections: 21(2), 278 (2010), and 26(4), 623 (2015). (Cited on page 26.)

Vazquez-Bare, G. (2023): "Identification and Estimation of Spillover Effects in Randomized Experiments," *Journal of Econometrics*, 237(1), 105237. https://doi.org/10.1016/j.jeconom.2021.10.014. (Cited on page 5.)

Wu, P., and Z. Xiao (2026): "Causal Mediation Analysis for Network Data with Graph Neural Network," working paper, arXiv:2608.13274v1, August 13. (Cited on page 6.)

Yu, A., L. Ge, and F. Elwert (2025): "Detecting and Understanding the Difference between Natural Mediation Effects and Their Randomized Interventional Analogues," working paper, arXiv:2407.02671v2. https://arxiv.org/abs/2407.02671v2. (Cited on pages 15 and 5.)

Supplemental Appendix to:

Mediated Interference

Table of Contents

Appendix A. Proofs for Identification 1
A.1. Randomization and mediator-law alignment 1
A.2. The conditional-independence argument 2
A.3. Interventional and natural means 4
Appendix B. Proofs for Estimation 5
Appendix C. Proofs for Inference 14
C.1. A bounded dependency-graph central limit theorem 14
C.2. Covariance concentration and fitted scores 17
C.3. Comparing variance estimates 20
Appendix D. Dependence graphs and covariance kernels 22
D.1. Variance and neighborhood growth 23
D.2. A fixed causal target 24
D.3. Covariance kernels 27
Appendix E. Alternative Identification Strategy via Difference-in-Difference 28
Appendix F. Exposures beyond immediate neighbors 31
Appendix G. Additional empirical results 37
G.1. Data and construction 37
G.2. Implementation of the algorithms 39
G.3. Regime contrasts and the original regressions 40
G.4. Within-saturation and spillover decompositions 41
G.5. Exposure and adjustment sensitivity 42
G.6. Inference sensitivity and scope 44
Appendix H. A finite-dimensional structural mean model and g-estimation 46
H.1. The response model and its rank condition 47
H.2. Estimation, stability, and its network limit 50
Appendix I. Identification counterexamples and coupling 52
Appendix J. Cross-fitting with separated training sets 52
Appendix K. An extension under absolute regularity 56
Appendix L. Unbounded outcomes 60
Appendix M. Regression coefficients and assignment-regime effects 63
M.1. Specifications and conditions for probability limits 63
M.2. The within-village treatment coefficient 65
M.3. Assigned baseline friends: control and treated farmers 66
M.4. Saturation contrasts 67
M.5. Proofs 67

APPENDIX A. PROOFS FOR IDENTIFICATION

Let P_n^{prim} denote the joint law of the pretreatment random element $\mathcal{W}_n$ and the response shocks ε^n. This law includes observed baseline covariates, the baseline graph, latent formation variables, and any dependence between them. The observed law is induced by the primitive distribution, the assignment design, and the measurable maps producing $(B_i, Z_i, G_i, Y_i)_{i=1}^n$. Thus $\mathbb{E}_D$ integrates assignments (T, P) holding primitives fixed. An expectation without that subscript integrates primitives and the assignment intervention in question; a target-population expectation also averages an independent uniform index I.

The baseline variable $B_i = (X_i, N_i^{\mathrm{pre}})$ collects the covariates and baseline network information selected for adjustment. It need not contain the entire graph or all formation primitives. Write $F_B = \mathcal{L}(B_I)$ and $Q_i(\mathrm{d}e \mid x) = \mathcal{L}(\varepsilon_i \mid B_i = x)$. For the natural intervention, $D_{i,s}(\mathrm{d}h)$ integrates the other villages' saturations and all peer assignments conditional on focal saturation s; the focal assignment is fixed at b. For the interventional mean, the conditional law is instead the product of $q_b^\star(\mathrm{d}g \mid x)$ and $K(\mathrm{d}i, \mathrm{d}e \mid x) = \mathcal{L}(I, \varepsilon_I \mid B_I = x)$. These two integrations differ precisely in whether exposure retains its dependence on the response primitives.

These conventions apply to the identification and formation appendices. The estimation, dependence, and inference appendices use the induced observed law conditional on the complete supported saturation profile; their $\mathbb{P}$ and $\mathbb{E}$ suppress that conditioning. An additional conditional law, such as conditioning on a training sample, is written explicitly. Throughout this appendix all conditional kernels refer to the target law in the result being proved. Integrals are well defined by absolute integrability; boundedness is used only in results that expressly impose it. For continuous exposures, assertions about conditional responses mean assertions about jointly measurable versions on the relevant kernel support, as in Assumption 3.3.

The measure-theoretic tools are the existence and disintegration of regular conditional distributions for standard Borel random variables, the tower property and conditional Jensen's inequality; see Kallenberg (2002, Chapter 6). The conditional product identities used below are verified directly, so a conditional-independence conclusion is tied to its specified σ-fields.

A.1. **Randomization and mediator-law alignment.**

Proof of Proposition 3.1. Fix the primitive population. Let $C_i = \{T_i = s, P_i = a\}$ and $p_i = \mathbb{P}_D(C_i) > 0$. Consistency gives $Y_i = Y_i(T, P)$ at the realized assignment, including every induced graph change. Therefore

$$\mathbb{E}_D\left[\frac{\mathbb{1}\{C_i\}Y_i}{p_i}\right] = \frac{\mathbb{P}_D(C_i)}{p_i}\mathbb{E}_D[Y_i(T,P) \mid C_i] = \mathbb{E}_D[Y_i(T,P) \mid C_i]\,.$$

Sum this equality and divide by n. Linearity does not require independence across outcomes, villages, or graph edges. In the Bernoulli design, $p_i = \mathbb{P}_D(T_{j(i)} = s)s^a(1-s)^{1-a}$ on supported cells. Conditional on $T_{j(i)} = s$, own assignment is independent of all other assignment draws, so the conditional external-saturation law remains $Q_{j(i),s}$. The conditional peer assignment law is exactly the law in Definition 2.1. Thus averaging the finite-population identity over primitives produces (2.11). The same argument applied to any linear contrast proves design identification of TE and S. ∎

Lemma A.1 (Peer-law alignment)**.** *For a fixed individual, under Bernoulli assignment and the external law $Q_{j(i),s}$,*

$$\mathcal{L}(G_i^s(b) \mid B_i) = \mathcal{L}(G_i \mid T_i = s, P_i = b, B_i)\,.$$

Proof. Write the complete peer draw as H, with design kernel $D_{i,s}(\mathrm{d}h)$. Randomization implies that conditional on $B_i = x$ the primitive conditional law is unchanged by $\{T_i = s, P_i = b\}$. Conditional independence of the

Bernoulli coordinates gives the same $D_{i,s}$ under either focal assignment. For every bounded measurable v,

$$\mathbb{E}[v(G_i) \mid T_i = s, P_i = b, B_i = x] = \int\int v\big(e_{i,n}(\mathcal{A}_n(w, s, b, h), s, h, X^n(w))\big) D_{i,s}(\mathrm{d}h)\, \mathbb{P}(\mathrm{d}w \mid B_i = x)$$
$$= \mathbb{E}[v(G_i^s(b)) \mid B_i = x] .$$

The notation (s, b, h) includes the external profile recorded in h. Equality for bounded indicators proves equality of kernels.

■

Lemma A.2 (Target-population standardization)**.** *Under Assumptions 2.1 and 2.2, the law $P_{s,a}^{\mathrm{tar}}$ in (3.2) is a probability law with baseline marginal F_B, and*

$$\int y\, P_{s,a}^{\mathrm{tar}}(\mathrm{d}x, \mathrm{d}g, \mathrm{d}y) = \int \mathbb{E}[Y \mid T_i = s, P_i = a, B = x] F_B(\mathrm{d}x) = R_{s,n}(a) .$$

Proof. Fix a supported (s, a) and write $C_i = \{T_i = s, P_i = a\}$ and $p_i = p_i^D(s, a) > 0$. For every bounded measurable v on the baseline space, randomization gives

$$\mathbb{E}\left[\frac{\mathbb{1}\{C_i\}}{p_i} v(B_i)\right] = \mathbb{E}\left[v(B_i) \frac{\mathbb{P}(C_i \mid \mathcal{W}_n)}{p_i}\right] = \mathbb{E}[v(B_i)] .$$

Taking $v = 1$ gives unit mass for each summand defining $P_{s,a}^{\mathrm{tar}}$. Positivity and countable additivity follow by monotone convergence. Averaging the same identity over $i \in [n]$ gives baseline marginal $n^{-1} \sum_{1 \le i \le n} \mathcal{L}(B_i) = F_B$. By consistency and the peer-law calculation in Lemma A.1,

$$\int y\, \mathrm{d}P_{s,a}^{\mathrm{tar}} = \frac{1}{n} \sum_{1 \le i \le n} \mathbb{E}\left[\frac{\mathbb{1}\{C_i\}}{p_i} Y_i\right]$$
$$= \frac{1}{n} \sum_{1 \le i \le n} \mathbb{E}\left[\mathbb{E}_D[Y_i(T, P) \mid C_i, \mathcal{W}_n, \varepsilon^n]\right]$$
$$= \frac{1}{n} \sum_{1 \le i \le n} \mathbb{E}[Y_i^s(a, G_i^s(a))] = R_{s,n}(a) .$$

The second equality integrates the conditional design law while holding the complete primitives fixed; the third uses its equality to the intervention law, including the external saturation draw. Absolute integrability under Assumption 2.2 permits the iterated expectations. Disintegrating $P_{s,a}^{\mathrm{tar}}$ over its baseline marginal proves the remaining equality. For equal inclusion probabilities, Bayes' formula shows that (3.2) coincides with ordinary assignment-cell conditioning for the uniform index. For unequal probabilities, the factor p_i^{-1} is needed to preserve the uniform target weights. ■

Joint target law. To express the cell laws on one probability space, choose fixed $\omega_z > 0$, with $\sum_{z \in \mathsf{Z}} \omega_z = 1$, and define

$$P^{\mathrm{tar}}(\mathrm{d}z, \mathrm{d}x, \mathrm{d}g, \mathrm{d}y) = \sum_{z_0 \in \mathsf{Z}} \omega_{z_0} \delta_{z_0}(\mathrm{d}z) P_{z_0}^{\mathrm{tar}}(\mathrm{d}x, \mathrm{d}g, \mathrm{d}y) .$$

Here δ_{z_0} is a point mass. The conditional law given $Z = z$ is P_z^{tar}. Because every cell has baseline marginal F_B, $P^{\mathrm{tar}}(Z = z \mid B) = \omega_z$. These probabilities combine the cell laws without changing their conditional quantities or the mediation functional.

A.2. **The conditional-independence argument.**

Proof of Proposition 3.4. Fix i, s, a, b and write $W = \mathcal{W}_n$. Let D contain the assignment random variables for the observed experiment and the common-peer intervention. They can be constructed jointly with D independent of (W, ε^n) by Assumption 2.1. Set $\mathcal{F} = \sigma(W, \mathsf{D})$. Both B_i and every exposure appearing in the proposition are $\mathcal{F}$-measurable.

First consider a bounded measurable function u of the shock. BNS and the measurability of B_i in W give

$$\mathbb{E}[u(\varepsilon_i) \mid W] = \int u(e) Q_i(\mathrm{d}e \mid B_i) =: a_u(B_i) \, .$$

For bounded measurable $v(W)$ and $w(\mathsf{D})$, independence of D gives

$$\begin{aligned} \mathbb{E}[u(\varepsilon_i) v(W) w(\mathsf{D})] &= \mathbb{E}[u(\varepsilon_i) v(W)] \mathbb{E}[w(\mathsf{D})] \\ &= \mathbb{E}[a_u(B_i) v(W)] \mathbb{E}[w(\mathsf{D})] \\ &= \mathbb{E}[a_u(B_i) v(W) w(\mathsf{D})] \, . \end{aligned}$$

Products of such functions generate $\mathcal{F}$. The monotone-class theorem therefore extends the last equality to every bounded $\mathcal{F}$-measurable test function. By the defining property of conditional expectation,

$$\mathcal{L}(\varepsilon_i \mid \mathcal{F}) = Q_i(\cdot \mid B_i) \, . \tag{A.1}$$

Equality of kernels can be chosen simultaneously on a countable determining class because the shock space is standard Borel.

This kernel identity also permits a random exposure argument. For any $\mathcal{F}$-measurable exposure H and any bounded jointly measurable $v(g, e)$, disintegration gives

$$\mathbb{E}[v(H, \varepsilon_i) \mid \mathcal{F}] = \int v(H, e) Q_i(\mathrm{d}e \mid B_i) \, .$$

One verifies this first for $v(g, e) = v_1(g) v_2(e)$ using (A.1), and then for bounded measurable v by another monotone-class argument. Positive and negative truncation extends it to the integrable structural responses in the proposition.

For the natural response, take $H = G_i^s(b)$ and $v(g, e) = f_{i,n}(s, a, g, e)$. Exposure sufficiency yields

$$\begin{aligned} \mathbb{E}[Y_i^s(a, G_i^s(b)) \mid B_i, G_i^s(b)] &= \mathbb{E}\left[\mathbb{E}[f_{i,n}(s, a, G_i^s(b), \varepsilon_i) \mid \mathcal{F}] \mid B_i, G_i^s(b)\right] \\ &= \mathbb{E}[h_{i,a}(G_i^s(b), B_i) \mid B_i, G_i^s(b)] \\ &= h_{i,a}(G_i^s(b), B_i) \, . \end{aligned}$$

When $b = a$ this is the same-assignment statement. When $b \neq a$, the exposure changes but remains $\mathcal{F}$-measurable; the identical shock kernel in (A.1) gives the cross-world statement. Thus neither case requires independent formation draws across worlds.

For the observed response, let $C_i = \{T_i = s, P_i = a\}$. On C_i, consistency and exposure sufficiency give $Y_i = f_{i,n}(s, a, G_i, \varepsilon_i)$. Since $C_i \in \mathcal{F}$, the same calculation, now conditional on (C_i, B_i, G_i), proves (3.11) wherever the cell has positive conditional probability. At baseline values outside that cell's support the equality is not used; Assumption 3.2 specifies the support needed for identification.

Finally let I be the independently selected uniform target index. The following averaging calculation can be performed under the observed design law or its target reweighting. Weighting the assignment cell does not change the conditional response kernel at a fixed individual. Given (B_I, G_I, C_I), the posterior probability of $I = i$ may

depend on all three conditioning variables. Nevertheless, if all $h_{i,a}$ equal h_a,

$$\begin{aligned}\mathbb{E}[Y_I \mid B_I, G_I, C_I] &= \sum_{1\le i\le n} \mathbb{P}(I=i \mid B_I, G_I, C_I) h_{i,a}(G_I, B_I) \\ &= h_a(G_I, B_I) \sum_{1\le i\le n} \mathbb{P}(I=i \mid B_I, G_I, C_I) \\ &= h_a(G_I, B_I)\,.\end{aligned}$$

The same summation conditional on $(B_I, G_I^s(b))$ proves the natural identity. Applying (A.1) to a fixed exposure argument also supplies the jointly measurable conditional-response versions in Assumptions 3.1(ii) and 3.3. The common-mean premise is used precisely in this last averaging step. The aligned laws ensure that the selected observed and intervention variables are those in the identification statement. ■

A.3. **Interventional and natural means.**

Proof of Theorem 3.2. Fix a, b. Let F_B be the target baseline law and let $K(\mathrm{d}i, \mathrm{d}e \mid x)$ be the regular conditional law of the augmented response primitive (I, ε_I) given $B_I = x$. By construction, the exposure draw is independent of the response primitives given $B = x$, with conditional joint distribution $K(\mathrm{d}i, \mathrm{d}e \mid x) q_b(\mathrm{d}g \mid x)$. Integrating the jointly measurable scalar map $f_{i,n}(s, a, g, e)$ against that product gives

$$\begin{aligned}\theta_{s,n}^{\mathrm{int}}(a,b) &= \int \left[\int \mathbb{E}[Y^s(a,g) \mid B = x] q_b(\mathrm{d}g \mid x) \right] F_B(\mathrm{d}x) \\ &= \int \left[\int m_a(g,x) q_b(\mathrm{d}g \mid x) \right] F_B(\mathrm{d}x) = \Psi_{s,ab,n}\,.\end{aligned}$$

The first equality uses mediator-law alignment, the second controlled-response exchangeability and consistency. Absolute continuity $q_b \ll q_a$ ensures that the observed version of m_a, defined q_a-almost everywhere, is determined at every exposure argument relevant to the integral. A change to m_a on a q_a-null set then leaves the integral unchanged. Without this condition the observed law does not determine the response on all intervention exposure values.

For $b = a$, apply iterated expectation under the observed arm.

$$\begin{aligned}\Psi_{s,aa,n} &= \int\int \mathbb{E}[Y \mid T_i = s, P_i = a, G = g, B = x] q_a(\mathrm{d}g \mid x) F_B(\mathrm{d}x) \\ &= \int \mathbb{E}[Y \mid T_i = s, P_i = a, B = x] F_B(\mathrm{d}x) = R_{s,n}(a)\,.\end{aligned}$$

The last equality follows from Lemma A.2, using the target law (3.2). This proves the regime-mean identity under the maintained assumptions. Substitution in (2.12) gives (2.13). ■

Proof of Theorem 3.3. Use the conditional kernel in Assumption 3.3 to evaluate the natural response at $G^s(b)$.

$$\begin{aligned}\theta_{s,n}^{\mathrm{nat}}(a,b) &= \mathbb{E}\big[\mathbb{E}(Y^s(a, G^s(b)) \mid B, G^s(b))\big] \\ &= \mathbb{E}[h_a(G^s(b), B)] \\ &= \int\int h_a(g,x) q_b(\mathrm{d}g \mid x) F_B(\mathrm{d}x) \\ &= \int\int m_a(g,x) q_b(\mathrm{d}g \mid x) F_B(\mathrm{d}x)\,.\end{aligned}$$

The second equality is exactly the cross-world mean condition, the third uses mediator-law alignment, and the fourth uses within-world exchangeability on the overlap support. The final expression is the interventional functional already identified. There is no assertion that the experiment observes a natural cross-world outcome. ■

Proof of the weighted representation (3.12). Write $\mathbb{E}_{\mathrm{tar}}$ for expectation under the joint target law and let $e_a(x) = P^{\mathrm{tar}}(Z = (s,a) \mid B = x) = \omega_{(s,a)} > 0$. Conditioning on $B = x$ gives

$$\mathbb{E}_{\mathrm{tar}}\left[\frac{\mathbb{1}\{Z=(s,a)\}}{e_a(B)} r_{ab}(G,B)Y \;\middle|\; B = x\right] = \frac{P^{\mathrm{tar}}(Z=(s,a) \mid B=x)}{e_a(x)} \int r_{ab}(g,x) m_a(g,x) q_a(\mathrm{d}g \mid x)$$
$$= \int m_a(g,x) q_b(\mathrm{d}g \mid x)\,.$$

The ratio of cell probabilities is one, and the last equality uses the defining Radon–Nikodym identity. Integrating over F_B yields

$$\Psi_{s,ab,n} = \mathbb{E}_{\mathrm{tar}}\left[\frac{\mathbb{1}\{Z=(s,a)\}}{e_a(B)} r_{ab}(G,B)Y\right] = \int r_{ab}(g,x)\, y\, P^{\mathrm{tar}}_{s,a}(\mathrm{d}x, \mathrm{d}g, \mathrm{d}y)\,.$$

The cell probability $\omega_{(s,a)}$ cancels e_a, so the result does not depend on the chosen cell probabilities. Definition (3.2) gives (3.12) under the actual experiment. For unbounded integrable summands, apply the bounded-function identity to positive and negative truncations and then use the absolute-integrability premise of the weighted representation. ■

For a finite exposure space, the relation to the covariance decomposition in Yu et al. (2025, Proposition 2) can be seen directly. Suppose each $Y^s(a,g)$ is integrable, and write $M = G^s(b)$ and $q_b(g \mid B) = \mathbb{P}(M = g \mid B)$. Conditional linearity gives

$$\mathbb{E}[Y^s(a,M) \mid B] = \sum_{g \in \mathsf{G}} \mathbb{E}[Y^s(a,g)\mathbb{1}\{M=g\} \mid B]$$
$$= \sum_{g \in \mathsf{G}} \big\{ h_a(g,B) q_b(g \mid B) + \mathrm{Cov}(Y^s(a,g), \mathbb{1}\{M=g\} \mid B) \big\}\,.$$

Subtracting the conditional mean under the independent exposure draw and integrating over B yields

$$\theta^{\mathrm{nat}}_{s,n}(a,b) - \theta^{\mathrm{int}}_{s,n}(a,b) = \sum_{g \in \mathsf{G}} \mathbb{E}\big[\mathrm{Cov}(Y^s(a,g), \mathbb{1}\{G^s(b)=g\} \mid B)\big]\,. \tag{A.2}$$

Assumption 3.3 makes each covariance zero, by taking the conditional expectation given (B, M) inside its defining product. Thus it is sufficient for equality of every cell. It is not necessary for equality of a particular cell or contrast, since nonzero covariance contributions may cancel. Differencing (A.2) gives the analogous statement for direct, mediated, and interaction components. The finite sum avoids a density assumption; for a general standard Borel exposure space the preceding proof uses conditional distributions and jointly measurable conditional means instead.

Appendix B. Proofs for Estimation

Condition throughout on the complete supported saturation profile. The number of villages and the number of score coordinates are fixed. All constants below are independent of the village sizes. I write the calculation for one cell (z, z') and suppress its subscripts where their meaning is fixed; within this appendix θ denotes the finite-n cell functional $\theta_{zz',n}$. Here $B_i = (X_i, N_i^{\mathrm{pre}})$ collects the selected baseline covariates and network information, $\mathsf{P}_i = \mathcal{L}(O_i)$ is the conditional marginal observation law, and $\mathsf{P} = n^{-1}\sum_{1 \le i \le n} \mathsf{P}_i$ is the law of a uniform target

index. For any candidate functions η, define $\mathsf{P}_i\psi_\eta = \int \psi(o;\eta)\mathsf{P}_i(\mathrm{d}o)$. If η is estimated, this integral is evaluated at the realized functions and remains random. Expectations involving the original array, its copy, independent evaluation, and algorithm randomization integrate their joint distribution, as specified in Assumption 4.4.

Let $O = O_I$ be the uniformly selected observation from Section 4. For a measurable baseline set A,

$$\begin{aligned}\mathbb{P}(B \in A, Z = (s,a)) &= \frac{1}{n}\sum_{1\le i\le n} \mathbb{P}(B_i \in A, Z_i = (s,a)) \\ &= \frac{1}{n}\sum_{1\le i\le n} \mathbb{1}\{t_{j(i)} = s\} s^a (1-s)^{1-a} F_B(A) \\ &= e_{(s,a)} F_B(A)\,.\end{aligned}$$

The second equality uses both the common baseline law and randomized individual assignment in Assumption 4.1. Hence the conditional cell probability for O is (4.1). The common conditional laws also give its m_z and q_z.

For a fixed site i, condition the second term of (4.3) on (B_i, G_i, Z_i). On $Z_i = z$ its residual has expectation zero. On $Z_i \neq z$ its indicator is zero. Condition the third term on (B_i, Z_i). On $Z_i = z'$ its mean is $\int m_z(g, B_i) q_{z'}(\mathrm{d}g \mid B_i) - \nu(B_i) = 0$, and otherwise its indicator is zero. Therefore

$$\mathbb{E}\psi(O_i;\eta_0) = \mathbb{E}\nu(B_i) = \int \nu \,\mathrm{d}F_B = \theta\,. \tag{B.1}$$

This sitewise identity uses the common baseline law. The cancellation for a candidate nuisance below concerns the average of site means; the candidate's individual site means need not vanish.

The augmented mediation score and its residual cancellations are established ingredients in Tchetgen Tchetgen and Shpitser (2012, Theorem 1) and Farbmacher et al. (2022). Their first-order cancellation is also an instance of the locally robust moment construction in Chernozhukov et al. (2022). The following calculation retains every term, including the distinct errors from estimating the two assignment probabilities.

Proof of Proposition 4.1. Fix a candidate collection obeying Assumption 4.3; the algebra below holds for every such collection and can then be evaluated at estimated functions. For this proof let $E_c v = \int v(g,x) q_c(\mathrm{d}g \mid x)$ at a fixed $B = x$. Write $\delta m = \widetilde{m} - m$, $\delta r = \widetilde{r} - r$, $\delta\nu = \widetilde{\nu} - \nu$ and $\kappa_c = e_c/\widetilde{e}_c$.

Term 1. The outer regression. Its conditional expectation, less the target conditional mean, is $\widetilde{\nu}(x) - \nu(x) = \delta\nu(x)$.

Term 2. The outcome residual. Iterated expectation, the observed outcome conditional mean, and the definition of e_z give

$$\begin{aligned}&\mathbb{E}\left[\frac{\mathbb{1}\{Z=z\}}{\widetilde{e}_z(x)}\widetilde{r}(G,x)\{Y - \widetilde{m}(G,x)\} \,\middle|\, B = x\right] \\ &= \frac{e_z(x)}{\widetilde{e}_z(x)}\int \widetilde{r}(g,x)\{\mathbb{E}[Y \mid G=g, Z=z, B=x] - \widetilde{m}(g,x)\} q_z(\mathrm{d}g \mid x) \\ &= \kappa_z E_z[\widetilde{r}(m - \widetilde{m})] = -\kappa_z E_z[r\delta m] - \kappa_z E_z[\delta r \delta m]\,.\end{aligned}$$

Term 3. The exposure residual. Conditioning first on (B, Z) gives

$$\begin{aligned}&\mathbb{E}\left[\frac{\mathbb{1}\{Z=z'\}}{\widetilde{e}_{z'}(x)}\{\widetilde{m}(G,x) - \widetilde{\nu}(x)\} \,\middle|\, B = x\right] \\ &= \kappa_{z'}\{E_{z'}[m + \delta m] - \nu - \delta\nu\} = \kappa_{z'}E_{z'}\delta m - \kappa_{z'}\delta\nu\,.\end{aligned}$$

Here $E_{z'}m = \nu$ by definition. Since $E_z[r\delta m] = E_{z'}\delta m$, the sum of the three expressions equals

$$(1-\kappa_{z'})\delta\nu + (\kappa_{z'}-\kappa_z)E_{z'}\delta m - \kappa_z E_z[\delta r \delta m]\,.$$

Integrating over F_B proves (4.4). Taking $\widetilde{\eta} = \eta_0$ proves the average mean identity; (B.1) proves its sitewise version.

I next compute the nuisance derivatives term by term. Let u be a real path parameter and perturb the named nuisance by u times a bounded direction, retaining all other nuisances at their true values. Work in an interval on which the candidate bounds and positive denominator condition hold.

Term 1. The outcome-regression direction h_m. Only the two residual terms change, linearly in u.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}\psi(O;e,m+uh_m,r,\nu)\Big|_{u=0} &= -\mathbb{E}\left[\frac{\mathbb{1}\{Z=z\}}{e_z}r(G,B)h_m(G,B)\right] + \mathbb{E}\left[\frac{\mathbb{1}\{Z=z'\}}{e_{z'}}h_m(G,B)\right]\\ &= -\int\!\!\int r(g,x)h_m(g,x)q_z(\mathrm{d}g\mid x)F_B(\mathrm{d}x)\\ &\quad + \int\!\!\int h_m(g,x)q_{z'}(\mathrm{d}g\mid x)F_B(\mathrm{d}x) = 0\,.\end{aligned}$$

The last equality uses the Radon–Nikodym identity on the overlap support.

Term 2. The ratio direction h_r. Its derivative is

$$\mathbb{E}\left[\frac{\mathbb{1}\{Z=z\}}{e_z}h_r(G,B)\{Y-m(G,B)\}\right] = \mathbb{E}\left[\frac{\mathbb{1}\{Z=z\}}{e_z}h_r(G,B)\mathbb{E}[Y-m(G,B)\mid B,G,Z]\right] = 0\,.$$

The inner expectation is zero on $Z=z$; the prefactor is zero on the other cells.

Term 3. The outer-regression direction h_ν. The outer term and the third score term yield

$$\mathbb{E}\left[h_\nu(B) - \frac{\mathbb{1}\{Z=z'\}}{e_{z'}(B)}h_\nu(B)\right] = \int h_\nu(x)\left\{1-\frac{\mathbb{P}(Z=z'\mid B=x)}{e_{z'}(x)}\right\}F_B(\mathrm{d}x) = 0\,.$$

Term 4. The assignment-probability directions. For the outcome-arm denominator,

$$\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}\left[\frac{\mathbb{1}\{Z=z\}r(G,B)\{Y-m(G,B)\}}{e_z(B)+uh_{e,z}(B)}\right]\Bigg|_{u=0} = -\mathbb{E}\left[\frac{\mathbb{1}\{Z=z\}h_{e,z}(B)r(G,B)}{e_z(B)^2}\{Y-m(G,B)\}\right] = 0\,.$$

For the exposure-arm denominator the derivative equals

$$-\mathbb{E}\left[\frac{\mathbb{1}\{Z=z'\}h_{e,z'}(B)}{e_{z'}(B)^2}\{m(G,B)-\nu(B)\}\right] = -\int\frac{h_{e,z'}(x)}{e_{z'}(x)}\{E_{z'}m-\nu(x)\}F_B(\mathrm{d}x) = 0\,.$$

When $z=z'$, a perturbation of the shared probability has both contributions and their sum is zero. The same case also has $r_{zz}=1$; admissible paths for $z=z'$ retain that identity. These observations cover coincident as well as distinct cells.

All displayed differentiations pass through the expectation by dominated convergence. In a sufficiently small admissible path interval, the denominator is at least a fixed positive constant, the residuals are bounded by $2M$, and each direction is bounded. For a simultaneous norm-differentiable path, the exact remainder provides a second justification. Each error norm is $\mathrm{O}(|u|)$, so the remainder divided by u tends to zero.

Finally I give the constants in its bound. The denominator restriction gives

$$|1-\kappa_c| \le e_*^{-1}|\delta e_c|\,, \qquad |\kappa_{z'}-\kappa_z| \le e_*^{-1}(|\delta e_{z'}|+|\delta e_z|)\,, \qquad |\kappa_z| \le e_*^{-1}\,.$$

By conditional Jensen's inequality and overlap,

$$\|E_{z'}\delta m\|_B^2 \le \|\delta m\|_{z'}^2 = \int\int (\delta m)^2 r \, \mathrm{d}q_z \, \mathrm{d}F_B \le R\|\delta m\|_z^2 .$$

Apply the Cauchy–Schwarz inequality separately to the three remainder terms to obtain

$$\begin{aligned} |\mathbb{E}_B(1-\kappa_{z'})\delta\nu| &\le e_*^{-1}\|\delta e\|_B\|\delta\nu\|_B , \\ |\mathbb{E}_B(\kappa_{z'}-\kappa_z)E_{z'}\delta m| &\le 2\sqrt{R}e_*^{-1}\|\delta e\|_B\|\delta m\|_z , \\ |\mathbb{E}_B\kappa_z E_z[\delta r\delta m]| &\le e_*^{-1}\|\delta r\|_z\|\delta m\|_z . \end{aligned}$$

Their sum is $\mathcal{B}(\widetilde{\eta})$. With exact probabilities the first two expressions vanish and $\kappa_z = 1$, giving the sharper constant one for the remaining product. ■

Lemma B.1 (Score prediction bound)**.** *Under Assumptions 4.1 and 4.3, the population evaluation norm of any admissible candidate collection obeys, for each marginal law* P_i *and for* P*,*

$$\begin{aligned} \|\psi(\cdot;\widetilde{\eta}) - \psi(\cdot;\eta_0)\|_{L_2(\mathsf{P}_i)} \le (1+e_*^{-1})\|\delta\nu\|_B + \frac{R+\sqrt{R}}{e_*}\|\delta m\|_z + \frac{2M}{e_*}\|\delta r\|_z \\ + \frac{2M(R+1)}{e_*^2}\|\delta e\|_B . \end{aligned} \tag{B.2}$$

For a random candidate, inequality (B.2) holds almost surely between the corresponding population integrals. If O_i^* is independent of the fitted parameters and has law P_i, then

$$\left\{\mathbb{E}\left[|\psi(O_i^*;\widehat{\eta}) - \psi(O_i^*;\eta_0)|^2 \mid \widehat{\eta}\right]\right\}^{1/2} = \|\psi(\cdot;\widehat{\eta}) - \psi(\cdot;\eta_0)\|_{L_2(\mathsf{P}_i)} .$$

Minkowski's inequality gives the unconditional bound by replacing each random nuisance norm on the right of (B.2) with the square root of its expected square.

Proof. Write $\pi_i(c) = \mathbb{P}(Z_i = c \mid B_i)$. Under the common conditional laws, $\pi_i(c) = \mathbb{1}\{t_{j(i)} = s_c\}s_c^{a_c}(1-s_c)^{1-a_c} \le 1$ for $c = (s_c, a_c)$. Therefore, for any measurable h with finite norm,

$$\begin{aligned} \mathsf{P}_i[\mathbb{1}\{Z_i = c\}h(G_i,B_i)^2] &= \int \pi_i(c) \int h(g,x)^2 q_c(\mathrm{d}g \mid x) F_B(\mathrm{d}x) \\ &\le \|h\|_c^2 . \end{aligned}$$

For a function of B_i alone the same calculation gives an upper bound $\|h\|_B^2$. These inequalities also hold for the average law P.

Subtract the oracle score from the candidate score. Add and subtract the terms with the candidate denominators and oracle residuals to obtain the exact decomposition

$$\begin{aligned} \psi(O_i;\widetilde{\eta}) - \psi(O_i;\eta_0) = \delta\nu(B_i) &- \frac{\mathbb{1}\{Z_i = z'\}}{\widetilde{e}_{z'}(B_i)}\delta\nu(B_i) \\ &- \frac{\mathbb{1}\{Z_i = z\}\widetilde{r}(G_i,B_i)}{\widetilde{e}_z(B_i)}\delta m(G_i,B_i) + \frac{\mathbb{1}\{Z_i = z'\}}{\widetilde{e}_{z'}(B_i)}\delta m(G_i,B_i) \\ &+ \frac{\mathbb{1}\{Z_i = z\}}{\widetilde{e}_z(B_i)}\delta r(G_i,B_i)\{Y_i - m(G_i,B_i)\} \\ &+ \mathbb{1}\{Z_i = z\}r(G_i,B_i)\{Y_i - m(G_i,B_i)\}(\widetilde{e}_z^{-1} - e_z^{-1})(B_i) \\ &+ \mathbb{1}\{Z_i = z'\}\{m(G_i,B_i) - \nu(B_i)\}(\widetilde{e}_{z'}^{-1} - e_{z'}^{-1})(B_i) . \end{aligned}$$

Call these seven terms $A_1, \dots, A_7$ in their displayed order. Each norm below integrates over P_i, with the candidate functions held at their chosen values.

For the outer regression,

$$\|A_1\|_2 = \|\delta\nu\|_B, \qquad \|A_2\|_2^2 \le e_*^{-2}\mathsf{P}_i[\mathbb{1}\{Z_i = z'\}\delta\nu(B_i)^2] \le e_*^{-2}\|\delta\nu\|_B^2 .$$

For the two outcome-regression terms,

$$\begin{aligned}\|A_3\|_2^2 &\le R^2 e_*^{-2}\mathsf{P}_i[\mathbb{1}\{Z_i = z\}\delta m(G_i, B_i)^2] \le R^2 e_*^{-2}\|\delta m\|_z^2 ,\\ \|A_4\|_2^2 &\le e_*^{-2}\|\delta m\|_{z'}^2 = e_*^{-2}\int\!\!\int \delta m(g,x)^2 r(g,x) q_z(\mathrm{d}g \mid x) F_B(\mathrm{d}x)\\ &\le R e_*^{-2}\|\delta m\|_z^2 .\end{aligned}$$

For the ratio error, bounded Y_i and m give

$$\|A_5\|_2^2 \le 4M^2 e_*^{-2}\mathsf{P}_i[\mathbb{1}\{Z_i = z\}\delta r(G_i, B_i)^2] \le 4M^2 e_*^{-2}\|\delta r\|_z^2 .$$

Finally, for either required cell c,

$$|\widetilde{e}_c^{-1} - e_c^{-1}| = \frac{|\widetilde{e}_c - e_c|}{\widetilde{e}_c e_c} \le e_*^{-2}|\delta e_c| .$$

Hence

$$\|A_6\|_2^2 \le 4M^2R^2 e_*^{-4}\|\delta e_z\|_B^2 , \qquad \|A_7\|_2^2 \le 4M^2 e_*^{-4}\|\delta e_{z'}\|_B^2 .$$

Minkowski's inequality, followed by these seven square-root bounds and $\|\delta e\|_B = \max_{c=z,z'}\|\delta e_c\|_B$, proves (B.2).

For an estimated collection, the same algebra holds at each realized nuisance estimate. If $\mathcal{T}$ is the σ-field generating that nuisance estimate and O_i^* is independent of $\mathcal{T}$, disintegration gives

$$\mathbb{E}[|\psi(O_i^*; \widehat{\eta}) - \psi(O_i^*; \eta_0)|^2 \mid \mathcal{T}] = \int |\psi(o; \widehat{\eta}) - \psi(o; \eta_0)|^2 \mathsf{P}_i(\mathrm{d}o) .$$

Apply the seven inequalities conditionally and then Minkowski's inequality over the training law. This proves the unconditional mean-square assertion. For example the contribution of the outer-regression error is at most $(1 + e_*^{-1})(\mathbb{E}\|\widehat{\nu} - \nu\|_B^2)^{1/2}$. A nuisance norm being $\mathrm{O}_{\mathbb{P}}(\varpi_n)$ does not by itself imply that its squared expectation is $\mathrm{O}(\varpi_n^2)$; the mean-square requirements in Assumption 4.4 supply the latter control where the proof uses it. Evaluation on an observation used to fit the nuisance is handled by neighborhood replacement, not by conditioning on the original nuisance estimate. ■

The preceding calculation varies nuisance parameters while holding the observation law fixed. To obtain a pathwise derivative, I instead vary that law and differentiate the identified functional. This distinction underlies the mediation influence function in Tchetgen Tchetgen and Shpitser (2012, Theorem 1). I follow the density-factorization and score-normalization argument of Chen and Ritzwoller (2023), with explicit domination conditions for each derivative.

A dominated one-observation submodel is a family $\{P_u : |u| < u_0\}$ of probability laws for a single generic observation $O = (Y, G, Z, B)$, with $P_0 = \mathsf{P}$, all absolutely continuous with respect to a fixed product of σ-finite measures on the outcome, exposure, assignment, and baseline spaces. The assignment factor uses counting measure. The family varies a marginal law to differentiate the functional; it is not an independence assumption about the n observed individuals. The next lemma imposes explicit smoothness and domination conditions on its factorized densities.

This marginal calculation applies here because the identified cell mean depends on the response regression, exposure kernel and baseline law of a generic observation. Differentiating these factors explains the three terms in the mediation score: the response residual, the exposure-law correction and the baseline average. Domination justifies interchanging differentiation and integration under the lemma's conditions. It does not assert that arbitrary marginal perturbations are generated by the structural network model, or establish an efficiency bound for its joint likelihood. The nuisance orthogonality calculation above and the dependency-graph arguments below provide the separate ingredients needed for estimation with dependent observations.

Lemma B.2 (Derivative of the mediation functional)**.** *Consider a dominated one-observation submodel with density $p_u(y \mid g,z,x)q_u(g \mid z,x)e_u(z \mid x)p_{B,u}(x)$ through the uniform-site law. Assume $|Y| \le M$ and common support for each component density on an interval $|u| < u_0$. On that support each density is differentiable in u and its log derivative is bounded in absolute value by a fixed K_S, uniformly in u and its arguments. Assume the overlap bounds hold throughout this interval. If S is the total score, then*

$$\frac{\mathrm{d}}{\mathrm{d}u}\int\int m_{z,u}(g,x)q_{z',u}(\mathrm{d}g \mid x)F_{B,u}(\mathrm{d}x)\bigg|_{u=0} = \mathbb{E}[(\psi_{zz'}(O;\eta_0)-\theta)S(O)]\,.$$

Proof. For any component density a_u, integration of its log derivative gives $a_u \le e^{K_S|u|}a_0$ and $|\partial_u a_u| \le K_S e^{K_S u_0}a_0$. Products of the relevant densities and their derivatives are therefore dominated by constants times their densities at zero. With $|Y| \le M$, dominated convergence justifies every differentiation and normalization step below. Write the component scores at zero as

$$S_Y = \partial_u \log p_u(Y \mid G,Z,B)\big|_0\,, \qquad S_G = \partial_u \log q_u(G \mid Z,B)\big|_0\,,$$
$$S_Z = \partial_u \log e_u(Z \mid B)\big|_0\,, \qquad S_B = \partial_u \log p_{B,u}(B)\big|_0\,.$$

Their sum is the total score $S = \partial_u \log(\mathrm{d}P_u/\,\mathrm{d}\lambda)|_0$. For example, differentiating $\int p_u(y \mid g,z,x)\lambda_Y(\mathrm{d}y) = 1$ gives

$$0 = \int p_0(y \mid g,z,x)S_Y(y,g,z,x)\lambda_Y(\mathrm{d}y) = \mathbb{E}[S_Y \mid G=g, Z=z, B=x]\,.$$

The same normalization calculation for the other three factors gives $\mathbb{E}[S_G \mid Z,B] = 0$, $\mathbb{E}[S_Z \mid B] = 0$, and $\mathbb{E}S_B = 0$.

Term 1. Variation of the outcome law. Differentiation under the integral yields

$$\begin{aligned}\dot m_z(g,x) &= \partial_u \int y p_u(y \mid g,z,x)\lambda_Y(\mathrm{d}y)\bigg|_0 \\ &= \int y S_Y(y,g,z,x)p_0(\mathrm{d}y \mid g,z,x) \\ &= \int \{y - m_z(g,x)\}S_Y(y,g,z,x)p_0(\mathrm{d}y \mid g,z,x)\,,\end{aligned}$$

where the last line subtracts $m_z(g,x)$ times the conditional normalization identity. Its contribution to the functional derivative is

$$\begin{aligned}A_Y &= \int\int \dot m_z(g,x)q_{z'}(\mathrm{d}g \mid x)F_B(\mathrm{d}x) \\ &= \int\int\int r(g,x)\{y - m_z(g,x)\}S_Y(y,g,z,x)p_0(\mathrm{d}y \mid g,z,x)q_z(\mathrm{d}g \mid x)F_B(\mathrm{d}x) \\ &= \mathbb{E}\left[\frac{\mathbb{1}\{Z=z\}}{e_z(B)}r(G,B)\{Y - m_z(G,B)\}S_Y\right]\,.\end{aligned}$$

The second line is the Radon–Nikodym change from $q_{z'}$ to q_z; the last inserts the observed assignment probability and cancels it with $e_z(B)$.

Term 2. Variation of the exposure law. The product rule gives

$$\begin{aligned} A_G &= \int\!\!\int m_z(g,x) S_G(g,z',x) q_{z'}(\mathrm{d}g \mid x) F_B(\mathrm{d}x) \\ &= \int\!\!\int \{m_z(g,x) - \nu(x)\} S_G(g,z',x) q_{z'}(\mathrm{d}g \mid x) F_B(\mathrm{d}x) \\ &= \mathbb{E}\left[\frac{\mathbb{1}\{Z=z'\}}{e_{z'}(B)}\{m_z(G,B) - \nu(B)\} S_G\right] . \end{aligned}$$

The subtraction uses $\int S_G(g,z',x) q_{z'}(\mathrm{d}g \mid x) = 0$ at each relevant x.

Term 3. Variation of the baseline law. Its contribution is

$$A_B = \int \nu(x) S_B(x) F_B(\mathrm{d}x) = \int \{\nu(x) - \theta\} S_B(x) F_B(\mathrm{d}x) = \mathbb{E}[(\nu(B) - \theta) S_B] .$$

Here $\mathbb{E}S_B = 0$. The functional does not integrate the assignment probabilities, so it has no separate S_Z derivative. Thus its derivative is $A_Y + A_G + A_B$. To verify every remaining inner product, write the three centered score terms as

$$U_Y = \frac{\mathbb{1}\{Z=z\}}{e_z(B)} r(G,B)\{Y - m_z(G,B)\}, \quad U_G = \frac{\mathbb{1}\{Z=z'\}}{e_{z'}(B)}\{m_z(G,B) - \nu(B)\}, \quad U_B = \nu(B) - \theta .$$

Their defining conditional means satisfy

$$\begin{aligned} \mathbb{E}[U_Y \mid G,Z,B] &= 0 , \\ \mathbb{E}[U_G \mid Z,B] &= \frac{\mathbb{1}\{Z=z'\}}{e_{z'}(B)} \left\{ \int m_z(g,B) q_{z'}(\mathrm{d}g \mid B) - \nu(B) \right\} = 0 , \\ \mathbb{E}U_B &= \int \nu(x) F_B(\mathrm{d}x) - \theta = 0 . \end{aligned}$$

For $j \in \{G,Z,B\}$, S_j is measurable given (G,Z,B), and therefore

$$\mathbb{E}[U_Y S_j] = \mathbb{E}[S_j \mathbb{E}(U_Y \mid G,Z,B)] = 0 .$$

The outcome-score normalization gives $\mathbb{E}[U_G S_Y] = \mathbb{E}[U_G \mathbb{E}(S_Y \mid G,Z,B)] = 0$. For $j \in \{Z,B\}$, conditioning on (Z,B) yields $\mathbb{E}[U_G S_j] = \mathbb{E}[S_j \mathbb{E}(U_G \mid Z,B)] = 0$. Finally, successive conditioning of the component-score normalization identities gives

$$\mathbb{E}[S_Y \mid B] = \mathbb{E}[\mathbb{E}(S_Y \mid G,Z,B) \mid B] = 0, \quad \mathbb{E}[S_G \mid B] = \mathbb{E}[\mathbb{E}(S_G \mid Z,B) \mid B] = 0, \quad \mathbb{E}[S_Z \mid B] = 0 .$$

Thus $\mathbb{E}[U_B S_j] = \mathbb{E}[U_B \mathbb{E}(S_j \mid B)] = 0$ for $j \in \{Y,G,Z\}$. Expanding all twelve products now gives

$$\begin{aligned} \mathbb{E}[(\psi - \theta)S] &= \mathbb{E}[(U_Y + U_G + U_B)(S_Y + S_G + S_Z + S_B)] \\ &= \mathbb{E}[U_Y S_Y] + \mathbb{E}[U_G S_G] + \mathbb{E}[U_B S_B] = A_Y + A_G + A_B . \end{aligned}$$

The same identities hold for $z = z'$. This verifies the derivative by its inner products with the factor scores, as in the mediation calculation of Tchetgen Tchetgen and Shpitser (2012) and the factorized-score arguments of Chen and Ritzwoller (2023). It concerns the one-observation functional; the sampling distribution under network dependence is established separately below. ■

The following bound specializes the proof of Cao and Leung (2025, Theorem 1) to a bounded mediation score and exact dependency-graph separation. Their argument already uses independent neighborhood replacement and centering of the score difference. The lemma is a specialization with explicit constants, rather than a new stability principle or a claim of a sharper general theorem.

Lemma B.3 (Neighborhood replacement bound)**.** *Under Assumptions 4.2, 4.3, and 4.4, define the following population-integral maps for each admissible η.*

$$d_i(o;\eta) = \psi(o;\eta) - \psi(o;\eta_0), \quad F_i(\eta) = \mathsf{P}_i d_i(\cdot;\eta), \quad f_i(o;\eta) = d_i(o;\eta) - F_i(\eta)\ .$$

Then

$$\left\| \frac{1}{\sqrt{n}} \sum_{1\le i\le n} f_i(O_i;\widehat{\eta}) \right\|_2 \le 2\sqrt{n}s_n + \left(\Delta_n \overline{\varpi}^2_{\psi,n} + 16 n s_n^2\right)^{1/2}, \qquad \overline{\varpi}_{\psi,n} = \min(\varpi_{\psi,n}, 2B_\psi)\ . \tag{B.3}$$

Proof. The map F_i integrates the evaluation argument against the marginal law of O_i. At a random nuisance estimate it retains the realized nuisance parameters, so $F_i(\widehat{\eta})$ is random. A separate independent evaluation O_i^* makes Jensen's inequality applicable.

$$\|F_i(\widehat{\eta}) - F_i(\widehat{\eta}^{(-i)})\|_2 \le \|\psi(O_i^*;\widehat{\eta}) - \psi(O_i^*;\widehat{\eta}^{(-i)})\|_2 \le s_n\ .$$

Combining this with the assumed training-observation bound and summing by the triangle inequality gives, for $U_i = f_i(O_i;\widehat{\eta}^{(-i)})$,

$$\left\| n^{-1/2} \sum_{1\le i\le n} \{f_i(O_i;\widehat{\eta}) - U_i\} \right\|_2 \le 2\sqrt{n}s_n\ . \tag{B.4}$$

To center U_i, define

$$\mathcal{A}_i = \sigma(O_l : l \notin N[i]) \vee \sigma(\widetilde{\boldsymbol{O}}, \mathsf{U})\ .$$

Assumption 4.2 separates the singleton $\{i\}$ from the entire retained original set. The copied array and U are jointly independent of the originals. Thus $O_i \perp\!\!\!\perp \mathcal{A}_i$ and the replacement learner is $\mathcal{A}_i$-measurable. Consequently

$$\begin{aligned} \mathbb{E}[U_i \mid \mathcal{A}_i] &= 0\ , \\ \mathbb{E}U_i^2 &= \mathbb{E}\left[\mathbb{E}[d_i(O_i;\widehat{\eta}^{(-i)})^2 \mid \mathcal{A}_i] - F_i(\widehat{\eta}^{(-i)})^2\right] \\ &\le \mathbb{E}[d_i(O_i^*;\widehat{\eta}^{(-i)})^2] \le \overline{\varpi}^2_{\psi,n}\ . \end{aligned}$$

The last bound also uses $|d_i| \le 2B_\psi$. There are at most $n\Delta_n$ ordered pairs (i,k) with $k \in N[i]$, including $i = k$. For each, the Cauchy–Schwarz inequality gives $|\mathbb{E}U_iU_k| \le \overline{\varpi}^2_{\psi,n}$.

For $k \notin N[i]$, put $Q_i = f_i(O_i;\widehat{\eta}^{(-i,-k)})$ and define Q_k in the same way. Because $i \notin N[k]$, Q_i is $\mathcal{A}_k$-measurable. Hence $\mathbb{E}[Q_iU_k] = \mathbb{E}[Q_i\mathbb{E}(U_k \mid \mathcal{A}_k)] = 0$. Reversing the roles of the indices gives $\mathbb{E}[U_iQ_k] = 0$. Also set

$$\mathcal{A}_{ik} = \sigma(O_l : l \notin N[i] \cup N[k]) \vee \sigma(\widetilde{\boldsymbol{O}}, \mathsf{U})\ .$$

The set $\{i,k\}$ has no edge to the retained original set, so its full observation pair is independent of $\mathcal{A}_{ik}$. Its two coordinates are independent of each other because i,k are nonadjacent. Conditional on $\mathcal{A}_{ik}$ the common estimate from double-neighborhood replacement observations is fixed. It follows that

$$\mathbb{E}[Q_iQ_k \mid \mathcal{A}_{ik}] = \mathbb{E}[Q_i \mid \mathcal{A}_{ik}]\mathbb{E}[Q_k \mid \mathcal{A}_{ik}] = 0\ .$$

Expanding the product now proves

$$\mathbb{E}[(U_i - Q_i)(U_k - Q_k)] = \mathbb{E}U_iU_k - \mathbb{E}U_iQ_k - \mathbb{E}Q_iU_k + \mathbb{E}Q_iQ_k = \mathbb{E}U_iU_k \, .$$

This step allows $N[i]$ and $N[k]$ to overlap. Their union, rather than their disjointness, determines the retained observations.

For the remaining terms, compare each estimate from replacement observations with the original nuisance estimate. The assumed bounds and the triangle inequality give

$$\|d_i(O_i; \widehat{\eta}^{(-i)}) - d_i(O_i; \widehat{\eta}^{(-i,-k)})\|_2 \leq 2s_n \, .$$

The independent-evaluation version and Jensen's inequality give the same bound for the difference between their F_i values. Thus $\|U_i - Q_i\|_2 \leq 4s_n$, and another Cauchy–Schwarz inequality gives $|\mathbb{E}U_iU_k| \leq 16s_n^2$ for every far pair. Write the normalized second moment as the sum over near and far ordered pairs.

$$K_{1,n} = \frac{1}{n} \sum_{1 \leq i \leq n} \sum_{k \in N[i]} \mathbb{E}U_iU_k, \qquad K_{2,n} = \frac{1}{n} \sum_{1 \leq i \leq n} \sum_{k \notin N[i]} \mathbb{E}U_iU_k \, .$$

The preceding inequalities give $|K_{1,n}| \leq \Delta_n \overline{\varpi}_{\psi,n}^2$ and $|K_{2,n}| \leq 16ns_n^2$. The first count includes all diagonal pairs, and the second includes any nonadjacent pair whose nuisance estimates computed from replacement observations share training observations. Summing gives

$$\mathbb{E}\left(n^{-1/2} \sum_{1 \leq i \leq n} U_i \right)^2 = \frac{1}{n} \sum_{1 \leq i,k \leq n} \mathbb{E}U_iU_k \leq \Delta_n \overline{\varpi}_{\psi,n}^2 + 16ns_n^2 \, . \tag{B.5}$$

Together with (B.4), this proves (B.3). ■

Proof of Theorem 4.2. Use the maps in Lemma B.3. The exact decomposition is

$$\widehat{\theta} - \theta = \frac{1}{n} \sum_{1 \leq i \leq n} \phi_i + \frac{1}{n} \sum_{1 \leq i \leq n} f_i(O_i; \widehat{\eta}) + \frac{1}{n} \sum_{1 \leq i \leq n} F_i(\widehat{\eta}) \, . \tag{B.6}$$

The last average is the mean remainder in Proposition 4.1, evaluated at the fitted functions. Its three components satisfy

$$|\mathbb{E}_B(1 - \widehat{\kappa}_{z'})\delta\widehat{\nu}| \leq e_*^{-1} \|\delta\widehat{e}\|_B \|\delta\widehat{\nu}\|_B = \mathrm{O}_{\mathbb{P}}(\varpi_e\varpi_\nu) \, ,$$
$$|\mathbb{E}_B(\widehat{\kappa}_{z'} - \widehat{\kappa}_z)E_{z'}\delta\widehat{m}| \leq 2\sqrt{R}e_*^{-1} \|\delta\widehat{e}\|_B \|\delta\widehat{m}\|_z = \mathrm{O}_{\mathbb{P}}(\varpi_e\varpi_m) \, ,$$
$$|\mathbb{E}_B\widehat{\kappa}_z E_z[\delta\widehat{r}\delta\widehat{m}]| \leq e_*^{-1} \|\delta\widehat{r}\|_z \|\delta\widehat{m}\|_z = \mathrm{O}_{\mathbb{P}}(\varpi_r\varpi_m) \, .$$

Each uses the candidate-function identity, which holds for every bounded nuisance estimate and therefore may be evaluated at a random nuisance estimate.

For consistency, the centered oracle average obeys

$$\mathbb{E}\left(\frac{1}{n} \sum_{1 \leq i \leq n} \phi_i \right)^2 = \frac{1}{n^2} \sum_{1 \leq i \leq n} \sum_{k \in N[i]} \mathbb{E}\phi_i\phi_k \leq B_\phi^2 \frac{\Delta_n}{n} \longrightarrow 0 \, .$$

Here site centering gives zero means, nonadjacent covariances vanish, and each remaining covariance is at most B_ϕ^2. Dividing (B.3) by $\sqrt{n}$ gives

$$\left\| \frac{1}{n} \sum_{1 \le i \le n} f_i(O_i; \widehat{\eta}) \right\|_2 \le 2 s_n + \left(\frac{4 B_\psi^2 \Delta_n}{n} + 16 s_n^2 \right)^{1/2} \longrightarrow 0 \, .$$

For either random average A_n, Markov's inequality applied to A_n^2 gives $\mathbb{P}(|A_n| > \varepsilon) \le \varepsilon^{-2} \mathbb{E} A_n^2 \to 0$. The mean remainder is $\mathrm{O}_{\mathbb{P}}(b_n) = \mathrm{o}_{\mathbb{P}}(1)$ because $b_n \to 0$. All three terms of (B.6) therefore tend to zero in probability. A fixed linear map C preserves this convergence. With known probabilities and bounded nuisance errors, consistency of either m or r makes the sole remaining product vanish, proving the doubly robust consistency conclusion.

For asymptotic linearity, multiply (B.6) by $\sqrt{n}$. Lemma B.3 bounds its second term by $2\sqrt{n} s_n + \left(\Delta_n \varpi_{\psi,n}^2 + 16 n s_n^2 \right)^{1/2} = \mathrm{o}(1)$ in L_2. Multiplying each of the three displayed mean-remainder bounds by $\sqrt{n}$ gives respectively $\mathrm{O}_{\mathbb{P}}(\sqrt{n} \varpi_e \varpi_\nu)$, $\mathrm{O}_{\mathbb{P}}(\sqrt{n} \varpi_e \varpi_m)$ and $\mathrm{O}_{\mathbb{P}}(\sqrt{n} \varpi_r \varpi_m)$, all tending to zero. The first term remains $n^{-1/2} \sum_{1 \le i \le n} \phi_i$, which proves (4.9). Applying the argument to the fixed number of coordinates proves the vector statement by a finite union bound. ∎

Appendix C. Proofs for Inference

The graph $\mathcal{H}_n$, closed neighborhoods $N[i]$ and size Δ_n cover the full observation $O_i = (B_i, Z_i, G_i, Y_i)$, including $B_i = (X_i, N_i^{\mathrm{pre}})$. Oracle expectations integrate that conditional observation law. Expectations involving fitted scores additionally integrate the fitting and replacement variables defined in Assumption 4.4.

All statements through the final remark condition on the entire supported village-saturation profile. The Cramér–Wold characterization, continuous-mapping theorem and Slutsky's lemma are used in their usual finite-dimensional forms (van der Vaart, 1998, Chapter 2). I give the dependence and covariance calculations explicitly.

C.1. **A bounded dependency-graph central limit theorem.** For a positive integer r, let $\mathcal{P}_r$ be the set of partitions of $[r]$ into nonempty disjoint blocks. For bounded variables $X_1, \dots, X_r$, define

$$\mathrm{cum}(X_1, \dots, X_r) := \sum_{\pi \in \mathcal{P}_r} (|\pi| - 1)! (-1)^{|\pi|-1} \prod_{A \in \pi} \mathbb{E} \prod_{a \in A} X_a \, , \qquad A_r := \sum_{\pi \in \mathcal{P}_r} (|\pi| - 1)! \, . \tag{C.1}$$

Write $\mathrm{cum}_r(X) = \mathrm{cum}(X, \dots, X)$. The first two cumulants are the mean and variance. Also define the odd double factorial by $(2k-1)!! := (2k-1)(2k-3) \cdots 3 \cdot 1 = (2k)!/(2^k k!)$ for $k \ge 1$, and $(-1)!! := 1$. It counts pair partitions of $2k$ labeled positions.

Lemma C.1 (Connected cumulants)**.** *Let $\mathbb{E} X_i = 0$, $|X_i| \le M_X$, and suppose the observation array has a dependency graph of maximum closed-neighborhood size Δ_n. If $n^{-1} \mathrm{Var}(\sum_{1 \le i \le n} X_i) \to v < \infty$ and $\Delta_n^4 / n \to 0$, then $n^{-1/2} \sum_{1 \le i \le n} X_i \xrightarrow{d} \mathcal{N}(0, v)$, allowing $v = 0$.*

Proof. The argument follows the cumulant method of Janson (1988), with the partition identities reviewed in Döring et al. (2022, Sections 2 and 3.4). Féray (2018, Section 4) extends dependency-graph reasoning by weighting weak interactions; Dousse and Féray (2019) obtain such weights from bounds on spin cumulants, and Eichelsbacher et al. (2015, Section 3) bound cumulant measures of stabilizing geometric functionals. I use the ordinary dependency graph and prove the needed bound directly. The local-dependence normal approximation in Chen and Shao (2004), by comparison, uses Stein's method and concentration inequalities. I first establish the cumulant identities and then enumerate their potentially nonzero terms.

There are also cumulant proofs outside the dependency-graph literature. Soshnikov (2002, Section 2, Lemmas 1–2) bound cumulants of linear statistics of determinantal point fields and deduce a normal limit when their normalized higher cumulants vanish. The bounds there come from integral kernels, not from a finite-range graph. Nourdin and Peccati (2010, Sections 4–5, especially Theorem 5.8) derive cumulants on Wiener space by repeated integration by parts and use contraction bounds to prove normal convergence on a fixed Wiener chaos. Their contraction condition replaces the connected-tuple count used below. In statistical mechanics, the spin cumulants studied by Dousse and Féray (2019) retain long-distance interactions through weights rather than declaring them independent. Thus the common proof strategy is to isolate connected contributions, bound higher cumulants after normalization, and recover the normal moments; the mechanism controlling those contributions is model-specific. Neither determinantal structure, Gaussian chaos nor an Ising expansion is assumed here. The elementary constants below are sufficient bounds and are not claimed to improve the sharper graph bounds in Féray (2018), Döring et al. (2022).

Two further comparisons illustrate the scope of this method. In homogeneous dynamics, Björklund and Gorodnik (2017, Sections 5–8) reorganize cumulants around separated groups of orbit points and control the remaining terms using exponential mixing of all orders. Their Theorem 1.1 also requires the group-growth, averaging-set and regularity conditions in Sections 1–2; pairwise mixing alone does not supply its argument. In enumerative probability, Féray (2019, Theorem 2.3 and Sections 3–4) obtain normal limits for pattern counts in multiset permutations and set partitions from weighted dependency graphs. Disjoint position sets can still be dependent there, so the weights account for interactions that an ordinary dependency graph would have to retain. These examples motivate checking higher-order joint dependence directly. The exact zero for disconnected tuples below comes instead from Assumption 4.2; none of those external mixing or combinatorial hypotheses is imported into the present theorem.

Step 1. Multilinearity and an absolute bound. In each product in (C.1), a particular variable X_a occurs in exactly one block moment and occurs there once. Linearity of that expectation therefore implies, for scalars c, d,

$$\operatorname{cum}(X_1, \dots, cU + dV, \dots, X_r) = c \operatorname{cum}(X_1, \dots, U, \dots, X_r) + d \operatorname{cum}(X_1, \dots, V, \dots, X_r) .$$

This proves multilinearity in each argument. Since the block sizes sum to r, each moment product has absolute value at most M_X^r, and hence $|\operatorname{cum}(X_1, \dots, X_r)| \le A_r M_X^r$.

Step 2. The independence cancellation. To verify the independence property, expand the log moment generating function

$$\log \mathbb{E} \exp \left(\sum_{1 \le a \le r} u_a X_a \right)$$

near zero. Boundedness makes the moment generating function analytic there and positive at zero. Expanding $\log(1 + x)$ and collecting the coefficient containing every u_a once gives (C.1). A partition with k blocks has $k!$ block orderings and receives the factor $(-1)^{k-1}/k$, giving $(k-1)!(-1)^{k-1}$. If two nonempty groups of variables are independent, the moment generating function is a product of their two generating functions. Its logarithm is a sum, each summand missing at least one of the variables. The derivative with respect to all $u_1, \dots, u_r$ is consequently zero.

Step 3. Enumeration of potentially nonzero terms. For an ordered tuple of sites $(i_1, \dots, i_r)$, make a graph on the labeled positions $[r]$. Positions a, b are connected if $i_a = i_b$ or if i_a, i_b are adjacent sites. If the position graph is disconnected, choose one component and its complement. Repeated site indices cannot lie in different components. The two corresponding site sets are therefore disjoint and have no dependence edge between them.

Setwise independence, rather than pairwise uncorrelatedness, gives the independent split needed in the preceding paragraph. Thus this tuple's cumulant is zero.

A connected position graph contains a spanning tree, a connected subgraph containing all positions and no cycle. Choose position 1 as its root. Orient every edge away from the root; each other position then has exactly one parent, the preceding position on its unique path to the root. There are at most r^{r-1} parent assignments, since each of the $r-1$ nonroot positions can name at most r parents. Some assignments are not trees, which only enlarges this upper bound. For a fixed rooted tree, choose the root's site in n ways. Visit each parent before its children. A child must lie in its parent's closed dependence neighborhood and therefore has at most Δ_n choices. It follows that

$$\#\{(i_1,\dots,i_r) : \text{connected position graph}\} \le r^{r-1} n \Delta_n^{r-1} . \tag{C.2}$$

The count includes repeated sites and counts a tuple more than once if several trees are compatible with it. For example, at $r=3$ there are at most $9n\Delta_n^2$ tuples by this bound, and $A_3 = 1+3+2 = 6$.

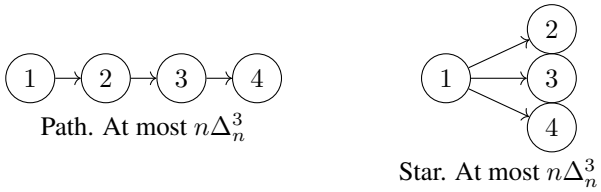


FIGURE 2. Two rooted spanning trees on four labeled positions. The root has n site choices and each arrow permits at most Δ_n choices for its child. Arrows describe the enumeration; the dependence graph itself is undirected.

Step 4. Normalized cumulants. Let $S_n = n^{-1/2}\sum_{1\le i\le n} X_i$. Applying multilinearity to the first argument expands it as a sum over i_1, applying it to the second introduces i_2, and repeating through the rth argument gives

$$\operatorname{cum}_r(S_n) = n^{-r/2} \sum_{1\le i_1\le n} \cdots \sum_{1\le i_r\le n} \operatorname{cum}(X_{i_1},\dots,X_{i_r}) .$$

Combining the zero cumulants for disconnected tuples, (C.2), and the bound $A_r M_X^r$ yields

$$\begin{aligned} |\operatorname{cum}_r(S_n)| &\le C_r n^{1-r/2}\Delta_n^{r-1} , \qquad C_r := r^{r-1} A_r M_X^r , \\ &= C_r(\Delta_n/n^{1/4})^{r-1} n^{(3-r)/4} \longrightarrow 0 \quad (r\ge 3) . \end{aligned} \tag{C.3}$$

In particular the order-three bound is $54 M_X^3 \Delta_n^2/\sqrt{n}$. The first cumulant is zero and the second tends to v. To display the first two nontrivial cancellations, centering gives

$$\begin{aligned} \operatorname{cum}(X_i,X_j,X_k) &= \mathbb{E}X_iX_jX_k , \\ \operatorname{cum}(X_i,X_j,X_k,X_l) &= \mathbb{E}X_iX_jX_kX_l - \mathbb{E}X_iX_j\mathbb{E}X_kX_l \\ &\quad - \mathbb{E}X_iX_k\mathbb{E}X_jX_l - \mathbb{E}X_iX_l\mathbb{E}X_jX_k . \end{aligned}$$

These formulas hold with repeated indices. At order four there are $1,7,6,1$ partitions into respectively one, two, three and four blocks, so $A_4 = 1+7+12+6 = 26$. Our uniform bound gives $|\operatorname{cum}_4(S_n)| \le 1664 M_X^4 \Delta_n^3/n$. Some connected tuples still have zero cumulant; counting all of them is a valid upper bound. Conversely, covariance between neighboring scores need not disappear. Those order-two terms remain in the limiting variance v; it is the normalized higher cumulants that vanish.

Step 5. Moments and the distributional limit. Expanding the exponential of the cumulant generating function gives the inverse partition identity $\mathbb{E}S_n^r = \sum_{\pi\in\mathcal{P}_r}\prod_{A\in\pi}\operatorname{cum}_{|A|}(S_n)$. Any block of size one contributes zero, and any block of size at least three tends to zero. For odd r no partition into pairs exists, so the limiting moment is zero. For $r = 2k$, choose the partner of the first remaining position in $2k-1$ ways, then choose the next in $2k-3$ ways, continuing to one. Exactly $(2k-1)!!$ pair partitions remain, each contributing v^k. For an explicit remainder bound at a fixed moment order $r \geq 3$, put

$$b_{r,n} = \max_{3\leq j\leq r} j^{j-1}A_j M_X^j n^{1-j/2}\Delta_n^{j-1}, \qquad K_{r,n} = \max(1, \mathbb{E}S_n^2, b_{r,n}) .$$

There are $|\mathcal{P}_r|$ partitions in total. Any contributing partition that is not a pair partition contains a block of size at least three, bounded by $b_{r,n}$; its other cumulant factors are bounded by $K_{r,n}$. Thus the absolute contribution of all such partitions is at most $|\mathcal{P}_r|b_{r,n}K_{r,n}^{r-1} \to 0$. This verifies the termwise limit with its finite combinatorial constant. The moment limits and the following determinacy argument prove convergence to $\mathcal{N}(0,v)$. If $v = 0$, $\mathbb{E}S_n^2 \to 0$ gives convergence directly by Chebyshev's inequality. ■

Remark C.1 (From moments to the normal law)**.** The moment argument is sufficient here without an additional tail assumption. Bounded second moments give tightness. Bounded $2r$th moments give uniform integrability of $|S_n|^r$, since $\mathbb{E}[|S_n|^r\mathbb{1}\{|S_n| > K\}] \leq K^{-r}\mathbb{E}|S_n|^{2r}$. Any subsequential distributional limit therefore has the moments calculated above. For a variable S with those moments, monotone convergence applied to the nonnegative series for $\cosh(uS)$ gives

$$\mathbb{E}\cosh(uS) = \sum_{k\geq 0}\frac{u^{2k}(2k-1)!!v^k}{(2k)!} = e^{vu^2/2} .$$

Thus $\mathbb{E}e^{|uS|} \leq 2\mathbb{E}\cosh(uS) < \infty$. The characteristic function can be expanded under the expectation and equals $e^{-vu^2/2}$, identifying the normal law. Every subsequential limit agrees, so the whole sequence converges. ♢

Proof of Theorem 5.1. For a fixed $c \in \mathbb{R}^{d_\theta}$ let $X_i = c^\top\phi_i$. Equation (B.1) gives $\mathbb{E}X_i = 0$ and the score bound gives $|X_i| \leq \|c\|_1 B_\phi$. Each X_i is measurable in O_i, so the setwise dependence graph is unchanged. Its normalized-sum variance equals $c^\top V_n c \to c^\top V c$. Lemma C.1 applies, including when the limiting variance is zero. Since this holds for every fixed c, Cramér–Wold yields the vector oracle limit. Theorem 4.2(ii) gives

$$\sqrt{n}(\widehat{\theta}_n - \theta_n) = n^{-1/2}\sum_{1\leq i\leq n}\phi_{i,n} + o_{\mathbb{P}}(1) \xrightarrow{d} \mathcal{N}(0,V) .$$

Under exact identification, its center is the finite-n natural or interventional causal vector. Lemma D.3 gives the additional deterministic bound $\sqrt{n}\|\theta_n - \theta\| \leq \sqrt{d_\theta n}\delta_{T,n} \to 0$. Adding this term to the expansion and applying Slutsky's lemma proves (5.3) centered at the fixed causal parameter. The fold-specific version has the same proof with the bounds and scores of Theorem J.1. ■

C.2. **Covariance concentration and fitted scores.** Write $\widetilde{V}_n = n^{-1}\sum_{1\leq i,k\leq n} w_{ik,n}\phi_i\phi_k^\top$ for the oracle quadratic form. The covariance error has the decomposition

$$\widehat{V}_n - V = (\widehat{V}_n - \widetilde{V}_n) + (\widetilde{V}_n - \mathbb{E}\widetilde{V}_n) + (\mathbb{E}\widetilde{V}_n - V_n) + (V_n - V) . \tag{C.4}$$

The next two calculations control each random term and the bias.

Proof of Proposition 5.2. Oracle step. Fix coordinates $a, b \in [d_\theta]$ and set $Q_n^{ab} = n^{-1} \sum_{1 \le i,k \le n} w_{ik,n} \phi_i^{(a)} \phi_k^{(b)}$. Its expectation satisfies

$$\begin{aligned} \mathbb{E}Q_n^{ab} - V_n^{ab} &= \frac{1}{n} \sum_{1 \le i,k \le n} w_{ik,n} \mathbb{E}[\phi_i^{(a)} \phi_k^{(b)}] - \frac{1}{n} \sum_{1 \le i,k \le n} \mathbb{E}[\phi_i^{(a)} \phi_k^{(b)}] \\ &= \frac{1}{n} \sum_{1 \le i,k \le n} (w_{ik,n} - 1) \mathbb{E}[\phi_i^{(a)} \phi_k^{(b)}] \longrightarrow 0 \end{aligned}$$

by the matrix bias condition in Assumption 5.2.

Let $\mathcal{D}_n = \{(i,k) : w_{ik,n} \neq 0\}$; it has at most nL_n ordered pairs. Expanding the variance gives

$$\mathrm{Var}(Q_n^{ab}) = \frac{1}{n^2} \sum_{(i,k) \in \mathcal{D}_n} \sum_{(j,l) \in \mathcal{D}_n} w_{ik,n} w_{jl,n} \mathrm{Cov}(\phi_i^{(a)} \phi_k^{(b)}, \phi_j^{(a)} \phi_l^{(b)}) \, .$$

For a fixed (i,k), if neither j nor l belongs to $N[i] \cup N[k]$, then the two site sets are disjoint and no edge joins them. Setwise dependence makes the products independent, so their covariance is zero. Otherwise count the possible (j,l) as follows. There are at most $2\Delta_n$ choices for an endpoint in $N[i] \cup N[k]$. That endpoint can be j or l, giving a factor two. Once it is chosen, the other endpoint must have a nonzero kernel weight, giving at most L_n choices, for either endpoint by symmetry of W_n. Hence at most $4\Delta_n L_n$ ordered second pairs can have nonzero covariance. Repeated indices and pairs counted twice are included in this upper bound.

Each product is bounded by B_ϕ^2, so

$$\begin{aligned} |\mathrm{Cov}(\phi_i^{(a)} \phi_k^{(b)}, \phi_j^{(a)} \phi_l^{(b)})| &\le \mathbb{E}|\phi_i^{(a)} \phi_k^{(b)} \phi_j^{(a)} \phi_l^{(b)}| + |\mathbb{E}\phi_i^{(a)} \phi_k^{(b)}| \, |\mathbb{E}\phi_j^{(a)} \phi_l^{(b)}| \\ &\le 2B_\phi^4 \, . \end{aligned}$$

Using $|w_{ik,n} w_{jl,n}| \le 1$ gives

$$\mathrm{Var}(Q_n^{ab}) \le \frac{(nL_n)(4\Delta_n L_n)(2B_\phi^4)}{n^2} = 8B_\phi^4 \frac{\Delta_n L_n^2}{n} \longrightarrow 0 \, . \tag{C.5}$$

By Chebyshev's inequality, for every $u > 0$ and coordinates a, b,

$$\mathbb{P}\{|\widetilde{V}_n^{ab} - \mathbb{E}\widetilde{V}_n^{ab}| > u\} \le \frac{8B_\phi^4 \Delta_n L_n^2}{nu^2} \longrightarrow 0 \, .$$

Since $\|A\|_{\mathrm{op}} \le d_\theta \max_{a,b} |A^{ab}|$, a union bound also gives

$$\mathbb{P}\{\|\widetilde{V}_n - \mathbb{E}\widetilde{V}_n\|_{\mathrm{op}} > u\} \le \frac{8d_\theta^4 B_\phi^4 \Delta_n L_n^2}{nu^2} \longrightarrow 0 \, .$$

The cross-coordinate formula can also be seen directly by polarization. For scalar arrays u, v, symmetry of W implies $\sum_{1 \le i,k \le n} w_{ik,n} u_i v_k = \sum_{1 \le i,k \le n} w_{ik,n} v_i u_k$, by interchanging i, k. Expanding $Q_n(u+v)$ gives

$$Q_n(u+v) = Q_n(u) + Q_n(v) + \frac{2}{n} \sum_{1 \le i,k \le n} w_{ik,n} u_i v_k \, .$$

Thus the cross entry is one half of the difference of these three scalar quadratic forms. This identity holds for the summed form, including its two cross-products. Since there are d_θ^2 entries and $\|A\|_{\mathrm{op}} \le \|A\|_F \le d_\theta \max_{a,b} |A^{ab}|$, a finite union bound gives convergence in operator norm. Combining the centered-entry bound, the bias bound and $\|V_n - V\|_{\mathrm{op}} \to 0$ proves the oracle conclusion. ∎

Proof of Proposition 5.2. Feasible step. Let $e_i = \widehat{\phi}_i - \phi_i$. For every pair (i,k) expand the difference of products.

$$\begin{aligned}\widehat{\phi}_i\widehat{\phi}_k^\top - \phi_i\phi_k^\top &= (\phi_i + e_i)(\phi_k + e_k)^\top - \phi_i\phi_k^\top \\ &= e_i\phi_k^\top + \phi_i e_k^\top + e_i e_k^\top \ . \end{aligned} \tag{C.6}$$

I bound the resulting sums without presuming independence of fitted scores.

First $\|W_n\|_{\mathrm{op}} \le L_n$. Indeed for $v \in \mathbb{R}^n$, the Cauchy–Schwarz inequality gives

$$\begin{aligned}\|W_n v\|^2 &\le \sum_{1\le i\le n}\left(\sum_{1\le k\le n}|w_{ik,n}|\right)\left(\sum_{1\le k\le n}|w_{ik,n}|v_k^2\right) \\ &\le L_n \sum_{1\le k\le n} v_k^2 \sum_{1\le i\le n}|w_{ik,n}| \le L_n^2\|v\|^2 \ ,\end{aligned}$$

where symmetry bounds column sums by the same L_n. For a scalar array u write $\|u\|_{2,n}^2 = n^{-1}\sum_{1\le i\le n} u_i^2$. For coordinates a, b, the three terms in (C.6) satisfy

$$\left|\frac{1}{n}\sum_{1\le i,k\le n} w_{ik,n}e_i^{(a)}\phi_k^{(b)}\right| \le L_n\|e^{(a)}\|_{2,n}\|\phi^{(b)}\|_{2,n} \ ,$$

$$\left|\frac{1}{n}\sum_{1\le i,k\le n} w_{ik,n}\phi_i^{(a)}e_k^{(b)}\right| \le L_n\|\phi^{(a)}\|_{2,n}\|e^{(b)}\|_{2,n} \ ,$$

$$\left|\frac{1}{n}\sum_{1\le i,k\le n} w_{ik,n}e_i^{(a)}e_k^{(b)}\right| \le L_n\|e^{(a)}\|_{2,n}\|e^{(b)}\|_{2,n} \ .$$

Since $\|\phi^{(a)}\|_{2,n} \le B_\phi$ and each coordinate error norm is at most the vector norm $\|e\|_{2,n}$,

$$\max_{a,b}|\widehat{V}_n^{ab} - \widetilde{V}_n^{ab}| \le L_n\{2B_\phi\|e\|_{2,n} + \|e\|_{2,n}^2\} \xrightarrow{\mathbb{P}} 0 \ .$$

The assumption $L_n\|e\|_{2,n} \xrightarrow{\mathbb{P}} 0$ also implies $\|e\|_{2,n} \xrightarrow{\mathbb{P}} 0$ because $L_n \ge 1$. The second product thus vanishes as well. The fixed-dimensional operator bound used in the oracle step proves $\|\widehat{V}_n - \widetilde{V}_n\|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0$. The four terms of (C.4) now prove the proposition.

It remains to verify the sufficient prediction rate. I follow the independent-neighborhood replacement argument of Cao and Leung (2025, Theorem 1), specialized to exact separation as in Lemma B.3. Let $\mathcal{D}_n = (O_1, \dots, O_n)$, let $\mathcal{D}_n' = (O_1', \dots, O_n')$ be the independent copy used in Assumption 4.4, and define

$$O_l^{(-i)} = \begin{cases} O_l', & l \in N[i] \ , \\ O_l, & l \notin N[i] \ , \end{cases} \qquad \widehat{\eta}^{(-i)} = \mathsf{L}_n((O_l^{(-i)})_{l=1}^n, \mathsf{U}) \ ,$$

where L_n is the complete nuisance estimation map, using the same independent auxiliary variable U in every nuisance estimate. For a fixed score coordinate, adding and subtracting its value at the nuisance estimate computed from these neighborhood replacement observations, we have

$$\begin{aligned} d_i &\coloneqq \psi(O_i;\widehat{\eta}) - \psi(O_i;\eta_0) \\ &= \{\psi(O_i;\widehat{\eta}) - \psi(O_i;\widehat{\eta}^{(-i)})\} + \{\psi(O_i;\widehat{\eta}^{(-i)}) - \psi(O_i;\eta_0)\} \ . \end{aligned}$$

The first difference has squared expectation at most s_n^2 by Assumption 4.4. For the second define $\mathcal{A}_i$ as in Lemma B.3. The independence proved there gives

$$\begin{aligned}&\mathbb{E}|\psi(O_i;\widehat{\eta}^{(-i)})-\psi(O_i;\eta_0)|^2\\&=\mathbb{E}\left[\int|\psi(o;\widehat{\eta}^{(-i)})-\psi(o;\eta_0)|^2\mathcal{L}(O_i)(\mathrm{d}o)\right]\\&=\mathbb{E}|\psi(O_i^*;\widehat{\eta}^{(-i)})-\psi(O_i^*;\eta_0)|^2\le\varpi_{\psi,n}^2\,.\end{aligned}$$

Here $\mathcal{A}_i=\sigma(O_l:l\notin N[i])\vee\sigma(\mathcal{D}_n',\mathsf{U})$, and $\widehat{\eta}^{(-i)}$ is $\mathcal{A}_i$-measurable. To make the conditional-law step explicit, for every bounded measurable u and $A\in\mathcal{A}_i$, independence gives

$$\mathbb{E}[\mathbb{1}_A u(O_i)]=\mathbb{P}(A)\int u(o)\mathcal{L}(O_i)(\mathrm{d}o)\,,\qquad\mathcal{L}(O_i\mid\mathcal{A}_i)=\mathcal{L}(O_i)\quad\text{a.s.}$$

For a nonnegative jointly measurable $v(o,\eta)$, approximation by simple functions and monotone convergence then imply

$$\mathbb{E}[v(O_i,\widehat{\eta}^{(-i)})\mid\mathcal{A}_i]=\int v(o,\widehat{\eta}^{(-i)})\mathcal{L}(O_i)(\mathrm{d}o)\,.$$

Apply this identity to the squared score difference in the preceding display. An independent O_i^* has the same conditional integral, which proves the second equality and licenses the prediction bound in Assumption 4.4. Using $(u+v)^2\le 2u^2+2v^2$ and averaging gives

$$\frac{1}{n}\sum_{1\le i\le n}\mathbb{E}[|d_i|^2]\le 2(s_n^2+\varpi_{\psi,n}^2)\,.$$

For the vector the right side is multiplied by d_θ. Markov's inequality at threshold $K(s_n+\varpi_{\psi,n})^2$ proves the empirical vector error is $\mathrm{O}_{\mathbb{P}}(s_n+\varpi_{\psi,n})$. Under Theorem 4.2(ii), the finite bound on V_n and Chebyshev's inequality give $\widehat{\theta}_n-\theta_n=\mathrm{O}_{\mathbb{P}}(n^{-1/2})$. Subtracting this common center gives

$$\|\widehat{\phi}-\phi\|_{2,n}\le\left(\frac{1}{n}\sum_{1\le i\le n}\|d_i\|^2\right)^{1/2}+\|\widehat{\theta}_n-\theta_n\|=\mathrm{O}_{\mathbb{P}}(s_n+\varpi_{\psi,n}+n^{-1/2})\,.$$

This proves the sufficient rate. The argument covers regime means using their simplified scores. For cross-fitting with separated training sets, independence follows from the separated training sets and the conditional calculation at the end of Appendix J supplies the corresponding empirical bound. ■

Proof of Corollary 5.1. Let $v=c^\top Vc>0$. The numerator $\sqrt{n}c^\top(\widehat{\theta}_n-\theta_n)$ converges to $\mathcal{N}(0,v)$, and the variance estimate converges to v. Continuity of the square root and division at $v>0$, followed by Slutsky's lemma, gives the standard-normal studentized limit. The endpoints $\pm z_{1-\alpha/2}$ have zero probability under that limit, proving interval coverage. Under the joint null, inversion is continuous at the positive-definite matrix $R_0VR_0^\top$. The Wald statistic consequently converges to $Z^\top Z$ for $Z\sim\mathcal{N}(0,I_r)$, which is χ_r^2. Under Assumption D.1, $\sqrt{n}(\theta_n-\theta)\to 0$, so Slutsky's lemma gives the same conclusions with the fixed target and its corresponding null. ■

C.3. **Comparing variance estimates.** A researcher may have two variance estimates that capture different sources of uncertainty. Network weights account for dependence extending across village boundaries, while a village-cluster calculation aggregates all contributions within a village. The distinction matters when training groups or relationships cross administrative boundaries, and when finite-sample covariance contributions are poorly approximated by one choice of weights. Leung (2025b) takes the larger of two variance estimates to retain their

complementary finite-sample properties; his analysis also distinguishes centering bias from covariance estimation. Here the same maximum construction protects coverage whenever at least one candidate is a valid variance benchmark for the correctly centered mediation contrast. For a scalar contrast write $\widehat{v}_1 = c^\top \widehat{V}_n^{\text{net}} c$ and let $\widehat{v}_2$ be a finite nonnegative candidate. For symmetric matrix candidates A, B, define

$$\mathcal{E}(A, B) = A + (B - A)_+ , \tag{C.7}$$

where the positive part replaces negative eigenvalues by zero. The construction gives $\mathcal{E}(A, B) \succeq A$ and $\mathcal{E}(A, B) \succeq B$.

Proposition C.1. *Suppose* $\sqrt{n}(c^\top \widehat{\theta}_n - c^\top \theta_n) \xrightarrow{d} \mathcal{N}(0, v)$ *for* $v > 0$ *and* $\widehat{v}_1 \xrightarrow{\mathbb{P}} v^+ \geq v$ *for a finite* v^+*. Let both* $\widehat{v}_1$ *and* $\widehat{v}_2$ *be finite and nonnegative. The interval with variance* $\max(\widehat{v}_1, \widehat{v}_2)/n$ *has asymptotic coverage at least* $1 - \alpha$ *for* $c^\top \theta_n$*. Under Assumption D.1, the same conclusion holds for* $c^\top \theta$*. Under the joint limit in Theorem 5.1, if* $R_0 A R_0^\top \xrightarrow{\mathbb{P}} R_0 V R_0^\top + D$ *with a finite* $D \succeq 0$ *and* $R_0 V R_0^\top \succ 0$*, the Wald test using* $R_0 \mathcal{E}(A, B) R_0^\top$ *and the* χ^2_r *cutoff has asymptotic size at most* α*.*

The scalar conclusion follows because the maximum-variance interval contains the benchmark interval in every sample. The matrix conclusion uses positive-semidefinite order and the corresponding inverse order for the covariance matrices of the contrasts in the joint test. Selecting the matrix with the larger trace does not ensure this property. Exact component identities can reduce the rank, so the test uses a fixed collection of linearly independent contrasts.

Remark C.2 (Conditional coverage and empirical variance comparisons)**.** For each contrast, one can report the network standard error, a village-cluster standard error motivated by saturation-design analyses such as Baird et al. (2018), and their maximum. Both candidates must estimate the variance of the same contrast, on the same scale and using the same target population. If the network estimate is consistent under Proposition 5.2, taking the maximum preserves its asymptotic coverage even if the cluster estimate has no probability limit. For the primary specification, I recommend reporting the network estimate when its geometry, bandwidth and score-error conditions can be justified. Report both candidates and their maximum when within-village aggregation provides a useful empirical comparison, following Leung (2025b). In that case label the maximum interval as conservative relative to the justified benchmark and show which estimate determines its width. A larger cluster estimate can widen the interval without supplying an independent consistency result.

These intervals are valid conditional on the realized saturation profile, and unconditionally when the profile support is fixed and finite and the conditions hold for every supported profile, using its corresponding standard error; see Appendix C. ♢

Proof of Proposition C.1. The scalar benchmark has limiting coverage $\mathbb{P}\{|\mathcal{N}(0, v)| \leq z_{1-\alpha/2}\sqrt{v^+}\} \geq 1 - \alpha$. In every sample its radius is at most the radius using $\max(\widehat{v}_1, \widehat{v}_2)$. Inclusion of these coverage events proves the assertion for any finite nonnegative second candidate.

For the matrix construction write $D_0 = B - A$. The spectral decomposition gives $D_0 = (D_0)_+ - (-D_0)_+$, so $\mathcal{E}(A, B) - A = (D_0)_+ \succeq 0$ and $\mathcal{E}(A, B) - B = (-D_0)_+ \succeq 0$. Congruence by R_0 preserves these inequalities. With probability tending to one, $R_0 A R_0^\top$ is positive definite, since its limit is $V_R + D$ with $V_R = R_0 V R_0^\top \succ 0$ and $D \succeq 0$. On that event the order-reversing property of matrix inversion (Bhatia, 2007, Exercise 1.2.12) gives

$$\{R_0 \mathcal{E}(A, B) R_0^\top\}^{-1} \preceq (R_0 A R_0^\top)^{-1} .$$

For completeness, if $H \succeq K \succ 0$, set $C = K^{-1/2} H K^{-1/2} \succeq I$. The spectral decomposition gives $C^{-1} \preceq I$ and hence

$$H^{-1} = K^{-1/2} C^{-1} K^{-1/2} \preceq K^{-1/2} I K^{-1/2} = K^{-1} .$$

This comparison extends the scalar interval-inclusion argument used for the variance maximum in Leung (2025b) to a joint test of several contrasts; it does not require the candidates to commute. The envelope's Wald statistic is therefore no larger than the benchmark's. Its limiting benchmark statistic is

$$Z^\top V_R^{1/2} (V_R + D)^{-1} V_R^{1/2} Z \le Z^\top Z , \qquad Z \sim \mathcal{N}(0, I_r) ,$$

by the same order argument. Acceptance at the χ^2_r cutoff has probability at least $1 - \alpha$. Samplewise inclusion completes the proof. The event on which the benchmark is singular has probability tending to zero; defining acceptance there leaves the asymptotic conclusion unchanged. ■

Remark C.3 (Removing profile conditioning)**.** Reintroduce the profile only in this remark. For a common deterministic target and a confidence set $C_n(t)$,

$$\mathbb{P}\{\theta \in C_n(T)\} = \sum_{t \in \mathfrak{T}} \mathbb{P}(T = t) \mathbb{P}\{\theta \in C_n(t) \mid T = t\} .$$

This is the law of total probability, or equivalently iterated conditional expectation (Kallenberg, 2002, Chapter 6). Write $p_n(t) = \mathbb{P}(T = t)$ and $a_n(t) = \mathbb{P}\{\theta \in C_n(t) \mid T = t\}$. If every conditional coverage tends to $1 - \alpha$, finiteness of $\mathfrak{T}$ gives

$$\begin{aligned} \big|\mathbb{P}\{\theta \in C_n(T)\} - (1 - \alpha)\big| &\le \sum_{t \in \mathfrak{T}} p_n(t) |a_n(t) - (1 - \alpha)| \\ &\le \max_{t \in \mathfrak{T}} |a_n(t) - (1 - \alpha)| \longrightarrow 0 . \end{aligned}$$

For conservative coverage, let $\epsilon_n = \max_t [(1 - \alpha) - a_n(t)]_+$. If $\liminf_{n \to \infty} a_n(t) \ge 1 - \alpha$ for every t, then $\epsilon_n \to 0$ and

$$\mathbb{P}\{\theta \in C_n(T)\} \ge \sum_{t \in \mathfrak{T}} p_n(t)(1 - \alpha - \epsilon_n) = 1 - \alpha - \epsilon_n .$$

Neither conclusion requires the profile probabilities to converge. Conditional randomization analyses likewise establish validity at the conditioning level before averaging (Athey et al., 2018); the calculation here averages the asymptotic coverage errors over a finite profile set. If profile probabilities converge to $p(t)$, the unconditional characteristic function of an unstudentized scalar contrast tends to $\sum_{t \in \mathfrak{T}} p(t) \exp(-u^2 v(t)/2)$. This is a centered normal mixture. Its fourth moment is $3 \sum_{t \in \mathfrak{T}} p(t) v(t)^2$, compared with $3\{\sum_{t \in \mathfrak{T}} p(t) v(t)\}^2$ for a normal with the same variance. Their difference is $3\mathrm{Var}_p(v(T))$. Conditional studentization removes this scale difference. When the target itself depends on the profile, averaging coverage concerns that profile-specific target; it does not establish coverage for a different common target. ♢

Appendix D. Dependence graphs and covariance kernels

All probabilities in this appendix condition on the entire supported village-saturation profile, as in Section 4. A dependence graph for the statistical analysis must cover every random input to an observation. This includes covariates and baseline summaries, as well as the exposure and outcome. Write $\mathcal{H}_n$ for the dependency graph for the observations, $N[i]$ for the closed neighborhood containing i and all its graph neighbors, and $\Delta_n = \max_i |N[i]|$. The baseline variable $B_i = (X_i, N_i^{\mathrm{pre}})$ is part of $O_i = (B_i, Z_i, G_i, Y_i)$ and therefore contributes to that graph. To construct it, list the independent primitive variables used to produce each complete observation. Observations

whose lists share an input are joined. The following lemma verifies that observations in two unconnected groups are independent as collections.

Lemma D.1. *Let $(\zeta_v)_{v\in\mathcal{V}_n}$ be mutually independent random elements, and let $\mathcal{R}_{i,n} \subseteq \mathcal{V}_n$ be deterministic finite sets. Suppose $O_i = h_{i,n}((\zeta_v)_{v\in\mathcal{R}_{i,n}})$ for measurable maps $h_{i,n}$. Join i, k whenever $\mathcal{R}_{i,n}\cap\mathcal{R}_{k,n} \neq \varnothing$. The resulting graph is a dependency graph for the complete observation array.*

Proof. Take disjoint site sets A, C with no edge between them. Their primitive-index unions $\mathcal{R}_A = \bigcup_{i\in A}\mathcal{R}_{i,n}$ and $\mathcal{R}_C = \bigcup_{k\in C}\mathcal{R}_{k,n}$ are disjoint. An index in both unions would belong to some $\mathcal{R}_{i,n}\cap\mathcal{R}_{k,n}$ and create an edge. Independence of the primitive family implies independence of $\sigma(\zeta_v : v \in \mathcal{R}_A)$ and $\sigma(\zeta_v : v \in \mathcal{R}_C)$. Every observation in A is measurable in the first field and every observation in C in the second. Their observation σ-fields are subfields of these independent fields and are therefore independent. This proves the required setwise property. The same argument covers any deterministic measurable score of each observation. ∎

For (2.3), a sufficient primitive family consists of mutually independent node marks $(U_i, X_i, \varepsilon_i)$, mutually independent dyad-mark pairs $(\eta^0_{ik}, \eta^1_{ik})$, and independent individual assignment uniforms, with the node and dyad families independent of one another. Dependence within a node mark and within a dyad pair is allowed. Conditioning on the entire profile removes the common first-stage assignment variation. Each $\mathcal{R}_i$ contains the union of the primitive indices used by B_i, Z_i, G_i, Y_i over all relevant graph and assignment configurations. This statistical construction is separate from BNS and the common conditional laws.

If an observation uses node marks within radius r in a deterministic eligibility graph and dyad marks with both endpoints in that ball, disjoint radius-r balls have disjoint inputs. Connecting all sites at distance in the eligibility graph at most $2r$ then gives a valid dependence graph. A dyad with an endpoint outside the ball, a partner-choice denominator, or a baseline statistic using the entire network must also be included among the inputs. Their inclusion can enlarge the dependence graph. Growing village size does not bound this enlargement.

D.1. **Variance and neighborhood growth.**

Lemma D.2 (Variance and operator bounds)**.** *Let $\mathbb{E}X_i = 0$, $|X_i| \le M_X$, and suppose X_i obey a dependency graph with maximum closed-neighborhood size Δ_n. Then*

$$\operatorname{Var}\left(\frac{1}{n}\sum_{1\le i\le n} X_i\right) \le \frac{M_X^2\Delta_n}{n}\,, \qquad \|\Gamma_n\|_{\mathrm{op}} \le M_X^2\Delta_n\,, \quad \Gamma_n = (\mathbb{E}X_iX_k)_{i,k}\,.$$

Proof. For $k \notin N[i]$, independence and centering give $\mathbb{E}X_iX_k = 0$. Every other entry is at most M_X^2 in absolute value by the Cauchy–Schwarz inequality. There are at most Δ_n such entries per row, hence at most $n\Delta_n$ in the sum for the variance. Dividing by n^2 proves the first inequality. For $v \in \mathbb{R}^n$, symmetry and $2|v_iv_k| \le v_i^2 + v_k^2$ give

$$\begin{aligned}|v^\top\Gamma_n v| &\le \frac{1}{2}\sum_{1\le i,k\le n}|\Gamma_{ik}|(v_i^2+v_k^2)\\ &= \sum_{1\le i\le n} v_i^2 \sum_{1\le k\le n}|\Gamma_{ik}| \le M_X^2\Delta_n\|v\|^2\,.\end{aligned}$$

Taking the supremum over unit vectors proves the operator bound. ∎

To state all growth conditions in one place, retain $J < \infty$, $n_j \to \infty$ and $n_j/n \to w_j > 0$. The main inference result imposes, for every supported profile,

$$\Delta_n\varpi_{\psi,n}^2 \to 0, \quad \sqrt{n}s_n \to 0, \quad \sqrt{n}(\varpi_e\varpi_\nu + \varpi_e\varpi_m + \varpi_r\varpi_m) \to 0\,,$$

$$\Delta_n^4/n \to 0, \quad \|V_n - V\|_{\mathrm{op}} \to 0\,,$$
$$\Delta_n L_n^2/n \to 0, \quad L_n(\varpi_{\psi,n} + s_n + n^{-1/2}) \to 0\,, \tag{D.1}$$
$$\left\| n^{-1} \sum_{1\le i,k\le n} (w_{ik,n} - 1)\mathbb{E}\phi_i\phi_k^\top \right\|_{\mathrm{op}} \to 0\,.$$

The score envelopes, positive cell probabilities, bounded density ratios, measurability and mean-square replacement bounds are those in Assumptions 4.3 and 4.4. For scalar inference, require $c^\top V c > 0$; for a joint test, require $R_0 V R_0^\top \succ 0$. These conditions are sufficient jointly. Exact separation in Assumption 4.2 makes the mixing coefficient zero beyond the dependence neighborhood; no additional mixing rate is used in that case. Appendix K gives the nonzero-mixing version and its additional rates.

In particular, large villages can contain an increasing number of independent pairs or local regions, while dependence neighborhoods stay bounded or grow slowly. A common stochastic village effect instead makes the observation dependence neighborhood of order $n_j \asymp n$. It then violates (D.1), even though the village assignment itself has been conditioned on. Including a village indicator among controls does not remove a random shock's contribution to the conditional observation law.

D.2. **A fixed causal target.** A triangular array can have a different causal mean at every n. Identification establishes the meaning of each mean; convergence to a fixed target also requires the intervention laws to stabilize. The following condition concerns those laws, independently of the nuisance estimates.

Let $F_B^\star$, $q_z^\star$ and $m_z^\star$ be, respectively, a baseline law, exposure kernels and response functions for a fixed target causal law on common standard Borel spaces, with $|m_z^\star| \le M$. The target vector is

$$\theta_{zz'} = \int\int m_z^\star(g,x) q_{z'}^\star(\mathrm{d}g \mid x) F_B^\star(\mathrm{d}x)\,.$$

For this law and each finite-n law, maintain Assumptions 2.1, 2.2, 3.1, and 3.2; add Assumption 3.3 for a natural interpretation. The finite-n observed laws are those of Assumption 4.1. Define, over the finite set of required cells,

$$a_{m,n} = \max_z \sup_{g,x} |m_{z,n}(g,x) - m_z^\star(g,x)|\,,$$
$$a_{q,n} = \max_z \sup_x \mathrm{TV}(q_{z,n}(\cdot \mid x), q_z^\star(\cdot \mid x))\,,$$
$$a_{B,n} = \mathrm{TV}(F_{B,n}, F_B^\star), \qquad \delta_{T,n} = a_{m,n} + 2M a_{q,n} + 2M a_{B,n}\,,$$

where $\mathrm{TV}(P,Q) = \sup_A |P(A) - Q(A)|$ and the versions are defined on the union of the relevant supports.

Assumption D.1 (Approximation to a fixed causal law)**.** The approximation error satisfies

$$\sqrt{n}\,\delta_{T,n} \longrightarrow 0 \qquad \text{as } n \to \infty\,, \tag{D.2}$$

uniformly over the finite set of supported saturation profiles.

This assumption requires changes in the causal laws to be negligible at the scale of the standard error. It controls the deterministic drift of the target as villages grow, independently of how well the nuisance functions are estimated. If the baseline distribution, exposure kernels and response functions are invariant in n, the approximation error is zero. When those laws vary, mere convergence to a limit is insufficient: their combined discrepancy must be $o(n^{-1/2})$ under the stated sufficient bound. The condition is not needed for inference centered at the finite-n causal

vector θ_n. Nor is it necessary for every fixed contrast, since changes in different cells can cancel. The lemma below makes the recentering bound explicit; the subsequent examples distinguish invariance, slow drift and cancellation.

Lemma D.3. *Under the bounded-response and common-space conditions preceding Assumption D.1,*

$$\|\theta_n - \theta\| \le \sqrt{d_\theta}\, \delta_{T,n} .$$

If only $\delta_{T,n} \to 0$ is imposed, consistency for θ_n implies consistency for θ. Under the stronger rate in that assumption, an asymptotic linear expansion centered at θ_n also holds centered at θ.

Proof. Fix (z, z'). Adding and subtracting two integrals yields

$$\begin{aligned}\theta_{zz',n} - \theta_{zz'} = &\int\int (m_{z,n} - m^\star_z)\, \mathrm{d}q_{z',n}\, \mathrm{d}F_{B,n} \\ &+ \int \left\{ \int m^\star_z\, \mathrm{d}(q_{z',n} - q^\star_{z'}) \right\} \mathrm{d}F_{B,n} \\ &+ \int \left\{ \int m^\star_z\, \mathrm{d}q^\star_{z'} \right\} \mathrm{d}(F_{B,n} - F^\star_B) =: A_{m,n} + A_{q,n} + A_{B,n} .\end{aligned}$$

The first term is at most $a_{m,n}$ in absolute value because both integrating measures are probability measures. For a signed difference of probability measures, $|\int f\ \mathrm{d}(P - Q)| \le 2\|f\|_\infty \mathrm{TV}(P, Q)$. Apply the positive and negative parts of that signed measure, whose masses both equal its total-variation distance. Hence

$$\begin{aligned}|A_{q,n}| &\le \int 2M\, \mathrm{TV}(q_{z',n}(\cdot \mid x), q^\star_{z'}(\cdot \mid x)) F_{B,n}(\mathrm{d}x) \le 2M a_{q,n} , \\ |A_{B,n}| &\le 2 \left\| \int m^\star_z\, \mathrm{d}q^\star_{z'} \right\|_\infty \mathrm{TV}(F_{B,n}, F^\star_B) \le 2M a_{B,n} .\end{aligned}$$

Thus every coordinate difference is bounded by $\delta_{T,n}$. Summing its square over the d_θ coordinates gives the asserted Euclidean bound. The triangle inequality proves the consistency assertion. Finally,

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \sqrt{n}(\widehat{\theta}_n - \theta_n) + \sqrt{n}(\theta_n - \theta), \qquad \|\sqrt{n}(\theta_n - \theta)\| \le \sqrt{d_\theta n}\, \delta_{T,n} = \mathrm{o}(1) .$$

This is the required centering step, with the approximation error bounded independently of the estimator. ∎

Remark D.1 (Primitive examples and approximate identification)**.** The condition holds with zero approximation error when expanding villages add independent copies of fixed finite network components, all individuals have the same baseline and conditional response laws, and the exposure law within each saturation is unchanged. For example, identical disjoint eligible pairs with identically distributed endpoint primitives and Bernoulli offers have this property when the formation and outcome maps are fixed and symmetric in endpoints. Villages can contain growing numbers of such pairs while J remains fixed. Positive conditional cell probabilities and a positive contrast variance remain requirements for estimation and studentization.

BNS alone does not imply (D.2). For example, take constant baseline and exposure variables, independent centered shocks $u_i \in \{-1/2, 1/2\}$ independent of assignment, and $Y^s_{i,n}(a, g) = a(\tau + n^{-1/4}) + u_i$ for fixed τ. The response ignores formation, so BNS, exposure sufficiency and the identification conditions hold. The direct effect is $\tau + n^{-1/4}$ and converges to τ, but its recentering error is $\sqrt{n}\, n^{-1/4} = n^{1/4}$. Thus a valid normal approximation around the finite-n effect need not remain valid around its limit. Exact invariance gives a sufficient zero-error case; the rate in Assumption D.1 is a sufficient bound, not a necessary condition if errors cancel in a particular contrast.

More generally, if the two exposure laws can be coupled conditional on x so that their draws disagree with probability at most $\rho_{q,n}$, then $a_{q,n} \le \rho_{q,n}$. Indeed for every measurable set A, the two membership indicators can differ only when the draws differ. An analogous coupling of baseline variables gives $a_{B,n} \le \rho_{B,n}$. Uniform response error $a_{m,n}$ and $\sqrt{n}(a_{m,n} + 2M\rho_{q,n} + 2M\rho_{B,n}) \to 0$ therefore suffice. These are sufficient restrictions on the expanding population, not consequences of a central limit theorem.

For an approximately identified array, retain the separate bound $b_{\mathrm{ID},n} = \|\theta_n - \theta_n^{\mathrm{causal}}\|$. If the finite-$n$ causal means themselves satisfy the preceding approximation bound to their fixed causal target, the triangle inequality gives $\|\theta_n - \theta\| \le b_{\mathrm{ID},n} + \sqrt{d_\theta}\delta_{T,n}$. It is therefore sufficient that both errors be $\mathrm{o}(n^{-1/2})$. The analogous calculation in Leung (2025b, Section 4.2) involves k_n growing clusters and an interference-approximation error of order $r_n^{-\gamma}$. There the additional centering restriction is $\sqrt{k_n} r_n^{-\gamma} \to 0$, alongside the variance conditions. Here the oracle standard error is of order $n^{-1/2}$ under Assumption 5.1, so the corresponding requirement is $\sqrt{n}\|\theta_n - \theta\| \to 0$. The analogy concerns the bias relative to the standard error; Leung's growing-cluster theorem is not an inference theorem for fixed J. $\diamondsuit$

Lemma D.4. *Suppose $\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow{d} \mathcal{N}(0, V)$, and let C_0 be a fixed matrix. If*

$$\theta_n - \theta = A\gamma_n + \rho_n, \qquad C_0 A = 0, \qquad \sqrt{n}\|C_0\rho_n\| \longrightarrow 0\,,$$

where A is fixed and all three sequences are deterministic, then $\sqrt{n}(C_0\widehat{\theta}_n - C_0\theta) \xrightarrow{d} \mathcal{N}(0, C_0 V C_0^\top)$.

Proof. Multiplying the centering decomposition by C_0 gives the exact equality

$$\begin{aligned}\sqrt{n}(C_0\widehat{\theta}_n - C_0\theta) &= C_0\sqrt{n}(\widehat{\theta}_n - \theta_n) + \sqrt{n}C_0 A\gamma_n + \sqrt{n}C_0\rho_n \\ &= C_0\sqrt{n}(\widehat{\theta}_n - \theta_n) + \sqrt{n}C_0\rho_n\,.\end{aligned}$$

To verify the cancellation, write $C_0 \in \mathbb{R}^{q\times d_\theta}$ and $A \in \mathbb{R}^{d_\theta \times p}$. For each coordinate $\ell \in [q]$,

$$(C_0 A\gamma_n)_\ell = \sum_{1\le j\le p} (C_0 A)_{\ell j}(\gamma_n)_j = 0\,.$$

Thus even a nonvanishing or diverging sequence γ_n makes no contribution to the contrast. Let $Z \sim \mathcal{N}(0, V)$. The map $x \mapsto C_0 x$ is continuous because $\|C_0 x - C_0 y\| \le \|C_0\|_{\mathrm{op}}\|x - y\|$. The continuous-mapping theorem gives $C_0\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow{d} C_0 Z$. For $u \in \mathbb{R}^q$, its limiting characteristic function is

$$\mathbb{E}e^{\mathrm{i}u^\top C_0 Z} = \exp(-u^\top C_0 V C_0^\top u/2)\,,$$

which identifies $\mathcal{N}(0, C_0 V C_0^\top)$, including a singular covariance. For every $\varepsilon > 0$, the deterministic rate on ρ_n implies $\mathbb{P}(\|\sqrt{n}C_0\rho_n\| > \varepsilon) = 0$ for all sufficiently large n. Adding this term preserves the distributional limit by Slutsky's lemma. No rate restriction on γ_n is needed. ■

For the decomposition matrix C in (4.7), $C\mathbf{1}_5 = 0$. If exposure and baseline laws are fixed and every response regression is shifted by the same c_n, then every cell mean shifts by c_n, so the four reported contrasts remain unchanged. Taking $c_n = n^{-1/4}$ violates Assumption D.1 for the cell vector but gives exact centering for $C\theta_n$. This example requires enough slack in the uniform outcome bound to accommodate the shift; a bounded response with an interior range provides it. It also preserves BNS because adding a deterministic intercept does not change shock independence. This is one reason to distinguish a sufficient condition for all means from a sufficient condition for the reported contrasts $C\theta_n$.

The relevant comparison across papers is the magnitude of the omitted term relative to the standard error. In Leung (2025b, Section 4.2), an interference tail of order $r_n^{-\gamma}$ must be negligible relative to $k_n^{-1/2}$, yielding $\sqrt{k_n} r_n^{-\gamma} \to 0$. Lemmas **??** and D.4 provide other ways to make the corresponding term negligible here. By contrast, Emmenegger et al. (2025, Theorem 2.2 and Assumption A4) retain the finite-population-size expected effect as the center. Part 4.1 is a nuisance-product restriction, and part 4.3 controls degree times squared nuisance error. Neither implies stabilization of that effect. The two types of approximation should therefore be checked separately.

Lemma D.5. *Suppose* $\sqrt{n}\, c^\top(\widehat{\theta}_n - \theta_n) \xrightarrow{d} \mathcal{N}(0, v)$ *with* $v > 0$, *and* $\widehat{v}_n \xrightarrow{\mathbb{P}} v$. *If a deterministic* $B_{c,n} \geq 0$ *satisfies* $|c^\top(\theta_n - \theta)| \leq B_{c,n}$, *then*

$$\left[c^\top \widehat{\theta}_n - z_{1-\alpha/2}\sqrt{\widehat{v}_n/n} - B_{c,n}, \quad c^\top \widehat{\theta}_n + z_{1-\alpha/2}\sqrt{\widehat{v}_n/n} + B_{c,n}\right]$$

has limiting lower coverage at least $1-\alpha$ *for* $c^\top\theta$, *for* $0 < \alpha < 1$. *Set* $\widehat{v}_n$ *to its nonnegative part if necessary.*

Proof. By the triangle inequality, the event $|c^\top(\widehat{\theta}_n - \theta_n)| \leq z_{1-\alpha/2}\sqrt{\widehat{v}_n/n}$ implies

$$|c^\top(\widehat{\theta}_n - \theta)| \leq |c^\top(\widehat{\theta}_n - \theta_n)| + |c^\top(\theta_n - \theta)| \leq z_{1-\alpha/2}\sqrt{\widehat{v}_n/n} + B_{c,n}\,.$$

The first event has probability tending to $1-\alpha$ by Slutsky's lemma and continuity of the normal distribution at the two critical values. The probability of the displayed inequality therefore has limit inferior at least $1-\alpha$, as required. ■

The argument uses inclusion of coverage events and therefore does not require $\sqrt{n}B_{c,n} \to 0$.

A justified choice is $B_{c,n} = \|c\|_1 \delta_{T,n}$ under the coordinate bound in Lemma D.3, or a smaller bound obtained by cancellation. This follows the use of explicit bias bounds in Armstrong and Kolesár (2020); their regression-specific critical values can be sharper than that from the triangle inequality proof here. The bound must be supplied by restrictions on the target laws or by a justified sensitivity range. It is not learned merely by estimating nuisance parameters. This alternative removes the need for negligible centering bias at the cost of interval width. The interval shrinks if $B_{c,n} \to 0$, but has the usual order-$n^{-1/2}$ width only if $B_{c,n} = \mathrm{O}(n^{-1/2})$.

D.3. **Covariance kernels.** Let d_n be a specified deterministic graph metric or pseudometric, and set $B_n(i,b) = \{k : d_n(i,k) \leq b\}$, including i. For bandwidth $h_n \geq 0$, let $H_{iu} = \mathbb{1}\{u \in B_n(i,h_n)\}$ and $D_{ii} = |B_n(i,h_n)|$. Define

$$w_{ik,n} = \frac{|B_n(i,h_n) \cap B_n(k,h_n)|}{\sqrt{|B_n(i,h_n)|\,|B_n(k,h_n)|}}, \qquad W_n = D^{-1/2} H H^\top D^{-1/2}\,. \tag{D.3}$$

Lemma D.6 (Kernel and bias bounds)**.** *The matrix in* (D.3) *satisfies* $W_n \succeq 0$, $w_{ii,n} = 1$, *and* $0 \leq w_{ik,n} \leq 1$. *If* $d_n(i,k) > 2h_n$, *then* $w_{ik,n} = 0$. *For scalar centered scores bounded by* M_ϕ *with dependence neighborhoods* $N[i]$, *its oracle covariance bias is bounded by*

$$\left|\frac{1}{n}\sum_{1\leq i,k\leq n}(w_{ik,n}-1)\mathbb{E}\phi_i\phi_k\right| \leq M_\phi^2 \Delta_n \max_{i\in[n],\ k\in N[i]} |1 - w_{ik,n}|\,.$$

Proof. For $v \in \mathbb{R}^n$, $v^\top W_n v = \|H^\top D^{-1/2} v\|^2 \geq 0$. Each diagonal entry is its ball size divided by itself. The numerator of $w_{ik,n}$ is the inner product of two incidence vectors; the Cauchy–Schwarz inequality bounds it by the product of their Euclidean norms, giving $w_{ik,n} \leq 1$. All entries are nonnegative. If the two balls shared u, the triangle inequality would give $d_n(i,k) \leq d_n(i,u) + d_n(u,k) \leq 2h_n$. Thus the entry is zero beyond that distance.

For the bias, nonadjacent scores are independent and centered. There are at most $n\Delta_n$ remaining ordered pairs and $|\mathbb{E}\phi_i\phi_k| \le M_\phi^2$. Therefore

$$\left|\frac{1}{n}\sum_{1\le i,k\le n}(w_{ik,n}-1)\mathbb{E}\phi_i\phi_k\right| \le \frac{1}{n}\sum_{1\le i\le n}\sum_{k\in N[i]}|w_{ik,n}-1|\,|\mathbb{E}\phi_i\phi_k| \le M_\phi^2\Delta_n \max_{i,k\in N[i]}|1-w_{ik,n}|\,.$$

For d_θ coordinates the same entry bound applies to every cross moment. Hence the operator-norm bias is at most d_θ times the displayed bound. This completes the fixed-dimensional statement used in Assumption 5.2. ■

Example D.1 (Growing villages on a lattice)**.** Let each of the fixed J villages be a d-dimensional periodic lattice whose side lengths diverge, with $n_j/n \to w_j > 0$. Use within-village coordinate sup-distance and assume no cross-village dependence in this example. Suppose every dependence edge has distance at most a fixed integer r. Choose integer $h_n \ge r$, with every lattice side longer than $4h_n + 2r + 1$. Every ball has size $(2h_n+1)^d$, and balls at dependence distance at most r share at least $(2h_n+1-r)^d$ vertices. Therefore

$$0 \le 1 - w_{ik,n} \le 1 - \left(1 - \frac{r}{2h_n+1}\right)^d \le \frac{dr}{2h_n+1} \quad (k \in N[i])\,.$$

The last inequality follows from $(1-x)^d \ge 1 - dx$ on $[0,1]$, proved by induction on d. Here $\Delta_n \le (2r+1)^d$ and $L_n \le (4h_n+1)^d$. The kernel bias vanishes if $h_n \to \infty$, covariance concentration holds if $h_n^{2d}/n \to 0$, and feasible-score error vanishes if $h_n^d(\varpi_{\psi,n} + s_n + n^{-1/2}) \to 0$. A logarithmic bandwidth satisfies these conditions for the categorical learner. The expanding lattice supplies more local regions within each village; the number of villages does not increase. This example verifies the geometry and bandwidth conditions. Common conditional laws and a finite nondegenerate variance limit must also hold for the scores to which it is applied.

A second kernel sets $w_{ik,n} = 1$ within prespecified independent blocks and zero otherwise. It is positive semidefinite because its quadratic form is the sum of squared block sums. Its bias is zero if every dependence edge lies within a block. Its maximum block size is L_n and must meet (D.1). Using a whole growing village as a single block generally fails the concentration rate when J is fixed.

APPENDIX E. ALTERNATIVE IDENTIFICATION STRATEGY VIA DIFFERENCE-IN-DIFFERENCE

Panel outcomes permit an alternative adjustment for exposure–outcome confounding when their structural form makes the confounding component cancel after differencing. This appendix defines the resulting post-period contrast and gives the conditional trend and support restrictions that identify it. The target averages over the same uniform individual index and baseline law as Appendix A; all expectations also integrate the assignment and response primitives. The means are defined at each finite n. Cell-conditioned expectations use the joint target law of Section 3.2; the same reweighting is applied to both periods and their primitives. Assumption D.1, applied to the differenced response functions and the corresponding exposure laws, gives sufficient rates for inference on a fixed limiting contrast.

Suppose outcomes are observed before and after assignment, and the post-period controlled response has the additive structure

$$Y_i^0 = c_0(B_i) + \alpha_i + u_i^0, \qquad Y_i^1(z,g) = c_1(B_i) + \alpha_i + h(z,g,B_i) + u_i^1\,. \tag{E.1}$$

The same α_i and u_i^1 enter every post-period intervention. The individual effect α_i can depend arbitrarily on formation primitives. Assume integrability, consistency, the correctly aligned mediator kernel $q_b = \mathcal{L}(G_I(b) \mid B_I)$, and

$$\mathbb{E}[u_i^1 - u_i^0 \mid Z_i = z, G_i = g, B_i] = r(B_i) \tag{E.2}$$

under the relevant target-standardized observed law. This is a conditional trend restriction *after conditioning on the endogenous exposure*. It is stronger than a reduced-form parallel-trends assertion. Assume $q_b \ll q_a$ and $q_d \ll q_c$ with covariate overlap for the two integrals below.

Theorem E.1 (Identification)**.** *Let* $d_z(g,x) = \mathbb{E}[Y^1 - Y^0 \mid Z = z, G = g, B = x]$. *Under* (E.1)–(E.2),

$$\mathbb{E}[Y_I^1(a, G_I(b)) - Y_I^1(c, G_I(d))] = \mathbb{E}_B\left[\int d_a(g,B) q_b(\mathrm{d}g \mid B) - \int d_c(g,B) q_d(\mathrm{d}g \mid B)\right] . \tag{E.3}$$

Proof. Write $\Delta Y = Y^1 - Y^0$ and $\Delta c(x) = c_1(x) - c_0(x)$. Consistency and (E.1) imply

$$\Delta Y_i = \Delta c(B_i) + h(Z_i, G_i, B_i) + (u_i^1 - u_i^0) .$$

Consequently, on the observed support,

$$\begin{aligned} d_z(g,x) &= \mathbb{E}[\Delta Y \mid Z = z, G = g, B = x] \\ &= \Delta c(x) + h(z,g,x) + \mathbb{E}[u^1 - u^0 \mid Z = z, G = g, B = x] && \text{(E.1)} \\ &= \Delta c(x) + h(z,g,x) + r(x) . && \text{(E.2)} \end{aligned}$$

The cancellation of α_i occurs before conditioning; no conditional independence of α_i and exposure is used.

Next consider the two post-period potential outcomes for the same selected individual I. Their common additive terms cancel for every realization.

$$\begin{aligned} &Y_I^1(a, G_I(b)) - Y_I^1(c, G_I(d)) \\ &\quad = \{c_1(B_I) + \alpha_I + h(a, G_I(b), B_I) + u_I^1\} \\ &\qquad - \{c_1(B_I) + \alpha_I + h(c, G_I(d), B_I) + u_I^1\} \\ &\quad = h(a, G_I(b), B_I) - h(c, G_I(d), B_I) . && \text{(E.1)} \end{aligned}$$

Since h is a common measurable response function, mediator alignment and iterated expectation give, for each required (z, v),

$$\begin{aligned} \mathbb{E}[h(z, G_I(v), B_I)] &= \int \mathbb{E}[h(z, G_I(v), x) \mid B_I = x] F_B(\mathrm{d}x) \\ &= \int\int h(z,g,x) q_v(\mathrm{d}g \mid x) F_B(\mathrm{d}x) . && \text{(mediator alignment)} \end{aligned}$$

Integrability permits these iterated integrals. Applying the equality to (a,b) and (c,d) identifies the expectation of the preceding difference in terms of h.

Finally substitute the observed difference regression. For F_B-almost every x,

$$\begin{aligned} &\int d_a(g,x) q_b(\mathrm{d}g \mid x) - \int d_c(g,x) q_d(\mathrm{d}g \mid x) \\ &\quad = \int h(a,g,x) q_b(\mathrm{d}g \mid x) - \int h(c,g,x) q_d(\mathrm{d}g \mid x) \end{aligned}$$

$$+\{\Delta c(x)+r(x)\}\left\{\int q_b(\mathrm{d}g\mid x)-\int q_d(\mathrm{d}g\mid x)\right\}$$
$$=\int h(a,g,x)q_b(\mathrm{d}g\mid x)-\int h(c,g,x)q_d(\mathrm{d}g\mid x)\,.$$

Both kernels have mass one. Covariate overlap identifies d_a and d_c on the target baseline support; $q_b \ll q_a$ and $q_d \ll q_c$ extend their observed versions to the exposure supports of the two integrals. Integrating the last display over F_B proves (E.3). The proof uses the structural cancellation and conditional trend restriction, rather than a cross-world independence assumption for the additive individual effect. ■

Lemma E.1 (Score for the identified difference-in-differences contrast)**.** *Maintain the hypotheses of Theorem E.1. For $\Delta Y = Y^1 - Y^0$, define*

$$\nu^{\Delta}_{ab}(x)=\int d_a(g,x)q_b(\mathrm{d}g\mid x)\,,$$
$$\psi^{\Delta}_{ab}=\nu^{\Delta}_{ab}(B)+\frac{\mathbb{1}\{Z=a\}}{e_a(B)}r_{ab}(G,B)\{\Delta Y-d_a(G,B)\}+\frac{\mathbb{1}\{Z=b\}}{e_b(B)}\{d_a(G,B)-\nu^{\Delta}_{ab}(B)\}\,.$$

Under the overlap and uniform bounds of Assumption 4.3 for ΔY and its nuisance parameters, $\Psi^{\Delta}=\psi^{\Delta}_{ab}-\psi^{\Delta}_{cd}$ has expectation equal to the target in (E.3). Its nuisance derivative is zero at the truth. For the dominated submodels of Lemma B.2, applied to $(\Delta Y, G, Z, B)$, the centered Ψ^{Δ} represents the pathwise derivative of the identified contrast.

Proof. Conditional on (Z,G,B), $\mathbb{E}[\Delta Y - d_a(G,B)\mid Z=a,G,B]=0$, so the first weighted residual has mean zero. For the second residual, conditioning on $B=x$ and summing over the assignment cell gives

$$\mathbb{E}\left[\frac{\mathbb{1}\{Z=b\}}{e_b(B)}\{d_a(G,B)-\nu^{\Delta}_{ab}(B)\}\mid B=x\right]=\int d_a(g,x)q_b(\mathrm{d}g\mid x)-\nu^{\Delta}_{ab}(x)=0\,.$$

It follows that $\mathbb{E}\psi^{\Delta}_{ab}=\int \nu^{\Delta}_{ab}\,\mathrm{d}F_B$. Subtracting the corresponding identity for (c,d) and using Theorem E.1 proves the mean assertion. For a candidate collection, put $\delta d_a=\widetilde{d}_a-d_a$ and define the other differences as in Proposition 4.1. Its exact bias for the first cell is

$$\mathbb{E}_B[(1-\kappa_b)\delta\nu^{\Delta}_{ab}+(\kappa_b-\kappa_a)\mathbb{E}_{q_b}(\delta d_a\mid B)-\kappa_a\mathbb{E}_{q_a}(\delta r_{ab}\delta d_a\mid B)]\,.$$

Every summand is a product of two errors. Along a differentiable admissible path through the truth, differentiating such a product leaves at least one zero error at the truth. Hence its derivative is zero; subtracting the same calculation for (c,d) proves nuisance orthogonality of Ψ^{Δ}. With known assignment probabilities, the exact bias is $-\mathbb{E}_B\int \delta r_{ab}\delta d_a\,\mathrm{d}q_a+\mathbb{E}_B\int \delta r_{cd}\delta d_c\,\mathrm{d}q_c$, whose absolute value is at most $\|\delta r_{ab}\|_a\|\delta d_a\|_a+\|\delta r_{cd}\|_c\|\delta d_c\|_c$ by the Cauchy–Schwarz inequality. Finally, the bounded-score domination in Lemma B.2 justifies differentiating both identified integrals. For any total submodel score S,

$$\dot\theta^{\Delta}_{ab}-\dot\theta^{\Delta}_{cd}=\mathbb{E}[(\psi^{\Delta}_{ab}-\theta^{\Delta}_{ab})S]-\mathbb{E}[(\psi^{\Delta}_{cd}-\theta^{\Delta}_{cd})S]=\mathbb{E}[(\Psi^{\Delta}-\mathbb{E}\Psi^{\Delta})S]\,.$$

This differentiates the observed-law functional after identification; it imposes no additional independence on the additive individual effect. The factor-by-factor derivative verification is the one in Appendix B, following Chen and Ritzwoller (2023). ■

This route is related to panel adjustment in Abadie (2005). Huber and Oberhänsli (2026) also study mediation with difference-in-differences and double machine learning; the network-exposure conditional trend and structural cancellation above are the conditions actually used. To estimate (E.3), replace Y by the observed difference in the

main mediation score and difference the two cell scores. This substitution produces exactly the same identified statistical functional, so Appendix B's remainder identity applies with $m = d$. For inference require bounded differences, the same common conditional laws and profile support conditions for these new nuisances, the dependency graph for the observations for both periods, the corresponding learning products and stability or buffers, and a nondegenerate long-run covariance for the differenced score. Under these explicit conditions the proofs of Theorems 4.2 and 5.1 apply to the substituted observation vector. This argument identifies zero-sum contrasts; it does not identify post-period levels by substituting a difference into a level formula.

APPENDIX F. EXPOSURES BEYOND IMMEDIATE NEIGHBORS

Higher-order exposure records peers reached through paths in the network. Changing one relationship can alter several such paths, so the exposure must be evaluated on the entire counterfactual graph. I first define the intervention and identify its mean. I then state how the number of primitive inputs changes the estimation and inference rates.

For an undirected graph ℓ on $[n]$, define its distance by

$$d_\ell(i,k) := \min\big(\{h \geq 0 : \exists i = i_0, i_1, \ldots, i_h = k \text{ with } \ell_{i_{v-1} i_v} = 1 \ (v \in [h])\}\big) ,$$

with distance infinity if there is no path and $d_\ell(i,i) = 0$. The order-h neighbors and radius-r ball are

$$\mathcal{S}_\ell(i,h) = \{k : d_\ell(i,k) = h\}, \qquad \mathcal{B}_\ell(i,r) = \{i\} \cup \bigcup_{h=1}^{r} \mathcal{S}_\ell(i,h) .$$

The shells are disjoint. A vertex reachable by several paths is counted once in its shortest-distance shell. For example, a count exposure is

$$G_i^{(r)} = \left(\sum_{k \in \mathcal{S}_A(i,h)} P_k, \sum_{k \in \mathcal{S}_A(i,h)} (1 - P_k) \right)_{h=1}^{r} .$$

More generally let $G_i^{(r)} = e_{i,n}^{(r)}(A, T, P_{-i}, X^n)$ be a jointly measurable map into a standard Borel space. It can preserve path weights, identities or peer covariates. The count example discards these distinctions and requires the associated response invariance.

Generate $A_i^s(b)$ using Definition 2.1 and set $G_i^{s,(r)}(b) = e_{i,n}^{(r)}(A_i^s(b), T^s, H_{i,P}^s, X^n)$. Let $Y_i^{s,(r)}(a,g) = f_{i,n}^{(r)}(s,a,g,\varepsilon_i)$ be the corresponding controlled response. Define the natural and interventional means as in (2.9)–(2.10), using the higher-order exposure. For a uniform target individual write their conditional response mean as $h_a^{(r)}(g,x)$, their observed regression as $m_a^{(r)}(g,x)$ and their aligned mediator law as $q_b^{(r)}(\mathrm{d}g \mid x)$.

Proposition F.1 (Higher-order interventions and identification)**.** *Under Assumptions 2.1, 2.2, 3.1, and 3.2,*

$$\theta_{s,n}^{\mathrm{int},(r)}(a,b) = \int\int m_a^{(r)}(g,x) q_b^{(r)}(\mathrm{d}g \mid x) F_B(\mathrm{d}x) .$$

If Assumption 3.3 also holds, the natural mean equals this same integral. The means with matching assignments equal the corresponding randomized regime means, and the components defined by (2.12) satisfy (2.13).

Proof. For each finite n, the graph space is finite and the shell map is measurable. More general exposure maps are measurable by hypothesis. Composition with the measurable formation and assignment maps therefore defines $G_i^{s,(r)}(b)$ as a random element in the specified standard Borel space, with a regular conditional distribution given B_i. Mediator alignment identifies the target conditional law with $q_b^{(r)}$.

Let I be the uniform target index and let $K(\mathrm{d}i, \mathrm{d}e \mid x)$ be the conditional law of (I, ε_I) given $B_I = x$. The exposure draw from the target-population conditional distribution is independent of these response primitives given B_I, so its joint conditional law is $K(\mathrm{d}i, \mathrm{d}e \mid x)q_b^{(r)}(\mathrm{d}g \mid x)$. Absolute integrability permits the iterated integrals

$$\begin{aligned}\theta_{s,n}^{\mathrm{int},(r)}(a,b) &= \int\int\int f_{i,n}^{(r)}(s,a,g,e)K(\mathrm{d}i,\mathrm{d}e \mid x)q_b^{(r)}(\mathrm{d}g \mid x)F_B(\mathrm{d}x) \\ &= \int\int h_a^{(r)}(g,x)q_b^{(r)}(\mathrm{d}g \mid x)F_B(\mathrm{d}x) \\ &= \int\int m_a^{(r)}(g,x)q_b^{(r)}(\mathrm{d}g \mid x)F_B(\mathrm{d}x)\ .\end{aligned}$$

The final equality uses controlled-response exchangeability. Since $q_b^{(r)}(\cdot \mid x) \ll q_a^{(r)}(\cdot \mid x)$, changing the observed regression on a $q_a^{(r)}$-null set cannot change this integral. Thus its value is identified on the required support.

For the natural mean, the simultaneous conditional-response version in Assumption 3.3 gives

$$\begin{aligned}\theta_{s,n}^{\mathrm{nat},(r)}(a,b) &= \mathbb{E}\left[\mathbb{E}(Y^{s,(r)}(a,G^{s,(r)}(b)) \mid B, G^{s,(r)}(b))\right] \\ &= \mathbb{E}[h_a^{(r)}(G^{s,(r)}(b),B)] \\ &= \int\int h_a^{(r)}(g,x)q_b^{(r)}(\mathrm{d}g \mid x)F_B(\mathrm{d}x)\ .\end{aligned}$$

The last equality uses mediator alignment again. Replacing $h_a^{(r)}$ by $m_a^{(r)}$ as above proves the natural identity.

When response and exposure assignments agree, the natural exposure is generated under the same focal assignment as the response. Consistency therefore gives its randomized regime mean. For the interventional mean with matching assignments, first integrate the observed conditional response over its own exposure law and then over F_B.

$$\begin{aligned}\theta_{s,n}^{\mathrm{int},(r)}(a,a) &= \int\int \mathbb{E}[Y \mid T_i = s, P_i = a, G^{(r)} = g, B = x]q_a^{(r)}(\mathrm{d}g \mid x)F_B(\mathrm{d}x) \\ &= \int \mathbb{E}[Y \mid T_i = s, P_i = a, B = x]F_B(\mathrm{d}x) = R_{s,n}(a)\ .\end{aligned}$$

The last equality follows from Lemma A.2. Substitution of these regime-mean identities in the component definitions gives the specified total. Both $r = 1$ and larger r use the same argument; only the exposure map and its identifying restrictions change. ■

Every graph shell is recomputed after the intervention. Retaining a realized peer-to-peer link while changing only the focal row would generally define a different intervention when formation outside that row can respond to the focal assignment.

For inference, condition on the full profile. Suppose all links lie in the deterministic eligibility graph K_n. Let $\mathcal{R}_i^{(r)}$ be the deterministic set of primitive inputs used by the complete observation $O_i^{(r)} = (B_i, Z_i, G_i^{(r)}, Y_i)$ over its possible configurations. Form a graph by intersecting these sets and let $\Delta_n^{(r)}$ be its maximum closed-neighborhood size. Define the score rates $\varpi_{\psi,n}^{(r)}, s_n^{(r)}, \varpi_m^{(r)}, \varpi_r^{(r)}, \varpi_\nu^{(r)}, \varpi_e^{(r)}$ and kernel row support $L_n^{(r)}$ for this enlarged observation and its complete learning procedure.

Proposition F.2 (Rates under higher-order dependence)**.** *Assume the independent-input construction of Lemma D.1 for $\mathcal{R}_i^{(r)}$. Suppose Assumptions 4.1, 4.3, and 4.4 hold for $O_i^{(r)}$, with a fixed number of reported cells and constants uniform in n and the selected radius $r = r_n$. Put $b_n^{(r)} = \varpi_e^{(r)}\varpi_\nu^{(r)} + \varpi_e^{(r)}\varpi_m^{(r)} + \varpi_r^{(r)}\varpi_m^{(r)}$. If*

$$\Delta_n^{(r)}(\varpi_{\psi,n}^{(r)})^2 \to 0, \qquad \sqrt{n}s_n^{(r)} \to 0, \qquad \sqrt{n}b_n^{(r)} \to 0\ ,$$

$$(\Delta_n^{(r)})^4/n \to 0, \qquad \|V_n^{(r)} - V_*\|_{\mathrm{op}} \to 0\,,$$
$$\Delta_n^{(r)}(L_n^{(r)})^2/n \to 0, \qquad L_n^{(r)}(\varpi_{\psi,n}^{(r)} + s_n^{(r)} + n^{-1/2}) \to 0\,, \tag{F.1}$$

and the covariance-weight and bias conditions of Assumption 5.2 hold, then the estimator is consistent, $\sqrt{n}(\widehat{\theta}_n^{(r)} - \theta_n^{(r)})$ converges to $\mathcal{N}(0, V_*)$, and its feasible covariance converges to V_*. Studentization holds for fixed contrasts with positive limiting variance.

Proof. A path of $h \le r$ realized edges is a path of h eligible edges because $A_{ik} \le K_{ik,n}$. Thus $\mathcal{B}_A(i,r) \subseteq \mathcal{B}_{K_n}(i,r)$. This establishes where exposed vertices can lie. The assumed $\mathcal{R}_i^{(r)}$ additionally contains the inputs to their edge-acceptance rules, baseline controls and outcomes. If two site collections have no intersection edge, their primitive unions are disjoint. Lemma D.1 then gives setwise independence of their complete higher-order observations.

For one score coordinate, use the enlarged closed neighborhoods $N^{(r)}[i]$. On a common independent copied array, let $\widehat{\eta}^{(r),(-i)}$ replace $N^{(r)}[i]$ and let $\widehat{\eta}^{(r),(-i,-k)}$ replace $N^{(r)}[i] \cup N^{(r)}[k]$. Define

$$d_i^{(r)}(o;\eta) = \psi^{(r)}(o;\eta) - \psi^{(r)}(o;\eta_0^{(r)})\,,$$
$$F_i^{(r)}(\eta) = \int d_i^{(r)}(o;\eta)\mathcal{L}(O_i^{(r)})(\mathrm{d}o), \qquad f_i^{(r)}(o;\eta) = d_i^{(r)}(o;\eta) - F_i^{(r)}(\eta)\,,$$
$$U_i^{(r)} = f_i^{(r)}(O_i^{(r)};\widehat{\eta}^{(r),(-i)})\,.$$

The nuisance estimate $\widehat{\eta}^{(r),(-i)}$ is measurable in the original observations outside $N^{(r)}[i]$, the copied array and the independent fitting variable. Conditional on these variables, $O_i^{(r)}$ has its marginal law by the setwise independence just proved. Hence

$$\mathbb{E}U_i^{(r)} = 0, \qquad \mathbb{E}[(U_i^{(r)})^2] \le (\varpi_{\psi,n}^{(r)})^2\,.$$

The stability bound applies both at the training observation and at an independent observation. The integral term is bounded by Jensen's inequality, giving

$$\|f_i^{(r)}(O_i^{(r)};\widehat{\eta}^{(r)}) - U_i^{(r)}\|_2$$
$$\le \|\psi^{(r)}(O_i^{(r)};\widehat{\eta}^{(r)}) - \psi^{(r)}(O_i^{(r)};\widehat{\eta}^{(r),(-i)})\|_2 + \|F_i^{(r)}(\widehat{\eta}^{(r)}) - F_i^{(r)}(\widehat{\eta}^{(r),(-i)})\|_2 \le 2s_n^{(r)}\,.$$

For a far pair $k \notin N^{(r)}[i]$, put $Q_i^{(r)} = f_i^{(r)}(O_i^{(r)};\widehat{\eta}^{(r),(-i,-k)})$ and define $Q_k^{(r)}$ with the same nuisance estimate. Conditional on the original observations outside the union and the copied array and fitting variable, $O_i^{(r)}$ and $O_k^{(r)}$ are independent with their marginal laws. Both Q terms are conditionally centered. Moreover $Q_i^{(r)}$ is measurable in the variables with respect to which $U_k^{(r)}$ is conditionally centered, and conversely. Therefore

$$\mathbb{E}[Q_i^{(r)}Q_k^{(r)}] = \mathbb{E}[Q_i^{(r)}U_k^{(r)}] = \mathbb{E}[U_i^{(r)}Q_k^{(r)}] = 0\,.$$

Inserting the original nuisance estimate between the single and double estimates and applying the preceding two-term stability bound twice gives $\|U_i^{(r)} - Q_i^{(r)}\|_2 \le 4s_n^{(r)}$. The Cauchy–Schwarz inequality now yields

$$|\mathbb{E}[U_i^{(r)}U_k^{(r)}]| = |\mathbb{E}[(U_i^{(r)} - Q_i^{(r)})(U_k^{(r)} - Q_k^{(r)})]| \le 16(s_n^{(r)})^2\,.$$

There are at most $n\Delta_n^{(r)}$ near ordered pairs and at most n^2 far ordered pairs. Expanding the square and applying the Cauchy–Schwarz inequality to near pairs gives

$$\mathbb{E}\left[\left(n^{-1/2}\sum_{1\le i\le n} U_i^{(r)}\right)^2\right] \le \frac{1}{n}\{n\Delta_n^{(r)}(\varpi_{\psi,n}^{(r)})^2 + 16n^2(s_n^{(r)})^2\}$$
$$= \Delta_n^{(r)}(\varpi_{\psi,n}^{(r)})^2 + 16n(s_n^{(r)})^2 .$$

Combining this bound with the sum of the replacement errors by Minkowski's inequality proves

$$\left\| n^{-1/2}\sum_{1\le i\le n} f_i^{(r)}(O_i^{(r)};\widehat{\eta}^{(r)})\right\|_2 \le 2\sqrt{n}s_n^{(r)} + \{\Delta_n^{(r)}(\varpi_{\psi,n}^{(r)})^2 + 16n(s_n^{(r)})^2\}^{1/2} \to 0 .$$

This is the neighborhood-replacement argument of Lemma B.3, with every neighborhood enlarged to cover the higher-order exposure; the treatment of neighborhood replacement observations follows Cao and Leung (2025), Chen et al. (2022).

For the mean term, write $E_z^{(r)}v = \int v \, \mathrm{d}q_z^{(r)}$. The exact score expansion is

$$\frac{1}{n}\sum_{1\le i\le n} F_i^{(r)}(\widehat{\eta}^{(r)}) = \mathbb{E}_B\big[(1-\widehat{\kappa}_{z'}^{(r)})\delta\widehat{\nu}^{(r)} + (\widehat{\kappa}_{z'}^{(r)} - \widehat{\kappa}_z^{(r)})E_{z'}^{(r)}\delta\widehat{m}^{(r)}$$
$$- \widehat{\kappa}_z^{(r)}E_z^{(r)}(\delta\widehat{r}^{(r)}\delta\widehat{m}^{(r)})\big] .$$

For any $\epsilon > 0$, the nuisance rates give a finite K_ϵ such that, with probability at least $1-\epsilon$ for all sufficiently large n, each realized nuisance norm is at most K_ϵ times its specified deterministic rate. On this event the absolute scaled mean error is bounded by

$$\frac{K_\epsilon^2\sqrt{n}}{e_*}\{\varpi_e^{(r)}\varpi_\nu^{(r)} + 2\sqrt{R}\varpi_e^{(r)}\varpi_m^{(r)} + \varpi_r^{(r)}\varpi_m^{(r)}\} \le K_\epsilon^2 e_*^{-1}\max(1, 2\sqrt{R})\sqrt{n}b_n^{(r)} \to 0 .$$

Letting $\epsilon \downarrow 0$ proves convergence in probability. Adding the centered and mean terms proves asymptotic linearity. The oracle average has variance at most $B_\phi^2\Delta_n^{(r)}/n \to 0$, so the expansion also proves consistency.

At each fixed cumulant order $k \ge 3$, the connected-tuple count becomes $k^{k-1}n(\Delta_n^{(r)})^{k-1}$. Uniform boundedness gives

$$|\operatorname{cum}_k(n^{-1/2}\sum_{1\le i\le n} c^\top\phi_i^{(r)})| \le k^{k-1}A_k(\|c\|_1 B_\phi)^k n^{1-k/2}(\Delta_n^{(r)})^{k-1} \to 0 .$$

Together with the variance limit this proves the joint normal approximation by Lemma C.1.

For covariance estimation, retain the radius superscript and fix coordinates $a, b \in [d_\theta]$. Let

$$\widetilde{V}_n^{(r),ab} = \frac{1}{n}\sum_{1\le i,k\le n} w_{ik,n}^{(r)}\phi_i^{(r),a}\phi_k^{(r),b} , \qquad \mathcal{D}_n^{(r)} = \{(i,k)\in[n]^2 : w_{ik,n}^{(r)} \ne 0\} .$$

There are at most $nL_n^{(r)}$ pairs in this set. For a fixed first pair (i,k), a second pair (j,l) has zero product covariance unless an endpoint is in $N^{(r)}[i] \cup N^{(r)}[k]$. To see this, if both endpoints are outside the union, the sets $\{i,k\}$ and $\{j,l\}$ are disjoint and have no connecting dependence edge; their full observation collections, and therefore their score products, are independent. There are at most $2\Delta_n^{(r)}$ possible endpoints in the union, two choices for its position in (j,l), and at most $L_n^{(r)}$ nonzero weights for the other endpoint. Symmetry gives the same row and column bound. Thus there are at most $4\Delta_n^{(r)}L_n^{(r)}$ potentially contributing second pairs. Since every score

coordinate is bounded by B_ϕ,

$$\left|\mathrm{Cov}(\phi_i^{(r),a}\phi_k^{(r),b}, \phi_j^{(r),a}\phi_l^{(r),b})\right| \le 2B_\phi^4 ,$$
$$\mathrm{Var}(\widetilde{V}_n^{(r),ab}) \le \frac{(nL_n^{(r)})(4\Delta_n^{(r)}L_n^{(r)})(2B_\phi^4)}{n^2} = \frac{8B_\phi^4\Delta_n^{(r)}(L_n^{(r)})^2}{n} .$$

In particular, Chebyshev's inequality and a union bound over the d_θ^2 coordinates give, for every $u > 0$,

$$\mathbb{P}\left(\max_{a,b\in[d_\theta]} |\widetilde{V}_n^{(r),ab} - \mathbb{E}\widetilde{V}_n^{(r),ab}| > u\right) \le \frac{8d_\theta^2 B_\phi^4\Delta_n^{(r)}(L_n^{(r)})^2}{nu^2} \to 0 .$$

The expectation error is exactly

$$\mathbb{E}\widetilde{V}_n^{(r)} - V_n^{(r)} = \frac{1}{n}\sum_{1\le i,k\le n}(w_{ik,n}^{(r)} - 1)\mathbb{E}[\phi_i^{(r)}\phi_k^{(r)\top}] ,$$

whose operator norm tends to zero by the assumed covariance bias condition.

For the feasible error define, coordinatewise,

$$d_i^{(r)} = \psi^{(r)}(O_i^{(r)}; \widehat{\eta}^{(r)}) - \psi^{(r)}(O_i^{(r)}; \eta_0^{(r)}) .$$

Insert $\psi^{(r)}(O_i^{(r)}; \widehat{\eta}^{(r),(-i)})$. The squared first difference has expectation at most $(s_n^{(r)})^2$ by stability. The single replacement estimate is independent of $O_i^{(r)}$, so the squared second difference has expectation at most $(\varpi_{\psi,n}^{(r)})^2$ by independent-evaluation accuracy. Using $(x+y)^2 \le 2x^2 + 2y^2$ in each coordinate and summing gives

$$\mathbb{E}\|d^{(r)}\|_{2,n}^2 \le 2d_\theta\{(s_n^{(r)})^2 + (\varpi_{\psi,n}^{(r)})^2\} .$$

For any $K > 0$, Markov's inequality consequently bounds $\mathbb{P}\{\|d^{(r)}\|_{2,n} > K(s_n^{(r)} + \varpi_{\psi,n}^{(r)})\}$ by $2d_\theta/K^2$ whenever the denominator is positive; if it is zero, the error is zero almost surely. The oracle normal approximation and asymptotic linearity also give $\|\widehat{\theta}_n^{(r)} - \theta_n^{(r)}\| = \mathrm{O}_\mathbb{P}(n^{-1/2})$. For $e_i^{(r)} = \widehat{\phi}_i^{(r)} - \phi_i^{(r)}$, the triangle inequality therefore yields

$$\|e^{(r)}\|_{2,n} \le \|d^{(r)}\|_{2,n} + \|\widehat{\theta}_n^{(r)} - \theta_n^{(r)}\| = \mathrm{O}_\mathbb{P}(\varpi_{\psi,n}^{(r)} + s_n^{(r)} + n^{-1/2}) .$$

Expand each feasible-minus-oracle product as $e_i^{(r)}\phi_k^{(r)\top} + \phi_i^{(r)}e_k^{(r)\top} + e_i^{(r)}e_k^{(r)\top}$. The deterministic weight matrix has operator norm at most $L_n^{(r)}$, by the row-and-column sum argument in Appendix C. The Cauchy–Schwarz inequality bounds its three normalized quadratic forms by, respectively, $L_n^{(r)}\|e^{(r),a}\|_{2,n}\|\phi^{(r),b}\|_{2,n}$, $L_n^{(r)}\|\phi^{(r),a}\|_{2,n}\|e^{(r),b}\|_{2,n}$, and $L_n^{(r)}\|e^{(r),a}\|_{2,n}\|e^{(r),b}\|_{2,n}$. Consequently,

$$\max_{a,b\in[d_\theta]} |\widehat{V}_n^{(r),ab} - \widetilde{V}_n^{(r),ab}| \le L_n^{(r)}\{2B_\phi\|e^{(r)}\|_{2,n} + \|e^{(r)}\|_{2,n}^2\} = \mathrm{o}_\mathbb{P}(1) .$$

Here $L_n^{(r)} \ge 1$, so the assumed rate also implies $\|e^{(r)}\|_{2,n} = \mathrm{o}_\mathbb{P}(1)$ and controls the squared-error term. The inequality $\|H\|_{\mathrm{op}} \le \|H\|_F \le d_\theta \max_{a,b}|H^{ab}|$ converts entrywise bounds to operator bounds. Combining the feasible error, centered oracle error, covariance bias and $V_n^{(r)} - V_*$ proves covariance consistency.

Finally, for a fixed contrast c with $v_c = c^\top V_* c > 0$, covariance consistency implies $\mathbb{P}\{c^\top\widehat{V}_n^{(r)}c \ge v_c/2\} \to 1$. The square root and division are continuous on this event, so Slutsky's lemma proves the studentized standard-normal limit. For a fixed full-row-rank matrix R_0 with $R_0V_*R_0^\top \succ 0$, operator convergence similarly bounds the smallest estimated eigenvalue away from zero with probability tending to one. Continuity of inversion then gives the joint Wald limit. Every constant above is independent of the chosen sequence $r = r_n$ by the uniform assumptions of the proposition. ∎

The proposition permits $r_n \to \infty$ when the displayed bounds continue to hold. For example, suppose all primitive inputs lie within radius in the eligibility graph $r + \ell_0$, including the endpoints of every dyad input. Then intersecting input sets imply a path in the eligibility graph of length at most $2(r + \ell_0)$, giving $\Delta_n^{(r)} \leq \max_i |\mathcal{B}_{K_n}(i, 2(r + \ell_0))|$. For maximum eligibility degree $d_K \geq 2$, breadth-first enumeration bounds a radius-h ball by $1 + d_K \sum_{1 \leq v \leq h} (d_K - 1)^{v-1}$. The root has at most d_K children and every later vertex at most $d_K - 1$ new children. Degree zero gives a singleton and degree one gives at most a pair. Thus the number of inputs can grow rapidly with radius even at fixed degree. On lattices the corresponding ball count grows polynomially. In either case the rate is evaluated against increasing individual sample size within the same fixed villages.

Lemma F.1 (Coloring bounds for enlarged observations)**.** *Let $\mathcal{H}_n^{(r)}$ be the deterministic dependence graph above and let $\chi_n = \chi(\mathcal{H}_n^{(r)})$ be its chromatic number, the minimum number of colors in a proper vertex coloring; then $\chi_n \leq \Delta_n^{(r)}$. If X_i is a centered measurable function of $O_i^{(r)}$ and $|X_i| \leq M_X$ for $M_X > 0$, then*

$$\operatorname{Var}\left(n^{-1} \sum_{1 \leq i \leq n} X_i \right) \leq \frac{\chi_n}{n^2} \sum_{1 \leq i \leq n} \mathbb{E} X_i^2 \leq \frac{\chi_n M_X^2}{n} ,$$

$$\mathbb{P}\left(\left| n^{-1} \sum_{1 \leq i \leq n} X_i \right| > u \right) \leq 2 \exp\left(-\frac{n u^2}{2 \chi_n M_X^2} \right), \qquad u > 0 . \tag{F.2}$$

For $K_{\mathrm{tab},n}$ such centered table averages, each with bound M_X, their maximum absolute error is at most $M_X \{2\chi_n \log(2K_{\mathrm{tab},n}/\delta)/n\}^{1/2}$ with probability at least $1 - \delta$, for $0 < \delta < 1$.

Proof. The degree bound is the greedy coloring bound in Janson (2004, equation (2.2)); the tail inequality is a consequence of that paper's Theorem 2.1. I give its ordinary-coloring proof with the constants used here. Order the vertices deterministically and assign each the first color absent from its already colored neighbors. At most $\Delta_n^{(r)} - 1$ colors are forbidden, producing a proper coloring with at most $\Delta_n^{(r)}$ colors. The optimal coloring can use fewer, so $\chi_n \leq \Delta_n^{(r)}$. For a connected component with maximum degree d, Brooks' theorem gives at most d colors unless the component is a complete graph or an odd cycle (Brooks, 1941); these exceptions require $d + 1$ colors. That refinement can improve the bound component by component, but is not required. Select a proper coloring attaining χ_n and let $C_1, \ldots, C_{\chi_n}$ be the color classes and $S_c = \sum_{i \in C_c} X_i$. Each class contains mutually independent observations. Any partition of that class has no cross-edge, so the setwise graph property implies independence of the two collections, and induction gives joint independence. In particular,

$$\mathbb{E}\left(\sum_{1 \leq c \leq \chi_n} S_c \right)^2 \leq \chi_n \sum_{1 \leq c \leq \chi_n} \mathbb{E} S_c^2 = \chi_n \sum_{1 \leq i \leq n} \mathbb{E} X_i^2 .$$

The first inequality is the pointwise Cauchy–Schwarz inequality, and the equality uses centering and independence within each class. Divide by n^2 to obtain the variance bound.

For the tail bound, Hoeffding's bounded-variable moment-generating-function inequality (Hoeffding, 1963) gives $\mathbb{E} e^{v X_i} \leq e^{v^2 M_X^2/2}$. One can obtain this constant directly by setting $g(v) = \log \mathbb{E} e^{v X_i}$. $g(0) = g'(0) = 0$ and $g''(v)$ is a variance under an exponentially tilted law supported on $[-M_X, M_X]$, hence at most M_X^2. Two integrations prove the inequality for every real v. Generalized Hölder's inequality and independence within each

color give

$$\mathbb{E}\exp\left(v\sum_{1\le c\le\chi_n} S_c\right) \le \prod_{1\le c\le\chi_n} (\mathbb{E}e^{\chi_n v S_c})^{1/\chi_n}$$
$$= \prod_{1\le c\le\chi_n}\prod_{i\in C_c} (\mathbb{E}e^{\chi_n v X_i})^{1/\chi_n} \le \exp(\chi_n v^2 n M_X^2/2)\,.$$

Markov's inequality bounds the upper tail by $\exp(-vnu + \chi_n v^2 n M_X^2/2)$. Its minimum over $v > 0$ occurs at $v = u/(\chi_n M_X^2)$ and equals the exponential in (F.2). Apply the same argument to $-X_i$ for the lower tail. A union bound over the $K_{\mathrm{tab},n}$ coordinates, with $u = M_X(2\chi_n \log(2K_{\mathrm{tab},n}/\delta)/n)^{1/2}$, proves the last assertion. Independence across different color classes is not used. If an available coloring has $c_n \ge \chi_n$ colors, replacing χ_n by c_n gives a valid, possibly weaker bound; computing an optimal coloring is unnecessary. Janson's fractional-cover bound can be sharper, but the ordinary chromatic number suffices here. ■

For empirical tables on fixed finite supports, a centered count is bounded by one and a centered outcome sum by $2M$. The lemma therefore supplies an explicit simultaneous table bound with $M_X = 2M$ when $M \ge 1$. This bound quantifies the precision lost as the exposure radius enlarges the dependence graph. If the number of categories also grows, the displayed logarithm must be combined with the changes in the probability floors and the learner's Lipschitz constants. Coloring alone does not control those additional quantities. For a degree-d_K eligibility graph, the preceding ball bound also bounds χ_n by $1 + d_K \sum_{1\le v\le 2(r+\ell_0)} (d_K - 1)^{v-1}$. On a lattice with ball bound $C(1+h)^d$, it instead gives $\chi_n \le C\{1 + 2(r+\ell_0)\}^d$.

Remark F.1 (Training separation for larger exposures)**.** Nuisance estimates from separated training sets must remove these enlarged neighborhoods and meet Assumption J.1 again. Under weak dependence, the coefficients in Appendix K must be computed for the complete higher-order observation. Larger exposure sets can worsen both neighborhood growth and mixing, so changing the map requires checking both quantities. The fixed-target conclusion also requires Assumption D.1 for the higher-order laws; otherwise the normal approximation centers at their finite-n causal means. ♢

APPENDIX G. ADDITIONAL EMPIRICAL RESULTS

This appendix documents the endline-network analysis in Section 6. All estimates use the current `Exposure_Round2` outputs. The primary specification uses treated-friend count and the 16 baseline variables listed below, labeled L5 in the replication files. Unless indicated otherwise, parentheses contain network standard errors with household score aggregation and bandwidth four. Stars denote pointwise two-sided normal tests: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$. The qualifications in Section G.6 apply throughout.

G.1. **Data and construction.**

Outcomes and analysis population. Table 3 records the analyzed samples. Knowledge and self-reported adoption come from `knowledge_adoption_extract_final.dta`; audited adoption and leaf health from `audits_extract_final.dta`; and inputs, yields and revenue from the corresponding `inputs`, `yields` and `net_revenues` extracts. The indices are retained without further standardization. The leaf-health index is multiplied by -1 so that larger values represent healthier leaves. For audits, finite tree measurements are first averaged within household and round. Revenue excludes values below $-20{,}000$ RWF/tree and observations with

fertilizer costs above 140 RWF/tree, following the analysis. Only finite outcomes in the stated rounds enter each sample.

The estimators average available household–round observations. They therefore concern the observed-record population, with weights proportional to the number of available rounds per household. They do not give every randomized household equal weight or impute an outcome for a missing round. The latter targets require additional observation and weighting assumptions. The original study retains tree–round observations for audited outcomes, so its audit coefficients and these household–round estimates have different weights even before covariate adjustment.

TABLE 3. Outcome samples

Outcome	Rounds	Observations	Households	Missing network
Knowledge	6,8,9	4,666	1,571	10
Self-reported adoption	6,8,9	4,666	1,571	10
Audited adoption	6,8,9	4,666	1,571	10
Leaf health	6,8,9	4,650	1,566	10
Inputs	6,7,8,9	6,252	1,577	11
Yields (kg/tree)	6,7,8,9	6,221	1,568	10
Net revenue (RWF/tree)	6,7,8,9	6,219	1,568	10

Notes. Observations are household–round records. The last column counts analyzed households absent from both network waves; their exposure is coded zero. All samples include 27 villages.

Baseline information. The primary set contains trimmed baseline harvest per tree and total coffee trees; indicators for rejuvenation, removing dead branches, removing suckers, weeding, lime, NPK, pesticides, mulch, compost and record keeping; a female-head indicator; household size; the head's years of schooling; and the number of coffee friends outside the randomized roster. The last variable is the nonnegative difference between reported total friends and reported roster friends. In the replication files these are, respectively,

```
Bharv_ptree_trim, Btottrees_trim, Brej, Bprun1p, Bprun4p, Bweed,
Bferti1, Bferti3, Bpesti, Bmulch, Bcomp, Bbook, Bfemalehhh,
Bnumhhmembers, Bschoolhhh, nonroster_friends.
```

The imputed baseline values are used as supplied. The primary set includes neither missingness indicators nor roster degree, neighbor-covariate summaries or geographical controls. All baseline predictors are centered and scaled within the outcome's analysis sample. The same set enters the response, classifier, outer regressions and direct regime regression. No variable selection or round indicators enter these nuisance fits. This choice is a maintained adjustment specification: BNS requires the remaining formation information to be uninformative about the fixed-exposure response shock conditional on these variables.

The expanded comparison set, labeled L2, has 50 variables. It contains 17 own-household baseline variables, adding the proportion of unproductive trees and average household schooling to the 15 own-household variables above; 21 missingness flags; baseline roster degree and reported total friends; five neighbor sums of harvest per tree, trees, the proportion of unproductive trees, household size and the head's schooling; and the five corresponding neighbor sums of missingness flags. Neighbor sums use the symmetrized baseline roster graph. L2 replaces the nonroster-friend count rather than appending variables to L5, so these sets are not a nested BNS refinement. In the L2 comparison the direct regime regression still uses the primary 16 variables. The same regime estimates therefore anchor every exposure and adjustment comparison.

Networks and measurement. The network file contains $1{,}578$ household records per wave. The same 16 randomized households are absent from both waves. The implementation assigns them zero exposures and treats them as isolates in the working graph; this coding is a measurement convention, not evidence that they had no contacts. Of households with observed endline records, the treated-friend count averages 1.771, has maximum 12, and is zero for 22.9%. For 15 households it exceeds the treated nominations listed in the release by one. The count specification uses the total. The covariance graph uses listed roster nominations, mapped to anonymized household identifiers, with loops removed and ties symmetrized. Its $2{,}448$ edges include 241 cross-village edges incident to 355 randomized households. The baseline graph has $2{,}274$ edges and is used only to construct baseline summaries in this analysis.

Wave 2 denotes the endline network interview, not outcome survey round 2. The original study's Online Appendix C and Table A8 place this network measurement at the final survey, whereas the outcomes also include earlier endline rounds. A mediation interpretation therefore requires endline exposure to represent the outcome-relevant relationships over the analyzed period and excludes outcome-driven changes in reported friendships not absorbed by the maintained model. The timing of the two measurements alone does not establish that condition.

G.2. **Implementation of the algorithms.**

Pooled scores. Let $\mathcal{R}_i$ be the set of observed rounds for household i, $r_i = |\mathcal{R}_i|$, H the number of analyzed households, and $R = \sum_i r_i$ the number of analyzed records. The implemented assignment probabilities are

$$\widehat{\lambda}_s = \frac{1}{R}\sum_{i=1}^{H} r_i \mathbb{1}\{T_i = s\}, \qquad e_{(s,a)} = \widehat{\lambda}_s s^a (1-s)^{1-a}\,. \tag{G.1}$$

These are the nominal Bernoulli probabilities for the record mixture, not an estimate of actual village quotas. Algorithm 1's outcome surface pools all records on (G, B, Z). For each outcome assignment z, its predictions at that assignment become the generated response in a pooled outer regression on (B, Z). A separate pooled regression on (B, Z) supplies every diagonal cell. The exposure-density ratio is obtained from the classifier of Z on (G, B) through Bayes' identity, using the same mixture probabilities as in (G.1).

The implementation uses `ranger` forests with 300 trees, minimum node size five, and seed `20260911`, with fixed offsets for different nuisance fits. Each tree resamples H households with replacement and includes all available rows of each sampled household. Assignment-cell indicators are always candidate split variables in the outcome, direct regime and outer regressions. Other split candidates follow the square-root rule, subject to the available predictor count. Out-of-bag predictions exclude the household from that tree's fitting sample; the fallback, if no such tree exists, is the full-forest prediction. The outer fit uses the generated responses from the preceding stage, so household exclusion at one stage does not exclude all indirect uses of that household through earlier fits.

Raw classifier probabilities are floored at 0.01 and renormalized; ratios are capped at 100. Conditional-mean predictions are clipped to $[-M_Y, M_Y]$, where M_Y is the largest absolute outcome in the analysis sample. These numerical bounds keep fitted quantities finite. They establish neither population overlap nor a uniform bound on unobserved outcomes. Table 4 reports the two directional ratios needed at 75% saturation. No primary ratio reaches the cap. The reported effective sample size is $(\sum w)^2 / \sum w^2$ for the outcome-residual weights on records in the response cell; it measures weight concentration, not the number of independent households.

TABLE 4. Directional exposure support in the primary specification

Outcome	Ratio	99th pct.	Floor (%)	Cap (%)	Weight ESS	Max. weight
Knowledge	q_0/q_1	6.627	0.000	0.000	705.268	33.309
	q_1/q_0	4.604	0.455	0.000	282.260	61.272
Self-reported adoption	q_0/q_1	6.627	0.000	0.000	705.268	33.309
	q_1/q_0	4.604	0.455	0.000	282.260	61.272
Audited adoption	q_0/q_1	6.627	0.000	0.000	705.268	33.309
	q_1/q_0	4.604	0.455	0.000	282.260	61.272
Leaf health	q_0/q_1	5.917	0.000	0.000	700.907	37.942
	q_1/q_0	5.559	0.000	0.000	276.927	77.524
Inputs	q_0/q_1	7.203	0.000	0.000	934.727	35.722
	q_1/q_0	7.484	0.000	0.000	332.487	87.582
Yields (kg/tree)	q_0/q_1	5.699	0.000	0.000	970.119	29.542
	q_1/q_0	4.123	0.000	0.000	360.925	86.530
Net revenue (RWF/tree)	q_0/q_1	5.586	0.000	0.000	946.983	34.219
	q_1/q_0	6.088	0.000	0.000	351.316	65.429

Notes. At saturation 0.75, q_a is the exposure law in assignment arm a. The ratio q_0/q_1 enters $\theta_{0.75}(1,0)$ and q_1/q_0 enters $\theta_{0.75}(0,1)$. The 99th percentile refers to the uncapped ratio. Floor is the percentage of denominator-arm raw posterior predictions below 0.01; cap is the percentage of uncapped ratios above 100. The fitted diagnostics do not test population absolute continuity.

Household covariance. For a cell-score vector $\widehat{\psi}_{ir}$, the point estimate and centered household contribution are

$$\widehat{\theta} = \frac{1}{R}\sum_{i=1}^{H}\sum_{r\in\mathcal{R}_i}\widehat{\psi}_{ir}, \qquad \widehat{u}_i = \frac{H}{R}\sum_{r\in\mathcal{R}_i}(\widehat{\psi}_{ir}-\widehat{\theta})\,. \tag{G.2}$$

Thus $\sum_i \widehat{u}_i = 0$. This aggregation preserves the record-weighted estimator. In particular, $\widehat{u}_i = (H/R)\sum_r \widehat{\psi}_{ir} - (Hr_i/R)\widehat{\theta}$, so it is not obtained by subtracting $\widehat{\theta}$ once from an unequally sized household's raw score sum.

Distances are computed on the full $1{,}594$-household endline roster graph. For analyzed households, put $A_{ik}^{(h)} = \mathbb{1}\{d(i,k)\le h\}$ and $d_i^{(h)} = \sum_k A_{ik}^{(h)}$, restricting both indices to the outcome sample after computing distances. Paths through households without an outcome remain admissible. The weights are

$$W_{ik}^{(h)} = \frac{\sum_\ell A_{i\ell}^{(h)}A_{k\ell}^{(h)}}{\sqrt{d_i^{(h)}d_k^{(h)}}}, \qquad h^* = \max\big(\{h\in\{0,\dots,8\} : L_H(h)/H \le 0.5\}\big)\,, \tag{G.3}$$

where $L_H(h) = \max_i |\{k : W_{ik}^{(h)} \neq 0\}|$. Self-inclusion keeps every denominator positive. The Gram construction gives symmetric positive-semidefinite weights, unit diagonal and entries in $[0,1]$, as in Appendix D. The selected radius is four; maximum row support is 766–769, or 48.8–48.9% of the analyzed households. Covariance estimates and contrast standard errors are

$$\widehat{V}_H^{\mathrm{net}} = \frac{1}{H}\sum_{i,k} W_{ik}^{(h^*)}\widehat{u}_i\widehat{u}_k^\top\,, \qquad \widehat{\mathrm{se}}(c^\top\widehat{\theta}) = \sqrt{c^\top\widehat{V}_H^{\mathrm{net}}c/H}\,. \tag{G.4}$$

The village comparator replaces $W_{ik}^{(h^*)}$ by $\mathbb{1}\{j(i)=j(k)\}$. Its scale agrees exactly with aggregating the original centered row scores within villages. The primary estimates use the network standard error alone. Tables 14 and 15 also report the larger of the network and village standard errors, for comparison.

G.3. **Regime contrasts and the original regressions.** Table 5 distinguishes three comparisons with the original equation (2). The current OLS column uses the analyzed household–round sample, the primary controls, and village and round fixed effects. The pooled mediation total averages W_s by analyzed saturation shares. The last

column instead uses weights proportional to $\widehat{\lambda}_s s(1-s)$, the population treatment-variance weights discussed in Appendix M. It is a comparison of estimated regime means, not the exact residualized OLS coefficient in an unbalanced panel with covariates. The original yield regression is Poisson; its reported coefficient 0.061 is on the log-mean scale and is omitted from this table of level effects.

TABLE 5. Within-village comparisons and regression weights

Outcome	Original eq. (2)	Current OLS	Pooled W_s	Variance-weighted W_s
Knowledge	1.235	1.234	1.189	1.175
Self-reported adoption	0.323	0.341	0.331	0.329
Audited adoption	0.040	0.049	0.045	0.046
Leaf health	0.031	0.040	0.030	0.033
Inputs	0.053	0.079	0.069	0.062
Yields (kg/tree)	—	0.034	0.024	0.022
Net revenue (RWF/tree)	5.225	6.755	4.246	4.060

Notes. Original coefficients are from Duflo et al. (2026, Table 2, Panel A); audited outcomes there use tree–round records. Current OLS uses the primary controls and household–round records. The two W_s columns use observation-share and treatment-variance weights, respectively. All entries are point estimates; their proximity does not imply equal estimands. The original Poisson yield coefficient has a different scale.

Tables 6 and 7 report the regime means and the policy and control-spillover contrasts. With nominal saturation weights, $V(s) = sR_s(1) + (1-s)R_s(0)$ and $S_s = R_s(0) - R_{.25}(0)$. The estimated knowledge policy effect is 0.692 at 75%, compared with 0.687 in the published Table 3, Panel A. The corresponding control spillover is -0.122, compared with -0.108 in Panel B. The revenue spillover is -26.2 RWF/tree, compared with -23.9 in the original study. These comparisons preserve each study's estimand and scale. The original randomization-inference p-values for the knowledge and revenue control spillovers are 0.706 and 0.262, whereas the reported network-normal p-values here are 0.039 and 0.048. The latter do not constitute a replication of the former tests.

TABLE 6. Estimated assignment-regime means

Outcome	$R_{.25}(0)$	$R_{.25}(1)$	$R_{.50}(0)$	$R_{.50}(1)$	$R_{.75}(0)$	$R_{.75}(1)$
Knowledge	0.062**	1.066***	-0.012	1.063***	-0.060	1.361***
	(0.028)	(0.207)	(0.036)	(0.121)	(0.052)	(0.112)
Self-reported adoption	0.033	0.307***	-0.026	0.287***	-0.028	0.357***
	(0.061)	(0.086)	(0.043)	(0.060)	(0.068)	(0.061)
Audited adoption	0.044	0.016	0.006	0.060	-0.108*	-0.025
	(0.040)	(0.053)	(0.036)	(0.045)	(0.058)	(0.036)
Leaf health	-0.027	-0.043	0.006	0.059**	-0.065*	-0.028
	(0.025)	(0.049)	(0.029)	(0.026)	(0.034)	(0.024)
Inputs	0.055	0.040	0.017	0.024	-0.055	0.130*
	(0.057)	(0.090)	(0.045)	(0.061)	(0.062)	(0.071)
Yields (kg/tree)	0.644***	0.581***	0.557***	0.574***	0.531***	0.616***
	(0.039)	(0.047)	(0.046)	(0.041)	(0.057)	(0.049)
Net revenue (RWF/tree)	142.012***	131.984***	118.053***	121.427***	115.831***	129.976***
	(8.556)	(10.756)	(10.235)	(9.662)	(12.302)	(10.098)

Notes. Diagonal augmented-score estimates from the common direct regime regression. These are standardized fitted means, not unadjusted cell averages. Network standard errors in parentheses.

G.4. **Within-saturation and spillover decompositions.** Tables 8 and 9 apply the same decomposition at each saturation. The last column evaluates the exposure change for a treated recipient. At 75%, the input response differs

TABLE 7. Population-policy effects and spillovers on controls

Outcome	$V(.50) - V(.25)$	$V(.75) - V(.25)$	$S_{.50}$	$S_{.75}$
Knowledge	0.213**	0.692***	-0.074	-0.122**
	(0.083)	(0.102)	(0.046)	(0.059)
Self-reported adoption	0.029	0.159*	-0.059	-0.061
	(0.061)	(0.086)	(0.064)	(0.106)
Audited adoption	-0.004	-0.083*	-0.039	-0.152**
	(0.047)	(0.048)	(0.055)	(0.069)
Leaf health	0.064**	-0.006	0.033	-0.038
	(0.032)	(0.030)	(0.038)	(0.040)
Inputs	-0.031	0.032	-0.038	-0.110
	(0.064)	(0.065)	(0.066)	(0.072)
Yields (kg/tree)	-0.063	-0.033	-0.087	-0.113*
	(0.045)	(0.050)	(0.055)	(0.063)
Net revenue (RWF/tree)	-19.765**	-13.065	-23.958*	-26.181**
	(9.985)	(10.626)	(12.226)	(13.210)

Notes. $V(s) = sR_s(1) + (1-s)R_s(0)$; $S_s = R_s(0) - R_{.25}(0)$. The policy mean uses nominal saturation weights. Network standard errors in parentheses.

significantly by recipient assignment, but this difference is not significant in the pooled comparison ($I = 0.016$, $p = 0.762$). For leaf health, the pooled treated-recipient response is 0.051 ($p = 0.005$), while the interaction is imprecise. Significance of one recipient response and insignificance of the other do not establish a significant difference between them.

For a secondary refinement, define $\mu_{z,z'}$ by integrating the response in z against the exposure distribution in z', using the same baseline population. The saturation spillover can be written as

$$\begin{aligned} S_P &= \mu_{(.75,0),(.25,0)} - R_{.25}(0), & S_N &= \mu_{(.25,0),(.75,0)} - R_{.25}(0)\,, \\ S_I &= S - S_P - S_N, & \mathrm{TE} &= D + S_P + S_N + S_I + (N+I)\,. \end{aligned} \tag{G.5}$$

Here S_P changes saturation at the low-saturation exposure distribution, S_N changes exposure at the low-saturation control response, and S_I is their interaction. Table 10 reports the five terms. For knowledge, the negative spillover is concentrated in $S_P = -0.182$ ($p = 0.048$). For inputs, $S_P = -0.352$ ($p = 0.002$) is partly offset by positive S_N and S_I. The cross-saturation means require additional directional support. Their exposure changes include peers' assignment composition as well as changes in relationships, so S_N is not a pure link-formation effect.

G.5. **Exposure and adjustment sensitivity.** Tables 11 and 12 replace the treated-friend count by an indicator for any treated friend, the sum of sessions attended by listed treated friends, treated friends outside the household's training group, or the pair of within-group and outside-group treated counts. Missing attendance is coded zero; outside-group count equals the count minus the within-group count. All specifications share the primary regime estimates, so their totals and saturation spillovers agree exactly. The tables report only the components that change.

The binary exposure retains the count model's assignment arguments but imposes a coarser exposure-sufficiency restriction. The attendance measure incorporates peers' realized participation, while group-specific exposures use the recipient's assigned training group. These are alternative observed-law functionals. Interpreting them as the same link-only intervention would require an expanded model for attendance or group assignment, since these features are not fixed pretreatment weights in the binary graph of Section 2.2.

TABLE 8. Within-saturation decomposition: 25% and 50%

Outcome	Within total	D_s	N_s	I_s	$N_s + I_s$
25% saturation					
Knowledge	1.003***	1.451**	0.066	-0.513	-0.447
	(0.217)	(0.653)	(0.060)	(0.691)	(0.661)
Self-reported adoption	0.275***	0.479***	0.002	-0.206	-0.204
	(0.092)	(0.181)	(0.037)	(0.132)	(0.126)
Audited adoption	-0.029	0.344	-0.019	-0.354*	-0.373
	(0.073)	(0.288)	(0.053)	(0.215)	(0.236)
Leaf health	-0.016	-0.107	-0.016	0.106	0.091*
	(0.047)	(0.077)	(0.031)	(0.067)	(0.052)
Inputs	-0.015	0.085	0.087*	-0.187**	-0.100
	(0.087)	(0.103)	(0.053)	(0.079)	(0.066)
Yields (kg/tree)	-0.063	-0.049	0.019	-0.034	-0.014
	(0.053)	(0.101)	(0.029)	(0.071)	(0.072)
Net revenue (RWF/tree)	-10.028	-18.594	0.320	8.247	8.567
	(12.111)	(20.240)	(6.296)	(15.367)	(14.386)
50% saturation					
Knowledge	1.075***	0.994***	0.021	0.059	0.080
	(0.114)	(0.204)	(0.042)	(0.133)	(0.116)
Self-reported adoption	0.313***	0.286**	0.019	0.008	0.027
	(0.082)	(0.128)	(0.044)	(0.067)	(0.055)
Audited adoption	0.054	0.083	0.078*	-0.107*	-0.029
	(0.046)	(0.077)	(0.046)	(0.056)	(0.041)
Leaf health	0.053	0.007	0.002	0.044	0.046*
	(0.038)	(0.052)	(0.033)	(0.046)	(0.028)
Inputs	0.007	-0.004	0.026	-0.015	0.011
	(0.047)	(0.075)	(0.050)	(0.080)	(0.042)
Yields (kg/tree)	0.017	-0.021	-0.056**	0.094***	0.038**
	(0.043)	(0.044)	(0.023)	(0.031)	(0.019)
Net revenue (RWF/tree)	3.373	-4.992	-12.706**	21.071**	8.366
	(10.511)	(11.193)	(5.377)	(8.220)	(5.109)

Notes. The within total is $W_s = R_s(1) - R_s(0) = D_s + N_s + I_s$. N_s evaluates the exposure change at a control response and $N_s + I_s$ at a treated response. Network standard errors in parentheses.

For the group-count pair, every control has zero within-group friends by construction. If treated households have positive probability of a positive within-group count, their exposure law is not absolutely continuous with respect to the control law. Consequently N and I are not identified in this specification and are omitted, including when a fitted forest returns numbers. The reverse direction, needed for D and $N + I$, still requires overlap among the zero-within-group exposures; the table reports those fitted contrasts without asserting that the remaining support condition has been verified. For knowledge its outcome-weight effective sample size in that direction is only 59.2 records. Changing probability floors cannot repair this structural support failure.

Table 13 uses the expanded L2 set with the count mediator. The direct knowledge effect remains large and significant, while $N = -0.027$ and $I = 0.070$ are imprecise. For inputs, the direct component increases to 0.226 ($p = 0.006$), and the network component and interaction cease to be significant at 10%. These comparisons show sensitivity of the channel allocation to adjustment. Because L5 and L2 are not nested and no joint cross-specification covariance is reported, this comparison is not a formal test of equality across adjustment sets.

TABLE 9. Within-saturation decomposition: 75% and pooled

Outcome	Within total	D_s	N_s	I_s	$N_s + I_s$
75% saturation					
Knowledge	1.420***	1.574***	0.029	-0.183	-0.154
	(0.106)	(0.194)	(0.055)	(0.151)	(0.141)
Self-reported adoption	0.385***	0.345***	-0.054	0.094	0.040
	(0.061)	(0.071)	(0.052)	(0.070)	(0.038)
Audited adoption	0.083*	0.036	-0.055	0.102**	0.047
	(0.044)	(0.060)	(0.045)	(0.049)	(0.032)
Leaf health	0.037	0.005	0.025	0.007	0.032
	(0.040)	(0.047)	(0.038)	(0.046)	(0.026)
Inputs	0.185***	0.135*	-0.124**	0.174**	0.050
	(0.069)	(0.075)	(0.052)	(0.068)	(0.045)
Yields (kg/tree)	0.085**	0.064	0.024	-0.003	0.021
	(0.043)	(0.050)	(0.043)	(0.065)	(0.037)
Net revenue (RWF/tree)	14.145	11.019	3.929	-0.802	3.126
	(9.369)	(12.091)	(8.246)	(13.541)	(8.964)
Pooled across saturations					
Knowledge	1.189***	1.323***	0.035	-0.169	-0.134
	(0.077)	(0.205)	(0.032)	(0.194)	(0.181)
Self-reported adoption	0.331***	0.354***	-0.013	-0.011	-0.023
	(0.043)	(0.070)	(0.028)	(0.053)	(0.042)
Audited adoption	0.045	0.128	0.005	-0.087	-0.082
	(0.030)	(0.079)	(0.028)	(0.060)	(0.060)
Leaf health	0.030	-0.021	0.006	0.045	0.051***
	(0.023)	(0.032)	(0.021)	(0.029)	(0.018)
Inputs	0.069*	0.070	-0.016	0.016	-0.001
	(0.037)	(0.049)	(0.035)	(0.051)	(0.029)
Yields (kg/tree)	0.024	0.004	-0.008	0.027	0.019
	(0.027)	(0.035)	(0.022)	(0.035)	(0.024)
Net revenue (RWF/tree)	4.246	-2.188	-3.307	9.740	6.433
	(6.038)	(8.014)	(4.261)	(7.595)	(5.534)

Notes. As Table 8. The pooled panel weights every component by the analyzed household–round saturation shares, separately for each outcome.

G.6. **Inference sensitivity and scope.** Tables 14 and 15 compare network radii one, two and four with village aggregation. At the reported precision, the positive audited-adoption interaction is significant at 5% with the primary network standard error but only at 10% with the maximum comparator. The negative revenue spillover likewise loses 5% significance under the maximum. The negative input network component and positive interaction retain 5% significance with that comparator. These are comparisons between estimated uncertainty measures, not alternative proofs of coverage. Proposition C.1 justifies the maximum only when at least one candidate is a valid benchmark.

Assignment law. The original Section 4.1 describes allocation according to village proportions; the randomization-inference programs rank household uniforms within villages and impose percentile cutoffs. Every observed village treatment count is within one household of its nominal quota. These facts support a quota interpretation, although the inference programs are not the original assignment script. Under a uniform quota K_j, for distinct households i, k in village j,

$$\Pr(P_k = 1 \mid P_i = a, K_j) = \frac{K_j - a}{n_j - 1}, \qquad \operatorname{Cov}(P_i, P_k \mid K_j) = -\frac{p_j(1-p_j)}{n_j - 1}, \quad p_j = K_j/n_j \,. \tag{G.6}$$

TABLE 10. Five-term refinement of the headline total

Outcome	D	S_P	S_N	S_I	$N+I$
Knowledge	1.574***	-0.182**	0.039	0.021	-0.154
	(0.194)	(0.092)	(0.040)	(0.079)	(0.141)
Self-reported adoption	0.345***	-0.233*	0.011	0.162	0.040
	(0.071)	(0.126)	(0.080)	(0.111)	(0.038)
Audited adoption	0.036	-0.234*	-0.074	0.156	0.047
	(0.060)	(0.135)	(0.104)	(0.127)	(0.032)
Leaf health	0.005	0.032	-0.091	0.021	0.032
	(0.047)	(0.080)	(0.061)	(0.084)	(0.026)
Inputs	0.135*	-0.352***	0.127*	0.114	0.050
	(0.075)	(0.111)	(0.068)	(0.099)	(0.045)
Yields (kg/tree)	0.064	-0.058	0.026	-0.081	0.021
	(0.050)	(0.114)	(0.047)	(0.090)	(0.037)
Net revenue (RWF/tree)	11.019	-27.972	10.438	-8.647	3.126
	(12.091)	(21.717)	(11.315)	(18.111)	(8.964)

Notes. The five columns sum to the total in Table 1 before rounding. The three S components sum to its saturation spillover. Cross-saturation exposure contrasts require their own identification and support conditions. Network standard errors in parentheses.

Thus independent random ranks do not give the independent Bernoulli assignments in Assumption 2.1. Under quotas, conditioning on own assignment changes the peer-assignment law, and every within-village pair has assignment dependence. To transfer the causal interpretation, one must either identify the original design as Bernoulli or define quota-specific exposure interventions, use their actual inclusion probabilities and establish identification and inference for sampling without replacement. Keeping one peer draw fixed while changing own assignment cannot satisfy both fixed-quota worlds. The reported algebra remains valid, but that design-specific link-change interpretation is not supplied by the present Bernoulli theorem.

Dependence geometry and bandwidth. Cross-village covariance is permitted by the theory. The unresolved issue is whether the measured graph and selected weights satisfy its conditions. Assumption 5.2 uses deterministic weights at the theorem's conditioning level; (G.3) uses a treatment-responsive observed graph. A sufficient transfer argument would exhibit admissible deterministic weights W_H^0 with

$$\left\| H^{-1}\widehat{U}^\top (W^{(h^*)} - W_H^0)\widehat{U} \right\|_{\mathrm{op}} = \mathrm{o}_{\mathbb{P}}(1)\,, \tag{G.7}$$

where the rows of $\widehat{U}$ are the centered household scores. Together with a valid household-score theorem and consistency for W_H^0, this would establish consistency for the implemented weights. Alternatively, a conditional theory must establish score centering, dependence and identification after conditioning on the random geometry; merely conditioning on endline links can change the assignment law. The support rule $L_H/H \leq 0.5$ checks neither route and implies none of the vanishing-bias or growth-rate conditions in Assumption 5.2. Positive semidefiniteness ensures nonnegative variances, not correct coverage.

Household weights and learning. Equations (G.2)–(G.4) establish the implemented normalization exactly. Applying the individual-level limit theory additionally requires a household-vector version of the score argument for the unequal observed-round weights, including the observation law, record-mixture probabilities and nuisance rates. With fixed, balanced rounds and appropriate common laws, aggregation is a direct change of sampling unit; unequal, potentially informative missingness needs a separate argument. Full-size bootstrap forests also need their own accuracy and neighborhood-stability bounds for the complete generated-response procedure. Household out-of-bag prediction alone does not establish Assumption 4.4. The subsampled-bagging sufficient condition in Cao

TABLE 11. Exposure sensitivity: binary exposure and attendance

Outcome	D	N	I	$N+I$
Any treated friend				
Knowledge	1.555***	-0.020	-0.114	-0.134
	(0.204)	(0.059)	(0.185)	(0.160)
Self-reported adoption	0.337***	-0.066	0.114	0.048
	(0.087)	(0.059)	(0.091)	(0.053)
Audited adoption	0.062	0.055	-0.034	0.021
	(0.057)	(0.085)	(0.093)	(0.033)
Leaf health	0.065	0.076*	-0.104	-0.028
	(0.074)	(0.040)	(0.075)	(0.061)
Inputs	0.088	-0.162***	0.259***	0.097**
	(0.077)	(0.048)	(0.071)	(0.044)
Yields (kg/tree)	0.118**	-0.025	-0.007	-0.032
	(0.055)	(0.045)	(0.073)	(0.041)
Net revenue (RWF/tree)	17.121	-2.191	-0.785	-2.976
	(12.503)	(12.193)	(18.923)	(10.048)
Sessions attended by treated friends				
Knowledge	1.614***	-0.002	-0.192	-0.194
	(0.261)	(0.055)	(0.219)	(0.213)
Self-reported adoption	0.360***	0.012	0.013	0.025
	(0.086)	(0.072)	(0.091)	(0.049)
Audited adoption	0.026	-0.007	0.064	0.057
	(0.066)	(0.045)	(0.056)	(0.039)
Leaf health	-0.016	0.005	0.049	0.053**
	(0.050)	(0.042)	(0.048)	(0.025)
Inputs	0.090	-0.092**	0.188***	0.095**
	(0.081)	(0.045)	(0.066)	(0.046)
Yields (kg/tree)	0.063	-0.036	0.059	0.022
	(0.047)	(0.046)	(0.059)	(0.029)
Net revenue (RWF/tree)	6.177	-0.731	8.700	7.968
	(10.700)	(11.065)	(13.149)	(6.419)

Notes. Primary adjustment set and common regime means. The top panel uses $\mathbb{1}\{G_i \geq 1\}$; the bottom panel sums listed treated friends' sessions attended. Each exposure has its own sufficiency and overlap requirements. Network standard errors in parentheses.

and Leung (2025, Section 4.3, Proposition 4) does not cover a full-size bootstrap merely because both average trees.

Causal conditions. BNS, exposure sufficiency and directional overlap remain maintained restrictions. Common conditional laws in Assumption 4.1 must also hold across the required villages and profiles. A cross-village contact can make the exposure law depend on another village's saturation; conditioning on the full profile does not automatically make that law common across households. The primary set does not contain cross-arm contact composition. Missing network records and the timing of the endline mediator require the measurement interpretation stated above. These issues concern the causal interpretation and theorem applicability of the fitted functional. They do not change its decomposition identity or the numerical results reported here.

APPENDIX H. A FINITE-DIMENSIONAL STRUCTURAL MEAN MODEL AND G-ESTIMATION

A structural mean model can reduce the dimension of the response regression, but it adds a restriction on that regression. It does not remove mediator confounding, create overlap, or identify a saturation coefficient after the estimating equation has projected that coefficient away. This appendix develops a complete partially linear

TABLE 12. Exposure sensitivity: training-group composition

Outcome	D	N	I	$N+I$
Treated friends outside own training group				
Knowledge	1.195***	-0.097	0.323	0.226
	(0.227)	(0.149)	(0.228)	(0.182)
Self-reported adoption	0.316***	-0.307	0.377*	0.069
	(0.075)	(0.190)	(0.194)	(0.054)
Audited adoption	0.077	-0.063	0.069	0.006
	(0.065)	(0.144)	(0.156)	(0.058)
Leaf health	0.045	0.168*	-0.176	-0.008
	(0.061)	(0.100)	(0.115)	(0.050)
Inputs	0.332***	-0.186**	0.039	-0.147***
	(0.093)	(0.074)	(0.098)	(0.056)
Yields (kg/tree)	0.077	-0.061	0.070	0.009
	(0.055)	(0.049)	(0.060)	(0.043)
Net revenue (RWF/tree)	10.726	-17.977	21.395	3.419
	(9.850)	(14.177)	(15.279)	(6.135)
Treated friends within and outside own group				
Knowledge	1.249***	—	—	0.171
	(0.326)			(0.329)
Self-reported adoption	0.229	—	—	0.156
	(0.204)			(0.207)
Audited adoption	-0.043	—	—	0.126*
	(0.089)			(0.072)
Leaf health	-0.024	—	—	0.061
	(0.093)			(0.085)
Inputs	0.283	—	—	-0.098
	(0.174)			(0.165)
Yields (kg/tree)	-0.045	—	—	0.130**
	(0.070)			(0.056)
Net revenue (RWF/tree)	-20.120	—	—	34.265*
	(21.949)			(19.661)

Notes. Primary adjustment set and common regime means. Group assignment enters the exposure definition. Dashes indicate components not identified because controls cannot have positive within-group exposure. Remaining contrasts require the reverse support condition. Network standard errors in parentheses.

specification with a rank condition, a product remainder, and network asymptotics. Following the centered-moment construction in Vansteelandt and Joffe (2014, Section 3.2, equation (14)), we center both the response and its feature vector. The centering law here is the uniform-site mixture conditional on the realized profile; its identification and dependence conditions are those established in this paper.

H.1. **The response model and its rank condition.** Fix a supported profile t and suppress n, t where harmless. Let $b(z, g, x) \in \mathbb{R}^p$ be a specified bounded measurable feature vector, with fixed p. Assume the common conditional response function satisfies

$$m(z, g, x) = \beta(x) + b(z, g, x)^\top \psi_0 \,, \tag{H.1}$$

for a bounded baseline function β and a fixed finite vector ψ_0. The representation holds on all required intervention supports. For causal interpretation, impose the controlled-response identification conditions as well; then differences of m at common x are controlled-response differences. Natural effects require the additional cross-world

TABLE 13. Expanded baseline adjustment with treated-friend count

Outcome	D	N	I	$N+I$
Knowledge	1.377***	-0.027	0.070	0.043
	(0.123)	(0.054)	(0.088)	(0.059)
Self-reported adoption	0.331***	0.016	0.039	0.054*
	(0.067)	(0.047)	(0.061)	(0.030)
Audited adoption	0.073	-0.005	0.016	0.010
	(0.047)	(0.030)	(0.037)	(0.017)
Leaf health	0.021	-0.030	0.046*	0.016
	(0.043)	(0.019)	(0.025)	(0.017)
Inputs	0.226***	-0.052	0.011	-0.041
	(0.083)	(0.040)	(0.055)	(0.041)
Yields (kg/tree)	0.042	-0.029	0.072*	0.043**
	(0.042)	(0.031)	(0.041)	(0.019)
Net revenue (RWF/tree)	7.200	-6.878	13.824	6.945*
	(10.172)	(9.774)	(11.351)	(3.546)

Notes. L2 contains the 50 conditioning variables described in Section G.1. The direct regime regression retains the primary set, holding TE and S fixed at the estimates in Table 1. These are current fits, not the superseded empirical estimates. Network standard errors in parentheses.

TABLE 14. Standard-error sensitivity: knowledge, adoption and leaf health

Contrast	Estimate	Network $h=1$	Network $h=2$	Network $h=4$	Village	Maximum
Knowledge						
TE	1.298	0.109	0.107	0.107	0.144	0.144
S	-0.122	0.066	0.066	0.059	0.061	0.061
D	1.574	0.205	0.201	0.194	0.192	0.194
N	0.029	0.067	0.067	0.055	0.048	0.055
I	-0.183	0.162	0.159	0.151	0.141	0.151
Self-reported adoption						
TE	0.324	0.065	0.071	0.093	0.104	0.104
S	-0.061	0.074	0.081	0.106	0.115	0.115
D	0.345	0.081	0.082	0.071	0.055	0.071
N	-0.054	0.062	0.062	0.052	0.045	0.052
I	0.094	0.074	0.075	0.070	0.065	0.070
Audited adoption						
TE	-0.069	0.046	0.051	0.053	0.061	0.061
S	-0.152	0.058	0.062	0.069	0.086	0.086
D	0.036	0.062	0.061	0.060	0.054	0.060
N	-0.055	0.044	0.044	0.045	0.052	0.052
I	0.102	0.054	0.053	0.049	0.059	0.059
Leaf health						
TE	-0.001	0.039	0.039	0.032	0.038	0.038
S	-0.038	0.049	0.048	0.040	0.040	0.040
D	0.005	0.056	0.056	0.047	0.037	0.047
N	0.025	0.039	0.039	0.038	0.032	0.038
I	0.007	0.047	0.047	0.046	0.037	0.046

Notes. All columns use the same saved primary cell scores and estimates. Network columns vary only the overlap radius. Maximum is the larger of the radius-four network and village-cluster standard errors. No column is adjusted for multiple testing.

TABLE 15. Standard-error sensitivity: inputs, yields and revenue

Contrast	Estimate	Network $h = 1$	Network $h = 2$	Network $h = 4$	Village	Maximum
Inputs						
TE	0.075	0.072	0.073	0.069	0.069	0.069
S	-0.110	0.080	0.080	0.072	0.065	0.072
D	0.135	0.096	0.094	0.075	0.094	0.094
N	-0.124	0.059	0.057	0.052	0.056	0.056
I	0.174	0.079	0.079	0.068	0.074	0.074
Yields (kg/tree)						
TE	-0.028	0.045	0.048	0.054	0.083	0.083
S	-0.113	0.051	0.056	0.063	0.082	0.082
D	0.064	0.057	0.058	0.050	0.060	0.060
N	0.024	0.045	0.046	0.043	0.051	0.051
I	-0.003	0.056	0.059	0.065	0.062	0.065
Net revenue (RWF/tree)						
TE	-12.036	10.230	10.728	11.441	17.609	17.609
S	-26.181	11.759	12.235	13.210	17.003	17.003
D	11.019	13.985	13.803	12.091	10.239	12.091
N	3.929	8.590	8.563	8.246	8.650	8.650
I	-0.802	12.368	12.699	13.541	10.791	13.541

Notes. As Table 14. In the knowledge frame, the maximum nonzero row shares at radii 1, 2 and 4 are 0.028, 0.144 and 0.488; at radius five the share rises to 0.707, exceeding the selection cutoff. The other frames give similar values.

condition. The model can be understood as a linear-in-parameters blip with respect to any supported reference feature; the feature need not itself be linear in exposure.

Let $F_i = b(Z_i, G_i, B_i)$, and use the conditional uniform-site mixture to define

$$r(x) = \mathbb{E}[F_I \mid B_I = x], \qquad \ell(x) = \mathbb{E}[Y_I \mid B_I = x], \qquad V_i = F_i - r(B_i), \quad U_i = Y_i - \ell(B_i)\ .$$

Here V_i is a residualized feature, not the covariance matrix of Section 5. Equation (H.1) implies

$$\ell = \beta + r^\top \psi_0, \qquad U_i = V_i^\top \psi_0 + \vartheta_i, \qquad \mathbb{E}[\vartheta_i \mid Z_i, G_i, B_i] = 0$$

at every site under the common response-law premise. Define the deterministic mixture matrix

$$\Gamma_n = \frac{1}{n} \sum_{1 \le i \le n} \mathbb{E}[V_i V_i^\top] = \mathbb{E}(\mathrm{Var}(F_I \mid B_I))\ . \tag{H.2}$$

Assume $\Gamma_n \to \Gamma \succ 0$ for each supported profile. The second equality uses mixture conditional variance; it is not an average of variances conditional on the raw village identifier.

Proposition H.1 (Identification and loss of saturation rank)**.** *Under* (H.1)*, the positive-definite matrix in* (H.2) *identifies* ψ_0 *from* $\mathbb{E}[V_I(U_I - V_I^\top \psi)] = 0$*. If instead a feature column is a function only of village saturation and the features are residualized within each village given* t*, that column is identically zero.*

Proof. The conditional mean of ϑ_i is zero, so $\mathbb{E}[V_i \vartheta_i] = 0$. Averaging gives the moment equation $\Gamma_n(\psi_0 - \psi) = 0$. A positive-definite Γ_n has trivial null space, proving uniqueness for all sufficiently large n. Conversely a nonzero null vector a satisfies $\mathbb{E}[(a^\top V_I)^2] = 0$, hence $a^\top V_I = 0$ almost surely; moving ψ along a leaves the centered regression unchanged. If a column is $h(t_{j(i)})$, conditioning on village and t makes both it and its conditional expectation equal to the same constant. Their difference is zero for every observation and every sample size, regardless of the number of distinct constants across villages. ∎

Combining villages permits between-village saturation variation under the common conditional law; it does not guarantee a positive eigenvalue. Collinearity of exposure features or lack of their conditional variation can still violate (H.2). Conducting inference within an identified subspace requires that subspace to be specified and its rank established.

For deterministic candidate functions r', ℓ', define

$$\rho_i(\psi; r', \ell') = (F_i - r'(B_i))\{Y_i - \ell'(B_i) - (F_i - r'(B_i))^\top \psi\} .$$

All candidates are bounded by fixed constants. Write $\delta_r = r' - r$ and $\delta_\ell = \ell' - \ell$.

Proposition H.2 (Product bias and orthogonality)**.** *At ψ_0 the exact average moment bias is*

$$\frac{1}{n} \sum_{1 \le i \le n} \mathbb{E}[\rho_i(\psi_0; r', \ell')] = \mathbb{E}[\delta_r(B_I)\{\delta_\ell(B_I) - \delta_r(B_I)^\top \psi_0\}] . \tag{H.3}$$

It has norm at most $\|\delta_r\|_2\{\|\delta_\ell\|_2 + \|\psi_0\|\|\delta_r\|_2\}$ and its derivative at (r, ℓ) in every bounded direction is zero.

Proof. Since $F_i - r' = V_i - \delta_r$ and the second factor at ψ_0 is $\vartheta_i - \delta_\ell + \delta_r^\top \psi_0$, their product is

$$V_i\vartheta_i - \delta_r\vartheta_i - V_i\delta_\ell + V_i\delta_r^\top\psi_0 + \delta_r\delta_\ell - \delta_r\delta_r^\top\psi_0 .$$

The first two terms have zero mean by the sitewise conditional response restriction. The next two have zero *mixture* mean because $\mathbb{E}[V_I \mid B_I] = 0$. These cancellations need not hold separately within each fixed village. The last two terms give (H.3). Apply the triangle inequality, the Cauchy–Schwarz inequality, and $|\delta_r^\top \psi_0| \le \|\delta_r\|\|\psi_0\|$ to obtain the bound. Replacing (δ_r, δ_ℓ) by (uh_r, uh_ℓ) makes the bias quadratic in u, proving the zero derivative. Correct r eliminates the bias for any ℓ'; correct ℓ alone leaves $-\mathbb{E}[\delta_r\delta_r^\top]\psi_0$. Thus the displayed identity, rather than an unrestricted claim of double robustness in (r, ℓ), states the exact robustness property. ■

H.2. **Estimation, stability, and its network limit.** Fit $(\widehat{r}, \widehat{\ell})$ from the full array, with all generated stages included in one measurable algorithm. Let

$$\widehat{V}_i = F_i - \widehat{r}(B_i), \quad \widehat{U}_i = Y_i - \widehat{\ell}(B_i), \quad \widehat{\Gamma} = n^{-1} \sum_{1 \le i \le n} \widehat{V}_i\widehat{V}_i^\top, \qquad \widehat{\psi} = \widehat{\Gamma}^{-1} n^{-1} \sum_{1 \le i \le n} \widehat{V}_i\widehat{U}_i .$$

For a measurable all-samples definition, use $\widehat{\psi} = 0$ when the smallest eigenvalue of $\widehat{\Gamma}$ is below $c_\Gamma/2$, where $c_\Gamma > 0$ is smaller than the limiting minimum eigenvalue for every supported profile. The event has probability tending to zero by the consistency argument below. Use the same guarded inverse (zero on the safeguard event) in every displayed sandwich covariance. This safeguard makes no assertion on unsupported or rank-deficient populations.

Theorem H.3 (G-estimator expansion)**.** *Assume the preceding common response model and rank condition; bounded Y, F and candidates; and the dependency graph for the observations. For the vector score $\rho(\psi_0; r, \ell)$ impose the same actual-row and independent-evaluation single/double hybrid consistency and stability inequalities as Assumption 4.4, with score rates $a_{\rho,n}, s_{\rho,n}$. Suppose the nuisance norms are $\mathrm{O}_{\mathbb{P}}(a_{r,n})$ and $\mathrm{O}_{\mathbb{P}}(a_{\ell,n})$, with deterministic sequences tending to zero, and*

$$\Delta_n a_{\rho,n}^2 \to 0, \qquad \sqrt{n} s_{\rho,n} \to 0, \qquad \sqrt{n} a_{r,n}(a_{\ell,n} + a_{r,n}) \to 0 . \tag{H.4}$$

Assume also $\|\widehat{r} - r\|_{2,n} = \mathrm{o}_{\mathbb{P}}(1)$. If $\Delta_n^4/n \to 0$ and $\Omega_n = n^{-1}\operatorname{Var}(\sum_{1 \le i \le n} V_i\vartheta_i) \to \Omega$ is finite, then

$$\sqrt{n}(\widehat{\psi} - \psi_0) = \Gamma_n^{-1} n^{-1/2} \sum_{1 \le i \le n} V_i\vartheta_i + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1}\Omega\Gamma^{-\top}) . \tag{H.5}$$

With kernel assumptions for the centered oracle scores $V_i\vartheta_i$ and $L_n\|\widehat{\rho}-\rho\|_{2,n} = \mathrm{o}_{\mathbb{P}}(1)$, where $\widehat{\rho}_i = \widehat{V}_i(\widehat{U}_i - \widehat{V}_i^{\top}\widehat{\psi})$, the corresponding sandwich covariance is consistent. Wald inference uses only nondegenerate contrast covariances.

Proof. First prove consistency of $\widehat{\Gamma}$. The oracle matrix $n^{-1}\sum_{1\le i\le n} V_iV_i^{\top}$ has expectation Γ_n and componentwise variance bounded by $16M_F^4\Delta_n/n$, where M_F bounds each coordinate of F and its conditional mean because each product is bounded by $4M_F^2$, each product covariance is at most $16M_F^4$ in absolute value, and there are at most $n\Delta_n$ nonzero covariance terms. Its dimension is fixed, and the bound vanishes under the degree condition. For $\delta_i = \widehat{r}(B_i) - r(B_i)$,

$$\widehat{V}_i\widehat{V}_i^{\top} - V_iV_i^{\top} = -V_i\delta_i^{\top} - \delta_iV_i^{\top} + \delta_i\delta_i^{\top} .$$

The operator norm of its average is bounded by $2\|V\|_{2,n}\|\delta\|_{2,n} + \|\delta\|_{2,n}^2 = \mathrm{o}_{\mathbb{P}}(1)$. Thus $\widehat{\Gamma} - \Gamma_n = \mathrm{o}_{\mathbb{P}}(1)$, and $\widehat{\Gamma} \to \Gamma \succ 0$. For every unit vector a, $|a^{\top}(\widehat{\Gamma} - \Gamma)a| \le \|\widehat{\Gamma} - \Gamma\|_{\mathrm{op}}$, so the safeguard is inactive with probability tending to one and $\widehat{\Gamma}^{-1} - \Gamma_n^{-1} = \mathrm{o}_{\mathbb{P}}(1)$, by the inverse identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ and bounded inverse norms.

For the score sum, set the deterministic site functional $F_i(r', \ell') = \mathbb{E}[\rho_i(\psi_0; r', \ell') - \rho_i(\psi_0; r, \ell)]$. The imposed single/double hybrid inequalities permit the proof in Appendix B to be applied to each component. The replacement cost in L^2 is at most $2\sqrt{n}s_{\rho,n}$, and the centered hybrid sum has second moment at most $\Delta_n a_{\rho,n}^2 + 16ns_{\rho,n}^2$. Both vanish. The average deterministic functional evaluated at the fitted functions is bounded by Proposition H.2; its normalized sum is $\mathrm{O}_{\mathbb{P}}(\sqrt{n}a_{r,n}(a_{\ell,n} + a_{r,n})) = \mathrm{o}_{\mathbb{P}}(1)$. Consequently

$$n^{-1/2}\sum_{1\le i\le n}\rho_i(\psi_0; \widehat{r}, \widehat{\ell}) = n^{-1/2}\sum_{1\le i\le n} V_i\vartheta_i + \mathrm{o}_{\mathbb{P}}(1) .$$

On the nonsingular event the estimating equation is exactly linear.

$$\widehat{\Gamma}\sqrt{n}(\widehat{\psi} - \psi_0) = n^{-1/2}\sum_{1\le i\le n}\rho_i(\psi_0; \widehat{r}, \widehat{\ell}) .$$

Each oracle score has sitewise mean zero. The bounded-score cumulant theorem proves the joint CLT of its normalized sum and in particular makes that sum $\mathrm{O}_{\mathbb{P}}(1)$. Multiplying by the consistent inverse proves (H.5). Changing the estimator on an event of vanishing probability does not affect convergence in distribution.

The kernel covariance proof applies directly to $\rho_i = V_i\vartheta_i$. Its oracle pair-product variance, weighted bias, and empirical replacement bound have precisely the main theorem's form. Thus $\widehat{\Omega} = n^{-1}\sum_{1\le i,k\le n} w_{ik}\widehat{\rho}_i\widehat{\rho}_k^{\top} \to \Omega$. Continuity of the finite-dimensional product gives $\widehat{\Gamma}^{-1}\widehat{\Omega}\widehat{\Gamma}^{-\top} \to \Gamma^{-1}\Omega\Gamma^{-\top}$. The CLT and a positive limiting contrast variance complete studentization. ∎

The actual-row consistency premise for $\widehat{r}$ follows, for example, from its own independent-evaluation consistency and single-replacement L^2 stability by the replacement argument and Markov's inequality in Appendix C. It is specified explicitly to cover a score-stability certificate that does not separately control this nuisance. Deterministic cross-fitting with separated training sets is another route. Define each fold's r, ℓ using its evaluation mixture, train outside its full dependence buffer, impose that theorem's training-law support and exact bias conditions, and use its conditional variance proof. The corresponding fold oracle and averaged Γ then replace the displayed full-sample quantities. Independent ordinary random folds, without this buffer or a proved stability bound, are not justified by these results.

APPENDIX I. IDENTIFICATION COUNTEREXAMPLES AND COUPLING

Expectations in these examples average the response primitives, assignment interventions, and uniform target index from Appendix A. The baseline variable is $B_i = (X_i, N_i^{\mathrm{pre}})$; where it is constant, adjustment supplies no further information.

Example I.1 (An interventional draw does not automatically preserve the regime mean)**.** Take two eligible individuals with independent $U_1, U_2 \sim \mathrm{Bernoulli}(1/2)$, constant baseline controls, an empty baseline network, and the treatment-invariant edge $A_{12} = U_1 U_2$. Randomize assignments independently. Let $G_i = A_{12}$ and set $Y_i^s(a, g) = gU_i$. Observed outcomes equal A_{12}, since $U_i^2 = U_i$, so every observed randomized mean equals $1/4$. The exposure drawn from the target-population distribution is $\mathrm{Bernoulli}(1/4)$ independent of U_i, giving

$$\mathbb{E}[Y_i^s(a, \widetilde{G}_i^s(a))] = \mathbb{E}[\widetilde{G}_i^s(a)]\mathbb{E}[U_i] = \frac{1}{4}\frac{1}{2} = \frac{1}{8}\,.$$

In the alternative model $Y_i^s(a, g) = g$, all observed assignments, graphs and outcomes have exactly the same joint law, while the interventional mean with matching assignments is $1/4$. Controlled-response exchangeability fails in the first model. No randomized-design argument can distinguish these interventional values from the observed law alone.

Example I.2 (Cross-saturation exposure changes without any link changes)**.** Let an individual have one permanent friend and $Y_i = G_i = P_k$. Under saturation s, $\mathbb{E}G_i = s$. Hence a cross-saturation mediator contrast equals $s_H - s_L$ while every counterfactual graph is identical. This example does not establish nonidentification of the saturation contrast: randomization identifies it. It establishes that a nonzero cross-saturation exposure effect need not reflect any change in relationships. Here the entire effect comes from changing the probability that the same friend receives an offer. Separating formation and transmission labels across saturations would require an additional edge intervention with its own joint law, support, and controlled-response assumptions.

A marginal exposure law alone determines the interventional target but need not determine a natural target. The latter also uses its coupling with the outcome shock. Holding peers fixed across focal assignments gives the intended link-change comparison; the choice of coupling of baseline and post-treatment formation noises remains structural. There is no unique inverse-cdf construction, and consistency can be implemented with any valid joint construction that agrees with the observed intervention. Under BNS, natural means depend only on the conditional exposure marginals, because Theorem 3.3 proves equality to (3.7). This is the restricted sense in which a coupling-invariance claim is valid.

APPENDIX J. CROSS-FITTING WITH SEPARATED TRAINING SETS

The construction follows the network cross-fitting principle in Emmenegger et al. (2025, Section 2.4, equation (6)). Remove an evaluation set and its dependency-graph neighbors from nuisance training. It extends the training–evaluation separation used in double machine learning (Chernozhukov et al., 2018, Farbmacher et al., 2022) to dependent observations. Separation permits a conditional variance calculation with the fitted parameters held fixed; the required graph covers every input to the score. With a fixed number of growing villages, a useful split must retain large regions within the observed saturation arms. Leaving out one entire village does not generally provide such a split. Their dependency graph is defined in Section 2.3; equation (6) removes its neighbors when constructing each training sample. The partition here is fixed conditional on the profile and geometry, and the theorem explicitly requires both evaluation and retained training samples to have positive arm shares and order-n

sizes. These conditions need not follow from an arbitrary random partition when neighborhoods grow. Ordinary independent-data mediation cross-fitting uses the same score cancellation (Farbmacher et al., 2022); independence of retained and evaluation collections supplies the additional step here. Balkus et al. (2026) discuss circumstances in which correlated evaluation and training samples can still be analyzed, while Cao and Leung (2025) use neighborhood stability to analyze learning without splitting. The theorem below establishes the separated-training route under its explicit premises.

Condition on the complete saturation profile as in Section 4. Let $K \geq 2$ be a fixed integer, distinct from the eligibility graph K_n. For a partition $I_1, \ldots, I_K$ of $[n]$, define

$$n_k = |I_k|, \quad \mathcal{R}_k = [n] \setminus \bigcup_{i \in I_k} N[i], \quad \lambda_{s,k} = \frac{|\{i \in I_k : t_{j(i)} = s\}|}{n_k}, \quad e_{(s,a),k} = \lambda_{s,k} s^a (1-s)^{1-a} .$$

The set $\mathcal{R}_k$ is the training set after removing the evaluation set and its dependence neighbors. The probabilities $e_{z,k}$ describe a uniformly selected evaluation observation in fold k. Training-set arm shares can differ and are used when fitting conditional distributions from that training set.

Assumption J.1 (Separated and supported folds)**.** The following conditions hold uniformly over supported profiles. Individual offers follow the conditionally independent Bernoulli design in Assumption 2.1.

(i) The partition is a deterministic function of the profile and the specified deterministic dependence graph. For constants $c_I, c_R, c_s > 0$ independent of n,

$$\min_k n_k \geq c_I n, \qquad \min_k |\mathcal{R}_k| \geq c_R n ,$$

$$\min_{k,s} \lambda_{s,k} \geq c_s, \qquad \min_{k,s} \frac{|\{i \in \mathcal{R}_k : t_{j(i)} = s\}|}{|\mathcal{R}_k|} \geq c_s$$

for all required saturations, and the required own-assignment probabilities are positive. There is a fixed $e_{*,F} > 0$ with $e_{z,k} \geq e_{*,F}$.

(ii) Every nuisance used in fold k, including generated responses, preprocessing and tuning, is measurable with respect to $\mathcal{T}_k = \sigma(O_i : i \in \mathcal{R}_k) \vee \sigma(\mathsf{U})$, where U is independent of the observation array. The fitted bounds in Assumption 4.3 hold, using $e_{*,F}$ for evaluation weights.

(iii) The common conditional laws in Assumption 4.1 hold on each deterministic evaluation and training subset. Let $\eta_{0,k}$ contain their common m, q, r, ν and the fold's evaluation probabilities, and let $\widehat{\eta}_k$ be its estimate. Define

$$\varpi^2_{\psi,k} = \max_{i \in I_k} \mathbb{E}\left[|\psi(O_i^*; \widehat{\eta}_k) - \psi(O_i^*; \eta_{0,k})|^2 \mid \mathcal{T}_k \right] ,$$

where O_i^* is independent of $\mathcal{T}_k$ with the law of O_i. Define $\varpi_{m,k}, \varpi_{r,k}, \varpi_{\nu,k}, \varpi_{e,k}$ to be the realized nuisance-error norms of Assumption 4.3. Write $b_k = \varpi_{e,k}\varpi_{\nu,k} + \varpi_{e,k}\varpi_{m,k} + \varpi_{r,k}\varpi_{m,k}$.

In part (iii), restriction to a deterministic subset preserves the common site laws; no additional conditioning on outcomes or exposures is performed. If the partition depends on those variables, its conditional laws need not agree and the assumption must be verified for that selection rule. With known fold probabilities $\varpi_{e,k} = 0$. To verify assignment-cell support, let $\mathcal{Z}_0$ be the finite set of required assignment cells, let $p_{\min} = \min_{(s,a) \in \mathcal{Z}_0} s^a (1-s)^{1-a} > 0$, and write $N^{\mathrm{tr}}_{z,k} = \sum_{i \in \mathcal{R}_k} \mathbb{1}\{Z_i = z\}$. Conditional Bernoulli assignment implies

$$\mu_{z,k} := \mathbb{E} N^{\mathrm{tr}}_{z,k} \geq c_R c_s p_{\min} n , \qquad \mathrm{Var}(N^{\mathrm{tr}}_{z,k}) \leq n/4 .$$

Chebyshev's inequality and a union bound therefore give the explicit bound

$$\mathbb{P}\left\{\min_{k,z} N^{\mathrm{tr}}_{z,k} < \frac{1}{2}c_R c_s p_{\min} n\right\} \le \sum_{1\le k\le K}\sum_{z\in\mathcal{Z}_0} \frac{4\mathrm{Var}(N^{\mathrm{tr}}_{z,k})}{\mu^2_{z,k}} \le \frac{K|\mathcal{Z}_0|}{c_R^2 c_s^2 p^2_{\min} n} \longrightarrow 0\,.$$

For supported endpoint saturations the own assignments are deterministic, so the same bound remains valid. This calculation provides assignment-cell support; learning within rare baseline or exposure cells is covered by the nuisance-error conditions in part (iii).

Define

$$\widehat{\theta}_F = \frac{1}{n}\sum_{1\le k\le K}\sum_{i\in I_k}\psi(O_i;\widehat{\eta}_k)\,, \qquad \phi_{i,F} = \psi(O_i;\eta_{0,k}) - \theta_n \quad (i\in I_k)\,.$$

Theorem J.1 (Consistency and asymptotic linearity for separated folds)**.** *Maintain Assumptions 4.1, 4.2, 4.3, and J.1.*

(i) If $\Delta_n/n \to 0$ *and* $\max_k b_k \xrightarrow{\mathbb{P}} 0$, *then* $\widehat{\theta}_F - \theta_n \xrightarrow{\mathbb{P}} 0$.

(ii) If

$$\Delta_n \max_k \varpi^2_{\psi,k} \xrightarrow{\mathbb{P}} 0, \qquad \sqrt{n}\max_k b_k \xrightarrow{\mathbb{P}} 0\,,$$

then

$$\sqrt{n}(\widehat{\theta}_F - \theta_n) = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}\phi_{i,F} + \mathrm{o}_{\mathbb{P}}(1)\,, \qquad \mathbb{E}\phi_{i,F} = 0\,.$$

Proof. Fix k. The sets I_k and $\mathcal{R}_k$ are disjoint and no dependence edge joins them. Assumption 4.2 gives independence of their entire observation collections. Since U is independent of the array, O_{I_k} is independent of $\mathcal{T}_k$. In particular, for every bounded measurable u on the evaluation-array space and $A\in\mathcal{T}_k$,

$$\mathbb{E}[\mathbb{1}_A u(O_{I_k})] = \mathbb{P}(A)\mathbb{E}[u(O_{I_k})], \qquad \mathbb{E}[u(O_{I_k})\mid\mathcal{T}_k] = \mathbb{E}[u(O_{I_k})] \quad \text{a.s.}$$

Thus the conditional joint law equals the marginal joint law, including the independence of any two unconnected evaluation subsets. This is the separation used by Emmenegger et al. (2025); here it is imposed on the entire evaluation collection and used with its own assignment-cell probabilities.

For $i\in I_k$, define $d_{i,k} = \psi(O_i;\widehat{\eta}_k) - \psi(O_i;\eta_{0,k})$ and $a_{i,k} = \mathbb{E}[d_{i,k}\mid\mathcal{T}_k]$. Then $\mathbb{E}[d_{i,k} - a_{i,k}\mid\mathcal{T}_k] = 0$ and

$$\mathbb{E}[(d_{i,k}-a_{i,k})^2\mid\mathcal{T}_k] = \mathbb{E}[d^2_{i,k}\mid\mathcal{T}_k] - a^2_{i,k} \le \varpi^2_{\psi,k}\,.$$

If $i,l\in I_k$ are nonadjacent, their conditional covariance is zero. Otherwise the conditional Cauchy–Schwarz inequality bounds its absolute value by $\varpi^2_{\psi,k}$. As in the dependency-degree control of Emmenegger et al. (2025, Assumption A4, part 4.3), nuisance accuracy is weighted by the number of possibly correlated evaluation terms. At most $n_k\Delta_n$ ordered pairs are adjacent or equal. Therefore

$$\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}\sum_{i\in I_k}(d_{i,k}-a_{i,k})\right\}^2 \,\middle|\, \mathcal{T}_k\right] \le \frac{n_k}{n}\Delta_n\varpi^2_{\psi,k}\,. \tag{J.1}$$

For any $\varepsilon,\delta>0$, conditional Chebyshev's inequality gives

$$\mathbb{P}\left(\left|n^{-1/2}\sum_{i\in I_k}(d_{i,k}-a_{i,k})\right| > \varepsilon\right) \le \mathbb{P}(\Delta_n\varpi^2_{\psi,k} > \delta) + \delta/\varepsilon^2\,.$$

Taking $n \to \infty$ and then $\delta \downarrow 0$ proves that the normalized centered sum is $o_{\mathbb{P}}(1)$ under part (ii). This step needs convergence in probability of the conditional variance bound, rather than an unstated unconditional moment-rate assumption.

For the mean term, a uniformly selected observation in I_k has baseline law F_B, conditional laws m_z, q_z, and cell probabilities $e_{z,k}$. The mediation-score cancellation of Tchetgen Tchetgen and Shpitser (2012), Farbmacher et al. (2022), established in Proposition 4.1, can therefore be applied with these probabilities, conditional on $\mathcal{T}_k$.

$$\frac{1}{n_k}\sum_{i\in I_k} a_{i,k} = \mathbb{E}_B\big[(1-\widehat{\kappa}_{z',k})\delta\widehat{\nu}_k + (\widehat{\kappa}_{z',k}-\widehat{\kappa}_{z,k})E_{z'}\delta\widehat{m}_k \\ - \widehat{\kappa}_{z,k}E_z(\delta\widehat{r}_k\delta\widehat{m}_k)\big] .$$

The three absolute values are bounded respectively by $e_{*,F}^{-1}\varpi_{e,k}\varpi_{\nu,k}$, $2\sqrt{R}e_{*,F}^{-1}\varpi_{e,k}\varpi_{m,k}$ and $e_{*,F}^{-1}\varpi_{r,k}\varpi_{m,k}$. With $C_F = e_{*,F}^{-1}\max(1, 2\sqrt{R})$, their total contribution after normalization is at most

$$\left| n^{-1/2}\sum_{i\in I_k} a_{i,k} \right| \le \frac{n_k}{n} C_F \sqrt{n}\, b_k = o_{\mathbb{P}}(1) .$$

This conditional product calculation is the mediation counterpart of the second-order remainder bound in Chernozhukov et al. (2018, Assumption 3.2). Both the centering and the product cancellation use evaluation probabilities. Replacing them by training or full-sample shares would in general leave a first-order mean error.

Now sum over k. Since K is fixed, if $Z_{k,n} = o_{\mathbb{P}}(1)$ for each fold then $\mathbb{P}(|\sum_{1\le k\le K} Z_{k,n}| > \varepsilon) \le \sum_{1\le k\le K}\mathbb{P}(|Z_{k,n}| > \varepsilon/K) \to 0$. This requires no independence between fitted folds. Adding and subtracting their oracle scores proves part (ii). Equation (B.1), with each fold's positive denominator, gives $\mathbb{E}\phi_{i,F} = 0$.

For part (i), let $B_{\psi,F} = M + 2M(R+1)/e_{*,F}$ and $B_{\phi,F} = B_{\psi,F} + M$. Bounded scores give $\varpi_{\psi,k} \le 2B_{\psi,F}$. Dividing (J.1) by n bounds the conditional variance of the unscaled average in fold k by $4B_{\psi,F}^2 n_k\Delta_n/n^2 \to 0$. Its mean error is at most $(n_k/n)C_F b_k \xrightarrow{\mathbb{P}} 0$. Finally the oracle average has variance at most $B_{\phi,F}^2\Delta_n/n \to 0$. Applying Chebyshev's inequality and summing the fixed number of folds proves consistency. ■

The fold-specific oracle scores remain deterministic functions of individual observations. The normal-approximation and covariance arguments of Section 5 consequently apply when their variance-limit and kernel conditions hold for these scores. For feasible covariance, conditional Markov's inequality applied to

$$\frac{1}{n}\sum_{i\in I_k}\mathbb{E}\big[|d_{i,k}|^2 \mid \mathcal{T}_k\big] \le \frac{n_k}{n}\varpi_{\psi,k}^2$$

shows that $\max_k \varpi_{\psi,k} = O_{\mathbb{P}}(a_n)$ implies empirical score error $O_{\mathbb{P}}(a_n)$. To see the probability step, restrict to $\varpi_{\psi,k}^2 \le Ca_n^2$, apply the conditional bound C/K_0 at threshold $K_0a_n^2$, and then let C and K_0 increase in that order. A finite union bound handles all folds. Under the oracle expansion and bounded aggregate variance the center error is $O_{\mathbb{P}}(n^{-1/2})$, so $L_n(a_n + n^{-1/2}) \to 0$ suffices for the fitted-score part of covariance consistency.

An expanding lattice village with bounded-range observation dependence admits such folds by dividing each village into a fixed number of large regions and removing boundary strips from the corresponding training sets. The removed fraction tends to zero if the strip width is small relative to the regions' linear dimensions. This is a geometric condition on the dependence graph. Increasing village size without that geometry does not guarantee a usable split. Repeated records of an individual stay together, and a fixed finite average of valid splits can be analyzed through the average of their oracle scores. A rule that chooses how many splits to average from their outcomes needs a separate argument, as distinct from the reproducibility analysis in Ritzwoller and Romano (forthcoming).

APPENDIX K. AN EXTENSION UNDER ABSOLUTE REGULARITY

This appendix allows dependence to persist beyond a specified neighborhood. The number of villages remains fixed and every village size diverges. All statements condition on the entire saturation profile, whose notation is suppressed. The common conditional laws, bounded scores, and target definitions are those in Section 4. Neighborhoods are defined using a deterministic metric d_n on the individual indices; the metric can include connections between villages.

For two σ-fields $\mathcal{A}, \mathcal{C}$, define

$$\beta(\mathcal{A}, \mathcal{C}) := \frac{1}{2} \sup \sum_{1 \le j \le J_0} \sum_{1 \le k \le K_0} |\mathbb{P}(A_j \cap C_k) - \mathbb{P}(A_j)\mathbb{P}(C_k)| ,$$

where the supremum is over finite measurable partitions (A_j) and (C_k) from the two fields. For standard Borel random elements this equals the total-variation distance between their joint law and the product of their marginals, with convention $\mathrm{TV}(P, Q) = \sup_A |P(A) - Q(A)|$. Define

$$\beta_n(r) := \sup_{A, C \subseteq [n]:\, d_n(A,C) > r} \beta(\sigma(O_i : i \in A), \sigma(O_i : i \in C)) ,$$

$$\mathcal{N}_n(i, r) := \{k : d_n(i, k) \le r\}, \qquad D_n(r) := \max_i |\mathcal{N}_n(i, r)| . \tag{K.1}$$

Here $d_n(A, C) = \min_{i \in A, k \in C} d_n(i, k)$, with distance infinity when a set is empty. Empty-set coefficients are zero. The coefficient controls arbitrary separated sets, so it is stronger than the singleton-versus-rest coefficient in Cao and Leung (2025). This version supports the setwise covariance and cumulant arguments below.

The elementary inequality used throughout is

$$|\mathbb{E}_P h - \mathbb{E}_Q h| \le 2\|h\|_\infty \, \mathrm{TV}(P, Q) . \tag{K.2}$$

For bounded h, integrate the level sets of $h + \|h\|_\infty$ over $[0, 2\|h\|_\infty]$ to prove it. Marginalization and measurable transformations cannot increase total variation. Adding a common independent auxiliary element also cannot increase it, by first integrating that element out. These facts allow replacement learners and their independent copied arrays to be included in the following comparisons.

Assumption K.1 (Mixing and neighborhood learning)**.** There are deterministic radii r_n such that $D_n := D_n(r_n)$ and $\beta_n := \beta_n(r_n)$ satisfy the rates specified in the result being used. The learning procedure obeys Assumption 4.4, replacing $N[i] \cup N[k]$ by $\mathcal{N}_n(i, r_n) \cup \mathcal{N}_n(k, r_n)$ in every nuisance estimate and bound. The copy array and the auxiliary algorithm randomization are independent of the observations and shared by the replacement comparisons. All functions obey Assumption 4.3.

Lemma K.1 (Replacement bound with mixing)**.** *Under Assumptions 4.1, 4.3, and K.1, with f_i defined in Lemma B.3,*

$$\left\| n^{-1/2} \sum_{1 \le i \le n} f_i(O_i; \widehat{\eta}) \right\|_2 \le 2\sqrt{n} s_n + \{D_n \overline{\varpi}_{\psi,n}^2 + 16 n s_n^2 + 160 B_\psi^2 n \beta_n\}^{1/2} , \quad \overline{\varpi}_{\psi,n} = \min(\varpi_{\psi,n}, 2B_\psi) . \tag{K.3}$$

Proof. Use the notation $U_i, Q_i, Q_k, \mathcal{A}_i, \mathcal{A}_{ik}$ from Lemma B.3, with the larger metric neighborhoods. That proof's initial triangle inequality and Jensen's inequality use only stability, so their contribution remains $2\sqrt{n} s_n$. Set $D_\psi = 4B_\psi$, which bounds each centered evaluation f_i in absolute value.

The joint law of O_i and its retained original collection differs from their independent product by at most β_n. Add the copied array and algorithm randomization to both laws. Under the independent product the single-replacement

learner is independent of the evaluation observation and the squared centered evaluation has expectation at most $\overline{\varpi}^2_{\psi,n}$. Since its squared value is bounded by D^2_ψ, (K.2) gives

$$\mathbb{E}U_i^2 \le \overline{\varpi}^2_{\psi,n} + 2D^2_\psi\beta_n \,.$$

For at most nD_n near ordered pairs, the Cauchy–Schwarz inequality gives this same upper bound on $|\mathbb{E}U_iU_k|$.

For a far pair $d_n(i,k) > r_n$, Q_i is measurable in $\mathcal{A}_k$. Under the product law separating O_k from that field, the centered U_k has conditional mean zero and hence $\mathbb{E}Q_iU_k = 0$. Comparing this product law with the original law yields $|\mathbb{E}Q_iU_k| \le 2D^2_\psi\beta_n$. The reversed comparison gives $|\mathbb{E}U_iQ_k| \le 2D^2_\psi\beta_n$. For Q_iQ_k, first separate the pair (O_i, O_k) from the retained original set outside both neighborhoods, at total-variation cost at most β_n. Then separate the two coordinates of the pair from each other, at cost at most another β_n. The copied array and algorithm randomization remain common independent auxiliaries. Under the resulting product law, conditional on the retained data the two centered evaluations are independent with mean zero. Thus $|\mathbb{E}Q_iQ_k| \le 4D^2_\psi\beta_n$.

Expanding the product as in the exact-dependence proof gives

$$\mathbb{E}U_iU_k = \mathbb{E}[(U_i - Q_i)(U_k - Q_k)] + \mathbb{E}Q_iU_k + \mathbb{E}U_iQ_k - \mathbb{E}Q_iQ_k \,.$$

The score and centering-map stability bounds still give $\|U_i - Q_i\|_2, \|U_k - Q_k\|_2 \le 4s_n$. Consequently

$$|\mathbb{E}U_iU_k| \le 16s_n^2 + 8D^2_\psi\beta_n \,.$$

Summing all ordered pairs, without requiring $\mathbb{E}U_i = 0$, gives

$$\begin{aligned}\mathbb{E}\left(n^{-1/2}\sum_{1\le i\le n} U_i\right)^2 &\le D_n(\overline{\varpi}^2_{\psi,n} + 2D^2_\psi\beta_n) + n(16s_n^2 + 8D^2_\psi\beta_n)\\ &\le D_n\overline{\varpi}^2_{\psi,n} + 16ns_n^2 + 10D^2_\psi n\beta_n \,.\end{aligned}$$

Use $D_n \le n$, substitute $D_\psi = 4B_\psi$, and add the initial replacement error. This proves (K.3). ■

Theorem K.1 (Consistency and linearity under mixing)**.** *Maintain the assumptions of Lemma K.1 and let* $b_n = \varpi_e\varpi_\nu + \varpi_e\varpi_m + \varpi_r\varpi_m$.

(i) If $D_n/n \to 0$, $\beta_n \to 0$, $s_n \to 0$, *and* $b_n \to 0$, *then* $\widehat{\theta}_n - \theta_n \xrightarrow{\mathbb{P}} 0$.

(ii) If $D_n\varpi^2_{\psi,n} \to 0$, $\sqrt{n}s_n \to 0$, $\sqrt{n}b_n \to 0$, *and* $n\beta_n \to 0$, *then the expansion* (4.9) *holds.*

Proof. Site centering follows from the common conditional laws and uses no independence. For two far sites, (K.2) applied to the product of centered scores gives $|\mathbb{E}\phi_i\phi_k| \le 2B^2_\phi\beta_n$. The nD_n near ordered pairs are bounded by B^2_ϕ. Hence

$$\mathbb{E}\left(n^{-1}\sum_{1\le i\le n}\phi_i\right)^2 \le B^2_\phi D_n/n + 2B^2_\phi\beta_n \longrightarrow 0$$

under part (i). Divide (K.3) by $\sqrt{n}$ and use $\overline{\varpi}_{\psi,n} \le 2B_\psi$ to show that the fitted centered average also vanishes. The exact three-product mean remainder is unchanged and is $\mathrm{O}_{\mathbb{P}}(b_n)$. This proves consistency using (B.6). Under part (ii), the right side of (K.3) tends to zero and the scaled mean remainder does too. This proves asymptotic linearity. ■

The requirement $n\beta_n(r_n) \to 0$ is the same rate form used by Cao and Leung (2025), here required for the joint information in any two separated sets of observations of arbitrary sizes. In particular, it controls two products of scores or two groups of cumulant arguments, rather than only one observation and a distant collection. It must be

balanced against $D_n(r_n)$ and the stability cost of replacing that many observations. The following stronger, explicit conditions give a self-contained inference extension.

Assumption K.2 (Mixing rates for normal approximation and covariance)**.** In addition to Theorem K.1(ii), assume

$$D_n^4/n \to 0, \qquad n^m \beta_n(r_n) \to 0 \quad \text{for every fixed integer } m \geq 1\,,$$
$$\|V_n - V\|_{\mathrm{op}} \to 0, \qquad V \succeq 0\,,$$

with $c^\top V c > 0$ for scalar inference or $R_0 V R_0^\top \succ 0$ for a joint test. The kernel obeys the symmetry, positivity and entry bounds in Assumption 5.2, and

$$\left\| n^{-1} \sum_{1 \leq i,k \leq n} (w_{ik,n} - 1)\mathbb{E}\phi_i \phi_k^\top \right\|_{\mathrm{op}} \to 0, \qquad D_n L_n^2/n + L_n^2 \beta_n(r_n) \to 0\,,$$
$$L_n\{s_n + \varpi_{\psi,n} + \sqrt{\beta_n(r_n)} + n^{-1/2}\} \to 0\,.$$

Theorem K.2 (Normal approximation and feasible covariance under mixing)**.** *Under Assumption K.2, $\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow{d} \mathcal{N}(0, V)$ and $\|\widehat{V}_n^{\mathrm{net}} - V\|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0$. The studentized intervals and joint tests in Section 5 have their coverage and size guarantees for the finite-n center θ_n. Under Assumption D.1, the same conclusions hold for the fixed causal center θ.*

Proof. For a scalar contrast write $X_i = c^\top \phi_i$ and $M_X = \|c\|_1 B_\phi$. Join positions in an ordered r-tuple when their site distance is at most r_n. The connected-tuple count is at most $r^{r-1} n D_n^{r-1}$ by (C.2). For a disconnected tuple, choose one position component and its complement. The corresponding two site groups have distance greater than r_n. Their joint law differs from its independent product by total variation at most β_n.

Under that product law the joint cumulant is zero. For a block A of a partition, the difference of its two moments is at most $2M_X^{|A|}\beta_n$ by (K.2). The difference of a product of k such moments is the sum of k terms obtained by replacing one factor at a time. Each term is at most $2M_X^r \beta_n$, so the product difference is at most $2kM_X^r\beta_n$. Multiplying by $(k-1)!$ and summing the partition formula gives

$$|\operatorname{cum}(X_{i_1}, \ldots, X_{i_r})| \leq 2M_X^r \widetilde{A}_r \beta_n\,, \qquad \widetilde{A}_r := \sum_{\pi \in \mathcal{P}_r} |\pi|!\,.$$

Thus multilinearity yields, for every fixed $r \geq 3$,

$$|\operatorname{cum}_r(n^{-1/2} \sum_{1 \leq i \leq n} X_i)| \leq r^{r-1} A_r M_X^r n^{1-r/2} D_n^{r-1} + 2M_X^r \widetilde{A}_r n^{r/2} \beta_n \longrightarrow 0\,.$$

This sum includes all n^r ordered tuples, with repetitions. Copies of one site belong to the same distance component. Connected tuples are bounded without a mixing reduction. Disconnected tuples need not have zero cumulant under the actual law, and there can be order n^r of them. Their aggregate contribution is the second term above. In particular, $A_3 = 6$, $A_4 = 26$, $\widetilde{A}_3 = 13$, and $\widetilde{A}_4 = 75$ give

$$|\operatorname{cum}_3(n^{-1/2}\textstyle\sum_{1\leq i\leq n} X_i)| \leq 54 M_X^3 D_n^2/\sqrt{n} + 26 M_X^3 n^{3/2}\beta_n\,,$$
$$|\operatorname{cum}_4(n^{-1/2}\textstyle\sum_{1\leq i\leq n} X_i)| \leq 1664 M_X^4 D_n^3/n + 150 M_X^4 n^2 \beta_n\,.$$

For general fixed $r \geq 3$,

$$n^{1-r/2} D_n^{r-1} = (D_n^4/n)^{(r-1)/4} n^{(3-r)/4} \longrightarrow 0\,.$$

Choose an integer $m \geq r/2$ to bound $n^{r/2}\beta_n \leq n^m\beta_n \to 0$. The decay required at every fixed order controls the distant tuples. The first cumulant is zero by site centering and the second converges to $c^\top Vc$. The moment–cumulant formula gives convergence of every fixed moment to its Gaussian counterpart. Applying that formula at higher even orders bounds the moments needed for uniform integrability. Tightness follows from the bounded second moments, and Gaussian moment determinacy identifies every subsequential limit, as in Remark C.1. This proves the scalar normal limit, including a degenerate limit when $c^\top Vc = 0$. Cramér–Wold and Theorem K.1 give the estimator limit.

For covariance concentration, let $\mathcal{E}_w = \{(i,k) : w_{ik,n} \neq 0\}$ and $Q_n^{ab} = n^{-1}\sum_{(i,k)\in\mathcal{E}_w} w_{ik,n}\phi_{i,a}\phi_{k,b}$. The exact expansion is

$$\mathrm{Var}(Q_n^{ab}) = \frac{1}{n^2}\sum_{(i,k),(\ell,m)\in\mathcal{E}_w} w_{ik,n}w_{\ell m,n}\mathrm{Cov}(\phi_{i,a}\phi_{k,b}, \phi_{\ell,a}\phi_{m,b}) \,.$$

There are at most nL_n ordered kernel pairs. For a fixed (i,k), a second pair is near if at least one of its endpoints is within distance r_n of i or k. Each of the four endpoint choices gives at most D_nL_n pairs, using the row bound and symmetry of the weights. The union bound therefore gives at most $4D_nL_n$ near pairs, including repeated indices and coincident pairs. Their product covariances are at most $2B_\phi^4$. Every remaining pair has $d_n(\{i,k\},\{\ell,m\}) > r_n$. Its covariance is bounded by $2B_\phi^4\beta_n$ using (K.2); no independence within either pair is required. Bounding their number by $(nL_n)^2$ and $|w_{ik,n}|$ by one gives

$$\mathrm{Var}(Q_n^{ab}) \leq 8B_\phi^4 D_nL_n^2/n + 2B_\phi^4 L_n^2\beta_n \longrightarrow 0 \,.$$

For $u > 0$, Chebyshev's inequality gives

$$\mathbb{P}(|Q_n^{ab} - \mathbb{E}Q_n^{ab}| > u) \leq \frac{8B_\phi^4 D_nL_n^2/n + 2B_\phi^4 L_n^2\beta_n}{u^2} \longrightarrow 0 \,.$$

There are a fixed number of entries. The assumed bias condition and $V_n \to V$ therefore give $\|Q_n - V\|_{\mathrm{op}} \xrightarrow{\mathbb{P}} 0$.

For a single-replacement evaluation, compare the joint evaluation/training law with their independent product. The squared score error is bounded by $4B_\psi^2$, so its expectation is at most $\varpi_{\psi,n}^2 + 8B_\psi^2\beta_n$. Inserting that nuisance estimate and using the same two-term squared-error inequality as in the feasible proof gives

$$\frac{1}{n}\sum_{1\leq i\leq n}\mathbb{E}\big[|\psi(O_i;\widehat{\eta}) - \psi(O_i;\eta_0)|^2\big] \leq 2s_n^2 + 2\varpi_{\psi,n}^2 + 16B_\psi^2\beta_n \,.$$

Summing coordinates and applying Markov's inequality bounds the empirical score error by $\mathrm{O}_{\mathbb{P}}(s_n + \varpi_{\psi,n} + \sqrt{\beta_n})$. The mean error is $\mathrm{O}_{\mathbb{P}}(n^{-1/2})$ by the expansion and finite variance limit. Write $\rho_n^2 = n^{-1}\sum_{1\leq i\leq n}\|\widehat{\phi}_i - \phi_i\|^2$. These bounds imply $\rho_n = \mathrm{O}_{\mathbb{P}}(s_n + \varpi_{\psi,n} + \sqrt{\beta_n} + n^{-1/2})$. The two cross products and the error product in (C.6), together with $\|W_n\|_{\mathrm{op}} \leq L_n$, give

$$\|\widehat{V}_n^{\mathrm{net}} - Q_n\|_{\mathrm{op}} \leq 2B_\phi\sqrt{d_\theta}L_n\rho_n + L_n\rho_n^2 \xrightarrow{\mathbb{P}} 0 \,.$$

The last rate in Assumption K.2 controls the linear term and, since $\rho_n \xrightarrow{\mathbb{P}} 0$, the quadratic term. Covariance consistency follows by the triangle inequality. These nondegeneracy conditions permit the continuous-mapping and Slutsky arguments in Section 5; Assumption D.1 permits recentering to θ. ■

The all-orders mixing condition is sufficient for this moment proof and is substantially stronger than the condition used for asymptotic linearity. To exhibit compatible growth rates, suppose the metric has polynomial ball growth $D_n(r) \leq C_D(1+r)^d$ and the setwise coefficient satisfies $\beta_n(r) \leq C_\beta n^A e^{-cr}$ for fixed positive constants.

Choose $r_n = (\log n)^2$. Then $D_n = \mathrm{O}((\log n)^{2d})$ and $n^m \beta_n(r_n) \le C_\beta \exp((m+A)\log n - c(\log n)^2) \to 0$ for each fixed m. Neighborhood radius grows much more slowly than the linear dimensions of expanding lattice villages with positive limiting shares. If a complete learner has $\varpi_{\psi,n} = \mathrm{O}(\sqrt{D_n/n + \beta_n})$, nuisance rates of the same order and $s_n = \mathrm{O}(D_n/n)$, with known assignment probabilities, these rates satisfy the estimation conditions. A polynomial-logarithmic kernel support meets the concentration and fitted-score rates. Its bias condition and the finite variance limit must still be verified for the specified process.

For the ball-intersection kernel, a sufficient scalar bias bound is

$$B_\phi^2 D_n \max_{d_n(i,k) \le r_n} |1 - w_{ik,n}| + 4B_\phi^2 n \beta_n(r_n) \, .$$

The first term counts the near pairs and the second bounds all far pairs using $|1 - w_{ik,n}| \le 2$. On a lattice with polynomial ball growth, choosing a logarithmic-power bandwidth h_n with $h_n/(D_n r_n) \to \infty$ makes the first term vanish while keeping L_n of logarithmic-power order. This verifies a joint bandwidth and mixing regime. With polynomially decaying dependence, Theorem K.1 can still apply when its displayed rates hold; the stronger all-orders condition imposed for this particular normal-approximation proof must also be checked before using Theorem K.2. A persistent common village shock generally violates these mixing conditions as well as exact local dependence.

Appendix L. Unbounded outcomes

The bounded-outcome assumption simplifies the main normal-approximation proof. This appendix allows unbounded observations, including outcomes with Gaussian noise, while retaining the fixed number of villages and the dependency graph for the observations. All probabilities and expectations condition on the supported saturation profile. The observation $O_i = (B_i, Z_i, G_i, Y_i)$, baseline adjustment $B_i = (X_i, N_i^{\mathrm{pre}})$, neighborhoods $N[i]$, and size Δ_n are those in Section 4. Population integrals of fitted functions retain the interpretation given at the start of Appendix B.

Assumption L.1 (Conditional moments)**.** For some fixed $p > 4$ and $M \ge 1$,

$$\max_{i \in [n]} \mathbb{E}[|Y_i|^p \mid B_i, G_i, Z_i] \le M^p \quad \text{almost surely} \, .$$

The true and fitted functions m, ν are bounded by M, assignment probabilities are at least $e_* > 0$, and true and fitted density ratios lie in $[0, R]$. These restrictions hold for the complete nuisance estimates computed from replacement observations as well. Retain the measurability, common conditional laws, and exact dependence requirements of Assumptions 4.1 and 4.2. The mean-square prediction and replacement conditions in Assumption 4.4 concern the score using the original, untruncated outcome.

Conditional Jensen's inequality implies the true regression bound. The separate fitted bound is imposed by clipping predictions and does not truncate the observed outcome. For example, $Y_i = m_{Z_i}(G_i, B_i) + \sigma_i U_i$, with bounded m and σ_i, and a conditional law $U_i \mid (B_i, G_i, Z_i) \sim \mathcal{N}(0,1)$, has conditional moments of every fixed finite order. The conditional Gaussian statement alone does not specify dependence across individuals; that dependence must still obey the observation dependency graph.

Let

$$H_\psi = M + \frac{2M(R+1)}{e_*}, \qquad H_\phi = H_\psi + M \, .$$

For any fitted or fixed candidate obeying the function bounds, the pathwise inequality

$$|\psi(O_i;\eta)| \le M + \frac{R}{e_*}(|Y_i| + M) + \frac{2M}{e_*}$$

and Minkowski's inequality give $\|\psi(O_i;\eta)\|_p \le H_\psi$. This bound holds even for a nuisance estimate using O_i. The oracle centered score has L_p norm at most H_ϕ.

Lemma L.1. *Under Assumption L.1, the mean identity, nuisance-product bound, and score-prediction inequality in Proposition 4.1 and Lemma B.1 continue to hold with their displayed constants. Lemma B.3 holds with* $\overline{\varpi}_{\psi,n} = \min(\varpi_{\psi,n}, 2H_\psi)$. *Consequently the consistency and asymptotic-linearity conclusions of Theorem 4.2 hold under its respective rates, replacing bounded outcomes by Assumption L.1.*

Proof. The exact mean calculation integrates Y only through its conditional mean. Conditional first moments are finite, so its three conditional expectations and the change of measure remain valid. The bound on the resulting three products depends on the function bounds and overlap, and therefore is unchanged.

For the score-prediction inequality, the only terms whose earlier bounds used the realized value of Y contain $Y - m$. Conditional Jensen's inequality and Minkowski's inequality give

$$\{\mathbb{E}[(Y_i - m(G_i, B_i))^2 \mid B_i, G_i, Z_i = z]\}^{1/2} \le 2M \, .$$

For a candidate fixed before the evaluation integral, therefore,

$$\mathbb{E}\left[\left\{\frac{\mathbb{1}\{Z_i = z\}}{\widetilde{e}_z(B_i)}\delta r(G_i, B_i)(Y_i - m(G_i, B_i))\right\}^2\right] \le \frac{4M^2}{e_*^2}\|\delta r\|_z^2 \, ,$$

$$\mathbb{E}\left[\left\{\mathbb{1}\{Z_i = z\} r(G_i, B_i)(Y_i - m(G_i, B_i))(\widetilde{e}_z^{-1} - e_z^{-1})(B_i)\right\}^2\right] \le \frac{4M^2R^2}{e_*^4}\|\delta e_z\|_B^2 \, .$$

The first inequality conditions on (B_i, G_i, Z_i) and uses a cell probability at most one. The second also integrates over G_i and uses $r \le R$. All remaining terms involve bounded nuisance parameters only. Taking square roots and adding their bounds proves the same prediction inequality. Conditioning on an independently evaluated nuisance estimate gives its random-function version.

In the replacement proof, independence of O_i and the retained training σ-field yields

$$\mathbb{E}U_i^2 \le \mathbb{E}|\psi(O_i^*; \widehat{\eta}^{(-i)}) - \psi(O_i^*; \eta_0)|^2 \le \min(\varpi_{\psi,n}^2, 4H_\psi^2) \, .$$

The two zero cross-expectations and the zero double-replacement expectation follow by the same conditional independence and centering. Their products are integrable by the Cauchy–Schwarz inequality. The bound $\|U_i - Q_i\|_2 \le 4s_n$ also uses only mean-square stability. Counting near and far pairs thus gives

$$\left\| n^{-1/2} \sum_{1 \le i \le n} f_i(O_i; \widehat{\eta}) \right\|_2 \le 2\sqrt{n}s_n + \{\Delta_n \min(\varpi_{\psi,n}^2, 4H_\psi^2) + 16ns_n^2\}^{1/2} \, .$$

For consistency, each oracle covariance is at most H_ϕ^2 by the Cauchy–Schwarz inequality, and is zero outside the dependence graph. Its average has variance at most $H_\phi^2\Delta_n/n$. Dividing the preceding display by $\sqrt{n}$ and combining with the three-product bias proves consistency. Under the stronger learning rates, the unscaled display tends to zero and the scaled bias tends to zero in probability, proving asymptotic linearity. ∎

A finite-moment normal approximation requires an additional balance between outcome tails and neighborhood growth. The truncation below is used only in the proof; the proposed estimator continues to use Y_i itself.

Lemma L.2. *Let $X_{i,n}$ be centered scalar variables with the dependence graph $\mathcal{H}_n$ and $\max_i \mathbb{E}|X_{i,n}|^p \le H^p$ for fixed $p > 4$ and $H < \infty$. Suppose $n^{-1}\mathrm{Var}(\sum_{1\le i\le n} X_{i,n}) \to v < \infty$. If a deterministic sequence $K_n \to \infty$, with $K_n \ge 1$, satisfies*

$$\Delta_n K_n^{-(p-2)} \to 0, \qquad \frac{\Delta_n^2 K_n^3}{\sqrt{n}} \to 0\,, \tag{L.1}$$

then $n^{-1/2}\sum_{1\le i\le n} X_{i,n} \xrightarrow{d} \mathcal{N}(0, v)$.

Proof. Suppress the array subscript and define

$$X_i^K = \mathrm{clip}_{K_n}(X_i) - \mathbb{E}\,\mathrm{clip}_{K_n}(X_i), \qquad R_i = X_i - X_i^K\,.$$

These variables are centered measurable functions of the original observations, so the same dependence graph applies. Since $|X_i - \mathrm{clip}_{K_n}(X_i)| \le |X_i|\mathbb{1}\{|X_i| > K_n\}$,

$$\begin{aligned}\mathbb{E}R_i^2 &= \mathrm{Var}(X_i - \mathrm{clip}_{K_n}(X_i)) \\ &\le \mathbb{E}[X_i^2\mathbb{1}\{|X_i| > K_n\}] \le K_n^{-(p-2)}\mathbb{E}|X_i|^p \le H^p K_n^{-(p-2)}\,.\end{aligned}$$

For $k \notin N[i]$ the covariance of R_i, R_k is zero. For the remaining at most $n\Delta_n$ ordered pairs, the Cauchy–Schwarz inequality bounds it by $H^p K_n^{-(p-2)}$. Thus

$$\left\| n^{-1/2} \sum_{1\le i\le n} R_i \right\|_2^2 \le H^p \Delta_n K_n^{-(p-2)} \longrightarrow 0\,. \tag{L.2}$$

Write $S_n = n^{-1/2}\sum_{1\le i\le n} X_i$ and $S_n^K = n^{-1/2}\sum_{1\le i\le n} X_i^K$. The inequality $|\|S_n^K\|_2 - \|S_n\|_2| \le \|S_n^K - S_n\|_2$ shows that $\mathrm{Var}(S_n^K) \to v$.

Each X_i^K is bounded by $2K_n$. The exact cumulant expansion and connected-tuple count in Lemma C.1 give, for each fixed integer $r \ge 3$,

$$\begin{aligned}|\,\mathrm{cum}_r(S_n^K)| &\le r^{r-1} A_r 2^r K_n^r n^{1-r/2}\Delta_n^{r-1} \\ &= r^{r-1} A_r 2^r \frac{\Delta_n^2 K_n^3}{\sqrt{n}} \left(\frac{\Delta_n K_n}{\sqrt{n}}\right)^{r-3} \longrightarrow 0\,.\end{aligned}$$

Indeed $\Delta_n K_n/\sqrt{n} \le \Delta_n^2 K_n^3/\sqrt{n} \to 0$ because $\Delta_n, K_n \ge 1$. The first cumulant is zero and the second converges to v. The inverse partition formula consequently gives convergence of every moment of S_n^K to the corresponding normal moment. Remark C.1 supplies tightness, uniform integrability and uniqueness of that moment limit. Hence $S_n^K \xrightarrow{d} \mathcal{N}(0, v)$. Equation (L.2) and Slutsky's lemma transfer the limit to S_n. If $v = 0$, the original variance convergence proves the result directly by Chebyshev's inequality. ∎

Theorem L.1. *Maintain Assumption L.1, the learning rates in Theorem 4.2(ii), covariance convergence $V_n \to V$, and (L.1). Then $\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow{d} \mathcal{N}(0, V)$. With Assumption D.1, the center can be the fixed causal parameter θ. If the covariance weights and fitted-score error satisfy Assumption 5.2, then $\widehat{V}_n^{\mathrm{net}} \to V$ in probability. Studentized contrasts and conservative variance combinations have the coverage conclusions in Section 5 for the corresponding center.*

Proof. For fixed $c \in \mathbb{R}^{d_\theta}$, Minkowski's inequality gives $\|c^\top \phi_i\|_p \le \|c\|_1 H_\phi$. Lemma L.2 therefore applies to every contrast. Cramér–Wold and Lemma L.1 give the estimator limit. The fixed-target claim follows from Lemma D.3, which only requires bounds on conditional means.

For covariance concentration, Hölder's inequality gives

$$\mathbb{E}|\phi_i^{(a)}\phi_k^{(b)}\phi_j^{(a)}\phi_l^{(b)}| \le \|\phi_i^{(a)}\|_4\|\phi_k^{(b)}\|_4\|\phi_j^{(a)}\|_4\|\phi_l^{(b)}\|_4 \le H_\phi^4 .$$

The same bound holds for the product of the two second moments. Consequently each possibly nonzero product covariance is at most $2H_\phi^4$. The pair enumeration in (C.5) is unchanged and yields $\mathrm{Var}(Q_n^{ab}) \le 8H_\phi^4\Delta_n L_n^2/n \to 0$. The covariance-bias condition controls its expectation. Also $\mathbb{E}\|\phi\|_{2,n}^2 \le d_\theta H_\phi^2$, so Markov's inequality gives $\|\phi\|_{2,n} = \mathrm{O}_\mathbb{P}(1)$. The feasible product expansion then gives the matrix bound

$$\|\widehat{V}_n - \widetilde{V}_n\|_{\mathrm{op}} \le L_n\{2\|\phi\|_{2,n}\|\widehat{\phi} - \phi\|_{2,n} + \|\widehat{\phi} - \phi\|_{2,n}^2\} = \mathrm{o}_\mathbb{P}(1) .$$

The bound follows by applying the operator norm of W_n to the data matrices of row scores and using their Frobenius norms divided by $\sqrt{n}$. The independent-replacement argument still gives empirical score error $\mathrm{O}_\mathbb{P}(\varpi_{\psi,n} + s_n + n^{-1/2})$. Thus the sufficient score rate in the main text remains valid. ■

Remark L.1 (Compatible rates)**.** If $\Delta_n = \mathrm{O}(n^a)$, choosing $K_n = n^\kappa$ in (L.1) is possible whenever

$$0 \le a < \frac{p-2}{4p-2}, \qquad \frac{a}{p-2} < \kappa < \frac{1/2 - 2a}{3} .$$

For $p = 6$, this permits $a < 2/11$. Conditional Gaussian outcomes with bounded means and variances supply every finite p, so any fixed $a < 1/4$ can be accommodated by taking a sufficiently large moment order. For a learner with $\varpi_{\psi,n}, \varpi_m, \varpi_r, \varpi_\nu = \mathrm{O}(\sqrt{\Delta_n/n})$ and $s_n = \mathrm{O}(\Delta_n/n)$, the original learning rates hold throughout these ranges. Kernel sizes $L_n = \mathrm{O}(n^b)$ with $0 \le b < (1-a)/2$ meet concentration and fitted-score rates; the covariance bias and variance limit remain additional conditions.

These rate calculations are conditional on verifying the complete learner's mean-square prediction and replacement bounds under the unbounded-outcome law. This appendix extends the exact-dependence theory; combining unbounded outcomes with the mixing extension also requires tail bounds for its coupling errors. ♢

APPENDIX M. REGRESSION COEFFICIENTS AND ASSIGNMENT-REGIME EFFECTS

This appendix studies only the linear specifications estimated by ordinary least squares in equations (2)–(5) of Duflo et al. (2026, Section 4.3). It characterizes the probability limits of the within-village treatment coefficient, the slopes on assigned baseline friends for control and treated farmers, and the saturation contrasts. The study also reports Poisson regressions; those estimators and their interpretation are outside the scope of this appendix. The results below concern outcome differences in levels.

M.1. **Specifications and conditions for probability limits.** Let $i \in [n] = \{1, \dots, n\}$ index individuals, $j(i) \in \{1, \dots, J\}$ their villages, and $r \in \{1, \dots, K\}$ survey rounds. Village j contains n_j individuals. The random village-saturation vector is $T = (T_1, \dots, T_J)$ and the own-assignment vector is $P = (P_1, \dots, P_n)$, with $P_i \in \{0, 1\}$. Condition on a supported realization $T = t$. Let J, K be fixed, let $n_j/n \to w_j > 0$, and suppose every compared saturation satisfies $0 < s < 1$. Write $\mathbb{E}_t$ for expectation over assignment and primitive randomness conditional on $T = t$; all probability limits below use this conditional law.

Observed and potential outcomes. Apply the outcome model of Section 2 to each survey round. The controlled potential outcome $Y_{ir}^s(a, g)$ is individual i's round-r outcome when own village saturation is s, own offer is a, and exposure is set to g. It is the round-specific version of $Y_i^s(a, g)$ in Section 2.4, with the same individual's primitive shocks retained across interventions. For a full assignment vector $p \in \{0, 1\}^n$, define

$$G_i(t, p) = e_{i,n}\{A(t, p), t, p_{-i}, X^n\}, \qquad Y_{ir}(t, p) = Y_{ir}^{t_{j(i)}}\{p_i, G_i(t, p)\}.$$

Here $A(t,p)$ is the entire graph regenerated under (t,p), $e_{i,n}$ is the exposure map in (2.5), X^n is the vector of baseline covariates, and p_{-i} contains all assignments other than i's. Thus $Y_{ir}(t,p)$ allows the network to respond to the assignment. If $p^{(i,a)}$ replaces coordinate i of p by a, write

$$Y_{ir}(t,a,p_{-i}) = Y_{ir}(t,p^{(i,a)}), \qquad \overline{Y}_i(t,p) = \frac{1}{K}\sum_{1\le r\le K} Y_{ir}(t,p).$$

The observed outcome satisfies $Y_{ir} = Y_{ir}(T,P)$ by consistency, and $\overline{Y}_i = K^{-1}\sum_r Y_{ir}$ is its observed round average. Comparing $Y_{ir}(t,1,P_{-i})$ and $Y_{ir}(t,0,P_{-i})$ holds the other offers and primitive shocks fixed and regenerates the graph in each world.

Averaging convention. Let I and R be independent uniform indices on $[n]$ and $\{1,\dots,K\}$, independent of all model variables. The record-average expectation is

$$\mathsf{E}_{t,n}h = \frac{1}{nK}\sum_{1\le i\le n}\sum_{1\le r\le K}\mathbb{E}_t[h(i,r)] = \mathbb{E}_t[h(I,R)].$$

These weights describe a complete household–round panel. Other observation units or missing rounds require the corresponding record weights throughout; in particular, the source study's tree-level audit regressions need not have these household weights. In conditional means below, unindexed Y and P denote Y_{IR} and P_I.

Let x_i contain prespecified pretreatment controls, and write

$$F_i^0 = \sum_{k\in\mathcal{R}\setminus\{i\}} A_{ik}^{\mathrm{pre}}P_k, \qquad d_i^0 = \sum_{k\in\mathcal{R}\setminus\{i\}} A_{ik}^{\mathrm{pre}}, \qquad H_{is} = \mathbb{1}\{t_{j(i)} = s\},$$

where $\mathcal{R}$ is the randomized roster and A_{ik}^{pre} is the baseline nomination indicator. Thus F_i^0 counts assigned baseline friends and d_i^0 counts all baseline friends in the roster; unindexed F^0, d^0 denote F_I^0, d_I^0. Set $\mathcal{S} = \{0.25, 0.50, 0.75\}$, $s_L = 0.25$, $s_H = 0.75$, and $\mathcal{S}_+ = \mathcal{S}\setminus\{s_L\}$. The four linear specifications are

(M.1) $$Y_{ir} = a_{j(i)}^{(2)} + b_r^{(2)} + x_i^\top\eta^{(2)} + \beta^{(2)}P_i + \varepsilon_{ir}^{(2)},$$

(M.2) $$Y_{ir} = a_{j(i)}^{(3)} + b_r^{(3)} + x_i^\top\eta^{(3)} + \alpha^{(3)}P_i + \beta_0^{(3)}F_i^0 + \delta_0^{(3)}d_i^0 + \beta_1^{(3)}P_iF_i^0 + \delta_1^{(3)}P_id_i^0 + \varepsilon_{ir}^{(3)},$$

(M.3) $$Y_{ir} = a_0^{(4)} + b_r^{(4)} + x_i^\top\eta^{(4)} + \sum_{s\in\mathcal{S}_+}\beta_s^{(4)}H_{is} + \varepsilon_{ir}^{(4)},$$

(M.4) $$Y_{ir} = a_0^{(5)} + b_r^{(5)} + x_i^\top\eta^{(5)} + \alpha^{(5)}P_i + \sum_{s\in\mathcal{S}_+}\{\beta_s^{(5)}H_{is} + \gamma_s^{(5)}P_iH_{is}\} + \varepsilon_{ir}^{(5)}.$$

The superscript $q\in\{2,3,4,5\}$ is the source equation number. The coefficients $a_j^{(q)}$ and $b_r^{(q)}$ are village and round intercepts, $\eta^{(q)}$ multiplies the baseline controls, and $\varepsilon_{ir}^{(q)}$ is a regression residual. Village and round indicators use a full-rank normalization; village effects are omitted from the saturation regressions because they would absorb saturation. Controls can differ across specifications. A hat denotes an estimated coefficient, and the corresponding unaccented symbol below denotes its probability limit.

Assumption M.1 (Convergence of regression moments)**.** For each linear specification, let $X_{ir}^{(q)}$ be its fixed-dimensional regressor vector after redundant columns are removed. Outcomes and regressors have finite second moments, and

$$\widehat{\Gamma}_{q,n} = \frac{1}{nK}\sum_{1\le i\le n}\sum_{1\le r\le K} X_{ir}^{(q)}X_{ir}^{(q)\top} \xrightarrow{\mathbb{P}} \Gamma_q \succ 0,$$

$$\widehat{h}_{q,n} = \frac{1}{nK} \sum_{1 \le i \le n} \sum_{1 \le r \le K} X_{ir}^{(q)} Y_{ir} \xrightarrow{\mathbb{P}} h_q.$$

The expectations of these empirical moments converge to the same Γ_q and h_q.

This is the law-of-large-numbers requirement that connects the regression calculations to the estimators. It holds, for example, when the per-individual regressor products and regressor–outcome products have uniformly bounded second moments, obey a dependency graph with $\Delta_n/n \to 0$, and their mean moments converge with a nonsingular limiting Gram matrix. The relevant graph must cover baseline friend counts and every regressor as well as outcomes. The source study selects controls by post-double-selection LASSO; applying these limits to selected specifications additionally requires that selection have the same limiting coefficient behavior as the prespecified regression under consideration.

Natural regime means and components. For comparison with Section 2.4, use its natural exposure $G_i^s(b)$. This is the exposure generated by setting individual i's village saturation to s and own offer to b, drawing the other village saturations from the design law conditional on that village receiving s, drawing peer offers under that profile, and regenerating the full graph. The same peer draws and primitive shocks are used across the two values of b. Define the round-averaged natural means by

$$\theta_{s,n}^{\mathrm{nat}}(a,b) = \frac{1}{nK} \sum_{1 \le i \le n} \sum_{1 \le r \le K} \mathbb{E}[Y_{ir}^s\{a, G_i^s(b)\}], \qquad R_{s,n}(a) = \theta_{s,n}^{\mathrm{nat}}(a,a).$$

The expectation here averages the primitives and the stated regime intervention, including the external saturation draw; it does not hold the entire profile at t. The two arguments a, b respectively set own assignment in the controlled response and in the network-generating world. Let $\theta_s(a,b)$ and $R_s(a)$ denote the limits of these finite-n means, assumed to exist when used, and set

$$\begin{aligned} D_s &= \theta_s(1,0) - \theta_s(0,0), \\ N_s &= \theta_s(0,1) - \theta_s(0,0), \\ I_s &= \theta_s(1,1) - \theta_s(1,0) - \theta_s(0,1) + \theta_s(0,0). \end{aligned}$$

These are the direct, network-mediated, and treatment–network interaction components. Their definitions imply

$$\text{(M.5)} \qquad W_s = R_s(1) - R_s(0) = D_s + N_s + I_s, \qquad S_s = R_s(0) - R_{s_L}(0), \qquad V(s) = R_s(0) + sW_s.$$

Here W_s is the within-saturation total, S_s the saturation spillover on untreated individuals, and $V(s)$ the saturation-policy mean. The headline total is $\mathrm{TE} = R_{s_H}(1) - R_{s_L}(0)$. The diagonal equality and the telescoping identity for W_s follow from the definitions. Identifying these natural components from observed outcomes is a separate question addressed in Section 3.

M.2. **The within-village treatment coefficient.** The coefficient read from source equation (2) compares treated and control farmers within villages. Its probability limit is a treatment-variance-weighted average of those differences.

Proposition M.1 (Within-village coefficient)**.** *Under Assumptions 2.1 and M.1, define*

$$r_{ja} = \lim_{n \to \infty} \frac{1}{n_j} \sum_{i : j(i) = j} \mathbb{E}_t[\overline{Y}_i \mid P_i = a], \qquad v_j = t_j(1 - t_j),$$

and suppose these means exist. Then

$$\widehat{\beta}^{(2)} \xrightarrow{\mathbb{P}} \beta^{(2)} = \frac{\sum_{j=1}^{J} w_j v_j (r_{j1} - r_{j0})}{\sum_{j=1}^{J} w_j v_j}. \tag{M.6}$$

If $r_{ja} = R_{t_j}(a)$, let $\lambda_s = \sum_{j:t_j=s} w_j$. The limit becomes

$$\beta^{(2)} = \sum_{s\in\mathcal{S}} \omega_s W_s = \sum_{s\in\mathcal{S}} \omega_s (D_s + N_s + I_s), \qquad \omega_s = \frac{\lambda_s s(1-s)}{\sum_{u\in\mathcal{S}} \lambda_u u(1-u)}. \tag{M.7}$$

For a complete balanced panel without additional controls, the same coefficient also has the exact sample representation

$$\widehat{\beta}^{(2)} = \frac{\sum_{j=1}^{J} n_j \widehat{p}_j (1-\widehat{p}_j)(\overline{Y}_{j1} - \overline{Y}_{j0})}{\sum_{j=1}^{J} n_j \widehat{p}_j (1-\widehat{p}_j)}, \qquad \widehat{p}_j = \frac{1}{n_j} \sum_{i:j(i)=j} P_i. \tag{M.8}$$

Here $\overline{Y}_{ja}$ is the village–arm mean of $\overline{Y}_i$, and the denominator must be positive. A pure-arm village contributes zero before any empty-arm mean is evaluated.

The equality $r_{ja} = R_{t_j}(a)$ connects the realized villages to the common target population; the common response, exposure, and baseline laws in Assumption 4.1, together with Assumptions 2.1 and 2.2, supply this equality when applied to the round-averaged outcome. Without it, (M.6) still gives the probability limit for the realized village population.

The source paper describes this coefficient as the within-village treatment–control difference. Under the Bernoulli assignment model here, its total-effect interpretation lets the network respond to own assignment. It therefore contains D_s, N_s, and I_s. For equal population shares in the three saturation arms, the weights are $(0.30, 0.40, 0.30)$. Additional controls preserve the probability-limit formula under the stated assumptions, although they generally change the exact sample formula.

M.3. **Assigned baseline friends: control and treated farmers.** For source equation (3), the reported diffusion coefficients are $\widehat{\beta}_0^{(3)}$ for control farmers and $\widehat{\beta}_0^{(3)} + \widehat{\beta}_1^{(3)}$ for treated farmers. To interpret their limits as changes in mean outcomes, the linear friend-count specification must describe those conditional means.

Proposition M.2 (Friend-count coefficients)**.** *Suppose Assumption M.1 holds and, at each n, the conditional mean m_n defined below is correctly specified by (M.2). Let $C = (j(I), R, x_I)$ collect village, round, and the included pretreatment controls, and define, on the compared support,*

$$m_n(a, f, d, c) = \mathsf{E}_{t,n}[Y_{IR} \mid P_I = a, F_I^0 = f, d_I^0 = d, C = c], \qquad m(a, f, d, c) = \lim_{n\to\infty} m_n(a, f, d, c).$$

For $a = 0, 1$, define

$$\Delta_a = m(a, f+1, d, c) - m(a, f, d, c)$$

at supported adjacent counts. Then Δ_a is constant across these (f, d, c), and

$$\widehat{\beta}_0^{(3)} \xrightarrow{\mathbb{P}} \Delta_0, \qquad \widehat{\beta}_0^{(3)} + \widehat{\beta}_1^{(3)} \xrightarrow{\mathbb{P}} \Delta_1, \qquad \widehat{\beta}_1^{(3)} \xrightarrow{\mathbb{P}} \Delta_1 - \Delta_0. \tag{M.9}$$

The fitted own-assignment contrast converges, at fixed supported (f, d, c), to

$$\widehat{\alpha}^{(3)} + \widehat{\beta}_1^{(3)} f + \widehat{\delta}_1^{(3)} d \xrightarrow{\mathbb{P}} m(1, f, d, c) - m(0, f, d, c). \tag{M.10}$$

These are comparisons along the baseline network at fixed baseline degree. Under Assumption 2.1 and consistency, $m(a, f, d, c)$ averages the full potential outcome with own offer set to a under the peer-assignment law

conditional on $F^0 = f, d^0 = d, C = c$. Thus increasing f changes the assignment mix among existing friends. It does not isolate the effect of relationships created by own treatment, and neither Δ_0 nor Δ_1 is the component N_s. The own-assignment intercept $\alpha^{(3)}$ concerns baseline isolates when $f = d = 0$ is supported; otherwise it is an extrapolation.

Correct specification is needed for the common increment in (M.9). Without it, Assumption M.1 still gives the projection limit $\Gamma_3^{-1} h_3$, whose corresponding entries and sums are exactly the limits of the reported coefficients. They need not equal a common conditional-mean increment.

M.4. **Saturation contrasts.** Source equation (4) compares all farmers across saturation arms. Equation (5) separates this comparison by own assignment, using untreated farmers in the 25% arm as its reference cell.

Proposition M.3 (Saturation coefficients)**.** *Suppose Assumptions 2.1 and M.1 hold. Suppose the observed assignment-cell means converge to $R_s(a)$ and the mean vectors of the included round and baseline controls converge to a common vector across compared cells. Then*

$$\widehat{\beta}_s^{(4)} \xrightarrow{\mathbb{P}} V(s) - V(s_L) = S_s + sW_s - s_L W_{s_L}, \qquad s \in \mathcal{S}_+. \tag{M.11}$$

For equation (M.4),

$$\widehat{\alpha}^{(5)} \xrightarrow{\mathbb{P}} W_{s_L}, \qquad \widehat{\beta}_s^{(5)} \xrightarrow{\mathbb{P}} S_s, \qquad \widehat{\gamma}_s^{(5)} \xrightarrow{\mathbb{P}} W_s - W_{s_L}. \tag{M.12}$$

Consequently, the reported sums have limits

$$\begin{aligned} \widehat{\alpha}^{(5)} + \widehat{\gamma}_s^{(5)} &\xrightarrow{\mathbb{P}} W_s, \\ \widehat{\beta}_s^{(5)} + \widehat{\gamma}_s^{(5)} &\xrightarrow{\mathbb{P}} R_s(1) - R_{s_L}(1), \\ \widehat{\alpha}^{(5)} + \widehat{\beta}_s^{(5)} + \widehat{\gamma}_s^{(5)} &\xrightarrow{\mathbb{P}} R_s(1) - R_{s_L}(0). \end{aligned} \tag{M.13}$$

The last limit is TE *at $s = s_H$.*

Thus $\beta_s^{(4)}$ is the policy contrast, $\beta_s^{(5)}$ is the saturation spillover on controls, and $\gamma_s^{(5)}$ is the change in the within-saturation treatment effect. Substituting $W_s = D_s + N_s + I_s$ gives their decomposition. In particular, the regression interaction $\gamma_s^{(5)}$ is a difference of total effects, rather than the treatment–network interaction I_s alone.

The common-mean premise is a population condition, not exact balance in a realized sample. Assumption 4.1 with common round weights and controls drawn from B supplies it and the cell-mean alignment. If mean controls differ across arms, the probability limits instead compare adjusted cell means. Specifically, for each saturation regression let C collect its round and baseline-control columns, excluding the intercept and displayed assignment indicators. With $S^\circ = t_{j(I)}$, define

$$\begin{aligned} m_s &= \lim_n \mathsf{E}_{t,n}[Y_{IR} \mid S^\circ = s], \quad m_{sa} = \lim_n \mathsf{E}_{t,n}[Y_{IR} \mid S^\circ = s, P_I = a], \\ c_s &= \lim_n \mathsf{E}_{t,n}[C \mid S^\circ = s], \qquad c_{sa} = \lim_n \mathsf{E}_{t,n}[C \mid S^\circ = s, P_I = a]. \end{aligned}$$

Assume these limits exist when used. If ζ_q is the limiting coefficient vector on C in specification q, define

$$d_s^{(4)} = m_s - c_s^\top \zeta_4, \qquad d_{sa}^{(5)} = m_{sa} - c_{sa}^\top \zeta_5. \tag{M.14}$$

The same coefficient differences hold with $V(s)$ replaced by $d_s^{(4)}$ and $R_s(a)$ by $d_{sa}^{(5)}$. These are projection limits even when the outcome model is misspecified.

M.5. **Proofs.**

From sample moments to coefficient limits. For each specification, least squares gives $\widehat{b}_q = \widehat{\Gamma}_{q,n}^{-1}\widehat{h}_{q,n}$ whenever the Gram matrix is invertible. Put $c = \lambda_{\min}(\Gamma_q) > 0$. On the event $\|\widehat{\Gamma}_{q,n} - \Gamma_q\|_{\mathrm{op}} < c/2$, which has probability tending to one,

$$\lambda_{\min}(\widehat{\Gamma}_{q,n}) \geq c/2, \qquad \|\widehat{\Gamma}_{q,n}^{-1}\|_{\mathrm{op}} \leq 2/c.$$

The inverse identity and triangle inequality give

$$\widehat{\Gamma}_{q,n}^{-1} - \Gamma_q^{-1} = \widehat{\Gamma}_{q,n}^{-1}(\Gamma_q - \widehat{\Gamma}_{q,n})\Gamma_q^{-1},$$
$$\|\widehat{b}_q - \Gamma_q^{-1}h_q\| \leq \frac{2}{c}\|\widehat{h}_{q,n} - h_q\| + \frac{2\|h_q\|}{c^2}\|\widehat{\Gamma}_{q,n} - \Gamma_q\|_{\mathrm{op}} = o_{\mathbb{P}}(1).$$

The population moment limits in Assumption M.1 imply that the finite-n population projection coefficients have the same limit. It therefore suffices to compute those population coefficients. For the stated dependency-graph sufficient condition, each centered scalar moment contribution Z_i with $\sup_i \mathbb{E}_t Z_i^2 \leq C$ satisfies

$$\mathrm{Var}_t\left(\frac{1}{n}\sum_i Z_i\right) \leq \frac{1}{n^2}\sum_i \sum_{k\in N[i]} |\mathrm{Cov}_t(Z_i, Z_k)| \leq \frac{C\Delta_n}{n} \longrightarrow 0.$$

Chebyshev's inequality and the fixed number of moment entries give the required convergence.

Proof of Proposition M.1. Under Assumption 2.1,

$$\mathbb{E}_t[P_i \mid x_i, j(i), r] = t_{j(i)}.$$

Randomization and consistency also give the potential-outcome representation

$$\mathbb{E}_t[\overline{Y}_i \mid P_i = a] = \frac{1}{K}\sum_{1\leq r\leq K} \mathbb{E}_t[Y_{ir}(t, a, P_{-i})].$$

The fitted value $t_{j(i)}$ is in the village-indicator span. The residual $P_i - t_{j(i)}$ is orthogonal to all village, round, and pretreatment-control columns. For i in village j, writing $s = t_j$ and splitting over the two assignment values yields

$$\mathbb{E}_t[(P_i - s)\overline{Y}_i] = s(1-s)\{\mathbb{E}_t[\overline{Y}_i \mid P_i = 1] - \mathbb{E}_t[\overline{Y}_i \mid P_i = 0]\},$$
$$\mathbb{E}_t[(P_i - s)^2] = s(1-s) = v_j.$$

Consequently, the finite-n population treatment coefficient is

$$\frac{\sum_j (n_j/n)v_j(r_{j1,n} - r_{j0,n})}{\sum_j (n_j/n)v_j}, \qquad r_{ja,n} = \frac{1}{n_j}\sum_{i:j(i)=j} \mathbb{E}_t[\overline{Y}_i \mid P_i = a].$$

This is the Frisch–Waugh–Lovell residualization calculation (Frisch and Waugh, 1933, Lovell, 1963). Taking limits and applying the preceding moment argument proves (M.6). Under $r_{ja} = R_{t_j}(a)$,

$$\sum_j w_j v_j (r_{j1} - r_{j0}) = \sum_s \lambda_s s(1-s)\{R_s(1) - R_s(0)\},$$
$$\sum_j w_j v_j = \sum_s \lambda_s s(1-s) > 0.$$

Division and the defining identity for W_s give (M.7).

For the sample identity, let $\mathcal{I}_j = \{i : j(i) = j\}$ and $\widetilde{P}_{ir} = P_i - \widehat{p}_{j(i)}$. The residual is orthogonal to village and round indicators because

$$\sum_{1\le r\le K}\sum_{i\in\mathcal{I}_j}\widetilde{P}_{ir} = K\left(\sum_{i\in\mathcal{I}_j}P_i - n_j\widehat{p}_j\right) = 0,$$

$$\sum_{1\le i\le n}\widetilde{P}_{ir} = \sum_j\left(\sum_{i\in\mathcal{I}_j}P_i - n_j\widehat{p}_j\right) = 0.$$

For $0 < \widehat{p}_j < 1$, split the outcome sum by assignment arm:

$$\begin{aligned}\sum_{i\in\mathcal{I}_j}(P_i-\widehat{p}_j)\overline{Y}_i &= (1-\widehat{p}_j)n_j\widehat{p}_j\overline{Y}_{j1} - \widehat{p}_j n_j(1-\widehat{p}_j)\overline{Y}_{j0}\\ &= n_j\widehat{p}_j(1-\widehat{p}_j)(\overline{Y}_{j1}-\overline{Y}_{j0}),\\ \sum_{i\in\mathcal{I}_j}(P_i-\widehat{p}_j)^2 &= n_j\widehat{p}_j(1-\widehat{p}_j)^2 + n_j(1-\widehat{p}_j)\widehat{p}_j^2\\ &= n_j\widehat{p}_j(1-\widehat{p}_j).\end{aligned}$$

When $\widehat{p}_j$ is zero or one, every residual in that village is zero. Its contribution vanishes without evaluating an empty-arm mean. Finally,

$$\widehat{\beta}^{(2)} = \frac{\sum_i\sum_r\widetilde{P}_{ir}Y_{ir}}{\sum_i\sum_r\widetilde{P}_{ir}^2} = \frac{K\sum_j n_j\widehat{p}_j(1-\widehat{p}_j)(\overline{Y}_{j1}-\overline{Y}_{j0})}{K\sum_j n_j\widehat{p}_j(1-\widehat{p}_j)}.$$

Cancelling K proves (M.8). ■

Proof of Proposition M.2. Correct conditional-mean specification and nonsingularity make the population projection coefficients equal to the conditional-mean coefficients. By the moment argument, their limits give

$$m(a,f,d,c) = \ell(c) + \alpha^{(3)}a + \beta_0^{(3)}f + \delta_0^{(3)}d + \beta_1^{(3)}af + \delta_1^{(3)}ad,$$

where ℓ is the limiting predictor from the village, round, and baseline controls. Holding d, c fixed gives

$$\begin{aligned}m(0,f+1,d,c) - m(0,f,d,c) &= \beta_0^{(3)},\\ m(1,f+1,d,c) - m(1,f,d,c) &= \beta_0^{(3)} + \beta_1^{(3)},\\ m(1,f,d,c) - m(0,f,d,c) &= \alpha^{(3)} + \beta_1^{(3)}f + \delta_1^{(3)}d.\end{aligned}$$

Coefficient convergence and subtraction prove both assertions.

For the potential-outcome interpretation, use the round-specific outcome $Y_{ir}(t,a,P_{-i})$ defined above. Own Bernoulli assignment is independent of the joint collection of other assignment draws, baseline quantities, and potential outcomes. Conditional on the village in C, its probability is constant across the individuals in the stratum. Consistency therefore gives, for each supported stratum,

$$\begin{aligned}&\mathsf{E}_{t,n}[Y \mid P = a, F^0 = f, d^0 = d, C = c]\\ &\quad= \mathsf{E}_{t,n}[Y_{IR}(t,a,P_{-I}) \mid F^0 = f, d^0 = d, C = c].\end{aligned}$$

Both sides use the same independently selected individual I and survey round R. The conditional peer-assignment distribution is the same for the two values of own offer, and varies when the conditioned friend count changes. Taking the limits in the proposition gives the stated interpretation. ■

Proof of Proposition M.3. Represent equation (M.3) with one intercept for each saturation and equation (M.4) with one intercept for each saturation–assignment cell. Give each cell mean and control mean in (M.14) an n subscript to denote its finite-n version. For finite-n population projection coefficients, their normal equations are

$$\begin{aligned} 0 &= \lambda_{s,n}\{m_{s,n} - d^{(4)}_{s,n} - c^{\top}_{s,n}\zeta_{4,n}\}, \\ 0 &= \lambda_{s,n}s^{a}(1-s)^{1-a}\{m_{sa,n} - d^{(5)}_{sa,n} - c^{\top}_{sa,n}\zeta_{5,n}\}. \end{aligned}$$

All compared cell probabilities are positive. Solving and taking limits proves (M.14). The coefficient parameterization is

$$\begin{aligned} d^{(4)}_{s_L} &= a^{(4)}_0, & d^{(4)}_{s} &= a^{(4)}_0 + \beta^{(4)}_s, \\ d^{(5)}_{s_L0} &= a^{(5)}_0, & d^{(5)}_{s_L1} &= a^{(5)}_0 + \alpha^{(5)}, \\ d^{(5)}_{s0} &= a^{(5)}_0 + \beta^{(5)}_s, & d^{(5)}_{s1} &= a^{(5)}_0 + \beta^{(5)}_s + \alpha^{(5)} + \gamma^{(5)}_s. \end{aligned}$$

Subtracting these intercepts gives

$$\begin{aligned} &\beta^{(4)}_s = d^{(4)}_s - d^{(4)}_{s_L}, \qquad \alpha^{(5)} = d^{(5)}_{s_L1} - d^{(5)}_{s_L0}, \qquad \beta^{(5)}_s = d^{(5)}_{s0} - d^{(5)}_{s_L0}, \\ &\gamma^{(5)}_s = (d^{(5)}_{s1} - d^{(5)}_{s0}) - (d^{(5)}_{s_L1} - d^{(5)}_{s_L0}). \end{aligned}$$

Under the common-mean premise, write c_q for the limiting control mean in specification q. The assignment law and cell-mean alignment imply

$$\begin{aligned} m_s &= sR_s(1) + (1-s)R_s(0) = V(s), \\ d^{(4)}_s &= V(s) - c_4^{\top}\zeta_4, \qquad d^{(5)}_{sa} = R_s(a) - c_5^{\top}\zeta_5. \end{aligned}$$

The common offsets cancel. Hence

$$\begin{aligned} \beta^{(4)}_s &= V(s) - V(s_L) = \{R_s(0) + sW_s\} - \{R_{s_L}(0) + s_LW_{s_L}\} \\ &= S_s + sW_s - s_LW_{s_L}, \\ \alpha^{(5)} &= R_{s_L}(1) - R_{s_L}(0) = W_{s_L}, \qquad \beta^{(5)}_s = R_s(0) - R_{s_L}(0) = S_s, \\ \gamma^{(5)}_s &= \{R_s(1) - R_s(0)\} - \{R_{s_L}(1) - R_{s_L}(0)\} = W_s - W_{s_L}. \end{aligned}$$

The moment argument transfers these identities to the probability limits of the sample coefficients. The three reported sums follow from

$$\begin{aligned} \alpha^{(5)} + \gamma^{(5)}_s &= W_s, \\ \beta^{(5)}_s + \gamma^{(5)}_s &= R_s(0) - R_{s_L}(0) + W_s - W_{s_L} = R_s(1) - R_{s_L}(1), \\ \alpha^{(5)} + \beta^{(5)}_s + \gamma^{(5)}_s &= R_s(0) - R_{s_L}(0) + W_s = R_s(1) - R_{s_L}(0). \end{aligned}$$

■

701 TAPPAN AVENUE, ROSS SCHOOL OF BUSINESS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109
Email address: rexhsieh@umich.edu