Yechan Park
Department of Economics, Harvard University
Littauer Center, 1805 Cambridge Street
Cambridge, MA 02138
Phone: (617) 256-3147
Email: yechanpark@fas.harvard.edu

September 15, 2026

Professors Emily Breza, Arun Chandrasekhar, and Davide Viviano
Organizers, NBER Workshop on Development Economics and Networks
for Early-Career Researchers
National Bureau of Economic Research
Cambridge, MA

Dear Professors Breza, Chandrasekhar, and Viviano,

I am writing to apply to the NBER Workshop on Development Economics and Networks for Early-Career Researchers, to be held in Cambridge on December 5, 2026. I am a second-year PhD student in economics at Harvard University. My research develops econometric methods for causal inference in the presence of spillovers, with an emphasis on network and spatial interference in development settings.

I would like to present my paper "A Design-Based Approach to Testing and Inference in (Quasi-)Experiments with Spillovers." The paper studies how to choose and validate the exposure measures that applied researchers use to summarize who else was treated, and it applies the method to two large-scale anti-poverty programs. The most recent version, dated August 18, 2026, is included in this packet and is also available on arXiv (arXiv:2607.08640). The paper is currently under journal review.

This packet contains, in order: this cover letter, my curriculum vitae with a list of current and completed PhD courses, a one-page statement of motivation and research interests, and the working paper.

If I am shortlisted, the following faculty members could provide a recommendation:

- Davide Viviano, Assistant Professor of Economics, Harvard University
  dviviano@fas.harvard.edu
- Elie Tamer, Louis Berkman Professor of Economics, Harvard University
  elietamer@fas.harvard.edu

Thank you for considering my application.

Sincerely,

Yechan Park

Yechan Park

Curriculum Vitae

Department of Economics, Harvard University
Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138
yechanpark@fas.harvard.edu · (617) 256-3147

Education

- **2025–Present**: PhD in Economics, *Harvard University* (second year)
- **April–July 2023**: Master's Program in Statistics, *Graduate School of Economics, The University of Tokyo* (graduate coursework; left to join Opportunity Insights)
- **March 2023**: B.A. in Economics, *Faculty of Economics, Department of Finance, The University of Tokyo*
  Thesis: *Three Essays in Causal Inference with Combined Experimental and Observational Data*
  Advised by Hidehiko Ichimura; Distinguished Thesis Award

PhD Coursework

Harvard University (Doctor of Philosophy, Economics)

- **Fall 2025 (completed)**: ECON 2010A Economic Theory (Microeconomics); ECON 2010C Economic Theory (Macroeconomics); ECON 2340 Spatial Mobility and Development: Evidence and Quantitative Models
- **Spring 2026 (completed)**: ECON 2010B Economic Theory (Microeconomics); ECON 2010D Economic Theory (Macroeconomics); ECON 2140 Econometric Methods
- **Fall 2026 (current)**: ECON 2610 Industrial Organization I; ECON 2810A Labor Market Analysis; STAT 210 Probability I; STAT 211 Statistical Inference I

The University of Tokyo, Graduate School of Economics (Master's Program, April–July 2023; completed)

- Microeconomics I; Econometrics II; Advanced Mathematical Statistics I; Probability Theory I; Advances in Economic Theory; Advanced Time Series Analysis; Macroeconomic Dynamics; Empirical Analysis of Dynamic Macroeconomic Models; Topics in Macroeconomic Theory; Economics of Education; Machine Learning for Applied Economics; Data Science and Machine Learning; Data Science for Practical Economic Research; Topics in Labor Economics; Quantitative Economic Analysis; Tutorial on Real Analysis I; Advanced Lectures in Biostatistics and Bioinformatics VII
- *Graduate-level courses taken as an undergraduate (2021–2023)*: Mathematics for Economics; Advanced Econometrics II; Advanced Microeconomics I; Advanced Labor Economics; Advanced Time Series Analysis; Data Science and Machine Learning; Probability for Finance I; Risk Management for Financial Institutions

Employment

- **August 2023–July 2025**: Predoctoral Research Fellow, *Opportunity Insights, Harvard University*

– **Mentors**: Raj Chetty (main), Susan Athey, Isaiah Andrews, and Guido Imbens

Research Interests

- Econometric theory; causal inference with spillovers, networks, and spatial interference; applied econometrics in development and labor economics

Working Papers

- **Park, Y.** (2026). "A Design-Based Approach to Testing and Inference in (Quasi-)Experiments with Spillovers." *Under review.* Link.
- **Park, Y., and Yang, X.** (2026). "Decomposition of Spillover Effects Under Misspecification: Pseudo-true Estimands and a Local–Global Extension." *arXiv preprint.* Link.
- **Park, Y., and Sasaki, Y.** (2024). "A Bracketing Relationship for Long-Term Policy Evaluation with Combined Experimental and Observational Data." *Revise and resubmit at Journal of Applied Econometrics* (revision submitted August 2026). Link.
- **Lim, C., and Park, Y.** (2026). "Dynamically Consistent Statistical Decisions." *arXiv preprint.* Link.
- **Park, Y., and Sasaki, Y.** (2026). "Choosing A Headline Estimand from Matching, DID, and Hybrid Designs: A Minimax-Regret Approach." *Under review.* Link.

Published Papers

- **Park, Y., and Sasaki, Y.** (2026). "Informativeness of Combined Experimental and Observational Data under Dynamic Selection." *Journal of Econometrics*, 254(B), 106219. Link.
- **Park, Y., and Yoshida, N.** (2024). "Asymptotic Expansion for Batched Bandits." *Statistical Inference for Stochastic Processes* (conditionally accepted). Link.
- **Park, Y., Yoshida, N., and Zhan, R.** (2022). "Beyond Central Limit Theorem for Higher Order Inference in Batched Bandits." *Neural Information Processing Systems 2022 Causal Machine Learning for Real World Impact Workshop.* Link.

Ongoing Research (Drafts Available Upon Request)

- **Andrews, I., Chetty, R., Park, Y., and Sanders, B.** "Weighing Experimental vs. Observational Evidence: Decision-Relevant Summaries of Treatment Effect Heterogeneity."

Presentations

- **EcoSta 2026** (10th International Conference on Econometrics and Statistics), Ryukoku University, Kyoto, August 2026
- **New England Economics PhD Conference (NEEPC) 2026**, Brown University, May 2026
- **Harvard Econometrics Workshop**, Harvard University, 2026
- **Neural Information Processing Systems (NeurIPS)**, Poster Presentation, December 2022

Awards & Grants

- **2026**: Harvard Griffin Fund in Economics ($9,500)
- **2023**: Scholarship for Short-Term Studies, Friends of UTokyo, Inc. ($8,000)

- **2023**: Distinguished Senior Thesis, The University of Tokyo
- **2020**: Champion, Asian Parliamentary Debating Tournament Asian Bridge (Individual Fourth Best Speaker)
- **2019**: Champion, Japan British Parliamentary Debate Novice Tournament Japan East (Individual Fourth Best Speaker)

Teaching and Professional Experience

- **Core Econometrics 2 (Graduate) / Advanced Econometrics 2 (Undergraduate), The University of Tokyo**
  Role: Teaching Assistant; Instructor: Hidehiko Ichimura
  Year: 2022 (Summer Semester)
- **Matsuo Lab, School of Engineering, The University of Tokyo**
  Role: Core Developer of the PyTorch library Pixyz (deep generative models)
  Year: 2020
- **Mainichi Newspapers, Social Research Center**
  Role: Data Scientist (survey design, big data analytics)
  Year: 2020 (Summer)
- **Referee service**: Journal of Applied Econometrics

Skills

- **Programming:** Python (Advanced), R (Advanced), Matlab (Intermediate), JavaScript (Intermediate), Stata (Intermediate)
- **Languages:** Japanese (Native), Korean (Native), English (Fluent)

Statement of Motivation and Research Interests

Yechan Park, Department of Economics, Harvard University

Research interests. My research develops econometric tools for settings in which a policy affects people other than its direct recipients. In development economics such spillovers are the rule rather than the exception: cash transfers change local prices and labor markets, information travels through social networks, and program benefits diffuse across villages. Empirical work handles this by summarizing who else was treated with an exposure measure, such as the share of treated neighbors within a radius or the number of treated network contacts. My work asks what these summaries identify, how the data can be used to choose among them, and how to conduct valid inference once dependence across units is taken seriously. Each of my projects is built around an existing field experiment and aims to change what an applied researcher would report.

Design-based testing and inference with spillovers. The paper I propose to present starts from the observation that economic theory rarely pins down the functional form of an exposure measure or its parameters, such as the radius of a ring. I show that a correctly specified exposure measure implies orthogonality conditions between the exposure and the assignment mechanism, and that these conditions can be used both to estimate the parameters and to test the specification. I establish consistency and asymptotic normality of the resulting estimator under spatial and network dependence in a design-based framework, in which the only source of randomness is the treatment assignment, and I characterize the efficient set of moment conditions. Applied to the general-equilibrium cash-transfer experiment of Egger et al. (2022) in Kenya and the program of Muralidharan et al. (2023) in India, the framework supports some of the previously reported radii and rejects others, and in the latter case the revised radius changes the estimated policy effects substantively. The procedure adds little to a standard ring analysis, and the inference accounts for the dependence that networks and geography create.

Local and global spillovers under misspecification. A companion paper with Xiaodong Yang asks what a researcher recovers when the exposure summary is wrong. We show that the estimand is the best approximation to the true policy effect among all functionals of the chosen summary, and that the conventional direct and spillover coefficients are exactly the components of that approximation. We then study a setting common in development applications: treatment spills over both globally, through market equilibrium, and locally, through network externalities, while the researcher models only one channel. Under conditions we characterize, the network estimator still recovers the network spillover and the equilibrium estimator still recovers the equilibrium spillover, so that partial models remain interpretable.

Motivation for attending. My methods are only as useful as the field studies they serve, and the questions I find hardest are the ones that arise when a design meets a real village or a real network: which channel of spillover a program plausibly operates through, what network data can and cannot be collected, and which estimands matter for the policy at hand. The workshop brings together exactly the researchers who confront these questions, and its emphasis on networks matches the direction in which my work is moving, from spatial exposure toward network exposure and its interaction with market equilibrium. I would value the chance to present my paper to a development audience, to learn from the experimental and network work of other participants, and to find collaborators for the applied side of this agenda.

arXiv:2607.08640v3 [econ.EM] 19 Aug 2026

A Design-Based Approach to Testing and Inference in (Quasi-)Experiments with Spillovers*

Yechan Park†

First version: July 9, 2026
This version: August 18, 2026

Abstract

Economic policies rarely affect only their direct targets. To study these spillovers, researchers summarize who else was treated with a simple exposure measure, such as the share of treated neighbors within a radius. But for many settings, economic theory provides little guidance on choosing the functional form (e.g., ring) of that measure or its parameters (e.g., radius). We show that the data can inform both choices. Correctly specified exposure measures imply orthogonality conditions that can be used for both estimation and testing. We establish consistency and asymptotic normality of the resulting estimator under spatial and network dependence in a design-based framework, with all randomness arising from treatment assignment. We then characterize the efficient moment conditions. Applied to two large-scale anti-poverty programs, the framework supports some prior radius estimates but rejects others. In the latter case, the revised radius yields substantively different policy-effect estimates.

Keywords: spillovers; exposure mappings; design-based inference; specification testing; spatial and network dependence; general equilibrium effects

*I thank Isaiah Andrews, Jesse Shapiro, Elie Tamer, and Davide Viviano for guidance and generous support. I also thank Raj Chetty for early feedback and encouragement. I am grateful to the authors of Egger et al. (2022) and Muralidharan et al. (2023), especially Ted Miguel and Paul Niehaus, for providing additional data and guidance for replication. I also benefited from helpful comments by Aristotelis Epanomeritakis, Tilman Graff, Gabriel Kreindler, Andreas Petrou-Zeniou, Yuya Sasaki, Neil Shephard, Rahul Singh, and Shengbin Wei, as well as from discussions with participants at the Harvard Econometrics Workshop and NEEPC 2026. I acknowledge support from the Harvard Griffin Fund in Economics. All remaining errors are mine.

†Department of Economics, Harvard University. Littauer Center, 1805 Cambridge St, Cambridge, MA 02138. Email: yechanpark@fas.harvard.edu

1 Introduction

Economic policies can generate effects far beyond their direct targets. Because economic agents are embedded in spatial, market, and social networks, shocks may propagate through prices, local labor markets, commuting and shopping patterns, and social interactions. Classical analyses of social multipliers in economics emphasize that such spillovers can amplify or attenuate policy impacts in important ways (e.g., Glaeser et al., 2003; Mani et al., 2013). More recently, a growing body of experimental and quasi-experimental work in development and spatial economics has made spillovers the primary object of interest, including studies of large-scale cash transfers, public works, and transport infrastructure (e.g. Miguel and Kremer, 2004; Donaldson and Hornbeck, 2016; Egger et al., 2022; Muralidharan et al., 2023; Franklin et al., 2024; Walker et al., 2024).

Despite this interest, spillovers pose econometric challenges. Unlike the standard no-interference setting (where each unit's outcome depends only on its own treatment), spillovers mean that unit i's outcome can depend on many, or even all units' treatments. Whereas under the standard assumption each unit has only a small number of potential outcomes (e.g., two in the binary-treatment case), with spillovers the number of potential outcomes can be very large (e.g., 2^n in the binary case), since each of the possible assignments can in principle produce a different outcome for that unit.

Applied work typically makes the problem tractable by replacing the full assignment vector W with a low-dimensional exposure mapping $g(W; X_i, \theta)$ (Manski, 2013; Aronow and Samii, 2017)[1]. The vector W collects all units' treatment statuses, X_i contains baseline covariates such as networks or commuting flows, and θ parameterizes features of the mapping, such as a distance radius or elasticity parameter. Examples include ring or neighborhood averages (Miguel and Kremer, 2004; Egger et al., 2022) and smooth spatial-decay kernels such as gravity or market-access measures (Redding and Venables, 2004; Donaldson and Hornbeck, 2016; Franklin et al., 2024).

This approach raises two key challenges. The first is the choice of the map structure $g\left(W; X_i, \theta\right)$. Often economic theory does not provide an exact functional form for this map beyond broad guidance, such as the idea that spillovers should decay with distance (in space, in a network, or along trade and commuting links). Reflecting such ambiguity, even in closely related empirical settings, different researchers make very different choices among these constructions. For example, in large-scale anti-poverty programs in low-income countries, researchers studying settings with very similar institutional features choose different exposure map functional forms (e.g. Egger et al., 2022; Franklin et al., 2024; Walker et al., 2024).[2] Even within a single paper, researchers often

[1] An alternative is to restrict interference to small, known groups (partial interference; e.g., households or dorm rooms), the special case in which the exposure map reads only own-group assignments. We focus on the general case because our applications feature market- and space-mediated spillovers that may cross such groupings.

[2] Egger et al. (2022) study large cash transfers in rural Kenya and proxy local spillovers by per-capita transfers within distance bands around each market ("distance buffers"). Franklin et al. (2024) evaluate an urban public-works program in Addis Ababa and, to capture spillovers across neighborhoods, combine a spatial-equilibrium model with

report multiple exposure constructions and informally compare them (Donaldson and Hornbeck, 2016; Muralidharan et al., 2023).[3]

Second, even given a particular functional form for the exposure mapping, empirical researchers must still select θ and account for the uncertainty this choice introduces into downstream policy estimates. For example, Egger et al. (2022) note that they "had no a priori knowledge of the relevant distances over which general equilibrium effects might operate…". In practice, applied papers often report results for several plausible values of θ, or select a preferred specification using model-selection criteria or informal diagnostics (e.g. Currie et al., 2015; Donaldson and Hornbeck, 2016; Egger et al., 2022). Yet standard errors and confidence intervals typically condition on this choice, treating θ as fixed rather than as a source of estimation uncertainty.[4]

To address these challenges in choosing and calibrating exposure mappings, this paper develops a design-based framework for testing and estimating exposure mappings in settings with spillovers. A key difficulty in such settings is that correctly specifying the outcome model is especially challenging given the complex dependence induced by spillovers. We adopt a design-based approach that sidesteps this by treating potential outcomes as fixed for the experimental sample and letting all randomness come from the known assignment mechanism; this randomization serves as the key assumption needed to test and conduct inference on the exposure mapping. Starting from a proposed exposure map $g(W; X_i, \theta)$, we develop tools to (i) test whether the implications of this map are consistent with observed outcomes under the known assignment mechanism and (ii) estimate the tuning parameters θ. To quantify uncertainty, we develop a design-based law of large numbers and central limit theorem for GMM estimators under spatial and network dependence, characterize an efficiency bound and optimal moments within the class of design-based estimators, and propagate the resulting exposure-map uncertainty into downstream policy estimands.

commuting data so that exposure is mediated by the commuting network rather than by geographic distance alone. In related Kenyan data, Walker et al. (2024) construct gravity-style "shopping network" exposure indices that weight village-level activity by baseline shopping flows from villages to markets, and compare the performance of these network-based measures to buffer-style measures of spatial exposure in explaining cross-market inflation (Walker et al., 2024, Figure A.17).

[3]For instance, Muralidharan et al. (2023) take as their main specification a ring-based exposure measure, given by the treated share of units within a 20 km radius of each location, but also show in their Appendix B that similar results obtain when exposure is measured using a market-access-style kernel. Likewise, Donaldson and Hornbeck (2016) summarize the impact of the nineteenth-century U.S. railroad network through a market-access index, but Appendix Table 3 reports estimates under a range of alternative exposure maps: they redefine market access using different measures of "market size" (population versus county wealth) and restrict the set of trading partners (for example, only counties beyond certain distance thresholds, only urban counties, only major cities, or only New York City).

[4]In Egger et al. (2022), local price changes are regressed on exposure measures defined over multiple distance buffers, with the outer radius chosen by a Schwarz/Bayesian information criterion; the selected 0–2 km band is then used as the main exposure definition in the reduced-form analysis (Egger et al., 2022, Section 4.2). Currie et al. (2015) estimate effects within several distance bands around polluting manufacturing plants and report alternative distances, but focus on a preferred near/far comparison, 0–1 versus 1–2 miles, guided by atmospheric dispersion evidence (Currie et al., 2015, pp. 278, 299–312). Donaldson and Hornbeck (2016) recompute market access under alternative trade elasticities as robustness checks (Donaldson and Hornbeck, 2016, pp. 843–844, Appendix Table 3). In each case, variation across θ is treated as specification analysis rather than propagated into sampling uncertainty.

Formally, we work with a known randomized (or quasi-randomized) assignment design $\mathcal{D}_n$ and a proposed exposure map $g(W; X_i, \theta)$. A key observation is that a correctly specified exposure mapping implies an *exposure sufficiency* property: there exists a true parameter θ_0 such that potential outcomes depend on the full treatment assignment vector only through the exposure mapping $g(W; X_i, \theta_0)$, so that conditional on $g(W; X_i, \theta_0)$ the remaining details of W carry no additional outcome-relevant information. This property implies a large set of orthogonality conditions between outcomes and functions of W, which we construct via what we call *design-side residuals*. Given any integrable outcome transformation ϕ and design function ψ, we orthogonalize $\psi(W)$ with respect to the exposure mapping by subtracting its conditional expectation given $g(W; X_i, \theta)$ (in the spirit of Borusyak and Hull, 2023), and for each unit i define the design-side residual $R_{i,\theta}\psi(W) := \psi(W) - \mathbb{E}\big[\psi(W) \mid g(W; X_i, \theta)\big]$, where the conditional expectation is taken with respect to $\mathcal{D}_n$, and thus can be computed by repeatedly sampling from the known design. Under exposure sufficiency and (quasi-)random assignment, we show that at the true parameter these design-side residuals are orthogonal to any function of unit i's outcome: $\mathbb{E}\big[\phi(Y_i)\, R_{i,\theta_0}\psi(W)\big] = 0$ for every ϕ and ψ. Varying (ϕ, ψ) generates a family of moment conditions that form the basis for GMM estimation; when the system is overidentified, the associated J-test assesses whether the proposed exposure mapping is consistent with the design. In this way, a single design-side residualization strategy supports both of our key objectives: evaluating the fit of a proposed exposure map and learning about the tuning parameter θ_0 from the data.

For inference, we establish consistency and asymptotic normality of the design-based GMM estimator under spatial and network dependence. We work in the affinity-set dependence framework of Chandrasekhar et al. (2023), which nests classical structures such as mixing random fields, m-dependence, and dependency graphs as special cases. Existing results in this framework are pointwise; to our knowledge we are the first to establish uniform consistency and asymptotic normality in it, and we give primitive conditions that cover both smooth exposure maps (gravity and market-access kernels) and non-smooth ones (the ring radius). We also characterize efficiency within this moment class: there is an asymptotic-variance lower bound for any moment the framework constructs, and a feasible sieve attains it.

Finally, we carry the resulting uncertainty about $g(W; X_i, \theta)$ into downstream policy estimands (e.g., average effects under counterfactual assignment rules) so that inference reflects both the experimental variation and uncertainty about the exposure specification itself.

We illustrate the framework in two applications to large-scale anti-poverty programs in development economics: the large-scale public-works reform in rural India analyzed by Muralidharan et al. (2023), and the GiveDirectly cash-transfer experiment in rural Kenya studied by Egger et al. (2022) together with the follow-up structural analysis of Walker et al. (2024). Both study large interventions in low-income economies, are explicitly motivated by general-equilibrium spillovers, involve spatially linked local markets, and rely primarily on ring-based exposure mappings.

The two applications yield contrasting conclusions. In the Muralidharan et al. (2023) application, the design-based overidentification tests do not reject their ring specification, and the estimated radii broadly support the original 20 km labor-market scale for the main income outcomes. In the Egger et al. (2022) application, by contrast, the ring specification is rejected for several core outcomes, and our support diagnostics reject the original 2 km radius for all outcomes, with the smallest non-rejected supports often being 4–6 km or larger. This pattern is consistent with the follow-up analysis of Walker et al. (2024), which argues that very local rings can miss market-level ambient spillovers that are common to households within local markets. Consistent with this narrative, we obtain a local fiscal transfer multiplier of 1.60, which is smaller than Egger et al. (2022)'s original 2.5, and much closer to the model-predicted multiplier of 1.54 by Walker et al. (2024).

Together, these applications illustrate the value of treating exposure mappings as objects of design-based inference rather than fixed researcher choices: the framework can support an economically motivated map, detect and guide revisions when it does not, and propagate the resulting uncertainty into policy-relevant estimates.

1.1 Related literature

Design-based uncertainty. We adopt a finite-population, design-based perspective in which $\{Y_i(\cdot), X_i\}_{i=1}^n$ are fixed and randomness arises only from the assignment mechanism. This tradition traces to Neyman (1923) and motivates modern distinctions between design-based and superpopulation uncertainty in regression and experimental analyses (Abadie et al., 2020). Beyond fully randomized designs, Rambachan and Roth (2025) develop a design-based theory of uncertainty for canonical quasi-experimental estimators when treatment propensities can vary across units, and Li and Ding (2017) provide finite-population central limit theorems yielding $\sqrt{n}$-Gaussian approximations under randomization-based dependence. Our contribution operates within this design-based framework but targets moment conditions implied by exposure-sufficiency restrictions for parameterized exposure mappings, under the cross-unit dependence induced by interference.

Interference and exposure mappings. Foundational potential-outcomes treatments of interference include Sobel (2006) and Hudgens and Halloran (2008). Building on this tradition, Aronow and Samii (2017) formalize *exposure mappings* and develop randomization-based estimators for causal effects under general (but specified) interference structures, while Sävje et al. (2021) clarify large-sample behavior of conventional estimators under unknown interference and show that standard variance formulas can fail. Leung (2022) allows treatments assigned to increasingly distant units to have smaller but nonzero effects under approximate neighborhood interference, a class that no finite-dimensional exposure map characterizes and so falls outside our framework. Recent work also studies what can and cannot be learned about interference structure from the design itself: Gao

et al. (2026) show that no specification test can have uniform power against any richer exposure-mapping alternative to an exposure-mapping model[5], while Zhong (2025) develops finite-sample randomization tests for the existence and extent of interference (for example, whether spillovers vanish beyond a given distance) under minimal assumptions on the network. Complementary work emphasizes robustness to unknown spillover mechanisms within broad classes and studies estimators (and sometimes designs) with minimax or neighborhood-adaptive guarantees; see Belloni et al. (2022) and Faridani and Niehaus (2024). Our approach is complementary to these robust and testing-based methods: rather than remaining fully agnostic about structure or testing a partially sharp no-spillover null, we take the parametric exposure maps used in applied work (e.g., ring, gravity, market-access) as maintained and use the known design to (i) estimate the tuning parameter θ_0 and (ii) test the associated exposure-sufficiency restrictions via overidentifying, design-implied orthogonality conditions.[6]

Testing models with trusted shocks. A line of work going back to Lucas (1980) views structural models as objects to be disciplined by their responses to shocks with well-understood sources and exogeneity properties. In quantitative trade, Adão et al. (2023) propose IV-based goodness-of-fit statistics that compare a model's predicted response to quasi-experimental tariff shocks with the observed response. We bring this "trusted shocks" logic to spillovers and exposure mappings: the parametric exposure family $g(W; X_i, \theta)$ plays the role of the structured model, a structural restriction on how treatments propagate even in otherwise reduced-form designs, the (quasi-)experimental assignment $W \sim \mathcal{D}_n$ (or $W \sim G_n(\cdot \mid X)$) provides the trusted shock, and the design-implied orthogonality conditions deliver both specification tests for the maintained exposure class and design-based estimators of the exposure parameter θ_0.

Orthogonalization and recentering. Our design-side residualization builds on a long tradition of orthogonalization in semiparametric and GMM settings, including the partially linear model of Robinson (1988) and the construction of "orthogonal instruments" in IO (Ackerberg and Crawford, 2009; Ackerberg et al., 2011; Andrews et al., 2025). Closest in spirit is the recentering approach of Borusyak and Hull (2023), who study settings where a known shock design generates a constructed regressor (a formula instrument) and recenter it by subtracting its conditional mean given covariates to obtain orthogonal moments for a downstream causal or structural parameter. Extensions to optimality are provided by Borusyak and Hull (2026), and to nonlinear settings by Borusyak et al.

[5] See Remark 3.7 for detailed discussion.

[6] A further complementary literature asks how exposure-based estimands should be interpreted when the researcher's chosen exposure mapping is misspecified. Sävje (2024) separates the use of an exposure mapping to define an estimand from the assumption that the mapping captures the complete causal structure, and establishes conditions under which exposure effects remain estimable under misspecification. Park and Yang (2026) instead take the marginal policy effect as primitive and show that a researcher-chosen exposure mapping induces a pseudo-true outcome model and a corresponding decomposition into direct and spillover components.

(2025), whose application is demand estimation rather than spillovers. Whereas that work uses recentering to identify and estimate a downstream parameter given a fixed formula or exposure, we apply the same orthogonalization logic upstream to discipline and test the exposure map itself: we treat $g(W; X_i, \theta)$ as the object of interest and use the known (quasi-)experimental design to identify, estimate, and test θ_0. Relatedly, Ritzwoller (2025) develops reweighting procedures for proximity-exposure regressions using residualized proximity measures to isolate variation orthogonal to alternative mediating channels.

The rest of the paper is organized as follows. Sections 2–4 develop the framework: identification via design-based orthogonality (Section 2), a large-sample theory under spatial and network dependence (Section 3), and the efficient choice of moments (Section 4). Section 5 propagates the estimated map into downstream estimands, and Section 6 presents the two applications.

2 Setup and Moment Conditions

2.1 Design, outcomes, and exposure maps

We observe units $i = 1, \ldots, n$ (regions, households, or network nodes) and a treatment assignment vector $W = (W_1, \ldots, W_n)^\top$ drawn from a known experimental (or quasi-experimental) design $\mathcal{D}_n$ (e.g., complete, Bernoulli, stratified, cluster-randomized, shock-based, etc.).

The assignment takes values in a known support $\mathcal{W}$. Binary treatment, $\mathcal{W} = \{0, 1\}^n$, is the leading case, but nothing in what follows requires it: the individual assignment W_i may be multi-valued or continuous, as with the per-village transfer amounts in our second application. For each assignment $w \in \mathcal{W}$, unit i has a potential outcome $Y_i(w)$, a function of the *entire* assignment vector w rather than unit i's own assignment w_i alone; this is what allows unit i's outcome to depend on other units' treatments. The realized outcome is $Y_i = Y_i(W)$.

Let X_i denote observed unit-level information, taking values in a set $\mathcal{X}$, that may be relevant for how assignment affects unit i, such as location, network links, strata, or baseline covariates. An *exposure map* is a function

$$g : \mathcal{W} \times \mathcal{X} \times \Theta \to \mathbb{R}^k,$$

with parameter $\theta \in \Theta \subset \mathbb{R}^p$ and $k \ll n$, intended to summarize the aspects of the assignment that are relevant for Y_i through

$$g(W; X_i, \theta).$$

When there is no risk of confusion, we use the shorthand notation

$$g_i(W; \theta) := g(W; X_i, \theta),$$

and write $g_i(w; \theta)$ for the exposure induced by a nonrandom assignment $w \in \mathcal{W}$.

Throughout, we adopt a finite-population, design-based perspective: the potential-outcome schedule $\{Y_i(\cdot)\}_{i=1}^n$ and covariates $\{X_i\}_{i=1}^n$ for the experimental (or quasi-experimental) sample are treated as fixed (or, equivalently, conditioned upon), and all randomness comes from $W \sim \mathcal{D}_n$.

We write $\mathbb{E}$ and $\mathbb{P}$ for expectation and probability with respect to $\mathcal{D}_n$, conditioning implicitly on this fixed schedule of potential outcomes and covariates. This perspective, as in Neyman (1923); Abadie et al. (2020); Li and Ding (2017), anchors inference to the specific network, market, or spatial environment actually exposed to the policy, rather than positing a hypothetical superpopulation experiment in which entire economies, including their equilibrium prices, networks, and cross-unit dependence, are repeatedly resampled and subjected to new assignments W. Such an experiment is least credible exactly when spillovers are system-wide, since it would then require strong assumptions about how the joint distribution of $\{Y_i(w)\}_{i,w}$ and the broader equilibrium environment vary coherently across draws. The design-based framework is correspondingly most appealing here: taking the realized economy as fixed and the known randomization or shock design as the sole source of uncertainty, it delivers a transparent basis for inference without additional assumptions on the population-generating process.

2.2 Examples of exposure maps

As a concrete illustration, we document three families of exposure maps that recur in applied work. Let $d(i,j)$ denote a fixed, nonstochastic distance between units i and j, where distance may be geographic, travel-time, or network distance.

Example 2.1 (Ring exposure)**.** A ring exposure map imposes a hard spatial cutoff: only treated units within distance θ of unit i contribute to exposure. Define the row-normalized weights

$$a_{ij}(\theta) := \frac{\mathbf{1}\{d(i,j) \le \theta\}\,\mathbf{1}\{j \neq i\}}{\sum_{k=1}^n \mathbf{1}\{d(i,k) \le \theta\}\,\mathbf{1}\{k \neq i\}}, \qquad a_{ii}(\theta) := 0,^{7} \tag{2.1}$$

and let

$$g^{\text{ring}}(W; X_i, \theta) := \sum_{j=1}^n a_{ij}(\theta) W_j.$$

Thus $g^{\text{ring}}(W; X_i, \theta)$ is the average treatment status among units lying within radius θ of i. This is the logic behind the distance-buffer specifications in Egger et al. (2022). It is also the main exposure design in Muralidharan et al. (2023), where the baseline specification uses the treated share of locations within a fixed 20 km radius.

Example 2.2 (Smooth spatial decay)**.** Smooth spatial-decay maps replace the hard cutoff in a ring design with weights that decline continuously with distance. Let $L_j > 0$ denote a pre-treatment

[7] We use the convention $a_{ij}(\theta) := 0$ for all j when no other unit lies within distance θ of i (empty denominator).

measure of the economic size or attractiveness of location j, such as population, employment, or market size. Define

$$a_{ij}(\theta) := \frac{K(d(i,j), L_j; \theta)\, \mathbf{1}\{j \neq i\}}{\sum_{k=1}^{n} K(d(i,k), L_k; \theta)\, \mathbf{1}\{k \neq i\}}, \qquad a_{ii}(\theta) := 0, \tag{2.2}$$

and

$$g^{\text{ssd}}(W; X_i, \theta) := \sum_{j=1}^{n} a_{ij}(\theta) W_j. \tag{2.3}$$

This formulation nests several familiar kernels. One example is a gravity-style exponential kernel,

$$K_{\text{grav}}(d, L; \theta) := L \, \exp(-\theta d),$$

under which exposure decays smoothly with distance at rate θ, used in Franklin et al. (2024); Walker et al. (2024). Another important example is a market-access kernel with power decay,

$$K_{\text{ma}}(d, L; \theta) := L(1 + \alpha d)^{-\theta},$$

which is closely related to market-access measures in Redding and Venables (2004) and Donaldson and Hornbeck (2016). This is the alternative exposure class considered in Appendix B of Muralidharan et al. (2023), where the authors use the form $(1 + \alpha d)^{-\theta}$ with $\alpha = 1/100$ and fix $\theta = 8$ based on Donaldson and Hornbeck (2016).

Example 2.3 (Network exposure)**.** When spillovers propagate through a known baseline network, distance may be measured in graph steps rather than kilometers. The ring exposure in Example 2.1 also covers network spillovers as a special case if distance is interpreted as graph distance rather than geographic distance. Let $G = (V, E)$ be a fixed graph on nodes $V = \{1, \ldots, n\}$, and let $\text{dist}_G(i, j)$ denote shortest-path distance on G. For $\theta \in \{1, 2, \ldots\}$, define the neighborhood

$$\mathcal{N}_{\leq \theta}(i) := \{\, j \neq i : \text{dist}_G(i, j) \leq \theta \,\}.$$

A natural scalar exposure map is then

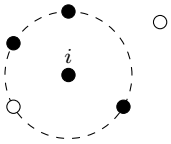
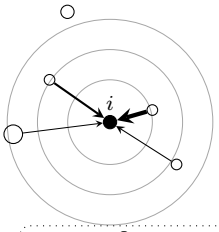
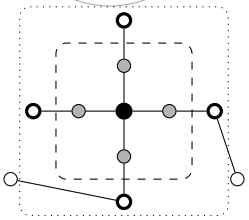
$$g^{\text{net}}(W; X_i, \theta) := \frac{1}{|\mathcal{N}_{\leq \theta}(i)|} \sum_{j \in \mathcal{N}_{\leq \theta}(i)} W_j, \tag{2.4}$$

with the convention $g^{\text{net}}(W; X_i, \theta) = 0$ if $|\mathcal{N}_{\leq \theta}(i)| = 0$.

In social-network applications, researchers often further decompose this into the share of treated friends, the share of treated friends of friends, and so on up to distance θ.[8] When $\theta = 1$, (2.4)

[8] In the network-interference literature this is often described as a K-hop restriction. We write the truncation parameter as θ here to keep notation consistent across the three exposure families.

Table 1: Illustration of canonical exposure maps

	Illustration	Informal description
Ring		Exposure is the average treatment status among units lying within distance θ of unit i. This is the main specification of Muralidharan et al. (2023) and the distance-buffer logic used in Egger et al. (2022).
Smooth spatial decay		All locations can affect i, but their influence declines smoothly with distance. Gravity-style exponential kernels and market-access kernels of the form $(1+\alpha d)^{-\theta}$ are leading examples; the latter is the alternative exposure class considered in Appendix B of Muralidharan et al. (2023).
Network		Exposure is averaged over nodes within graph distance θ of i. In social-network applications this often corresponds to the share of treated friends, or more generally treated friends, friends of friends, and so on, up to distance θ.

reduces to the familiar share of treated friends. This type of restriction is common in empirical work on peer effects and network spillovers; see, for example, Cai et al. (2015) and Bramoullé et al. (2009).

Across these examples, the role of θ is always the same: it indexes how far spillovers reach or how quickly they decay. Table 1 illustrates the three families.

2.3 Exposure sufficiency

We now state the two primitive conditions that generate the design-based moment restrictions. The first condition is the usual finite-population randomization condition: the potential-outcome schedule is fixed, and all randomness comes from the known assignment law.

Assumption 2.4 (Randomized assignment)**.** *The assignment W is drawn from a known law $\mathcal{D}_n$ that does not depend on the potential outcomes. Formally, for every collection of potential outcomes $\{Y_i(\cdot)\}_{i=1}^n$,*

$$W \,\big|\, \{Y_i(\cdot), X_i\}_{i=1}^n \sim \mathcal{D}_n.$$

The second condition is the exposure-map hypothesis. It states that, at the correct value θ_0, the candidate exposure map contains all information in the assignment vector that is relevant for unit i's outcome.

Hypothesis 2.5 (Exposure map)**.** *The exposure mapping $g(W; X_i, \theta)$ is* well specified *if there exist $\theta_0 \in \Theta$ and measurable functions $\widetilde{Y}_i : \mathbb{R}^k \to \mathbb{R}$ such that, for every assignment w,*

$$Y_i(w) = \widetilde{Y}_i\big(g(w; X_i, \theta_0)\big), \qquad i = 1, \ldots, n.$$

Thus Hypothesis 2.5 makes the exposure $g(W; X_i, \theta_0)$ a summary of the assignment for unit i: the map is *well specified* when the exposure captures everything in W relevant to Y_i. The goal is to test this restriction and learn θ_0 using the known design $\mathcal{D}_n$.

Our main result is that, combined with the known design, well specification has an observable consequence. Because the assignment law is known and does not depend on the potential outcomes, conditioning on the exposure removes all dependence between Y_i and the residual variation in W. We call this consequence exposure sufficiency. It is what makes the map testable: the design leaves assignment variation that a well-specified exposure must render irrelevant to Y_i, and any leftover dependence is evidence against the map.

Theorem 2.6 (Exposure sufficiency)**.** *Suppose Assumption 2.4 and Hypothesis 2.5 hold. Then, for every unit i,*

$$Y_i \perp\!\!\!\perp W \,\big|\, g(W; X_i, \theta_0)$$

under the design distribution.

To see what the theorem requires, and why it can be tested, consider the ring map of Example 2.1, where $g(W; X_i, \theta)$ is the treated share of units within radius θ of unit i. Applied researchers already probe this choice informally: one draws rings of half a mile, a mile, two miles, and looks for the radius at which estimated spillovers level off. The design-based restriction makes that intuition precise. Well specification at radius θ_0 says unit i's outcome depends on the assignment only through the within-θ_0 treated share: two assignments with the same share produce the same outcome for i, no matter which of those neighbors are treated and regardless of the treatment of any unit beyond θ_0. This is a substantive economic restriction, and it can fail in two ways: a treated unit just outside the radius still affects i, or the identity rather than the count of treated neighbors matters. The known randomization turns each failure into something observable. If the ring is correct, then under the design the treatments of units outside radius θ_0 are uncorrelated with Y_i once we condition on the within-ring share; leftover correlation is evidence that spillovers reach past θ_0. This is exactly the equal-to-zero condition that the moments of the following subsection formalize, both to test a candidate radius and to select θ_0.

2.4 Orthogonal moments

The ring discussion tested one exposure map with one natural statistic, the treatments of units outside the radius. The conditional independence in Theorem 2.6 says much more: *any* function

of the assignment, once purged of what the candidate exposure explains, must be unrelated to *any* function of the outcome. We encode this with a residual that strips from a design function the part predictable from the exposure. To construct the moments, let $\psi : \mathcal{W} \times \mathcal{X} \to \mathbb{R}$ be any integrable design function, possibly depending on the unit's covariates X_i.[9] For each unit i and candidate parameter θ, define the design residual

$$R_{i,\theta}\psi(W) := \psi(W; X_i) - \mathbb{E}[\psi(W; X_i) \mid g(W; X_i, \theta)], \tag{2.5}$$

where the conditional expectation is taken under the known design $\mathcal{D}_n$. This residual removes from $\psi(W)$ the component explained by the candidate exposure $g(W; X_i, \theta)$.

Corollary 2.7 (Orthogonal moments)**.** *Under the conditions of Theorem 2.6, for any outcome transformation $\phi : \mathbb{R} \to \mathbb{R}$ and design function ψ with $\mathbb{E}[\phi(Y_i)^2] < \infty$ and $\mathbb{E}[\psi(W)^2] < \infty$,*

$$\mathbb{E}[\phi(Y_i)\, R_{i,\theta_0}\psi(W)] = 0. \tag{2.6}$$

Equivalently,

$$\mathbb{E}[\phi(Y_i)\,\{\psi(W) - \mathbb{E}[\psi(W) \mid g(W; X_i, \theta_0)]\}] = 0.$$

Two features of the corollary matter for what follows. First, it is agnostic about the map: nothing in (2.6) is special to the ring, so the same construction disciplines gravity, market-access, or network exposures, with only $g(W; X_i, \theta)$ changing. Second, the equalities hold at θ_0 for *every* admissible (ϕ, ψ), which is what lets a single strategy serve two ends: matching them identifies and estimates θ_0, while a candidate map that violates them is detectable through the resulting overidentification.

For estimation, fix a finite dictionary of moment-generating pairs

$$\{(\phi_m, \psi_m) : m = 1, \ldots, M\},$$

where each $\phi_m : \mathbb{R} \to \mathbb{R}$ is an outcome transformation and each $\psi_m : \mathcal{W} \times \mathcal{X} \to \mathbb{R}$ is a design function. Define the i-specific moment contribution

$$\eta_{m,i,n}(\theta) := \phi_m(Y_i)\, R_{i,\theta}\psi_m(W), \tag{2.7}$$

[9] The design residual $R_{i,\theta}$ below evaluates the design function at the unit's covariates X_i, so the unit-centering enters through the same index i that $R_{i,\theta}$ already carries. This covers the leading applied case, in which ψ aggregates the assignment in a neighborhood of X_i (for example, population-weighted treatment averages over distance bands around unit i). To avoid clutter we continue to write $\psi(W)$, and $\psi_m(W)$ for dictionary elements, each understood to be evaluated at the relevant unit's covariates inside $R_{i,\theta}$.

and the corresponding sample moment

$$\eta_{m,n}(\theta) := \frac{1}{n}\sum_{i=1}^{n} \eta_{m,i,n}(\theta).$$

Stacking the M moments gives

$$\eta_n(\theta) := \big(\eta_{1,n}(\theta), \ldots, \eta_{M,n}(\theta)\big)^\top.$$

A design-based GMM estimator is then

$$\hat{\theta}_n \in \arg\min_{\theta\in\Theta} Q_n(\theta), \qquad Q_n(\theta) := \eta_n(\theta)^\top \Lambda_n \eta_n(\theta), \tag{2.8}$$

where Λ_n is a positive semidefinite weight matrix.

Remark 2.8 (Practical implementation). The conditional expectations

$$\mathbb{E}[\psi_m(W) \mid g(W; X_i, \theta)]$$

are design-side objects: because the assignment law $\mathcal{D}_n$ is known (Assumption 2.4), each is a functional of the design alone and involves neither the potential outcomes nor any outcome model, so it can be computed without additional assumptions. When enumeration of the support of $\mathcal{D}_n$ is feasible, the projection is evaluated directly. Otherwise, one draws assignments $W^{(b)} \sim \mathcal{D}_n$, $b = 1, \ldots, B$, and approximates $\mathbb{E}[\psi_m(W) \mid g(W; X_i, \theta) = s]$ by regressing $\psi_m(W^{(b)})$ on a flexible function of the simulated exposure $g(W^{(b)}; X_i, \theta)$: for a discrete exposure, by averaging ψ_m within its realized values; for a continuous exposure, by a low-order polynomial or kernel smoother in s, with B and the smoother chosen so that the residual approximation error is first-order negligible. In the applications we use a quadratic in the exposure index across a few hundred placebo draws, fit by cross-validated ridge regression in the Egger et al. (2022) application. △

Remark 2.9 (Overidentification and testing exposure maps). When $M > \dim(\theta)$, the system is overidentified. The associated design-based J-statistic can therefore be used to test the maintained exposure specification or to compare alternative exposure maps. Section 3 derives the large-sample behavior of $\hat{\theta}_n$, and Section 4 studies the efficient choice of moments. △

Remark 2.10 (Multi-unit restrictions). Hypothesis 2.5 also implies orthogonality restrictions for any finite set of units. For any finite $S \subset \{1, \ldots, n\}$, let

$$Y_S := (Y_i)_{i\in S}, \qquad g_S(W; \theta) := \big(g(W; X_i, \theta)\big)_{i\in S}.$$

If the exposure map is well specified, then

$$Y_S \perp\!\!\!\perp W \mid g_S(W;\theta_0),$$

and hence, for any square-integrable ϕ_S and ψ,

$$\mathbb{E}[\phi_S(Y_S)\{\psi(W) - \mathbb{E}[\psi(W) \mid g_S(W;\theta_0)]\}] = 0.$$

These restrictions could generate moments based on pairs or larger groups of units, such as moments involving (Y_i, Y_j) and the joint exposure vector $(g(W;X_i,\theta_0), g(W;X_j,\theta_0))$. We do not pursue the full multi-unit moment system here: it greatly expands the class of possible moments, and it is less clear how to choose among them. Instead, the paper focuses on the unit-level moments, which nest the moments used in applied work and already deliver tractable estimators, specification tests, and efficiency analysis. Developing practical procedures that exploit the broader cross-unit restrictions is left for future work. △

Remark 2.11 (Quasi-experimental shock designs and relation to recentering). Many applications of exposure mappings are based on *quasi-experimental* variation rather than literal randomized assignment. In such settings, the researcher observes a realized shock vector W (e.g., line openings in a transport network, sectoral demand shocks, or other environmental shocks), treats the baseline covariates $X = (X_1, \ldots, X_n)$ as fixed, and specifies a *shock design* $\mathcal{D}_n^{\text{shock}}$ that captures the as-good-as-random component of W. Provided this design is known and satisfies the same structural property as Assumption 2.4,

$$W \mid \{Y_i(\cdot)\}_{i=1}^n \sim \mathcal{D}_n^{\text{shock}} \quad \text{for all potential-outcome schedules,}$$

all of the constructions above go through after replacing $\mathcal{D}_n$ by $\mathcal{D}_n^{\text{shock}}$. In practice, $\mathcal{D}_n^{\text{shock}}$ is implemented via permutations, placebo networks, or other simulation schemes, and the re-randomization step used to approximate $\mathbb{E}[\psi_m(W) \mid g(W;X_i,\theta)]$ is carried out by simulating W from this shock design.

This design-side residualization is analogous in spirit to the recentering approach of Borusyak and Hull (2023): both start from a known shock or assignment distribution and construct transformed instruments or moments that are orthogonal, by design, to certain components of the assignment. However, the goals and maintained structures are different. Borusyak and Hull (2023) assume a particular linear homogeneous treatment-effect model and a given formula instrument, and their objective is to identify and estimate the resulting coefficient β.[10] By contrast, in the present framework the exposure mapping itself is the primary object of interest: we use the ran-

[10]The recentering logic is not intrinsically tied to a linear model: recently, Borusyak et al. (2025) apply it to nonlinear (nested and mixed logit) demand estimation.

domized or quasi-random shock design to learn about θ_0 in $g(W; X_i, \theta_0)$ and to test whether a proposed exposure class is consistent with the design-implied orthogonality conditions. That said, we provide a general procedure to conduct formal inference on policy functionals in a downstream stage, discussed in Section 5. △

3 Design-Based Consistency and Asymptotic Normality

Section 2 derived design-based moment restrictions implied by exposure sufficiency. We now study the large-sample behavior of GMM estimators and tests constructed from those moments. Our goal is to establish design-based consistency and asymptotic normality of the GMM estimator, and to develop the associated overidentification tests of the exposure-map specification. See Remark 3.4 for discussion of exact randomization tests.

To separate the asymptotic argument from any particular choice of moments, we work with a generic finite-dimensional moment contribution

$$\Psi_i(\theta) \in \mathbb{R}^q.$$

In the exposure-mapping application, this vector is obtained by stacking finitely many residualized design moments,

$$\Psi_i(\theta) = \big(\eta_{1,i,n}(\theta), \ldots, \eta_{q,i,n}(\theta)\big)^\top, \qquad \eta_{m,i,n}(\theta) = \phi_m(Y_i) R_{i,\theta} \psi_m(W).$$

The generic notation lets us state the LLN, consistency, CLT, and asymptotic-normality results once, in terms of the dependence structure induced by the assignment design. These results are then used directly for exposure-map estimation here and for the optimal-moment analysis in Section 4.

3.1 Finite-population setup and GMM criterion

Following the design-based literature (e.g., Freedman, 2008; Aronow and Samii, 2017), we work with a triangular array of experiments $\{(\mathcal{U}_n, \mathcal{D}_n)\}_{n\geq 1}$, where both the population size $N_n := |\mathcal{U}_n|$ and the assignment design $\mathcal{D}_n$ are allowed to change with n.

For each n, let $\mathcal{U}_n = \{1, \ldots, N_n\}$ denote the finite experimental population. For each unit $i \in \mathcal{U}_n$ and parameter $\theta \in \Theta \subset \mathbb{R}^p$, let $\Psi_i(\theta) \in \mathbb{R}^q$ be a q-dimensional moment vector, with true parameter value θ_0. The array $\{\Psi_i(\theta)\}_{i\in\mathcal{U}_n}$ is generated by the known assignment design $\mathcal{D}_n$, while the potential outcomes and baseline attributes are treated as fixed. The design fixes the distribution of these moments in any given population; the large-sample results below require, in addition, regularity conditions on the dependence the design induces across units, which we impose through the affinity-set structure introduced next.

The sample moment is the normalized finite-population average

$$\bar{\Psi}_n(\theta) := \frac{1}{N_n}\sum_{i=1}^{N_n}\Psi_i(\theta), \qquad \mu_n(\theta) := \frac{1}{N_n}\sum_{i=1}^{N_n}\mathbb{E}[\Psi_i(\theta)].$$

Let Λ_n be a symmetric positive semidefinite $q \times q$ weight matrix with $\Lambda_n \xrightarrow{p} \Lambda \succeq 0$, and define the quadratic GMM criterion

$$Q_n(\theta) := \bar{\Psi}_n(\theta)^\top \Lambda_n \bar{\Psi}_n(\theta), \qquad \hat{\theta}_n \in \arg\min_{\theta\in\Theta} Q_n(\theta).$$

In the limit, write

$$\begin{aligned} \mu(\theta) &:= \lim_{n\to\infty} \mu_n(\theta), \\ Q(\theta) &:= \mu(\theta)^\top \Lambda\, \mu(\theta), \end{aligned} \tag{3.1}$$

whenever the limit exists. In particular, we assume that the limit is centered at the true parameter, $\mu(\theta_0) = 0$.[11]

3.2 Affinity sets and consistency

To accommodate design-induced spatial or network dependence, for each $i \in \mathcal{U}_n$ we fix an *affinity set* $A_i \subseteq \mathcal{U}_n$ with $i \in A_i$ collecting units whose assignments may have non-negligible covariance with that of unit i. We write $|A_i|$ for its cardinality and allow arbitrary dependence within A_i. In the spatial examples, A_i can be read as a growing geographic or travel-time ball around unit i; in the market-access and gravity examples, it collects locations whose assignment shocks receive non-negligible kernel weight for i; and in the network examples, it corresponds to a local graph neighborhood whose radius may grow slowly with N_n. Outside A_i we do not impose conditional independence; instead, we only assume that the aggregate covariance contribution from $\{j \notin A_i\}$ is asymptotically negligible relative to the contribution from within A_i.

Define centered variables

$$Z_{i,n}(\theta) := \Psi_i(\theta) - \mathbb{E}[\Psi_i(\theta)], \qquad \bar{Z}_n(\theta) := \frac{1}{N_n}\sum_{i=1}^{N_n} Z_{i,n}(\theta) = \bar{\Psi}_n(\theta) - \mu_n(\theta),$$

[11] This centering follows directly from the design-based orthogonality result in Corollary 2.7.

and the *within-affinity covariance matrix*[12]

$$\Omega_n(\theta) := \sum_{i=1}^{N_n} \sum_{j \in A_i} \mathrm{Cov}\big(Z_{i,n}(\theta),\, Z_{j,n}(\theta)\big) \in \mathbb{R}^{q \times q}.$$

We first state the high-level ULLN and identification requirements. The following subsection then explains how the ULLN is verified under primitive conditions in the two leading cases.

Assumption 3.1 (Design-based ULLN and identification)**.**

AS–LLN1 ***Deterministic stabilization.*** *There exists a deterministic $\mu : \Theta \to \mathbb{R}^q$ such that*

$$\sup_{\theta \in \Theta} \|\mu_n(\theta) - \mu(\theta)\| \;\to\; 0.$$

AS–LLN2 ***Uniform LLN for the design-based moments.*** *The sample moments converge uniformly in probability to their limits:*

$$\sup_{\theta \in \Theta} \big\|\bar{\Psi}_n(\theta) - \mu_n(\theta)\big\| \;\xrightarrow{p}\; 0.$$

AS–LLN3 ***Parameter space and identification of the population criterion.*** *The parameter space $\Theta \subset \mathbb{R}^p$ is compact and $\theta_0 \in \Theta$. The limit map $\mu : \Theta \to \mathbb{R}^q$ from* AS–LLN1 *is continuous on Θ. Given* AS–LLN1*, define $Q(\theta) = \mu(\theta)^\top \Lambda\, \mu(\theta)$ on Θ, and assume*

$$Q(\theta) = 0 \quad \Longleftrightarrow \quad \theta = \theta_0.$$

AS–LLN4 ***Weight convergence.*** $\Lambda_n \xrightarrow{p} \Lambda \succeq 0$.

Assumption 3.1 is stated at a high level in terms of the population map $\mu(\theta)$ and the sample map $\bar{\Psi}_n(\theta)$. The next result shows that these high-level conditions are sufficient for consistency.[13]

Remark 3.2 (Interpretable necessary conditions for identification). We impose identification at the criterion level in AS–LLN3. Two necessary conditions specific to the moments (2.7) are worth mentioning. First, if $w \mapsto g(W; X_i, \theta)$ is injective at some θ, then conditioning on $g(W; X_i, \theta)$ is equivalent to conditioning on W, so $R_{i,\theta}\psi \equiv 0$ for every ψ by (2.5) and the moments are uninformative about that θ; identification therefore requires strictly positive residual variation, $\mathbb{E}[(R_{i,\theta}\psi(W))^2] > 0$ for some admissible ψ. This is the degeneracy of a continuous exposure parameter paired with a

[12] We take the affinity relation to be symmetric ($j \in A_i \Leftrightarrow i \in A_j$), so that $\Omega_n(\theta)$ is symmetric; this holds in all our examples (metric or travel-time balls, symmetric kernels, undirected graph neighborhoods), and any directed relation may be replaced by its symmetrization $\{j : j \in A_i \text{ or } i \in A_j\}$.

[13] Genuinely discrete exposure parameters (for example, an integer hop-count) are handled by a separate finite-grid selection argument in Appendix B.2.

finely resolved assignment, such as a smooth market-access decay in (near-)continuous distances. Second, the outcomes must respond to the assignment: under the sharp null of no effect on any unit each $\phi(Y_i)$ is design-nonrandom, so $\mathbb{E}[\phi(Y_i)R_{i,\theta}\psi(W)] = \phi(Y_i)\,\mathbb{E}[R_{i,\theta}\psi(W)] = 0$ for *every* θ by the law of iterated expectations, and θ_0 is not separated. △

Theorem 3.3 (Design-based GMM consistency)**.** *Suppose Assumption 3.1 holds. Then*

$$\sup_{\theta\in\Theta}\big|Q_n(\theta) - Q(\theta)\big| \xrightarrow{p} 0,$$

and any sequence of minimizers $\hat{\theta}_n \in \arg\min_{\theta\in\Theta} Q_n(\theta)$ *satisfies* $\hat{\theta}_n \xrightarrow{p} \theta_0$.

3.3 Primitive conditions for the uniform law of large numbers

Consistency was established above under the high-level uniform law of large numbers assumed in AS–LLN2,

$$\sup_{\theta\in\Theta}\big\|\bar{\Psi}_n(\theta) - \mu_n(\theta)\big\| \xrightarrow{p} 0.$$

This subsection gives primitive conditions under which that uniform law holds. The smooth case follows from a standard argument; the ring needs care, because its sample path is a step function of the radius θ and the smooth-GMM argument does not apply. We verify it for the two exposure maps used in the applications, continuing Examples 2.2 and 2.1; Appendix B.2 states the primitive conditions more formally and gives the proofs.

Example 2.2 (smooth spatial decay), continued. When the weights $a_{ij}(\theta)$ vary smoothly with θ, as in the gravity and market-access kernels, the moment vector is Lipschitz in θ and a standard covering argument applies. The condition that is not automatic concerns the design projection $m_{i,m,\theta}(s) := \mathbb{E}[\psi_m(W) \mid g_i(W;\theta) = s]$, which must itself vary regularly in both arguments,

$$\big|m_{i,m,\theta}(s) - m_{i,m,\theta'}(s')\big| \le M_{i,n}\big(\|\theta-\theta'\| + \|s-s'\|\big), \qquad \frac{1}{N_n}\sum_{i=1}^{N_n}\mathbb{E}\,M_{i,n} = O(1).$$

Smoothness of the exposure map does not by itself deliver this, so we impose the displayed regularity condition directly as a primitive assumption in Appendix B.2.

Example 2.1 (ring exposure), continued. Raising the radius θ changes exposure only through the units whose distance to i crosses θ, so the sample path $\theta \mapsto \bar{\Psi}_n(\theta)$ moves in steps and is not differentiable; in place of a smoothness argument, the uniform law rests on a condition on the distances. That condition is the spatial counterpart of the bounded-density requirement familiar from regression-discontinuity and threshold-regression designs, with pairwise distance d_{ij} playing the role of the running variable and the radius θ that of the cutoff: as a band of radii shrinks, the

weighted mass of distances inside it must vanish. Formally, for an interval $I \subset \mathbb{R}_+$ of length $|I|$,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{|I| \le \delta} \frac{1}{N_n} \sum_{i=1}^{N_n} \sum_{j \neq i} a_{ij} \, \mathbb{1}\{d(i,j) \in I\} = 0,$$

with a_{ij} the exposure weights of Example 2.1. Moving the radius then shifts only a vanishing share of the moment, so the empirical criterion converges uniformly and its population limit is continuous in θ even though every sample path jumps.

Remark 3.4 (Relation to exact randomization tests). A common inferential approach in experiments with spillovers is the finite-sample exact randomization test. These tests are cleanest for a *sharp* null: one that pins down each unit's potential outcome under every assignment, so realized outcomes can be recomputed for any counterfactual assignment and the statistic has a known permutation distribution. Exposure sufficiency is not sharp: it restricts how outcomes depend on the assignment through the exposure map but leaves the potential outcomes otherwise unspecified. For such non-sharp nulls, exact tests remain available by choosing focal units and conditioning on the assignment cell within which the focal units' outcomes are invariant; see, for example, Athey et al. (2018). Our target, however, is not a single fixed null but estimation of the continuous tuning parameter θ, together with inference on the downstream regression coefficients through the GMM criterion $Q(\theta)$. Exact randomization inference is already awkward for the coefficient-based analyses in the applied work cited above, and more so for estimating θ or evaluating Q. We therefore take the asymptotic conditional-moment route developed in this section, which draws on the implications of exposure sufficiency directly. △

3.4 Asymptotic normality

We now strengthen the LLN conditions above to obtain a CLT and a GMM asymptotic normality result. We retain the finite-population, design-based setup and notation introduced above, and assume Assumption 3.1 holds so that $\hat{\theta}_n \xrightarrow{p} \theta_0$ by Theorem 3.3.

Asymptotic normality rests on two further ingredients, both imposed at the population level. We state them formally as Assumption B.11 in Appendix B.3 and describe their roles here. The first delivers a *pointwise* central limit theorem for the moment vector at the truth. Because the design-based moments are dependent across units through the affinity sets of Section 3.2, we invoke the affinity-set central limit theorem of Chandrasekhar et al. (2023): under bounded fourth moments, decay of the within-affinity covariances, and a stabilizing aggregate covariance $\Omega_n / N_n \to \Omega$ with Ω positive definite, the centered moments satisfy $\sqrt{N_n}\, \bar{\Psi}_n(\theta_0) \Rightarrow \mathcal{N}(0, \Omega)$.

The second ingredient carries this pointwise statement to asymptotic normality of the GMM estimator $\hat{\theta}_n$. It has two parts: *mean differentiability* of the population moment map $\mu(\theta)$ at θ_0, with full-rank Jacobian G, and *stochastic equicontinuity* of the centered empirical process

$\mathbb{G}_n(\theta) = \sqrt{N_n}\big(\bar{\Psi}_n(\theta) - \mu(\theta)\big)$ near θ_0. Both are restrictions on the population map and on the empirical process, not on the individual sample paths $\Psi_i(\theta)$. This is what lets the argument, a Z-estimator application of van der Vaart and Wellner (1996, Theorem 3.3.1) to the projected map $\theta \mapsto G^\top \Lambda_n \bar{\Psi}_n(\theta)$, cover the non-differentiable ring path alongside the smooth gravity and market-access kernels; the affinity-set conditions enter only through the pointwise CLT and the equicontinuity of $\mathbb{G}_n$.

Theorem 3.5 (Design-based GMM asymptotic normality)**.** *Under Assumption 3.1 and Assumption B.11,*

$$\sqrt{N_n}\,(\hat{\theta}_n - \theta_0) \;\Rightarrow\; \mathcal{N}\Big(0,\; (G^\top \Lambda\, G)^{-1} G^\top \Lambda\, \Omega\, \Lambda\, G\, (G^\top \Lambda\, G)^{-1}\Big),$$

where $G = G(\theta_0)$ and Ω is the design-based asymptotic covariance in Assumption B.11 AN–AFF3*. With the optimal weight $\Lambda = \Omega^{-1}$, the asymptotic variance simplifies to $(G^\top \Omega^{-1} G)^{-1}$. Moreover, if $\Lambda_n \xrightarrow{p} \Omega^{-1}$, then the efficient-weight case is obtained.*

Corollary 3.6 (Design-based overidentification test)**.** *Suppose Assumption 3.1 and Assumption B.11 hold with $q > p$, and let $\Lambda_n \xrightarrow{p} \Omega^{-1}$ as in Theorem 3.5. Under the maintained exposure specification (Hypothesis 2.5),*

$$N_n\, Q_n(\hat{\theta}_n) \;\Rightarrow\; \chi^2_{q-p}.$$

The statistic $N_n Q_n(\hat{\theta}_n)$ therefore provides a formal test of the maintained exposure map, as anticipated in Remark 2.9: rejection is evidence that residualized assignment variation left over after conditioning on $g(W; X_i, \theta)$ remains predictive of outcomes through the chosen moments. Different finite moment collections yield different J-statistics; Section 4 studies how to choose moments efficiently.

Remark 3.7 (Scope of the specification test). Contemporaneously, Gao et al. (2026) prove that specification testing of exposure-mapping models is impossible against any richer exposure-mapping alternative: every testing procedure, design included, has worst-case Type I and Type II errors summing to one, at every sample size. The alternative is simply too large in that it places no structure across units and so admits adversarial outcome schedules that are maximally separated from the null yet generate the same observed-data distribution under every assignment. The implication is that informative tests exist only against alternatives restricted beyond what exposure mappings alone impose, as in their consistent test against a linear-in-means model.

Our test naturally embodies the restriction their result requires, in the form applied work already adopts. We maintain a parametric exposure class $\{g(\cdot;\theta) : \theta \in \Theta\}$ and a finite set of residualized design moments; the null states that some $\theta_0 \in \Theta$ satisfies $\mu(\theta_0) = 0$. The J-test then has power against alternatives that keep the moment criterion bounded away from zero uniformly over Θ, under the convergence conditions of Assumption 3.1 maintained along the alternative sequence. Rejection therefore signals that assignment variation left over after conditioning on the

proposed exposure map remains predictive of outcomes. Non-rejection means the maintained class is consistent with the design through these moments, and nothing more: it does not certify exposure sufficiency against the unrestricted alternative. △

Remark 3.8 (Nonconservative design-based variance). The variance in Theorem 3.5 is design-based but not a Neyman-style conservative bound. Neyman-type bounds arise because the exact variance of a treatment-effect estimator involves co-moments of the same unit's potential outcomes across assignments, which are never jointly observed (Neyman, 1923; Abadie et al., 2020). No such term arises here: by Corollary 2.7, the score

$$\phi(Y_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid g(W; X_i, \theta_0)]\right\}$$

is mean zero unit by unit, whatever the outcome functions $\widetilde{Y}_i$ may be, so the centered score is observed for the realized assignment and Ω is identified as a second-moment functional of observed scores under the known design. Given the graph-HAC condition of Assumption 3.9, the sandwich therefore estimates the exact large-sample variance for the maintained exposure model. △

It remains to estimate the asymptotic design covariance matrix Ω. In our setting Ω captures spatial and network dependence through the affinity sets, so a natural object is a spatial/graph HAC estimator built from $\{Z_{i,n}\}_{i\le N_n}$. Rather than spell out primitive conditions for a particular estimator, we impose a high-level consistency requirement in the spirit of spatial and network HAC methods, and refer to the existing literature for sufficient conditions. For spatial dependence, see Conley (1999); Kim and Sun (2011). For general network dependence and ψ-dependent processes on graphs, Kojevnikov et al. (2021); Sasaki (2025) provide LLN, CLT, and consistency results for a closely related network HAC estimator. We formalize our requirement as follows.

Assumption 3.9 (Graph-HAC estimation of the asymptotic covariance)**.** *Let Ω be the design-based asymptotic covariance in Assumption B.11* AN–AFF3*. There exists a sequence of graph-HAC (Conley-type) estimators $\hat{\Omega}_n$ constructed from $\{Z_{i,n}\}_{i\le N_n}$ and the affinity sets $\{A_i\}_{i\le N_n}$ such that $\hat{\Omega}_n$ is consistent in operator norm for the normalized asymptotic covariance:*

$$\left\|\hat{\Omega}_n - \Omega_n/N_n\right\|_{\mathrm{op}} \xrightarrow{p} 0.$$

Evaluated at a preliminary consistent estimator $\hat{\theta}_n$ (Algorithm 4.2), the resulting plug-in error is $o_p(1)$: the design centering is computed from the known design rather than estimated, and $\hat{\theta}_n \xrightarrow{p} \theta_0$ under Assumption 3.1 (Theorem 3.3).

4 Efficient Moments

Sections 2–3 introduced a class of design-based moments parameterized by an outcome transformation ϕ and a design function ψ,

$$\eta_{\phi,\psi,n}(\theta) := \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \phi(Y_i)\, R_{i,\theta}\psi(W).$$

Two facts delimit what optimality within this class means. First, the class is exhaustive at the unit level: Appendix C.1 shows that the collection of these moments over all bounded measurable (ϕ, ψ) is equivalent to the exposure sufficiency $Y_i \perp\!\!\!\perp W \mid g(W; X_i, \theta_0)$. Second, a correctly-specified exposure map also implies cross-unit restrictions, involving pairs or larger sets of units (Remark 2.10), which lie outside this class and are not exploited. The bound below is therefore an efficiency bound within the unit-level moment class, the class containing the moments used in applied work and in our applications, not a semiparametric bound over all implications of an exposure map hypothesis.

Theorem 4.1 (Efficiency bound and sieve approximation; informal)**.** *Under the regularity conditions of Appendix C:*

(i) *There is a finite $V^\star > 0$ such that every GMM estimator of θ_0 built from finitely many moments in the unit-level class has design-based asymptotic variance at least $V^\star$*

(ii) *For a nested dictionary $\{(\phi_m, \psi_m) : m \geq 1\}$ whose span is dense in the unit-level moment class, the finite-dictionary oracle variances satisfy $V_M^\star \to V^\star$. If the fixed-dimensional feasible-GMM conditions hold for every fixed M and $b_n/N_n \to 0$, there exists a deterministic, nondecreasing sequence $M_n \to \infty$ such that $\frac{M_n^2 b_n}{N_n} \to 0$ and $\sqrt{N_n}(\hat{\theta}_{M_n,n} - \theta_0) \Rightarrow \mathcal{N}(0, V^\star)$*

For empirical work, fixing a finite dictionary and running two-step GMM with an estimated covariance matrix yields the optimal GMM variance within that dictionary. As the population dictionary spaces become dense, their oracle variances approach the within-class bound. Theorem C.16 further establishes feasible attainment along some sufficiently slowly growing deterministic dictionary sequence.

Algorithm 4.2 (Two-step GMM)**.** Given a dictionary $\{(\phi_m, \psi_m) : m = 1, \ldots, M\}$, stack $\eta_n^{(M)}(\theta) := \big(\eta_{1,n}(\theta), \ldots, \eta_{M,n}(\theta)\big)^\top$, $\eta_{m,n}(\theta) := \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \phi_m(Y_i) R_{i,\theta}\psi_m(W)$.

(a) Compute $\hat{\theta}_n^{(1)} \in \arg\min_{\theta \in \Theta} \eta_n^{(M)}(\theta)^\top \eta_n^{(M)}(\theta)$.

(b) Estimate the covariance of $\sqrt{N_n}\, \eta_n^{(M)}(\theta_0)$ by a graph-HAC estimator $\hat{\Omega}_{M,n}$ (Assumption 3.9) evaluated at $\hat{\theta}_n^{(1)}$.

(c) Compute $\hat{\theta}_n^{(2)} \in \arg\min_{\theta \in \Theta} \eta_n^{(M)}(\theta)^\top \hat{\Omega}_{M,n}^{-1} \eta_n^{(M)}(\theta)$; then

$$\sqrt{N_n}(\hat{\theta}_n^{(2)} - \theta_0) \Rightarrow \mathcal{N}\big(0, (G^{(M)\top} \Omega^{(M)-1} G^{(M)})^{-1}\big)$$

as in Theorem 3.5, where $G^{(M)}$ and $\Omega^{(M)}$ are the limiting Jacobian and covariance of the stacked moments (Appendix C.4).

5 Stage 2: Outcome regression given the exposure map

Stage 1 learns the exposure map $g(W; X_i, \hat{\theta}_n)$. In most applications, the object of ultimate interest is a downstream exposure–response or policy parameter, such as the coefficient from an outcome regression on the learned exposure index. This section states the main implication for such second-step inference.

The main case is a regular exposure-map parameter $\theta \in \Theta \subset \mathbb{R}^p$ estimated in Stage 1. Since $\hat{\theta}_n$ is an input into the downstream exposure–response equation, the sampling uncertainty in $\hat{\theta}_n$ must be propagated into second-step inference.

For concreteness, consider the linear exposure–response projection. Define

$$Z_i(\theta) := \begin{pmatrix} 1 \\ g(W; X_i, \theta) \end{pmatrix}.$$

The target (α_0, β_0) is the design-based projection coefficient satisfying

$$\lim_{n\to\infty} \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \mathbb{E}\Big[Z_i(\theta_0)\{Y_i - \alpha_0 - \beta_0^\top g(W; X_i, \theta_0)\} \Big] = 0. \tag{5.1}$$

A correctly specified linear conditional mean is sufficient for (5.1), but the projection interpretation does not require it.

Collect the regular parameters as

$$\zeta = (\theta^\top, \alpha, \beta^\top)^\top.$$

The two-step estimator is characterized by stacking the projected Stage 1 GMM equation with the Stage 2 exposure–response equation:

$$G_1^\top \Lambda_{1,n} \bar{\Psi}_{1,n}(\hat{\theta}_n) = o_p(N_n^{-1/2}), \qquad \bar{\Psi}_{2,n}(\hat{\theta}_n, \hat{\alpha}_n, \hat{\beta}_n) = o_p(N_n^{-1/2}).$$

Under a joint affinity-set CLT, mean differentiability of the stacked population map, and stochastic equicontinuity, Appendix D shows that

$$\sqrt{N_n}(\hat{\zeta}_n - \zeta_0) \Rightarrow \mathcal{N}(0, V_\zeta),$$

where V_ζ is the corresponding design-based sandwich covariance matrix. Inference for the exposure–response coefficient β_0 uses the (β, β) block of V_ζ. More general low-dimensional Stage 2 moments,

including nonlinear regressions or policy-functional estimating equations, are handled by the same stacked-system argument.

Remark 5.1 (Design-based interpretation of the Stage-2 standard errors). The Stage-2 projection moments, unlike the Stage-1 moments (Remark 3.8), are mean zero only in aggregate: (5.1) fixes the average moment at zero but leaves the unit-level means nonzero, and those means are not identified from the realized assignment (each involves $\widetilde{Y}_i$ at unrealized exposures). The feasible sandwich must therefore center at the sample mean, so it need not reproduce the exact design variance of $\hat{\beta}$. When the affinity sets are singletons or block-diagonal the discrepancy is positive semidefinite and the reported standard errors are conservative, with no assumption of correct specification; it vanishes when the linear exposure–response holds unit by unit; and it is of ambiguous sign under general overlapping dependence absent a local-alignment condition. Appendix D.3 makes this precise and states the condition. △

6 Applications

This section applies the framework to two large-scale anti-poverty programs in development economics that explicitly study general-equilibrium (GE) effects: (i) the Smartcards reform of India's National Rural Employment Guarantee Scheme analyzed by Muralidharan et al. (2023); and (ii) the GiveDirectly cash-transfer experiment in rural Kenya studied by Egger et al. (2022), together with the follow-up structural analysis of Walker et al. (2024).

As alluded to in the introduction, the two settings are chosen to be similar ex ante, but we find that they yield contrasting conclusions.

6.1 Revisiting Muralidharan et al. (2023)

We begin with the general-equilibrium effects of India's National Rural Employment Guarantee Scheme (NREGS) studied by Muralidharan et al. (2023). The program guarantees up to 100 days of public employment per year to rural households. Muralidharan et al. (2023) exploit a large randomized rollout of biometric "Smartcards" that improved the implementation of NREGS at the mandal level, and combine direct effects on treated mandals with spillovers to nearby untreated areas.[14]

[14] Administrative units nest as district ⊃ mandal ⊃ Gram Panchayat (GP): a mandal is a sub-district (average population roughly 62,500 in the study sample), and a GP is a village-cluster government comprising one or more census villages. Treatment was randomized at the mandal level, outcomes are measured at the GP (or household) level, and spatial exposure is computed from geocoded 2001 Census village locations (Muralidharan et al., 2023). Our Stage 1 below implements the exposure measure at the census-village level; see Appendix E.1.

Ring-based exposure measure. In their main specification (their equation (1)), Muralidharan et al. (2023) estimate regressions of the form

$$Y_{pmd} \;=\; \alpha \;+\; \beta_0 T_m \;+\; \beta_1 \mathrm{NR}^{20}_{pmd} \;+\; X'_{pmd}\gamma \;+\; \varepsilon_{pmd}, \tag{6.1}$$

where Y_{pmd} is an outcome such as NREGS earnings, wage-labor income, or total income for Gram Panchayat (GP) p in mandal m and district d, T_m is the Smartcard treatment indicator for mandal m, and X_{pmd} are controls. The key spillover variable NR^{20}_{pmd} is the share of GPs in other mandals within 20 km of p that were assigned to treatment; it is a ring-based neighborhood-treatment measure constructed at the GP level using geographic distance, an instance of the ring exposure map of Example 2.1 with radius $\theta = 20$ km.

The headline finding of Muralidharan et al. (2023) is that 86% of beneficiary income gains came from non-program earnings: from higher private-sector wages and employment rather than from NREGS payments themselves, with statistically significant spatial spillovers of wages and employment onto nearby areas. This finding is a decomposition of the total effect on the treated into a direct component (captured by β_0) and a neighborhood component operating through NR^{20}_{pmd}. The authors fix the radius entering that decomposition a priori at 20 km, motivated by commuting distances and the geographic scale of local labor markets, rather than estimate it. The decomposition therefore rests on a choice that is neither estimated nor tested against the design.

Stage 1: letting the data choose the radius. We apply the Stage 1 design-based GMM procedure of Sections 2 and 3, taking the ring radius as the tuning parameter θ. The exposure measure follows Muralidharan et al. (2023)'s construction: for each GP we compute the population-weighted fraction of non-same-mandal treated census villages within radius θ, using the same 6,662-village geometry as the original paper. The design functions are population-weighted annulus averages of the mandal treatment indicator in distance bands around GP i, built from the same village-level distance structure as the exposure. Because physical interaction distances vary continuously in space, we treat the radius as a continuous parameter and search over a fine 0.1 km grid. We report Wald confidence intervals justified by the asymptotic normality result and the design-based overidentification (J) test of Section 3. Full technical details (the two-step criterion, Wald interval construction, and overidentification test) are in Appendix E.1.

Table 2 reports the Stage-1 estimates: the selected radii are 23.7 km for total income, 14.1 km for NREGS earnings, and 25.8 km for wage-labor income. For each outcome the table also gives the Wald standard error and 95% confidence interval for $\hat{\theta}$, and the overidentification statistic $J(\hat{\theta})$ with its p-value. The J statistics are small and none is significant at conventional levels ($p = 0.42$, 0.54, and 0.40), so at the selected radii the design-based moments give no evidence against the ring specification.

Figure 1 plots the Stage-1 criterion against the candidate radius for each outcome, with the

dashed vertical line marking the selected radius $\hat{\theta}$; the criterion is small where the ring's implied design-based moments are close to zero. In all three panels the criterion is high at short radii, falls to a clear interior minimum at the selected radius, and rises beyond it, so the design identifies a finite interaction radius for every outcome. The total-income criterion also has local minima near the two component radii, consistent with total income aggregating a more local and a broader-reaching margin.

Table 2: Stage-1 radius estimates for Muralidharan et al. (2023) outcomes.

Outcome Y	$\hat{\theta}$ (km)	Wald SE (km)	95% Wald CI	$J(\hat{\theta})$ (p-value)
Total income	23.7	1.22	[21.30, 26.10]	6.01 (0.422)
NREGS earnings	14.1	0.64	[12.84, 15.36]	5.05 (0.538)
Wage-labor income	25.8	1.05	[23.75, 27.85]	6.26 (0.395)

Notes: Population-weighted village-level exposure and instruments (3 km bands, 0.1 km radius grid). Wald SE and CI treat the radius as a continuous parameter; see Appendix E.1 for details. The J-statistic is the overidentification test, with 6 degrees of freedom; p-values in parentheses.

Taken together, the exhibits broadly support the exposure choice that Muralidharan et al. (2023) make on institutional grounds: the ring specification is not rejected for any outcome, and the estimated radius for total income, the outcome underlying the decomposition, is 23.7 km, close to the 20 km they adopt a priori (first row of Table 2).

NREGS earnings and wage-labor income decompose total income into a more local and a broader-reaching component (rows two and three of Table 2), and they separate in the direction the authors' own mechanism predicts: NREGS earnings, the direct public-employment margin, localize to 14.1 km, whereas wage-labor income, which operates through private labor markets, reaches farther, to 25.8 km.[15] The total-income radius, 23.7 km, falls between them, and the 20 km choice lies inside this range.

Stage 2: the income-source decomposition. Re-estimating Muralidharan et al. (2023)'s outcome equations at the outcome-specific radii leaves their conclusions essentially intact (Table 3). Total income (9344.2) and wage-labor income (7627.9) are approximately unchanged from the 20 km specification; only NREGS earnings move materially, falling from 1294.6 to 758.7. Muralidharan et al. (2023) attribute 14% of the income gain to program earnings and 86% to non-program earnings. At the design-based radii the program share falls to 8.1% (758.7 of 9344), so non-program

[15] Muralidharan et al. (2023) emphasize that jobcard holders "could only do NREGS work in their own villages," so the public-employment margin should track local program access; the 14.1 km estimate has a tight 95% confidence interval of [12.8, 15.4] km. In their mechanism for wage-labor income, better NREGS implementation raises reservation wages and propagates through markets they describe as "spatially integrated beyond individual GPs or even mandals."

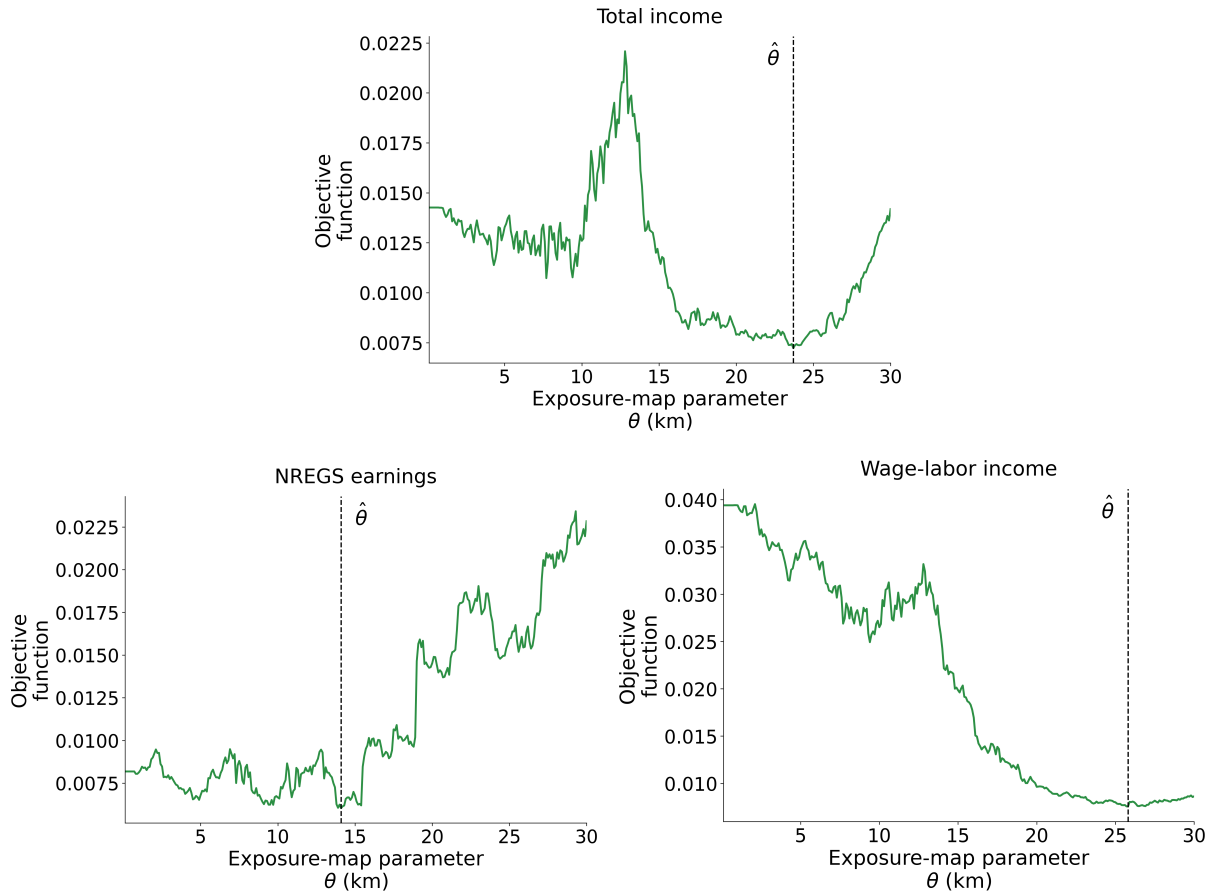


Figure 1: Stage-1 objective functions for total income (top), NREGS earnings (bottom left), and wage-labor income (bottom right). Dashed line: selected radius $\hat{\theta}$. Population-weighted village-level exposure and instruments (3 km bands), strict placebo design.

earnings account for roughly 91.9% of the gain.[16] Their central finding, that the income gains come predominantly from non-program (general-equilibrium) earnings rather than direct program payments, is thus robust to estimating the radius.

The design-based standard errors are close to their fixed-radius counterparts despite additionally propagating the sampling uncertainty in $\hat{\theta}$, so treating the radius as estimated rather than known does not materially inflate the uncertainty around the decomposition.

Table 3: Treatment effects for Muralidharan et al. (2023) outcomes: original 20 km specification vs. design-consistent radii.

Outcome	20 km spec. (MNS)	Design-based ($\hat{\theta}$)	$\hat{\theta}$ (km)	Δ vs. 20 km
Total income	9579.5 (4538.7)	9344.2 (4969.1)	23.7	−235.3
NREGS earnings	1294.6 (1061.0)	758.7 (624.6)	14.1	−535.9
Wage-labor income	7607.2 (2720.7)	7627.9 (2996.3)	25.8	+20.7

Notes: Entries are adjusted treatment effects with standard errors in parentheses. The "20 km spec." column is Muralidharan et al. (2023)'s original fixed-radius specification; the "Design-based" column re-estimates the same Equation (6.1)-style regression (treatment, neighborhood exposure, district fixed effects, strata controls; baseline total income included only for total income) at the outcome-specific radius $\hat{\theta}$ selected in Stage 1. Standard errors are clustered at the mandal level; the design-based standard error additionally propagates the Stage-1 Wald variance of $\hat{\theta}$ via the delta method.

6.2 Revisiting Egger et al. (2022)

Egger et al. (2022) study the GiveDirectly cash-transfer experiment in rural Kenya, in which treatment was randomized both across villages and, through a second-tier saturation design, across sublocations. The two-tier design generates experimental variation in indirect exposure to transfers at the village and local-market level; this is the variation our Stage 1 exploits below.[17]

Their main specification relates household outcomes to the amount transferred to the household's own village and to the amount transferred to other villages within 2 km. The 2 km outer radius, the distance beyond which transfers are assumed to have no effect, is selected by a Bayesian information criterion over concentric 2 km distance bands. We refer to the single included band as their scalar 0–2 km spillover measure; in the language of Section 2, it is a one-band member of the annular exposure-index class we estimate below.

The headline finding of Egger et al. (2022) is a local transfer multiplier of roughly 2.5 (their Table V), together with large positive spillovers onto non-recipient households. The spillover component of that accounting is estimated from the 0–2 km exposure, so the multiplier inherits the

[16] Estimating sources at different radii need not yield shares that sum to one. As an alternative accounting exercise, restricting to the two sources with outcome-specific radii (NREGS and wage labor) gives a program share of 9.0%, of the same small magnitude as the 8.1% in the text.

[17] See Egger et al. (2022) for details on the experimental design and data collection.

support choice; and the support is selected by in-sample predictive fit rather than estimated from, or tested against, the experimental design.

Stage 1. We follow the same two-stage template as in the Muralidharan et al. (2023) application. Let

$$A_v(W) = \left(T^{\neg v}_{v,0-2}(W), T^{\neg v}_{v,2-4}(W), \ldots, T^{\neg v}_{v,18-20}(W)\right)'$$

denote neighboring-village transfer amounts in the ten annuli up to 20 km. The first of these, $T^{\neg v}_{v,0-2}(W)$, is their scalar 0–2 km spillover measure defined above. For a candidate *support* $R_m = 2m$, the radius beyond which transfers are assumed to have no causal effect on a household, Stage 1 forms the linear spillover exposure index

$$g^{(m)}_v(W;\theta_m) = \sum_{j=1}^{m} \theta_{m,j} T^{\neg v}_{v,2(j-1)-2j}(W), \qquad \mathbf{1}'\theta_m = 1,$$

and estimates θ_m from the design-based exposure-sufficiency moments.[18] In this notation, the exposure map used in Stage 1 is

$$\left(T^{\text{own}}_v(W), g^{(m)}_v(W;\theta_m)\right).$$

Stage 2. Stage 2 returns to Egger et al. (2022)'s original IV specification, as in the Muralidharan et al. (2023) application, replacing their scalar 0–2 km spillover measure with the estimated index; standard errors account for the generated weights $\widehat{\theta}_m$ by the delta method. We keep the same outcome equation and instrumenting logic:

$$Y_{iv} = \alpha + \beta_{\text{own},m} T^{\text{own}}_v(W) + \gamma_m g^{(m)}_v(W;\widehat{\theta}_m) + X'_{iv}\delta + \varepsilon_{iv}.$$

Figure 2 traces the Stage-2 recipient and non-recipient effects as the support widens from the authors' 2 km to the fully flexible 20 km: the recipient point estimates are stable across the entire path, while the non-recipient effects attenuate and lose precision until, by 20 km, they are generally no longer distinguishable from zero.[19] The two endpoints anchor the comparison. At $R = 2$ km the normalized index collapses to their single 0–2 km ring ($\theta_{1,1} = 1$), so the leftmost point replicates the original Egger et al. (2022) specification; at $R = 20$ km the index imposes no support restriction across the ten annuli; each interior point imposes, ex ante, that transfers beyond R km do not

[18] The scale of θ_m is immaterial for exposure sufficiency: replacing θ_m by $c\theta_m$, with $c \neq 0$, only rescales the downstream coefficient on the index. We impose $\mathbf{1}'\theta_m = 1$ and let the Stage 2 coefficient absorb scale. When $m = 1$, this normalization implies $\theta_{1,1} = 1$. Thus the 2 km case coincides with the original Egger et al. one-ring exposure measure, and Stage 1 has no nontrivial shape parameter to estimate.

[19] The confidence intervals along both paths do not necessarily widen monotonically with the support; Appendix E.2, part D, discusses potential explanations, including collinearity among the annular transfer variables at particular supports.

affect the household. The contrast between the two paths is what one would expect: recipients respond mainly to the direct, own-village transfer, which does not depend on the support, whereas non-recipients respond only through spillovers, which do. The lower panel of each plot reports the p-value path of the support test introduced next.

Choosing the support by a design-based test. Neither endpoint can serve as the downstream specification: 2 km is an untested restriction, and the fully flexible 20 km index estimates the pure-spillover effects too imprecisely to be informative. We therefore select the smallest support the design does not reject. Exposure sufficiency makes the support a testable restriction: if the annuli beyond R_m carry no outcome-relevant assignment variation, zeroing their coefficients leaves the design-based moments satisfied, which is the restriction the overidentification test of Section 3 (Corollary 3.6) evaluates. We test the unrestricted 20 km index against the restricted index that zeroes all annuli beyond R_m, implemented as a likelihood-ratio statistic along the path (lower panels of Figure 2), and select the smallest non-rejected support. This is the same parsimony goal behind the Bayesian information criterion Egger et al. (2022) use to set their outer radius; the difference is that the BIC scores cross-sectional predictive fit, whereas the design-based test scores consistency with the randomization, the identifying assumption the downstream estimates rely on.[20]

This procedure rejects the original 2 km support for all three core outcomes: in every lower panel of Figure 2, the p-value path lies below 0.05 at 2 km and crosses above it by 4 km for assets and household income and by 6 km for household expenditure. The smallest non-rejected support is 6 km for household expenditure and 4 km for assets and household income; across the broader set of outcomes in Appendix E.2 the selected supports are 4–6 km, occasionally larger. The 2 km exposure is thus too local from the design's perspective.

Table 4 isolates what the support choice does, since the exposure measure is the only thing that changes between the two columns; the change bites on the non-recipient margin only. The recipient total effects survive the change of exposure map: all three point estimates move by less than half of Egger et al. (2022)'s own standard error. The non-recipient side is where the exposure choice bites: every pooled spillover attenuates by about one to one and a half of Egger et al. (2022)'s standard errors. Expenditure falls from 334.7 to 224.9 and remains significant; income falls from 225.0 to 81.1 and is no longer distinguishable from zero at the 5% level; assets, not significant in

[20] As a check on post-selection and spatial-dependence concerns, the diamonds in Figure 2 report a sublocation-level sample-splitting exercise. We split at the level of *sublocations* (the higher administrative tier at which the saturation design randomizes treatment intensity, each containing many villages and a local market) rather than at the village or household level: because cash-transfer spillovers operate through the local market (Walker et al., 2024), dependence is contained within a sublocation and is approximately negligible across them, so assigning whole sublocations to the selection and estimation subsamples makes the two approximately independent while leaving the within-sublocation spillovers the exposure map captures intact. We aggregate across repeated splits following the median procedure of Chernozhukov et al. (2025): we report the median point estimate across splits and form the band from the median lower and median upper confidence limits, yielding an approximately 90% interval. The split estimates are noisier but point the same way: 2 km is too local and the selected supports are modestly larger.

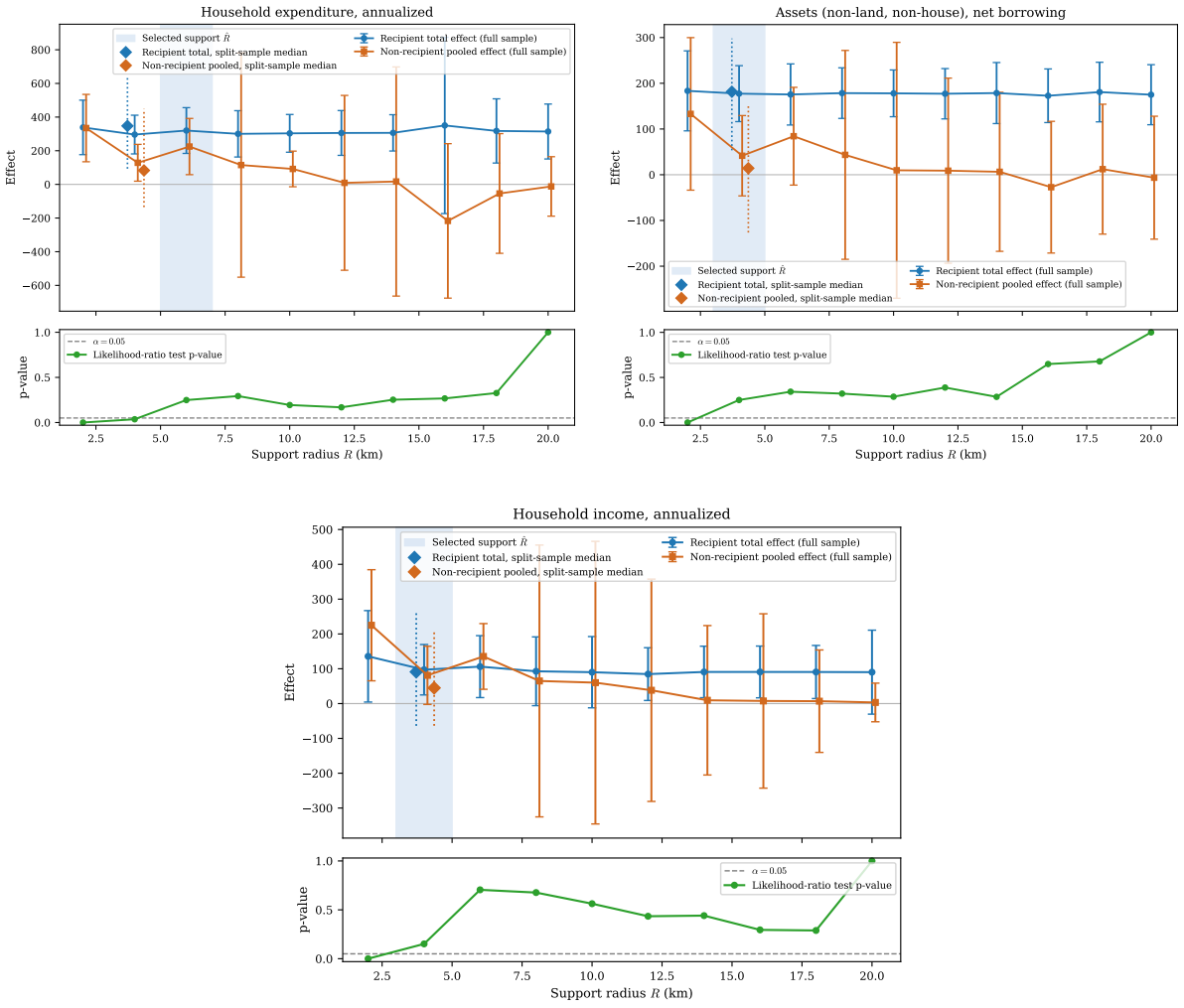


Figure 2: Recipient and non-recipient effects across candidate supports, Egger et al. (2022) application

Notes: The figure plots recipient total effects and non-recipient pooled spillover effects across candidate supports $R = 2, \ldots, 20$ km. The shaded band marks the selected support. The lower panel reports the likelihood-ratio test p-value path, and the dashed horizontal line marks $\alpha = 0.05$. Diamonds report sublocation-level split-sample medians with their median intervals (dotted whiskers). The appendix reports additional outcome diagnostics.

the original specification, falls from 133.1 to 41.6.[21] Relative to the Muralidharan et al. (2023) application, then, the exposure-map choice does matter here.

The last column of Table 4 reports the overidentification (J) test of the selected linear-index specification. The test rejects for expenditure and assets ($p = 0.007$ and 0.001) and does not reject for income ($p = 0.134$).

Table 4: Core IV components: Egger et al. vs. design-based estimates

	Egger IV	Design-based IV	$\Delta\beta$, ΔSE	J p-val
Household expenditure, annualized				
Recipients	338.6 (100.6)	320.0 (69.4)	−18.6, −31.2	0.007
Non-recipients	334.7 (131.1)	224.9 (85.2)	−109.8, −45.9	0.007
Assets (non-land, non-house), net borrowing				
Recipients	183.4 (55.2)	177.3 (31.2)	−6.1, −24.0	0.001
Non-recipients	133.1 (101.3)	41.6 (44.8)	−91.5, −56.5	0.001
Household income, annualized				
Recipients	135.7 (82.1)	97.2 (36.9)	−38.5, −45.1	0.134
Non-recipients	225.0 (100.1)	81.1 (42.6)	−143.8, −57.4	0.134

Notes: Entries are IV point estimates with standard errors in parentheses. Egger IV values re-estimate the original 2 km Table 1 specification on our replication sample (cluster standard errors); they closely match the published Table I values. Design-based IV values use each outcome's selected design-based support ($\hat{R} = 6.0$ km for household expenditure, 4.0 km for assets and income); the design-side conditional expectations entering the index are estimated by cross-validated ridge regression on a quadratic polynomial in the exposure index, using 200 placebo draws of the assignment. Recipient entries report total effects; non-recipient entries report pooled spillover effects, both with joint delta-method standard errors. J p-val is the design-GMM overidentification p-value for the selected support; the support-selection rule is the likelihood-ratio test.

The spillover attenuation and the residual J rejection are both consistent with a market-level "ambient" component in the spillovers, a component that a fit-based radius rule misses. Walker et al. (2024) argue that local GE responses include a market-level "ambient" or "slack" component:

$$Y_i \;=\; \alpha \;+\; \beta^{\text{loc}} g^{\text{loc}}(W; X_i, R_0) \;+\; \underbrace{A_{m(i)}(W)}_{\text{market-level ambient}} \;+ u_i,$$

where g^{loc} captures very local spillovers and $A_{m(i)}(W)$ is common to households in local market $m(i)$. A component that is approximately common within a market is largely absorbed by the intercept and controls, so outer annuli that mainly proxy $A_{m(i)}(W)$ barely reduce residual variance and the BIC prefers short radii, even when the beyond-2 km transfers generating that component move the total general-equilibrium response; and no annular index spans such a component at any support, which is why the J test still rejects at $\hat{R}$.[22] The design-based diagnostic instead scores

[21] The "Egger IV" column re-estimates the original 2 km specification on our replication sample; the estimates closely match the published Table 1 values (see the table notes).

[22] Egger et al. (2022) choose the radius r by minimizing $\text{BIC}(r) = N \log \hat{\sigma}_r^2 + k(r) \log N$, where $\hat{\sigma}_r^2$ is the residual variance from regressing Y_i on the radius-r ring exposures and controls, $k(r)$ is the number of parameters, and N is

the radius by consistency with the randomization rather than by residual fit, so it retains those annuli and extends the support. In addition, broader market-level forces such as inflation or price responses can offset part of the measured income and non-recipient gains. Together these explain why the support diagnostic moves beyond 2 km even as some downstream estimates become smaller or less precise.

Implications for multiplier accounting. Recomputing Egger et al. (2022)'s Table V multiplier accounting with the design-selected indexes gives a mean local transfer multiplier of about 1.60, well below the original headline value near 2.5 and close to the 1.54 implied by Walker et al. (2024)'s calibrated general-equilibrium model (Table 5). The expenditure and income components are about 1.63 and 1.57. We obtain standard errors by a delta method that propagates both the Stage 1 uncertainty in the index weights and the Table V coefficient uncertainty. With only 84 sublocation clusters these multipliers are imprecisely estimated, with standard errors comparable to Egger et al. (2022)'s own; the mean multiplier carries a standard error of 1.10, so although its point estimate lies well below 2.5, it is not statistically distinguishable from one. The Walker et al. (2024) benchmark is on the same real (deflated) basis as Egger et al. (2022)'s Table V. Appendix E.2, part E details the reparameterization and the delta-method inference.

Table 5: Local transfer multiplier: design-based estimates and benchmarks

	Design-based	Egger et al. (2022) Table V	Walker et al. (2024) model
Expenditure	1.63 (1.49)	2.58 (1.44)	–
Income	1.57 (0.97)	2.47 (1.71)	–
Mean	1.60 (1.10)	2.52 (1.39)	1.54

Notes: Entries are multiplier point estimates with standard errors in parentheses, all on a real basis: transfers and outcomes are deflated to January 2015 USD, matching Egger et al. (2022)'s Table V. Our design-based estimates recompute that deflated accounting after replacing each ring exposure with the design-selected index; their standard errors are delta-method values that propagate the Stage-1 uncertainty in the index weights and the Table V coefficient uncertainty (Appendix E.2, part E). The Egger et al. (2022) column reproduces their Table V estimates, whose standard errors come from a wild clustered bootstrap. The Walker et al. (2024) figure is a model-implied real multiplier from a calibrated general-equilibrium model, with no sampling standard error and reported only in aggregate (their nominal counterpart is 1.83).

The second application thus illustrates the revision use of the framework: when the design rejects the original exposure map, both the map and the headline policy estimate change materially.

the number of villages: the radius is scored by how much the rings reduce residual variance, traded off against the penalty.

7 Conclusion

Estimates of spillover effects depend on an exposure map that turns the realized assignment into an index of each unit's exposure. Both the functional form of that map (e.g., a ring) and its parameter (e.g., the radius) are usually fixed before estimation, with little guidance for either choice. This paper shows that the same randomization that identifies treatment effects can guide both choices. Specifically, a correctly specified exposure map implies orthogonality conditions that can be used for both estimation and testing: design-based GMM can estimate the mapping parameter, and the same conditions, when overidentified, can test the map itself.

We establish consistency and asymptotic normality of the resulting estimator for the maps used in applied work, with all randomness arising from the assignment, and we characterize the efficient moments. Finally, we carry the resulting uncertainty about $g(W; X_i, \theta)$ into downstream policy estimands (e.g., average effects under alternative assignment rules), so that inference reflects both the experimental variation and the uncertainty about the exposure specification itself.

We apply the method to two large-scale anti-poverty experiments. For the NREGS reform studied by Muralidharan et al. (2023), the 20 km radius chosen on institutional grounds is not rejected by the design, and the program's headline decomposition is robust: even at the data-chosen radius, the income gains come predominantly from non-program earnings. For the cash-transfer experiment of Egger et al. (2022), the 2 km support is rejected for every core outcome, and replacing it with the smallest support the design does not reject lowers the estimated local transfer multiplier (mean) from about 2.5 to 1.60, close to the value implied by the independent structural model of Walker et al. (2024).

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1):265–296.

Ackerberg, D. and Crawford, G. S. (2009). Estimating price elasticities in differentiated product demand models with endogenous characteristics. Working paper.

Ackerberg, D., Crawford, G. S., and Hahn, J. (2011). Orthogonal instruments: Estimating price elasticities in the presence of endogenous product characteristics. Presentation at CREST, Paris.

Adão, R., Costinot, A., and Donaldson, D. (2023). Putting quantitative models to the test: An application to Trump's trade war. NBER Working Paper 31321, National Bureau of Economic Research.

Andrews, I., Barahona, N., Gentzkow, M., Rambachan, A., and Shapiro, J. M. (2025). Structural

estimation under misspecification: Theory and implications for practice. *The Quarterly Journal of Economics*, page qjaf018.

Aronow, P. M. and Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947.

Athey, S., Eckles, D., and Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240.

Belloni, A., Fang, F., and Volfovsky, A. (2022). Neighborhood adaptive estimators for causal inference under network interference. *arXiv preprint arXiv:2212.03683*.

Borusyak, K., Bravo, M. C., and Hull, P. (2025). Estimating demand with recentered instruments. *arXiv preprint arXiv:2504.04056*.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Borusyak, K. and Hull, P. (2026). Optimal formula instruments. *Econometrica*. Forthcoming.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.

Cai, J., Janvry, A. D., and Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108.

Chandrasekhar, A. G., Jackson, M. O., McCormick, T. H., and Thiyageswaran, V. (2023). General covariance-based conditions for central limit theorems with dependent triangular arrays. *arXiv preprint arXiv:2308.12506*.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2025). Fisher–Schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. *Econometrica*, 93(4):1121–1164.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

Currie, J., Davis, L., Greenstone, M., and Walker, R. (2015). Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings. *American Economic Review*, 105(2):678–709.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., and Walker, M. (2022). General equilibrium effects of cash transfers: experimental evidence from kenya. *Econometrica*, 90(6):2603–2643.

Faridani, S. and Niehaus, P. (2024). Linear estimation of global average treatment effects. Technical report, National Bureau of Economic Research.

Franklin, S., Imbert, C., Abebe, G., and Mejia-Mantilla, C. (2024). Urban public works in spatial equilibrium: Experimental evidence from ethiopia. *American Economic Review*, 114(5):1382–1414.

Freedman, D. A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2):180–193.

Gao, C., Harshaw, C., Sävje, F., and Wang, Y. (2026). On the impossibility of specification testing of interference models based on exposure mappings. Version v1, May 12, 2026.

Glaeser, E. L., Sacerdote, B. I., and Scheinkman, J. A. (2003). The social multiplier. *Journal of the European Economic Association*, 1(2-3):345–353.

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

Kim, M.-j. and Sun, Y.-J. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation. *Econometric Theory*, 27(5):879–921.

Kojevnikov, D., Marmer, V., and Song, K. (2021). Limit theorems for network dependent random variables. *Journal of Econometrics*, 225(2):141–163.

Leung, M. P. (2022). Causal inference under approximate neighborhood interference. *Econometrica*, 90(1):267–293.

Li, X. and Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520):1759–1769.

Lucas, Robert E., J. (1980). Methods and problems in business cycle theory. *Journal of Money, Credit and Banking*, 12(4):696–715.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149):976–980.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23.

Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Muralidharan, K., Niehaus, P., and Sukhtankar, S. (2023). General equilibrium effects of (improving) public employment programs: Experimental evidence from India. *Econometrica*, 91(4):1261–1295.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. *Ann. Agricultural Sciences*, pages 1–51.

Park, Y. and Yang, X. (2026). Decomposition of spillover effects under misspecification: Pseudo-true estimands and a local-global extension.

Rambachan, A. and Roth, J. (2025). Design-based uncertainty for quasi-experiments. *Journal of the American Statistical Association*. Forthcoming; published online August 27, 2025.

Redding, S. and Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, 62(1):53–82.

Ritzwoller, D. M. (2025). Regression adjustments for disentangling spillover effects. Job market paper, Stanford Graduate School of Business.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: journal of the Econometric Society*, pages 931–954.

Sasaki, Y. (2025). Gmm and m estimation under network dependence. *arXiv preprint arXiv:2503.00290*.

Sävje, F. (2024). Causal inference with misspecified exposure mappings: Separating definitions and assumptions. *Biometrika*, 111(1):1–15.

Sävje, F., Aronow, P. M., and Hudgens, M. G. (2021). Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, 49(2):673–701.

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.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Walker, M. W., Shah, N., Miguel, E., Egger, D., Soliman, F. S., and Graff, T. (2024). Slack and economic development. Technical report, National Bureau of Economic Research.

Zhong, L. (2025). Unconditional randomization tests for interference. Version v3, July 15, 2025.

A Proofs for Section 2

Proof of Theorem 2.6 and Corollary 2.7. Fix a unit i, and write

$$G_i(\theta_0) := g(W; X_i, \theta_0).$$

By Hypothesis 2.5, there exists $\theta_0 \in \Theta$ and a measurable map $\widetilde{Y}_i : \mathbb{R}^k \to \mathbb{R}$ such that, for every assignment $w \in \mathcal{W}$,

$$Y_i(w) = \widetilde{Y}_i\big(g(w; X_i, \theta_0)\big).$$

Under the design $\mathcal{D}_n$, the realized outcome is $Y_i = Y_i(W)$. Hence

$$Y_i = \widetilde{Y}_i\big(g(W; X_i, \theta_0)\big) = \widetilde{Y}_i\big(G_i(\theta_0)\big) \quad \text{a.s.}$$

By Assumption 2.4, the potential-outcome schedule is fixed under $\mathcal{D}_n$, so each map $\widetilde{Y}_i$ is nonrandom. Thus Y_i is measurable with respect to $\sigma(G_i(\theta_0))$.

We first verify the conditional independence statement. Let $f : \mathbb{R} \to \mathbb{R}$ and $h : \mathcal{W} \to \mathbb{R}$ be bounded measurable functions. Since $Y_i = \widetilde{Y}_i(G_i(\theta_0))$ almost surely,

$$f(Y_i) = f\Big(\widetilde{Y}_i(G_i(\theta_0))\Big)$$

is $\sigma(G_i(\theta_0))$-measurable. Therefore,

$$\begin{aligned}\mathbb{E}\big[f(Y_i)h(W) \,\big|\, G_i(\theta_0)\big] &= f\Big(\widetilde{Y}_i(G_i(\theta_0))\Big)\,\mathbb{E}\big[h(W) \,\big|\, G_i(\theta_0)\big] \\ &= \mathbb{E}\big[f(Y_i) \,\big|\, G_i(\theta_0)\big]\,\mathbb{E}\big[h(W) \,\big|\, G_i(\theta_0)\big].\end{aligned}$$

This verifies

$$Y_i \;\perp\!\!\!\perp\; W \;\big|\; g(W; X_i, \theta_0)$$

under the design distribution.

We next verify the orthogonality condition. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a square-integrable outcome transformation, $\mathbb{E}[\phi(Y_i)^2] < \infty$, and let $\psi : \mathcal{W} \times \mathcal{X} \to \mathbb{R}$ be a square-integrable design function, $\mathbb{E}[\psi(W)^2] < \infty$. By definition,

$$R_{i,\theta_0}\psi(W) = \psi(W) - \mathbb{E}[\psi(W) \mid G_i(\theta_0)].$$

Hence

$$\mathbb{E}[R_{i,\theta_0}\psi(W) \mid G_i(\theta_0)] = 0 \quad \text{a.s.}$$

Moreover, since Y_i is $\sigma(G_i(\theta_0))$-measurable, $\phi(Y_i)$ is also $\sigma(G_i(\theta_0))$-measurable. Because conditional expectation is an L^2 contraction, $\psi(W) \in L^2$ implies $R_{i,\theta_0}\psi(W) = \psi(W) - \mathbb{E}[\psi(W) \mid G_i(\theta_0)] \in$

L^2; together with $\phi(Y_i) \in L^2$, the Cauchy–Schwarz inequality gives $\mathbb{E}\big[\,|\phi(Y_i)\, R_{i,\theta_0}\psi(W)|\,\big] \le \|\phi(Y_i)\|_{L^2}\, \|R_{i,\theta_0}\psi(W)\|_{L^2} < \infty$, so the product is integrable and the law of iterated expectations applies below. The law of iterated expectations gives

$$\begin{aligned}\mathbb{E}[\phi(Y_i)R_{i,\theta_0}\psi(W)] &= \mathbb{E}\big[\mathbb{E}\big[\phi(Y_i)R_{i,\theta_0}\psi(W)\,\big|\,G_i(\theta_0)\big]\big] \\ &= \mathbb{E}\big[\phi(Y_i)\mathbb{E}\big[R_{i,\theta_0}\psi(W)\,\big|\,G_i(\theta_0)\big]\big] \\ &= 0.\end{aligned}$$

Equivalently,

$$\mathbb{E}[\phi(Y_i)\,\{\psi(W) - \mathbb{E}[\psi(W) \mid g(W; X_i, \theta_0)]\}] = 0,$$

which is the desired unit-level design-based orthogonality condition.

Since i, ϕ, and ψ were arbitrary, the result holds for every unit and every admissible outcome transformation and design function. This completes the proof. □

B Additional Results and Proofs for Section 3

Throughout this appendix we work under the notation and conditions of Section 3. In particular, for each n and $i \in \mathcal{U}_n$ we define, for $\theta \in \Theta$,

$$Z_{i,n}(\theta) := \Psi_i(\theta) - \mathbb{E}[\Psi_i(\theta)], \qquad \bar{Z}_n(\theta) := \frac{1}{N_n}\sum_{i=1}^{N_n} Z_{i,n}(\theta),$$

and

$$\mu_n(\theta) := \frac{1}{N_n}\sum_{i=1}^{N_n} \mathbb{E}[\Psi_i(\theta)], \qquad \mu(\theta) := \lim_{n\to\infty} \mu_n(\theta),$$

whenever the limit exists. The GMM criterion and its limit are

$$Q_n(\theta) := \bar{\Psi}_n(\theta)^\top \Lambda_n\, \bar{\Psi}_n(\theta), \qquad Q(\theta) := \mu(\theta)^\top \Lambda\, \mu(\theta),$$

with $\Lambda_n \xrightarrow{p} \Lambda \succeq 0$. Assumption 3.1 in the main text encodes the high-level uniform LLN, identification, and weight convergence conditions needed for consistency.

The main text focuses on the regular continuous exposure-map case. This appendix gives the formal supporting material for the baseline regular continuous case and then records the finite-grid variants used in applications.

B.1 Consistency of the baseline regular case

Uniform convergence of the criterion. The key step is to show that the high-level uniform LLN for the moments implies uniform convergence of the quadratic criterion $Q_n(\theta)$ to $Q(\theta)$.

Lemma B.1 (Uniform convergence of the quadratic criterion)**.** *Suppose Assumption 3.1 holds. Then*

$$\sup_{\theta\in\Theta}\left|Q_n(\theta)-Q(\theta)\right| \xrightarrow{p} 0.$$

Proof. Define, for each n and $\theta\in\Theta$,

$$\Delta_n(\theta) := \bar{\Psi}_n(\theta)-\mu_n(\theta) = \bar{Z}_n(\theta).$$

Then

$$\bar{\Psi}_n(\theta) = \mu_n(\theta)+\Delta_n(\theta).$$

Substituting into $Q_n(\theta)$ yields

$$\begin{aligned} Q_n(\theta) &= \bar{\Psi}_n(\theta)^\top \Lambda_n \bar{\Psi}_n(\theta) \\ &= \big(\mu_n(\theta)+\Delta_n(\theta)\big)^\top \Lambda_n \big(\mu_n(\theta)+\Delta_n(\theta)\big) \\ &= \Delta_n(\theta)^\top \Lambda_n \Delta_n(\theta) + 2\,\mu_n(\theta)^\top \Lambda_n \Delta_n(\theta) + \mu_n(\theta)^\top \Lambda_n \mu_n(\theta). \end{aligned}$$

Subtracting $Q(\theta)=\mu(\theta)^\top\Lambda\mu(\theta)$ and adding and subtracting $\mu_n(\theta)^\top\Lambda\mu_n(\theta)$ gives

$$\begin{aligned} Q_n(\theta)-Q(\theta) = &\underbrace{\Delta_n(\theta)^\top \Lambda_n \Delta_n(\theta)}_{T_{1,n}(\theta)} + \underbrace{2\,\mu_n(\theta)^\top \Lambda_n \Delta_n(\theta)}_{T_{2,n}(\theta)} \\ &+ \underbrace{\mu_n(\theta)^\top(\Lambda_n-\Lambda)\mu_n(\theta)}_{T_{3,n}(\theta)} + \underbrace{\left\{\mu_n(\theta)^\top\Lambda\mu_n(\theta)-\mu(\theta)^\top\Lambda\mu(\theta)\right\}}_{T_{4,n}(\theta)}. \end{aligned} \tag{B.1}$$

We bound each term $T_{k,n}(\theta)$ uniformly in $\theta\in\Theta$.

First term. For each θ,

$$\left|\Delta_n(\theta)^\top \Lambda_n \Delta_n(\theta)\right| \le \|\Lambda_n\|_{\mathrm{op}}\,\|\Delta_n(\theta)\|^2,$$

hence

$$\sup_{\theta\in\Theta}\left|\Delta_n(\theta)^\top \Lambda_n \Delta_n(\theta)\right| \le \|\Lambda_n\|_{\mathrm{op}}\Big(\sup_{\theta\in\Theta}\|\Delta_n(\theta)\|\Big)^2.$$

By Assumption 3.1 (AS–LLN2),

$$\sup_{\theta\in\Theta}\|\Delta_n(\theta)\| = \sup_{\theta\in\Theta}\left\|\bar{\Psi}_n(\theta)-\mu_n(\theta)\right\| \xrightarrow{p} 0,$$

and by Assumption 3.1 (AS–LLN4), $\Lambda_n \xrightarrow{p} \Lambda$ implies $\|\Lambda_n\|_{\mathrm{op}} = O_p(1)$. Therefore

$$\sup_{\theta\in\Theta} \big|T_{1,n}(\theta)\big| = o_p(1).$$

Second term. Similarly,

$$\big|\mu_n(\theta)^\top \Lambda_n \Delta_n(\theta)\big| \le \|\mu_n(\theta)\|\, \|\Lambda_n\|_{\mathrm{op}}\, \|\Delta_n(\theta)\|,$$

so

$$\sup_{\theta\in\Theta} \big|T_{2,n}(\theta)\big| \le 2\Big(\sup_{\theta\in\Theta} \|\mu_n(\theta)\|\Big)\|\Lambda_n\|_{\mathrm{op}}\Big(\sup_{\theta\in\Theta} \|\Delta_n(\theta)\|\Big).$$

By Assumption 3.1 (AS–LLN1), $\mu_n(\theta) \to \mu(\theta)$ uniformly on Θ. Continuity of μ and compactness of Θ imply

$$\sup_{\theta\in\Theta} \|\mu(\theta)\| < \infty, \qquad \sup_{\theta\in\Theta} \|\mu_n(\theta)\| = O(1).$$

Combining with Assumption 3.1 (AS–LLN2) and (AS–LLN4), we obtain

$$\sup_{\theta\in\Theta} \big|T_{2,n}(\theta)\big| = o_p(1).$$

Third term. For each θ,

$$\big|\mu_n(\theta)^\top (\Lambda_n - \Lambda)\mu_n(\theta)\big| \le \|\mu_n(\theta)\|^2\, \|\Lambda_n - \Lambda\|_{\mathrm{op}},$$

hence

$$\sup_{\theta\in\Theta} \big|T_{3,n}(\theta)\big| \le \Big(\sup_{\theta\in\Theta} \|\mu_n(\theta)\|\Big)^2 \|\Lambda_n - \Lambda\|_{\mathrm{op}}.$$

By Assumption 3.1 (AS–LLN4), $\|\Lambda_n - \Lambda\|_{\mathrm{op}} \xrightarrow{p} 0$, and as above $\sup_{\theta\in\Theta} \|\mu_n(\theta)\| = O(1)$. Thus

$$\sup_{\theta\in\Theta} \big|T_{3,n}(\theta)\big| = o_p(1).$$

Fourth term. For each θ,

$$\begin{aligned}
&\mu_n(\theta)^\top \Lambda \mu_n(\theta) - \mu(\theta)^\top \Lambda \mu(\theta) \\
&\quad = \big(\mu_n(\theta) - \mu(\theta)\big)^\top \Lambda \mu_n(\theta) + \mu(\theta)^\top \Lambda \big(\mu_n(\theta) - \mu(\theta)\big),
\end{aligned}$$

so

$$\big|\mu_n(\theta)^\top \Lambda \mu_n(\theta) - \mu(\theta)^\top \Lambda \mu(\theta)\big|$$

$$\leq \|\mu_n(\theta) - \mu(\theta)\| \, \|\Lambda\|_{\mathrm{op}} \, \|\mu_n(\theta)\| + \|\mu(\theta)\| \, \|\Lambda\|_{\mathrm{op}} \, \|\mu_n(\theta) - \mu(\theta)\|.$$

Taking suprema over $\theta \in \Theta$ and using Assumption 3.1 (AS–LLN1), together with boundedness of $\sup_\theta \|\mu_n(\theta)\|$ and $\sup_\theta \|\mu(\theta)\|$, we obtain

$$\sup_{\theta \in \Theta} \left| T_{4,n}(\theta) \right| = \sup_{\theta \in \Theta} \left| \mu_n(\theta)^\top \Lambda \mu_n(\theta) - \mu(\theta)^\top \Lambda \mu(\theta) \right| \to 0.$$

Combining the four bounds with (B.1), we conclude that

$$\sup_{\theta \in \Theta} \left| Q_n(\theta) - Q(\theta) \right| \leq \sum_{k=1}^{4} \sup_{\theta \in \Theta} |T_{k,n}(\theta)| = o_p(1).$$

This establishes the desired uniform convergence. □

Argmin consistency and proof of Theorem 3.3. We now use Lemma B.1 and the identification part of Assumption 3.1 to show that any minimizer of $Q_n(\theta)$ is consistent.

Lemma B.2 (Argmin consistency)**.** *Suppose Assumption 3.1 holds and Lemma B.1 is in force. Then any $\hat{\theta}_n \in \arg\min_{\theta \in \Theta} Q_n(\theta)$ satisfies $\hat{\theta}_n \xrightarrow{p} \theta_0$.*

Proof. By Assumption 3.1 (AS–LLN3), Θ is compact and $\mu(\theta)$ is continuous in θ on Θ. Since $\Lambda \succeq 0$ is fixed, $Q(\theta) = \mu(\theta)^\top \Lambda \mu(\theta)$ is continuous on Θ. Assumption 3.1 (AS–LLN3) further implies that Q is uniquely minimized at θ_0.

Fix $\eta > 0$ and define

$$S_\eta := \{\theta \in \Theta : \|\theta - \theta_0\| \geq \eta\}.$$

Compactness of Θ implies S_η is compact. By continuity and uniqueness of the minimizer,

$$\delta_\eta := \inf_{\theta \in S_\eta} \left\{ Q(\theta) - Q(\theta_0) \right\} > 0.$$

By Lemma B.1,

$$\sup_{\theta \in \Theta} \left| Q_n(\theta) - Q(\theta) \right| \xrightarrow{p} 0.$$

Hence, for $\varepsilon = \delta_\eta / 4$,

$$\mathbb{P}\left(\sup_{\theta \in \Theta} \left| Q_n(\theta) - Q(\theta) \right| > \frac{\delta_\eta}{4} \right) \to 0.$$

For all large n, with probability at least $1 - o(1)$, the event

$$\mathcal{A}_{n,\eta} := \left\{ \sup_{\theta \in \Theta} \left| Q_n(\theta) - Q(\theta) \right| \leq \frac{\delta_\eta}{4} \right\}$$

occurs. On $\mathcal{A}_{n,\eta}$ we have, for every $\theta \in S_\eta$,

$$Q_n(\theta) \geq Q(\theta) - \frac{\delta_\eta}{4} \geq Q(\theta_0) + \delta_\eta - \frac{\delta_\eta}{4} = Q(\theta_0) + \frac{3\delta_\eta}{4},$$

while

$$Q_n(\theta_0) \leq Q(\theta_0) + \frac{\delta_\eta}{4}.$$

Therefore, on $\mathcal{A}_{n,\eta}$,

$$Q_n(\theta) > Q_n(\theta_0) \qquad \text{for all } \theta \in S_\eta,$$

so no minimizer of Q_n can lie in S_η. In particular,

$$\mathbb{P}\big(\hat{\theta}_n \in S_\eta\big) \leq \mathbb{P}(\mathcal{A}_{n,\eta}^c) \to 0.$$

Since $\eta > 0$ is arbitrary, this implies $\hat{\theta}_n \xrightarrow{p} \theta_0$. □

Proof of Theorem 3.3. Under Assumption 3.1 (AS–LLN1)–(AS–LLN4), Lemma B.1 gives

$$\sup_{\theta \in \Theta} \big|Q_n(\theta) - Q(\theta)\big| \xrightarrow{p} 0.$$

Assumption 3.1 (AS–LLN3) ensures that Q is continuous on Θ and uniquely minimized at θ_0. Applying Lemma B.2 then yields $\hat{\theta}_n \xrightarrow{p} \theta_0$ for any sequence of minimizers $\hat{\theta}_n \in \arg\min_{\theta \in \Theta} Q_n(\theta)$. This proves the consistency result stated in Theorem 3.3. □

B.2 Primitive LLN via affinity sets

This subsection verifies Assumption 3.1 (AS–LLN2) from primitive design-based conditions. The only substantive residual remains the Section 2 design residual

$$R_{i,\theta}\psi(W) = \psi(W) - \mathbb{E}[\psi(W) \mid g_i(W;\theta)].$$

For the proofs below, use the standard centered-moment notation

$$Z_{i,n}(\theta) := \Psi_i(\theta) - \mathbb{E}\Psi_i(\theta), \qquad \bar{Z}_n(\theta) := \frac{1}{N_n} \sum_{i=1}^{N_n} Z_{i,n}(\theta).$$

Thus the desired ULLN is simply

$$\sup_\theta \|\bar{Z}_n(\theta)\| \xrightarrow{p} 0.$$

All expectations and covariances in this subsection are with respect to the known design $\mathcal{D}_n$, conditional on the realized finite population.

B.2.1 Smooth exposure maps

The smooth case covers gravity and market-access exposure maps. We state projection regularity directly because smoothness of $g_i(W;\theta)$ alone does not imply smoothness of $\mathbb{E}[\psi(W) \mid g_i(W;\theta)]$ as θ varies.

Assumption B.3 (Smooth-kernel ULLN primitives)**.** *Let $\Theta \subset \mathbb{R}^p$ be compact and let the moment dictionary $\{(\phi_m, \psi_m) : m = 1, \dots, q\}$ be fixed. Write*

$$\Psi_{i,m}(\theta) := \phi_m(Y_i)\{\psi_m(W) - m_{i,m,\theta}(g_i(W;\theta))\}, \qquad m_{i,m,\theta}(s) := \mathbb{E}[\psi_m(W) \mid g_i(W;\theta) = s].$$

Assume the following conditions.

(S1) ***Bounded dictionary.*** *There are constants $C_\phi, C_\psi < \infty$ such that $|\phi_m(Y_i)| \le C_\phi$ and $|\psi_m(W)| \le C_\psi$ for all i, m, n.*

(S2) ***Smooth exposure and regular projection.*** *There exist nonnegative envelopes $G_{i,n}(W)$ and $M_{i,m,n}$ such that, for all $\theta, \theta' \in \Theta$,*

$$\|g_i(W;\theta) - g_i(W;\theta')\| \le G_{i,n}(W)\|\theta - \theta'\|,$$

and for all support points s, s',

$$|m_{i,m,\theta}(s) - m_{i,m,\theta'}(s')| \le M_{i,m,n}\{\|\theta - \theta'\| + \|s - s'\|\}.$$

The envelopes satisfy

$$\sup_n \frac{1}{N_n} \sum_{i=1}^{N_n} \max_{1 \le m \le q} M_{i,m,n}\{1 + \mathbb{E}G_{i,n}(W)\} < \infty.$$

(S3) ***Affinity covariance bound.*** *There exist affinity sets $A_i = A_{i,n}$, with $i \in A_i$, such that $b_n := \max_i |A_i| = o(N_n)$. Moreover, with*

$$\rho_n := \sup_{\theta \in \Theta} \frac{1}{N_n^2} \sum_{i=1}^{N_n} \sum_{j \notin A_i} \left\| \operatorname{Cov}\{Z_{i,n}(\theta), Z_{j,n}(\theta)\} \right\|_{\mathrm{op}},$$

we have $\rho_n \to 0$.

Lemma B.4 (Smooth-kernel ULLN)**.** *Under Assumption B.3,*

$$\sup_{\theta \in \Theta} \left\| \frac{1}{N_n} \sum_{i=1}^{N_n} \{\Psi_i(\theta) - \mathbb{E}\Psi_i(\theta)\} \right\| \xrightarrow{p} 0.$$

Proof. First, Assumption B.3 (S1) implies $\sup_{i,n,\theta} \mathbb{E}\|\Psi_i(\theta)\|^2 < \infty$ because $|m_{i,m,\theta}(s)| \le C_\psi$. Hence, for any i, j, θ,

$$\left\| \operatorname{Cov}\{Z_{i,n}(\theta), Z_{j,n}(\theta)\} \right\|_{\mathrm{op}} \le C$$

for a finite constant C. Therefore

$$\sup_{\theta\in\Theta} \mathbb{E}\|\bar{Z}_n(\theta)\|^2 \le C \left\{ \frac{b_n}{N_n} + \rho_n \right\} = o(1). \tag{B.1$'$}$$

Indeed,

$$\mathbb{E}\|\bar{Z}_n(\theta)\|^2 = \operatorname{tr}\{\operatorname{Var}(\bar{Z}_n(\theta))\} \le q \, \|\operatorname{Var}(\bar{Z}_n(\theta))\|_{\mathrm{op}},$$

and

$$\|\operatorname{Var}(\bar{Z}_n(\theta))\|_{\mathrm{op}} \le \frac{1}{N_n^2} \sum_{i=1}^{N_n} \sum_{j=1}^{N_n} \left\| \operatorname{Cov}\{Z_{i,n}(\theta), Z_{j,n}(\theta)\} \right\|_{\mathrm{op}}.$$

The within-affinity part has at most $N_n b_n$ terms and each term is bounded by the uniform second-moment bound; the outside-affinity part is exactly controlled by ρ_n. This gives (B.1$'$).

Second, Assumption B.3 (S2) implies that the moment vector is Lipschitz in θ. For each m,

$$\begin{aligned} &|m_{i,m,\theta}(g_i(W;\theta)) - m_{i,m,\theta'}(g_i(W;\theta'))| \\ &\quad \le M_{i,m,n}\{1 + G_{i,n}(W)\}\|\theta - \theta'\|, \end{aligned}$$

so

$$\|\Psi_i(\theta) - \Psi_i(\theta')\| \le L_{i,n}\|\theta - \theta'\|, \qquad L_{i,n} := \sqrt{q} C_\phi \max_m M_{i,m,n}\{1 + G_{i,n}(W)\}.$$

By the envelope condition,

$$\frac{1}{N_n} \sum_i \mathbb{E} L_{i,n} = O(1), \qquad \frac{1}{N_n} \sum_i L_{i,n} = O_p(1). \tag{B.2$'$}$$

Now let Θ_ε be a finite ε-net of the compact set Θ. For fixed ε, (B.1$'$), Chebyshev's inequality, and a union bound imply

$$\max_{\theta_r \in \Theta_\varepsilon} \|\bar{Z}_n(\theta_r)\| \xrightarrow{p} 0.$$

For arbitrary $\theta \in \Theta$, let $\pi_\varepsilon(\theta)$ be a nearest net point. By (B.2$'$),

$$\|\bar{Z}_n(\theta) - \bar{Z}_n(\pi_\varepsilon(\theta))\| \le \varepsilon \left\{ \frac{1}{N_n} \sum_i L_{i,n} + \frac{1}{N_n} \sum_i \mathbb{E} L_{i,n} \right\} = \varepsilon O_p(1).$$

Combining the net-point bound with the oscillation bound and then sending $\varepsilon \downarrow 0$ proves the result. □

For bounded assignments $|W_j| \leq 1$ (in particular binary treatment, as in our applications), the exposure-derivative bounds below hold as displayed. For normalized gravity weights $a_{ij}(\theta) \propto A_j \exp(-\theta d_{ij})$, the exposure derivative satisfies

$$\left| \frac{\partial}{\partial \theta} \sum_{j \neq i} a_{ij}(\theta) W_j \right| \leq 2 \sum_{j \neq i} a_{ij}(\theta) d_{ij}.$$

For market-access weights $a_{ij}(\theta) \propto A_j (1 + \alpha d_{ij})^{-\theta}$,

$$\left| \frac{\partial}{\partial \theta} \sum_{j \neq i} a_{ij}(\theta) W_j \right| \leq 2 \sum_{j \neq i} a_{ij}(\theta) \log(1 + \alpha d_{ij}).$$

Thus the exposure-smoothness part of Assumption B.3 (S2) follows from bounded weighted first moments of distance or log-distance, while the regularity of the design projection is the separate primitive condition stated above.

B.2.2 Ring exposures

The smooth proof above does not apply to a ring exposure because the sample path $\theta \mapsto \Psi_i(\theta)$ can be stepwise. Nevertheless, a continuum ULLN can be proved under a direct shell-regularity condition. The condition says that small changes in the cutoff affect only a vanishing average mass of pair distances and that the design projection changes only through this shell.

For this subsection the parameter $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$ is the ring radius, and set

$$Z_{i,n}(\theta) := \Psi_i(\theta) - \mathbb{E}\Psi_i(\theta), \qquad \bar{Z}_n(\theta) := \frac{1}{N_n} \sum_{i=1}^{N_n} Z_{i,n}(\theta).$$

Assumption B.5 (Ring ULLN primitives)**.** *The following conditions hold.*

(HR1) ***Bounded moment contributions.*** *There is a finite constant C such that*

$$\sup_n \sup_{i \leq N_n} \sup_{\theta \in \Theta} \|\Psi_i(\theta)\| \leq C \quad \textit{almost surely.}$$

(HR2) ***Pointwise affinity variance control.*** *There exist affinity sets $A_i = A_{i,n}$ with $i \in A_i$ and $b_n := \max_i |A_i| = o(N_n)$ such that*

$$\rho_n^{\text{ring}} := \sup_{\theta \in \Theta} \frac{1}{N_n^2} \sum_{i=1}^{N_n} \sum_{j \notin A_i} \left\| \operatorname{Cov}\{Z_{i,n}(\theta), Z_{j,n}(\theta)\} \right\|_{\text{op}} \to 0.$$

(HR3) ***Uniform shell oscillation.*** *For every interval $I \subset \Theta$, there are deterministic envelopes $H_{i,n}(I) \geq 0$ such that, whenever $\theta, \theta' \in I$,*

$$\|\Psi_i(\theta) - \Psi_i(\theta')\| \leq H_{i,n}(I) \quad \textit{almost surely.}$$

Define

$$\omega_n(\delta) := \sup_{I \subset \Theta:\ |I| \leq \delta} \frac{1}{N_n} \sum_{i=1}^{N_n} H_{i,n}(I).$$

Then

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \omega_n(\delta) = 0.$$

Lemma B.6 (Ring ULLN)**.** *Under Assumption B.5,*

$$\sup_{\theta \in \Theta} \left\| \bar{\Psi}_n(\theta) - \mu_n(\theta) \right\| = \sup_{\theta \in \Theta} \|\bar{Z}_n(\theta)\| \xrightarrow{p} 0.$$

Proof. First fix $\theta \in \Theta$. The same trace and covariance decomposition used in Lemma B.4 gives

$$\mathbb{E}\|\bar{Z}_n(\theta)\|^2 \leq C \left(\frac{b_n}{N_n} + \rho_n^{\text{ring}} \right),$$

where the constant can change from line to line. This bound holds uniformly in θ by Assumption B.5.

Fix $\delta > 0$ and cover Θ by intervals $I_1, \ldots, I_{M_\delta}$ of length at most δ, with $M_\delta \leq C/\delta$. Choose one grid point $r_\ell \in I_\ell$ for each interval. For fixed δ, Chebyshev's inequality and a union bound imply

$$\Pr\left(\max_{\ell \leq M_\delta} \|\bar{Z}_n(r_\ell)\| > \eta \right) \leq \frac{C M_\delta}{\eta^2} \left(\frac{b_n}{N_n} + \rho_n^{\text{ring}} \right) \to 0.$$

For any $\theta \in I_\ell$,

$$\begin{aligned} \|\bar{Z}_n(\theta) - \bar{Z}_n(r_\ell)\| &\leq \frac{1}{N_n} \sum_{i=1}^{N_n} \|\Psi_i(\theta) - \Psi_i(r_\ell)\| + \frac{1}{N_n} \sum_{i=1}^{N_n} \mathbb{E}\|\Psi_i(\theta) - \Psi_i(r_\ell)\| \\ &\leq 2\omega_n(\delta). \end{aligned}$$

Therefore

$$\sup_{\theta \in \Theta} \|\bar{Z}_n(\theta)\| \leq \max_{\ell \leq M_\delta} \|\bar{Z}_n(r_\ell)\| + 2\omega_n(\delta).$$

Taking $n \to \infty$ for fixed δ and then sending $\delta \downarrow 0$ proves the result. □

Lemma B.7 (Shell regularity implies continuity of the mean map)**.** *Suppose Assumption B.5* (HR3) *holds and Assumption 3.1* AS–LLN1 *holds for the continuous ring class $\Theta = [\underline{\theta}, \overline{\theta}]$. Then the limit*

mean map $\theta \mapsto \mu(\theta)$ is uniformly continuous on Θ. Hence the continuity part of Assumption 3.1 AS–LLN3 *follows from the shell primitive rather than being imposed separately.*

Proof. For $\theta, \theta' \in I$ with $|I| \le \delta$, Assumption B.5 (HR3) gives

$$\|\mu_n(\theta) - \mu_n(\theta')\| \le \frac{1}{N_n} \sum_{i=1}^{N_n} \mathbb{E}\|\Psi_i(\theta) - \Psi_i(\theta')\| \le \omega_n(\delta).$$

Therefore, using the uniform convergence in Assumption 3.1 AS–LLN1,

$$\sup_{|\theta-\theta'|\le\delta} \|\mu(\theta) - \mu(\theta')\| \le 2 \sup_{\theta\in\Theta} \|\mu_n(\theta) - \mu(\theta)\| + \omega_n(\delta).$$

Taking first $\limsup_{n\to\infty}$ and then $\delta \downarrow 0$ yields uniform continuity of μ. □

A useful primitive sufficient condition for the envelope condition (HR3) of Assumption B.5 comes from weighted shell mass. Suppose

$$g_i(W;\theta) = \sum_{j\neq i} a_{ij,n} W_j \mathbb{1}\{d_{ij} \le \theta\}, \qquad \sup_{i,n} \sum_{j\neq i} |a_{ij,n}| < \infty,$$

and define the weighted shell mass

$$S_{i,n}(I) := \sum_{j\neq i} |a_{ij,n}| \mathbb{1}\{d_{ij} \in I\}.$$

If the design projection satisfies the shell bound

$$\big|\mathbb{E}[\psi_m(W) \mid g_i(W;\theta)] - \mathbb{E}[\psi_m(W) \mid g_i(W;\theta')]\big| \le L S_{i,n}(I) \quad \text{whenever } \theta, \theta' \in I,$$

and $\phi_m(Y_i)$ is uniformly bounded, then one may take $H_{i,n}(I) = C S_{i,n}(I)$. The shell condition becomes

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{I\subset\Theta:\ |I|\le\delta} \frac{1}{N_n} \sum_{i=1}^{N_n} \sum_{j\neq i} |a_{ij,n}| \mathbb{1}\{d_{ij} \in I\} = 0.$$

This is an *atomless-shell* (no-mass-point) condition for continuous cutoffs: no single radius may carry a non-negligible share of the weighted pairwise mass. It is the spatial analogue of the familiar requirement, in threshold and regression-discontinuity models, that the running variable have no mass point at the cutoff, and the modulus $\omega_n(\delta)$ above is the asymptotic-equicontinuity device that turns it into a uniform law of large numbers. Equivalently, for the weighted shell measures

$$\nu_n(I) := \frac{1}{N_n} \sum_{i=1}^{N_n} \sum_{j\neq i} |a_{ij,n}| \mathbb{1}\{d_{ij} \in I\},$$

which record, on the radius axis, how much exposure weight sits at each distance (a weighted analogue of the pair-distance distribution, or Ripley's K function in spatial statistics), small intervals must have uniformly small mass. A sufficient way to obtain this is that ν_n converge uniformly over intervals in the relevant cutoff region to a limiting measure with no atoms and locally bounded interval mass. The condition can fail in regular lattice designs when many pairs lie exactly at the candidate cutoff radius; in that case the continuous-cutoff target may be ill posed at that radius. This observation is separate from finite VC dimension: VC controls the number of threshold sets, while the shell condition controls the size of the jumps under the design.

B.2.3 Finite-grid and K-hop maps

For genuinely discrete maps, the main primitive result is finite-grid selection. This is the natural formulation for network cutoffs $K \in \{1, \dots, K_{\max,n}\}$ and for any exposure problem that the researcher intentionally defines as a finite list of candidate specifications.

Assumption B.8 (Finite-grid ULLN primitives for non-smooth maps)**.** *Let Θ_n be a finite candidate set with cardinality $K_n = |\Theta_n|$. Assume*

$$\sup_n \sup_{i \le N_n} \sup_{\theta \in \Theta_n} \mathbb{E}\|\Psi_i(\theta)\|^2 < \infty.$$

There exist affinity sets $A_i = A_{i,n}$ with $i \in A_i$ and $b_n := \max_i |A_i|$ such that, with

$$\rho_n^{\mathrm{fg}} := \sup_{\theta \in \Theta_n} \frac{1}{N_n^2} \sum_{i=1}^{N_n} \sum_{j \notin A_i} \left\| \operatorname{Cov}\{Z_{i,n}(\theta), Z_{j,n}(\theta)\} \right\|_{\mathrm{op}},$$

we have

$$K_n \left(\frac{b_n}{N_n} + \rho_n^{\mathrm{fg}} \right) \to 0.$$

Lemma B.9 (Finite-grid ULLN)**.** *Under Assumption B.8,*

$$\sup_{\theta \in \Theta_n} \left\| \frac{1}{N_n} \sum_{i=1}^{N_n} \{\Psi_i(\theta) - \mathbb{E}\Psi_i(\theta)\} \right\| \xrightarrow{p} 0.$$

Proof. For each fixed $\theta \in \Theta_n$, the same trace and covariance decomposition as in the smooth proof gives

$$\mathbb{E}\|\bar{Z}_n(\theta)\|^2 \le C \left(\frac{b_n}{N_n} + \rho_n^{\mathrm{fg}} \right).$$

Therefore, by Chebyshev's inequality and a union bound,

$$\mathbb{P}\left(\sup_{\theta \in \Theta_n} \|\bar{Z}_n(\theta)\| > \eta \right) \le \frac{C K_n}{\eta^2} \left(\frac{b_n}{N_n} + \rho_n^{\mathrm{fg}} \right) \to 0.$$

□

Lemma B.10 (Finite-grid argmin consistency)**.** *Let Θ_n be finite and define*

$$Q_n(\theta) := \bar{\Psi}_n(\theta)^\top \Lambda_n \bar{\Psi}_n(\theta), \qquad Q_n^0(\theta) := \mu_n(\theta)^\top \Lambda \mu_n(\theta).$$

Suppose the finite-grid ULLN in Lemma B.9 holds, $\Lambda_n \xrightarrow{p} \Lambda$, and $\sup_{\theta\in\Theta_n} \|\mu_n(\theta)\| = O(1)$. Suppose also that there is a unique grid target $\theta_{0,n} \in \Theta_n$ and a constant $c > 0$ such that

$$\min_{\theta\in\Theta_n:\ \theta\neq\theta_{0,n}} \{Q_n^0(\theta) - Q_n^0(\theta_{0,n})\} \geq c$$

for all sufficiently large n. Then any finite-grid minimizer

$$\hat{\theta}_n \in \arg\min_{\theta\in\Theta_n} Q_n(\theta)$$

satisfies

$$\Pr(\hat{\theta}_n = \theta_{0,n}) \to 1.$$

Proof. The finite-grid ULLN implies

$$\sup_{\theta\in\Theta_n} \|\bar{\Psi}_n(\theta) - \mu_n(\theta)\| \xrightarrow{p} 0.$$

Together with $\Lambda_n \xrightarrow{p} \Lambda$ and $\sup_{\theta\in\Theta_n} \|\mu_n(\theta)\| = O(1)$, this gives

$$\sup_{\theta\in\Theta_n} |Q_n(\theta) - Q_n^0(\theta)| \xrightarrow{p} 0.$$

With probability tending to one, this supremum is smaller than $c/3$. On that event, for every $\theta \neq \theta_{0,n}$,

$$Q_n(\theta) - Q_n(\theta_{0,n}) \geq \{Q_n^0(\theta) - Q_n^0(\theta_{0,n})\} - 2c/3 \geq c/3 > 0.$$

Thus the unique minimizer of Q_n over Θ_n is $\theta_{0,n}$ with probability tending to one. □

B.3 Asymptotic normality of the baseline regular case

This subsection proves the regular asymptotic-normality result used in Section 3. The dependent triangular-array central limit theorem is needed only at the true exposure parameter, and the local behavior of the criterion is handled by the Z-estimator theorem of van der Vaart and Wellner (1996, Theorem 3.3.1). For ring exposures the sample moments need not be differentiable: the requirement is the projected estimating equation in Assumption B.11 AN–AFF5, a condition on the estimator rather than on the sample paths. When the radius is selected on a numerical grid, the grid mesh

enters the same first-order-negligibility requirement.

We use the notation of Section 3. Define

$$\bar{\Psi}_n(\theta) := \frac{1}{N_n}\sum_{i=1}^{N_n}\Psi_i(\theta), \qquad \mu_n(\theta) := \frac{1}{N_n}\sum_{i=1}^{N_n}\mathbb{E}[\Psi_i(\theta)],$$

and, at the true parameter,

$$Z_{i,n} := \Psi_i(\theta_0) - \mathbb{E}[\Psi_i(\theta_0)], \qquad \bar{Z}_n := \frac{1}{N_n}\sum_{i=1}^{N_n} Z_{i,n}.$$

The within-affinity covariance matrix is

$$\Omega_n := \sum_{i=1}^{N_n}\sum_{j\in A_i}\mathrm{Cov}\{Z_{i,n}, Z_{j,n}\}.$$

Assumption B.11 AN–AFF3 states that $\Omega_n/N_n \to \Omega$, where Ω is finite and positive definite.

The asymptotic-normality conditions summarized in Section 3 are the following.

Assumption B.11 (Design-based AN conditions)**.** *In addition to Assumption 3.1, suppose:*

AN–AFF1 ***Interiority.*** *θ_0 is an interior point of Θ.*

AN–AFF2 ***Mean differentiability and full rank.*** *There exists a neighborhood $\mathcal{N}(\theta_0)$ and a continuous map $G : \mathcal{N}(\theta_0) \to \mathbb{R}^{q\times p}$ such that*

$$\mu(\theta) = \mu(\theta_0) + G(\theta_0)(\theta-\theta_0) + r(\theta), \qquad \|r(\theta)\| \le C\|\theta-\theta_0\|^2 \quad \text{for } \theta\in\mathcal{N}(\theta_0),$$

for some finite C, and $G^\top\Lambda G$ is nonsingular at θ_0, where $G := G(\theta_0)$.

AN–AFF3 ***Affinity-set CLT and covariance stabilization.*** *The array $\{Z_{i,n}\}$ has uniformly bounded fourth moments. Let $Z_{i,n,\ell}$ denote the ℓth component of $Z_{i,n}$ for $\ell = 1,\dots,q$. For each scalar index (i,ℓ), define a scalar affinity set*

$$A_n(i,\ell) \subseteq \mathcal{U}_n \times \{1,\dots,q\} \quad \text{by} \quad A_n(i,\ell) := \big\{(j,\ell') : j\in A_i,\ \ell' = 1,\dots,q\big\},$$

and define the "outside-affinity" scalar sum

$$Z_{-i,\ell} := \sum_{(j,\ell')\notin A_n(i,\ell)} Z_{j,n,\ell'}.$$

The scalar array $\{Z_{i,n,\ell}\}$ then satisfies the three affinity-set covariance conditions of Chan-

drasekhar et al. (2023):

$$\sum_{i=1}^{N_n}\sum_{\ell=1}^{q}\sum_{(j,\ell'),(k,\ell'')\in A_n(i,\ell)}\mathbb{E}\big[\ |Z_{i,n,\ell}|\ Z_{j,n,\ell'}Z_{k,n,\ell''}\ \big]\ =\ o\big(\|\Omega_n\|_F^{3/2}\big), \tag{A1}$$

$$\sum_{i=1}^{N_n}\sum_{\ell=1}^{q}\sum_{j=1}^{N_n}\sum_{\ell'=1}^{q}\sum_{(k,\ell'')\in A_n(i,\ell)}\sum_{(l,\tilde{\ell})\in A_n(j,\ell')}\mathrm{Cov}\big(Z_{i,n,\ell}Z_{k,n,\ell''},\ Z_{j,n,\ell'}Z_{l,n,\tilde{\ell}}\big)\ =\ o\big(\|\Omega_n\|_F^2\big), \tag{A2}$$

$$\sum_{i=1}^{N_n}\sum_{\ell=1}^{q}\mathbb{E}\Big[\ \big|\mathbb{E}\big[\,Z_{i,n,\ell}\,Z_{-i,\ell}\ \big|\ Z_{-i,\ell}\big]\big|\ \Big]\ =\ o\big(\|\Omega_n\|_F\big). \tag{A3}$$

Moreover, the within-affinity covariance has uniformly bounded row sums,

$$\sup_n \max_{1\le i\le N_n}\max_{1\le \ell\le q}\sum_{(j,\ell')\in A_n(i,\ell)}\big|\,\mathrm{Cov}(Z_{i,n,\ell},Z_{j,n,\ell'})\big|<\infty,$$

and the aggregate covariance stabilizes:

$$\frac{\Omega_n}{N_n}\ \to\ \Omega,\qquad \Omega \text{ finite and positive definite.}$$

Under these conditions,

$$\sqrt{N_n}\,\bar{Z}_n\ \Rightarrow\ \mathcal{N}(0,\Omega).$$

In addition, the finite-population centering at the true parameter is negligible at the $\sqrt{N_n}$ scale:

$$\sqrt{N_n}\,\big\|\mu_n(\theta_0)-\mu(\theta_0)\big\|\ \to\ 0.$$

Consequently,

$$\sqrt{N_n}\,\big(\bar{\Psi}_n(\theta_0)-\mu(\theta_0)\big)\ \Rightarrow\ \mathcal{N}(0,\Omega).$$

In the exposure-mapping application, this condition holds exactly because the residualized design-based orthogonality moments satisfy $\mu_n(\theta_0)=\mu(\theta_0)=0$ for every n.

AN–AFF4 ***Empirical process equicontinuity.*** *Writing*

$$\mathbb{G}_n(\theta):=\sqrt{N_n}\big(\bar{\Psi}_n(\theta)-\mu(\theta)\big),$$

the process $\{\mathbb{G}_n(\theta):\theta\in\mathcal{N}(\theta_0)\}$ is stochastically equicontinuous at θ_0: for any sequence $h_n\downarrow 0$,

$$\sup_{\|\theta-\theta_0\|\le h_n}\big\|\mathbb{G}_n(\theta)-\mathbb{G}_n(\theta_0)\big\|\ \xrightarrow{p}\ 0.$$

AN–AFF5 ***Projected estimating equation.*** *The estimator satisfies the projected GMM estimating equation up to* $o_p(N_n^{-1/2})$*:*

$$\left\|G^\top \Lambda_n \bar{\Psi}_n(\hat{\theta}_n)\right\| = o_p(N_n^{-1/2}).$$

B.3.1 Pointwise affinity-set CLT

Lemma B.12 (Affinity-set CLT for the moment vector)**.** *Under Assumptions 3.1 and B.11* AN–AFF3*,*

$$\Omega_n^{-1/2} \sum_{i=1}^{N_n} Z_{i,n} \Rightarrow \mathcal{N}(0, I_q).$$

Consequently,

$$\sqrt{N_n}\{\bar{\Psi}_n(\theta_0) - \mu_n(\theta_0)\} \Rightarrow \mathcal{N}(0, \Omega).$$

Since Assumption B.11 AN–AFF3 *also contains the fixed-limit centering condition,*

$$\sqrt{N_n}\{\bar{\Psi}_n(\theta_0) - \mu(\theta_0)\} \Rightarrow \mathcal{N}(0, \Omega).$$

Proof. The first display is the multivariate affinity-set CLT of Chandrasekhar et al. (2023). Assumption B.11 AN–AFF3 verifies their three covariance conditions (their Assumptions 1–3) for the *stacked* scalar array $\{Z_{i,n,\ell}\}$ with the dimension-inclusive affinity sets $A_n(i,\ell)$ defined there, including the cross-component terms; their multivariate Theorem 1 then yields $\Omega_n^{-1/2}\sum_{i=1}^{N_n} Z_{i,n} \Rightarrow \mathcal{N}(0, I_q)$ directly. Since

$$\sum_{i=1}^{N_n} Z_{i,n} = N_n\{\bar{\Psi}_n(\theta_0) - \mu_n(\theta_0)\},$$

we can write

$$\sqrt{N_n}\{\bar{\Psi}_n(\theta_0) - \mu_n(\theta_0)\} = \Big(\frac{\Omega_n}{N_n}\Big)^{1/2}\Big(\Omega_n^{-1/2}\sum_{i=1}^{N_n} Z_{i,n}\Big).$$

The first factor converges to $\Omega^{1/2}$ and the second factor converges to $\mathcal{N}(0, I_q)$, so Slutsky's theorem gives the stated $\mathcal{N}(0,\Omega)$ limit. The last display follows from the additional condition $\sqrt{N_n}\|\mu_n(\theta_0) - \mu(\theta_0)\| \to 0$. □

B.3.2 The Z-estimator theorem used below

Lemma B.13 (Finite-dimensional specialization of van der Vaart and Wellner (1996), Theorem 3.3.1)**.** *Let* $\Theta \subset \mathbb{R}^p$*, let* $M_n : \Theta \to \mathbb{R}^p$ *be random maps, and let* $M : \Theta \to \mathbb{R}^p$ *be deterministic. Suppose that* $\hat{\theta}_n \xrightarrow{p} \theta_0$*,* $M(\theta_0) = 0$*, and*

$$\|M_n(\hat{\theta}_n)\| = o_p(N_n^{-1/2}).$$

Suppose further that M is differentiable at θ_0 with nonsingular derivative A, that

$$\sqrt{N_n}\{M_n(\theta_0) - M(\theta_0)\} \Rightarrow Z$$

for a tight random vector Z, and that for every sequence $\theta_n \xrightarrow{p} \theta_0$,

$$\sqrt{N_n}\big[\{M_n(\theta_n) - M(\theta_n)\} - \{M_n(\theta_0) - M(\theta_0)\}\big] = o_p\big(1 + \sqrt{N_n}\|\theta_n - \theta_0\|\big). \tag{Z}$$

Then

$$\sqrt{N_n}(\hat{\theta}_n - \theta_0) = -A^{-1}\sqrt{N_n}\{M_n(\theta_0) - M(\theta_0)\} + o_p(1),$$

and hence

$$\sqrt{N_n}(\hat{\theta}_n - \theta_0) \Rightarrow -A^{-1}Z.$$

Proof. This is the finite-dimensional version of van der Vaart and Wellner (1996, Theorem 3.3.1) with rate $\sqrt{N_n}$. In their notation, M_n and M correspond to the random and limiting maps Ψ_n and Ψ, condition (Z) is their local stochastic expansion condition, the approximate-zero condition is $M_n(\hat{\theta}_n) = o_p(N_n^{-1/2})$, and the nonsingular matrix A is the continuously invertible derivative of the limiting map. Since the parameter is finite-dimensional, continuous invertibility of the derivative is equivalent to nonsingularity of A. □

B.3.3 Proof of Theorem 3.5

Proof of Theorem 3.5. By Theorem 3.3, Assumption 3.1 implies $\hat{\theta}_n \xrightarrow{p} \theta_0$. We apply Lemma B.13 to the projected GMM maps

$$M_n(\theta) := G^\top \Lambda_n \bar{\Psi}_n(\theta), \qquad M(\theta) := G^\top \Lambda \mu(\theta),$$

where $G = G(\theta_0)$.

First, $M(\theta_0) = 0$ because $\mu(\theta_0) = 0$. By Assumption B.11 AN–AFF2,

$$\mu(\theta) = \mu(\theta_0) + G(\theta - \theta_0) + r(\theta), \qquad \|r(\theta)\| \le C\|\theta - \theta_0\|^2,$$

so M is differentiable at θ_0 with derivative

$$A := G^\top \Lambda G.$$

The matrix A is nonsingular by Assumption B.11 AN–AFF2.

Second, Lemma B.12 and $\Lambda_n \xrightarrow{p} \Lambda$ imply

$$\sqrt{N_n}\{M_n(\theta_0) - M(\theta_0)\} = G^\top \Lambda_n \sqrt{N_n}\{\bar{\Psi}_n(\theta_0) - \mu(\theta_0)\} \Rightarrow G^\top \Lambda Z,$$

where $Z \sim \mathcal{N}(0, \Omega)$.

Third, we verify the stochastic expansion condition (Z). Let $\theta_n \xrightarrow{p} \theta_0$. The random-sequence version of Assumption B.11 AN–AFF4 follows from its deterministic shrinking-ball formulation: choose a deterministic sequence $h_n \downarrow 0$ such that $\Pr(\|\theta_n - \theta_0\| > h_n) \to 0$ and apply the supremum bound on the event $\{\|\theta_n - \theta_0\| \le h_n\}$. Therefore

$$\sqrt{N_n}\big[\{\bar{\Psi}_n(\theta_n) - \mu(\theta_n)\} - \{\bar{\Psi}_n(\theta_0) - \mu(\theta_0)\}\big] = o_p(1).$$

Multiplying by $G^\top \Lambda_n$ preserves the $o_p(1)$ order. The difference between Λ_n and Λ contributes

$$G^\top(\Lambda_n - \Lambda)\sqrt{N_n}\{\mu(\theta_n) - \mu(\theta_0)\} = o_p(1)\, O\big(\sqrt{N_n}\|\theta_n - \theta_0\|\big),$$

where the $O(\|\theta_n - \theta_0\|)$ bound follows from local mean differentiability. Hence the total remainder is

$$o_p\big(1 + \sqrt{N_n}\|\theta_n - \theta_0\|\big),$$

which is exactly condition (Z).

Fourth, Assumption B.11 AN–AFF5 gives

$$\|M_n(\hat{\theta}_n)\| = \|G^\top \Lambda_n \bar{\Psi}_n(\hat{\theta}_n)\| = o_p(N_n^{-1/2}).$$

This is a projected estimating-equation condition. It is not a differentiability condition and is therefore compatible with ring exposures whose sample moments are step functions of the cutoff. In smooth cases it can be verified by an ordinary sample first-order condition and a Jacobian law of large numbers. In non-smooth continuous-cutoff cases it is maintained directly as part of the regular Z/GMM condition. If the continuous cutoff is computed by grid search, the grid approximation must be first-order negligible, for example a grid mesh $\Delta_n = o(N_n^{-1/2})$, so that the computed estimator satisfies the projected estimating equation of Assumption B.11 AN–AFF5.

All conditions of Lemma B.13 are now verified. Therefore

$$\sqrt{N_n}(\hat{\theta}_n - \theta_0) = -(G^\top \Lambda G)^{-1} G^\top \Lambda \sqrt{N_n}\{\bar{\Psi}_n(\theta_0) - \mu(\theta_0)\} + o_p(1).$$

Combining this expansion with Lemma B.12 gives

$$\sqrt{N_n}(\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}\Big(0, (G^\top \Lambda G)^{-1} G^\top \Lambda \Omega \Lambda G (G^\top \Lambda G)^{-1}\Big).$$

If $\Lambda = \Omega^{-1}$, the variance reduces to $(G^\top \Omega^{-1} G)^{-1}$. If $\Lambda_n = \hat{\Omega}_n^{-1}$ and $\hat{\Omega}_n \xrightarrow{p} \Omega$, the efficient-weight case follows by Slutsky's theorem. □

Remark B.14 (Continuous and finite-grid rings). A ring exposure can be regular even though

the sample path $\theta \mapsto \bar{\Psi}_n(\theta)$ is stepwise. The regularity needed for Wald inference is exactly the regular Z/GMM condition in Assumption B.11: differentiability of the mean map $\theta \mapsto \mu(\theta)$ at θ_0, a pointwise affinity-set CLT, local stochastic equicontinuity, and the projected estimating equation. The continuum ULLN in Lemma B.6 provides one primitive shell-regularity route for the corresponding uniform convergence. Thus the theorem does not require differentiability of the sample moments, but it does require the shell and empirical-process regularity stated in the assumptions.

A genuinely finite-grid radius or finite K-hop parameter is different. Under a fixed population separation gap, the selected discrete parameter is consistently selected and has a degenerate first-order limit. Wald inference for the discrete cutoff is therefore not the target. If a fine grid is used only as a numerical implementation of an underlying continuous-cutoff estimator, the continuous Wald limit applies provided the grid mesh Δ_n satisfies $\Delta_n = o(N_n^{-1/2})$ and the projected estimating-equation remainder remains $o_p(N_n^{-1/2})$. △

B.4 Finite-grid exposure-map variants

This subsection collects the finite-grid exception to the baseline regular continuous theory. It treats a finite discrete exposure-map component with an optional continuous component.

Corollary B.15 (Finite discrete exposure-map components)**.** *Suppose the Stage 1 exposure-map parameter can be written as $\theta = (\vartheta, \rho)$, where ϑ takes values in a finite set $\mathcal{V}$ and, conditional on ϑ, ρ lies in a compact set $\mathcal{B}(\vartheta) \subset \mathbb{R}^{p_\rho}$. For each $\vartheta \in \mathcal{V}$, define the profile population criterion*

$$Q^p(\vartheta) := \inf_{\rho \in \mathcal{B}(\vartheta)} Q(\vartheta, \rho), \qquad Q(\vartheta, \rho) := \mu(\vartheta, \rho)^\top \Lambda \mu(\vartheta, \rho).$$

Assume that Q^p has a unique minimizer ϑ_0 with a strict gap:

$$Q^p(\vartheta) - Q^p(\vartheta_0) \geq c > 0 \qquad \text{for all } \vartheta \neq \vartheta_0,$$

and let

$$\rho_0 \in \arg\min_{\rho \in \mathcal{B}(\vartheta_0)} Q(\vartheta_0, \rho).$$

For each ϑ, let

$$\hat{\rho}_n(\vartheta) \in \arg\min_{\rho \in \mathcal{B}(\vartheta)} Q_n(\vartheta, \rho),$$

and define the profiled finite-grid estimator

$$\hat{\vartheta}_n \in \arg\min_{\vartheta \in \mathcal{V}} Q_n\big(\vartheta, \hat{\rho}_n(\vartheta)\big), \qquad \hat{\rho}_n := \hat{\rho}_n(\hat{\vartheta}_n).$$

Assume that the uniform LLN and deterministic-stabilization conditions hold uniformly over $\{(\vartheta, \rho) :$

$\vartheta \in \mathcal{V}, \rho \in \mathcal{B}(\vartheta)\}$. *Hence the finite-population profile criteria*

$$Q_n^{0,p}(\vartheta) := \inf_{\rho \in \mathcal{B}(\vartheta)} \mu_n(\vartheta, \rho)^\top \Lambda \mu_n(\vartheta, \rho)$$

satisfy $\sup_{\vartheta \in \mathcal{V}} |Q_n^{0,p}(\vartheta) - Q^p(\vartheta)| \to 0$*, so the fixed limit-profile gap transfers to the finite-population profile for all sufficiently large* n*. Assume also that the asymptotic-normality conditions of Theorem 3.5 hold for the continuous component* ρ *conditional on* $\vartheta = \vartheta_0$*. Then*

$$\Pr(\hat{\vartheta}_n = \vartheta_0) \to 1.$$

Consequently,

$$\sqrt{N_n}\,(\hat{\rho}_n - \rho_0) = \sqrt{N_n}\big(\hat{\rho}_n(\vartheta_0) - \rho_0\big) + o_p(1),$$

so the continuous Stage 1 exposure-map parameter has the same first-order asymptotic distribution as the oracle estimator that fixes the correct discrete component ϑ_0*.*

Remark B.16 (What the finite-grid results do and do not say). Corollary B.15 is a Stage 1 result about exposure-map parameters. The continuous component ρ may depend on the selected discrete component $\hat{\vartheta}_n$ in finite samples. The conclusion is not that this dependence is absent, but that under a fixed strict profile gap the selected class equals the population class with probability approaching one. Hence the selected estimator equals the oracle fixed-class estimator with probability approaching one, which removes the discrete selection step from the first-order asymptotic variance.

This fixed-gap condition rules out local ties among discrete classes. If two candidate exposure maps have population criteria that are equal, or differ only at the $N_n^{-1/2}$ scale, the selected criterion can have a nonstandard post-selection distribution, analogous to the maximum of two noisy sample means. Such local-tie cases are not covered by the finite-grid corollary. △

Remark B.17 (Discrete rings and network radii). Corollary B.15 applies when a finite-grid component of the exposure map is point-identified by a strict profile gap. Examples include non-nested finite grids of scalar ring radii or K-hop network radii when the profile population criterion has a unique minimizer.

The ring case covered by Assumption B.11 is a regular continuous-parameter problem. In the scalar ring case, $g(W; X_i, R)$ is an average over units within radius R, and increasing R does not usually refine the smaller-radius exposure in the sense of sigma-fields; the regular continuous-parameter theory applies provided the population moment map is mean differentiable. △

Remark B.18 (Stage 1 versus Stage 2). The notation $\theta = (\vartheta, \rho)$ in this section refers only to the Stage 1 exposure-map parameter. Downstream causal or exposure–response parameters, such as regression coefficients or average treatment effects, are introduced in Section 5. If the Stage 1

discrete component is selected consistently, it contributes no first-order term to Stage 2 inference. Any regular continuous Stage 1 component ρ, however, must be propagated into Stage 2 by the stacked estimating-equation or delta-method argument described in Section 5 and Appendix D. $\triangle$

B.4.1 Proof of Corollary B.15

Let $\theta = (\vartheta, \rho)$ with $\vartheta \in \mathcal{V}$ finite and $\rho \in \mathcal{B}(\vartheta)$ compact. Define

$$Q^p(\vartheta) := \inf_{\rho \in \mathcal{B}(\vartheta)} Q(\vartheta, \rho), \qquad Q_n^p(\vartheta) := \inf_{\rho \in \mathcal{B}(\vartheta)} Q_n(\vartheta, \rho).$$

The sample estimator in Corollary B.15 is equivalently

$$\hat{\vartheta}_n \in \arg\min_{\vartheta \in \mathcal{V}} Q_n^p(\vartheta), \qquad \hat{\rho}_n = \hat{\rho}_n(\hat{\vartheta}_n).$$

Lemma B.19 (Selection consistency for finite profiled criteria)**.** *Suppose $\mathcal{V}$ is finite and*

$$\sup_{\vartheta \in \mathcal{V}} \sup_{\rho \in \mathcal{B}(\vartheta)} |Q_n(\vartheta, \rho) - Q(\vartheta, \rho)| \xrightarrow{p} 0.$$

If Q^p has a unique minimizer ϑ_0 and satisfies the strict profile gap

$$Q^p(\vartheta) - Q^p(\vartheta_0) \geq c > 0 \qquad \text{for all } \vartheta \neq \vartheta_0,$$

then $\Pr(\hat{\vartheta}_n = \vartheta_0) \to 1$.

Proof. The uniform convergence of Q_n implies uniform convergence of the profile criteria:

$$\sup_{\vartheta \in \mathcal{V}} |Q_n^p(\vartheta) - Q^p(\vartheta)| \xrightarrow{p} 0,$$

because taking infima over ρ is one-Lipschitz with respect to the sup norm. For any $\vartheta \neq \vartheta_0$,

$$Q_n^p(\vartheta) - Q_n^p(\vartheta_0) = \{Q^p(\vartheta) - Q^p(\vartheta_0)\} + o_p(1) \geq c + o_p(1),$$

uniformly over the finite set $\mathcal{V} \setminus \{\vartheta_0\}$. Hence, with probability approaching one, every $\vartheta \neq \vartheta_0$ has a larger profile sample criterion than ϑ_0, so $\Pr(\hat{\vartheta}_n = \vartheta_0) \to 1$. $\square$

Proof of Corollary B.15. Selection consistency follows from Lemma B.19. On the event $\{\hat{\vartheta}_n = \vartheta_0\}$,

$$\hat{\rho}_n = \hat{\rho}_n(\vartheta_0).$$

Therefore, for any $\varepsilon > 0$,

$$\begin{aligned}\Pr\Big(\Big\|\sqrt{N_n}(\hat{\rho}_n - \rho_0) - \sqrt{N_n}\{\hat{\rho}_n(\vartheta_0) - \rho_0\}\Big\| > \varepsilon\Big) \\ \leq \Pr(\hat{\vartheta}_n \neq \vartheta_0) \to 0.\end{aligned}$$

Thus

$$\sqrt{N_n}(\hat{\rho}_n - \rho_0) = \sqrt{N_n}\{\hat{\rho}_n(\vartheta_0) - \rho_0\} + o_p(1).$$

The conditional asymptotic-normality assumption then gives the oracle first-order distribution. □

C Details and Proofs for Section 4

This appendix formalizes the optimal-moment construction used in Section 4; Theorems C.11 and C.16 below together constitute the formal statement of Theorem 4.1. The main object is the class of product residualized moments

$$\phi(Y_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid g(W; X_i, \theta)]\right\}.$$

The appendix proceeds in five steps. First, Appendix C.1 shows that, for each fixed unit, the vanishing of all such product residualized moments is equivalent to the unit-level exposure-sufficiency condition. Second, Appendix C.2 constructs the averaged moment space used by the GMM estimator. Third, Appendix C.3 derives the oracle moment through the Riesz representation theorem and the Cauchy–Schwarz inequality. Fourth, Appendix C.4 shows how finite dictionaries and sieve GMM approximate the oracle moment. Finally, Appendix C.5 extends the argument to vector-valued exposure parameters.

We write

$$G_{i,\theta} := g(W; X_i, \theta), \qquad G_i := G_{i,\theta_0}.$$

The design-side residual operator is

$$R_{i,\theta}\psi(W) := \psi(W) - \mathbb{E}[\psi(W) \mid G_{i,\theta}].$$

C.1 Moment characterization of unit-level exposure sufficiency

The first result is a population characterization of the unit-level conditional-independence restriction. It is a statement about the joint law of (Y_i, W, G_i) under the design $\mathcal{D}_n$ for a fixed (i, n). It does not require independence across units.

Assumption C.1 (Measurability)**.** *For each fixed experiment n and unit i, the random variables Y_i, W, and $G_i = g(W; X_i, \theta_0)$ take values in standard Borel spaces. The maps $w \mapsto Y_i(w)$ and*

$w \mapsto g(w; X_i, \theta_0)$ *are measurable. Conditional expectations are understood as regular conditional expectations under the design* $\mathcal{D}_n$.

Lemma C.2 (Product-moment characterization)**.** *Fix an experiment* n *and a unit* i*. Under Assumption C.1, the following two statements are equivalent:*

1.

$$Y_i \perp\!\!\!\perp W \mid G_i.$$

2. For every bounded measurable ϕ *and every bounded measurable* ψ*,*

$$\mathbb{E}\left[\phi(Y_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid G_i]\right\}\right] = 0.$$

Proof. First suppose that

$$Y_i \perp\!\!\!\perp W \mid G_i.$$

Then, for every bounded measurable ψ,

$$\mathbb{E}[\psi(W) \mid Y_i, G_i] = \mathbb{E}[\psi(W) \mid G_i].$$

Hence, for every bounded measurable ϕ,

$$\begin{aligned}&\mathbb{E}\left[\phi(Y_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid G_i]\right\}\right] \\ &\quad = \mathbb{E}\left[\phi(Y_i)\left\{\mathbb{E}[\psi(W) \mid Y_i, G_i] - \mathbb{E}[\psi(W) \mid G_i]\right\}\right] = 0.\end{aligned}$$

Conversely, suppose that

$$\mathbb{E}\left[\phi(Y_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid G_i]\right\}\right] = 0$$

for every bounded measurable pair (ϕ, ψ). Fix a bounded measurable ψ and define

$$U_\psi(W) := \psi(W) - \mathbb{E}[\psi(W) \mid G_i].$$

We will show that

$$\mathbb{E}[U_\psi(W) \mid Y_i, G_i] = 0.$$

Let a and b be bounded measurable functions. Since G_i is a measurable function of W, the product

$$\widetilde{\psi}(W) := b(G_i)\psi(W)$$

is a bounded measurable function of W. Applying the assumed moment condition with $\phi(Y_i) =$

$a(Y_i)$ and design function $\widetilde{\psi}(W)$, we obtain

$$0 = \mathbb{E}\left[a(Y_i)\left\{b(G_i)\psi(W) - \mathbb{E}[b(G_i)\psi(W) \mid G_i]\right\}\right].$$

But

$$\mathbb{E}[b(G_i)\psi(W) \mid G_i] = b(G_i)\mathbb{E}[\psi(W) \mid G_i],$$

and therefore

$$0 = \mathbb{E}\left[a(Y_i)b(G_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid G_i]\right\}\right] = \mathbb{E}[a(Y_i)b(G_i)U_\psi(W)].$$

The preceding identity holds for every product function $a(Y_i)b(G_i)$. Since products of functions of Y_i and functions of G_i generate the bounded measurable functions of the pair (Y_i, G_i), the same identity extends, by a standard monotone-class argument, to every bounded measurable function $h(Y_i, G_i)$:

$$\mathbb{E}[h(Y_i, G_i)U_\psi(W)] = 0.$$

Hence

$$\mathbb{E}[U_\psi(W) \mid Y_i, G_i] = 0.$$

Equivalently,

$$\mathbb{E}[\psi(W) \mid Y_i, G_i] = \mathbb{E}[\psi(W) \mid G_i]$$

for every bounded measurable ψ. This is precisely

$$Y_i \perp\!\!\!\perp W \mid G_i.$$

□

Remark C.3 (Finite support is only a special case). Lemma C.2 does not require W to have finite support. If W has finite support $\mathcal{W}_n$, then every bounded measurable design function can be written as

$$\psi(W) = \sum_{w \in \mathcal{W}_n} \psi(w)\mathbf{1}\{W = w\}.$$

In that special case, it is enough to verify the moment restrictions for the atom indicators $\mathbf{1}\{W = w\}$. With continuous-support or large-support designs, the exact population characterization uses all bounded measurable design functions, while feasible implementation uses a countable determining dictionary. △

For feasible implementation it is useful to replace the full class of bounded measurable functions by countable dense dictionaries.

Assumption C.4 (Countable determining dictionaries)**.** *For each fixed (i,n), $L^2(Y_i)$ and $L^2(W)$ are separable under the design $\mathcal{D}_n$. Let*

$$\Phi_{i,n} = \{\phi_{\ell,i,n} : \ell \geq 1\}, \qquad \Psi_n = \{\psi_{m,n} : m \geq 1\}$$

be countable dense subsets of $L^2(Y_i)$ and $L^2(W)$, respectively.

Lemma C.5 (Countable moment characterization)**.** *Suppose Assumptions C.1 and C.4 hold. Then*

$$Y_i \perp\!\!\!\perp W \mid G_i$$

if and only if

$$\mathbb{E}\left[\phi_{\ell,i,n}(Y_i)\left\{\psi_{m,n}(W) - \mathbb{E}[\psi_{m,n}(W) \mid G_i]\right\}\right] = 0$$

for every $\ell, m \geq 1$.

Proof. The forward direction follows from Lemma C.2. For the converse, define the bilinear form

$$B_{i,n}(\phi,\psi) := \mathbb{E}\left[\phi(Y_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid G_i]\right\}\right].$$

By Cauchy–Schwarz and the fact that conditional expectation is an L^2-projection,

$$\begin{aligned} |B_{i,n}(\phi,\psi)| &\leq \|\phi(Y_i)\|_{L^2}\, \|\psi(W) - \mathbb{E}[\psi(W) \mid G_i]\|_{L^2} \\ &\leq \|\phi(Y_i)\|_{L^2} \|\psi(W)\|_{L^2}. \end{aligned}$$

Thus $B_{i,n}$ is continuous on $L^2(Y_i) \times L^2(W)$. If it vanishes on the dense product dictionary $\Phi_{i,n} \times \Psi_n$, it vanishes on all of $L^2(Y_i) \times L^2(W)$, and therefore on all bounded measurable (ϕ,ψ). Lemma C.2 then implies $Y_i \perp\!\!\!\perp W \mid G_i$. □

Remark C.6 (No cross-unit independence is used)*.* The equivalence in Lemmas C.2 and C.5 is unit-level. It concerns the joint law of (Y_i, W, G_i) under the assignment design $\mathcal{D}_n$ for fixed (i,n). It does not require Y_i and Y_j, or the moment kernels for different units, to be independent. Cross-unit dependence affects the variance and limiting distribution of averaged moments, not the validity of the unit-level moment characterization. △

C.2 Averaged product residualized moment space

Lemma C.2 gives a unit-level equivalence. The GMM estimator uses averaged versions of these unit-level restrictions. For an admissible pair (ϕ,ψ), define

$$\eta_{\phi,\psi,i,n}(\theta) := \phi(Y_i) R_{i,\theta}\psi(W) = \phi(Y_i)\left\{\psi(W) - \mathbb{E}[\psi(W) \mid G_{i,\theta}]\right\}.$$

The corresponding averaged product residualized moment is

$$\eta_{\phi,\psi,n}(\theta) := \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \eta_{\phi,\psi,i,n}(\theta).$$

At $\theta = \theta_0$, unit-level exposure sufficiency implies

$$\mathbb{E}[\eta_{\phi,\psi,i,n}(\theta_0)] = 0 \qquad \text{for every } i,$$

and hence

$$\mathbb{E}[\eta_{\phi,\psi,n}(\theta_0)] = 0.$$

Let $\mathcal{H}_0$ denote the linear span of averaged product residualized moment directions evaluated at θ_0. A generic direction $\eta \in \mathcal{H}_0$ is generated by finite linear combinations

$$\eta_{i,n} = \sum_{r=1}^{R} a_r \phi_r(Y_i) R_{i,\theta_0} \psi_r(W),$$

with experiment-level average

$$\eta_n = \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \eta_{i,n}.$$

For two such directions η, ξ, define the design-based asymptotic covariance bilinear form

$$\Omega(\eta, \xi) := \lim_{n \to \infty} N_n \operatorname{Cov}(\eta_n, \xi_n),$$

whenever the limit exists. Write

$$\|\eta\|_\Omega^2 := \Omega(\eta, \eta).$$

Assumption C.7 (Asymptotic covariance Hilbert space)**.** *The following conditions hold.*

1. *For every $\eta, \xi \in \mathcal{H}_0$, the limit $\Omega(\eta, \xi)$ exists and is finite.*
2. *Ω is symmetric and positive semidefinite on $\mathcal{H}_0$.*
3. *After quotienting out zero-variance directions*

$$\mathcal{N}_\Omega := \{\eta \in \mathcal{H}_0 : \|\eta\|_\Omega = 0\},$$

the completion of $\mathcal{H}_0/\mathcal{N}_\Omega$ under $\|\cdot\|_\Omega$ is a Hilbert space, denoted

$$(\mathcal{H}, \langle \cdot, \cdot \rangle_\Omega).$$

The covariance Ω is the covariance of averaged design-based moments. It contains all cross-unit

covariance terms:

$$N_n \operatorname{Var}(\eta_n) = \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \sum_{j \in \mathcal{U}_n} \operatorname{Cov}(\eta_{i,n}, \eta_{j,n}).$$

Thus cross-unit dependence is not ruled out. It is summarized by the asymptotic covariance metric and handled by the LLN, CLT, and graph-HAC conditions in Section 3.

C.3 Oracle moment for scalar exposure parameters

This subsection treats the scalar case $\theta_0 \in \mathbb{R}$. The vector case is given in Appendix C.5.

For a direction $\eta \in \mathcal{H}$, let $\eta_n(\theta)$ denote the corresponding averaged product residualized moment at candidate parameter θ, and define the population mean map

$$\mu_\eta(\theta) := \lim_{n \to \infty} \mathbb{E}[\eta_n(\theta)],$$

whenever the limit exists. At the true value,

$$\mu_\eta(\theta_0) = 0.$$

Define the local derivative functional

$$G(\eta) := \frac{\partial}{\partial \theta} \mu_\eta(\theta) \Big|_{\theta = \theta_0}.$$

For η in the linear span of the dictionary directions, $\eta_n(\theta)$ is the corresponding finite linear combination of sample moments and $\mu_\eta(\theta)$ is defined directly. A general $\eta \in \mathcal{H}$ is an Ω-norm limit of span elements and need not come with a canonical finite-sample moment path away from θ_0. Accordingly, off-θ_0 objects such as $\mu_\eta(\theta)$ and the just-identified estimator in part 3 below are used only for span directions; G is then extended from the span to all of $\mathcal{H}$ by continuity, using the boundedness in part 2, and statements for general $\eta \in \mathcal{H}$ are understood in this extension sense.

Assumption C.8 (Local regularity)**.** *The following conditions hold.*

1. *For every $\eta \in \mathcal{H}$, the map $\theta \mapsto \mu_\eta(\theta)$ is differentiable in a neighborhood of θ_0, with derivative $G(\eta)$ at θ_0.*

2. *The map $G : \mathcal{H} \to \mathbb{R}$ is a bounded linear functional under the Ω-norm. That is, there exists $C_G < \infty$ such that*

$$|G(\eta)| \le C_G \|\eta\|_\Omega \qquad \text{for all } \eta \in \mathcal{H}.$$

3. *For every $\eta \in \mathcal{H}$ with $G(\eta) \neq 0$, the just-identified estimator based on*

$$\eta_n(\theta) = 0$$

admits the local expansion

$$\sqrt{N_n}(\hat{\theta}_{\eta,n} - \theta_0) = -\frac{\sqrt{N_n}\eta_n(\theta_0)}{G(\eta)} + o_p(1),$$

and

$$\sqrt{N_n}\eta_n(\theta_0) \Rightarrow \mathcal{N}(0, \|\eta\|_\Omega^2).$$

4. ***Nonzero information.*** *The derivative functional does not vanish identically on $\mathcal{H}$: there exists $\eta \in \mathcal{H}$ with $G(\eta) \neq 0$. Equivalently, via the Riesz representation below, $\|\eta^\star\|_\Omega > 0$, so that $V^\star = 1/\|\eta^\star\|_\Omega^2$ is finite and the infimum in Theorem C.11 runs over a nonempty set. If instead $G \equiv 0$, the exposure parameter is unidentified at first order by this moment class and no $\sqrt{N_n}$-consistent estimator built from it exists; empirically this corresponds to a flat GMM criterion (weak identification).*

Lemma C.9 (Variance of a scalar moment direction)**.** *Under Assumptions C.7 and C.8, the asymptotic variance of the just-identified estimator based on a scalar direction $\eta \in \mathcal{H}$ with $G(\eta) \neq 0$ is*

$$V(\eta) = \frac{\|\eta\|_\Omega^2}{G(\eta)^2}.$$

Proof. The expansion in Assumption C.8 gives

$$\sqrt{N_n}(\hat{\theta}_{\eta,n} - \theta_0) = -\frac{\sqrt{N_n}\eta_n(\theta_0)}{G(\eta)} + o_p(1).$$

Since

$$\sqrt{N_n}\eta_n(\theta_0) \Rightarrow \mathcal{N}(0, \|\eta\|_\Omega^2),$$

the continuous mapping theorem implies

$$\sqrt{N_n}(\hat{\theta}_{\eta,n} - \theta_0) \Rightarrow \mathcal{N}\left(0, \frac{\|\eta\|_\Omega^2}{G(\eta)^2}\right).$$

□

By Assumption C.8, G is a bounded linear functional on $\mathcal{H}$. The Riesz representation theorem therefore implies that there exists a unique element $\eta^\star \in \mathcal{H}$ such that

$$G(\eta) = \langle \eta^\star, \eta \rangle_\Omega \qquad \text{for all } \eta \in \mathcal{H}.$$

We call $\eta^\star$ the oracle product residualized moment.

Lemma C.10 (Cauchy–Schwarz lower bound)**.** *Under Assumptions C.7 and C.8, for any $\eta \in \mathcal{H}$ with $G(\eta) \neq 0$,*

$$V(\eta) = \frac{\|\eta\|_\Omega^2}{G(\eta)^2} \geq \frac{1}{\|\eta^\star\|_\Omega^2}.$$

Equality holds if and only if η is proportional to $\eta^\star$.

Proof. By the Riesz representation,

$$G(\eta)^2 = \langle \eta^\star, \eta \rangle_\Omega^2.$$

By Cauchy–Schwarz,

$$\langle \eta^\star, \eta \rangle_\Omega^2 \leq \|\eta^\star\|_\Omega^2 \|\eta\|_\Omega^2.$$

Rearranging gives

$$\frac{\|\eta\|_\Omega^2}{G(\eta)^2} \geq \frac{1}{\|\eta^\star\|_\Omega^2}.$$

The equality condition is the equality condition in Cauchy–Schwarz: η must be proportional to $\eta^\star$ in $\mathcal{H}$. □

Theorem C.11 (Oracle variance bound)**.** *Suppose Assumptions C.1, C.7, and C.8 hold. Among scalar directions in the averaged product residualized moment space $\mathcal{H}$, the oracle variance is*

$$V^\star := \frac{1}{\|\eta^\star\|_\Omega^2}.$$

Equivalently,

$$\inf_{\eta \in \mathcal{H}: G(\eta) \neq 0} \frac{\|\eta\|_\Omega^2}{G(\eta)^2} = V^\star,$$

and equality is attained if and only if the moment direction is proportional to $\eta^\star$.

Proof. The moment class is generated by the product residualized moments characterized in Lemma C.2. Within the Hilbert space $\mathcal{H}$ induced by the asymptotic covariance metric, Lemma C.10 gives the sharp lower bound and the equality condition. □

Remark C.12 (Within-class optimality). Theorem C.11 is a within-class optimality result. It does not characterize the globally efficient estimator under all implications of exposure sufficiency. In particular, exposure sufficiency also implies multi-unit restrictions involving (Y_i, Y_j), joint exposure vectors, and more complicated kernels. The oracle variance $V^\star$ is the efficiency bound for the averaged product residualized moment space $\mathcal{H}$, which is the space used by the proposed GMM procedure. △

Remark C.13 (Instrument-only moments as a subspace). The instrument-only moments correspond to the special case

$$\phi(Y_i) = Y_i, \qquad \psi(W) = h(W).$$

Their span is generally a strict subspace of $\mathcal{H}$. Theorem C.11 optimizes over the larger product residualized moment space generated by arbitrary admissible (ϕ, ψ). $\triangle$

C.4 Finite dictionaries and sieve GMM

The oracle moment $\eta^\star$ is generally infinite-dimensional. The feasible estimator approximates it using a finite dictionary of product residualized moments.

Let

$$\{(\phi_m, \psi_m) : m \geq 1\}$$

be a countable dictionary of outcome transformations and design-side functions. For each m, define

$$\eta_{m,n}(\theta) := \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \phi_m(Y_i) \left\{ \psi_m(W) - \mathbb{E}[\psi_m(W) \mid G_{i,\theta}] \right\}.$$

For a finite dictionary of size M, stack

$$\eta_n^{(M)}(\theta) := (\eta_{1,n}(\theta), \ldots, \eta_{M,n}(\theta))^\top.$$

Let

$$G^{(M)} := (G(\eta_1), \ldots, G(\eta_M))^\top, \qquad \Omega^{(M)} := (\Omega(\eta_k, \eta_\ell))_{k,\ell \leq M}.$$

Lemma C.14 (Finite-dimensional oracle)**.** *Suppose $\Omega^{(M)}$ is nonsingular and $G^{(M)} \neq 0$. For scalar θ_0, the optimal linear combination*

$$a^\top \eta_n^{(M)}(\theta)$$

solves

$$\min_{a : a^\top G^{(M)} \neq 0} \frac{a^\top \Omega^{(M)} a}{(a^\top G^{(M)})^2}.$$

The solution is proportional to

$$a_M^\star = \Omega^{(M)-1} G^{(M)},$$

and the finite-dictionary oracle variance is

$$V_M^\star = \frac{1}{G^{(M)\top} \Omega^{(M)-1} G^{(M)}}.$$

Proof. Normalize $a^\top G^{(M)} = 1$. The problem becomes

$$\min_a a^\top \Omega^{(M)} a \quad \text{subject to} \quad a^\top G^{(M)} = 1.$$

The Lagrangian first-order condition gives

$$2\Omega^{(M)}a - \lambda G^{(M)} = 0,$$

so

$$a \propto \Omega^{(M)-1} G^{(M)}.$$

Substituting the normalized solution yields

$$V_M^\star = \left\{ G^{(M)\top} \Omega^{(M)-1} G^{(M)} \right\}^{-1}.$$

□

Equivalently, define the finite-dimensional Riesz representer

$$\eta_M^\star := \sum_{m=1}^{M} \gamma_{M,m}^\star \eta_m, \qquad \gamma_M^\star := \Omega^{(M)-1} G^{(M)}.$$

Then $\eta_M^\star$ is the Riesz representer of G restricted to

$$\mathcal{H}_M := \operatorname{span}\{\eta_1, \ldots, \eta_M\}.$$

The feasible finite-dictionary GMM estimator is

$$\hat{\theta}_{M,n} \in \operatorname*{argmin}_{\theta \in \Theta} \eta_n^{(M)}(\theta)^\top \Lambda_{M,n} \eta_n^{(M)}(\theta),$$

where the efficient choice is

$$\Lambda_{M,n} = \widehat{\Omega}_{M,n}^{-1}.$$

This yields the familiar two-step procedure: first obtain a preliminary consistent estimate of θ_0, then estimate the asymptotic covariance matrix of the stacked product residualized moments, and finally re-estimate using the inverse covariance weight.

We next combine population sieve approximation with the paper's fixed-dimensional feasible GMM theory. The latter is imposed only pointwise in the dictionary size; no growing-dimensional linearization or central limit theorem is assumed.

Assumption C.15 (Dense sieve and pointwise feasible inference)**.** *The following conditions hold.*

(S1) The linear span of $\{\eta_m : m \geq 1\}$ *is dense in* $\mathcal{H}$ *under the* Ω*-norm:*

$$\overline{\bigcup_{M \geq 1} \operatorname{span}\{\eta_1, \ldots, \eta_M\}}^{\|\cdot\|_\Omega} = \mathcal{H}.$$

In continuous-support cases, this dictionary may be constructed from countable dense classes of outcome transformations and design-side functions as in Assumption C.4.

(S2) For every fixed M, the consistency and asymptotic-normality conditions of Theorem 3.5 hold for $\eta_n^{(M)}(\theta)$, and the feasible two-step weight converges to $\{\Omega^{(M)}\}^{-1}$, as in Assumption 3.9. Thus, pointwise in M,

$$\sqrt{N_n}(\hat{\theta}_{M,n} - \theta_0) \Rightarrow \mathcal{N}(0, V_M^\star).$$

(S3) With $b_n := \max_{i \in \mathcal{U}_n} |A_i|$,

$$\frac{b_n}{N_n} \to 0.$$

Theorem C.16 (Oracle approximation and slow-sieve attainment)**.** *Suppose Assumptions C.7, C.8, and C.15 hold. Then:*

(i) *The finite-dimensional Riesz representers converge to the oracle representer:*

$$\|\eta_M^\star - \eta^\star\|_\Omega \to 0, \qquad V_M^\star \to V^\star.$$

(ii) *There exists a deterministic, nondecreasing sequence $M_n \to \infty$ such that*

$$\frac{M_n^2 b_n}{N_n} \to 0$$

and the corresponding feasible sieve estimator satisfies

$$\sqrt{N_n}(\hat{\theta}_{M_n,n} - \theta_0) \Rightarrow \mathcal{N}(0, V^\star).$$

Proof. For part (i), Assumption C.15(S1) implies that the orthogonal projection of $\eta^\star$ onto $\mathcal{H}_M$ converges to $\eta^\star$ in $\|\cdot\|_\Omega$. The finite-dimensional Riesz representer on $\mathcal{H}_M$ is precisely $\eta_M^\star$, so

$$\|\eta_M^\star - \eta^\star\|_\Omega \to 0.$$

Since

$$V_M^\star = \frac{1}{\|\eta_M^\star\|_\Omega^2}, \qquad V^\star = \frac{1}{\|\eta^\star\|_\Omega^2},$$

the variance convergence follows by continuity and the nonzero-information condition in Assumption C.8.

For part (ii), let d_{BL} be the bounded-Lipschitz metric on probability laws and write

$$X_{M,n} := \sqrt{N_n}(\hat{\theta}_{M,n} - \theta_0).$$

For every fixed M, Assumption C.15(S2) gives

$$d_{\mathrm{BL}}\left(\mathcal{L}(X_{M,n}), \mathcal{N}(0, V_M^\star)\right) \to 0.$$

By this convergence and Assumption C.15(S3), one can choose a strictly increasing sequence of integers n_M, with $n_M \geq M$, such that, for every $n \geq n_M$,

$$d_{\mathrm{BL}}\left(\mathcal{L}(X_{M,n}), \mathcal{N}(0, V_M^\star)\right) \leq \frac{1}{M}, \qquad M^2 \frac{b_n}{N_n} \leq \frac{1}{M}.$$

Define

$$M_n := \max\{M : n_M \leq n\},$$

with $M_n = 1$ before n_1. Then M_n is deterministic, nondecreasing, and tends to infinity. Moreover,

$$\frac{M_n^2 b_n}{N_n} \leq \frac{1}{M_n} \to 0.$$

Finally, the triangle inequality gives

$$\begin{aligned} &d_{\mathrm{BL}}\left(\mathcal{L}(X_{M_n,n}), \mathcal{N}(0, V^\star)\right) \\ &\quad \leq \frac{1}{M_n} + d_{\mathrm{BL}}\left(\mathcal{N}(0, V_{M_n}^\star), \mathcal{N}(0, V^\star)\right) \to 0, \end{aligned}$$

where the last step uses part (i). Because d_{BL} metrizes weak convergence, this proves the claim. □

Remark C.17 (Scope of the slow-sieve result). Theorem C.16(ii) is an existence result. It shows that the fixed-dimensional feasible theory, population sieve approximation, and the dimension restriction

$$M_n^2 b_n / N_n \to 0$$

are jointly attainable along some sufficiently slow deterministic sequence. It does not provide an explicit growth rule or claim that the displayed rate, by itself, validates every prespecified growing sequence. Such a result would additionally require uniform growing-dimensional linearization, a triangular-array central limit theorem for the efficient direction, and uniform control of covariance estimation and estimated design projections. △

C.5 Vector parameters and optimal linear aggregation

The previous subsections stated the oracle geometry for scalar θ_0. Suppose now that

$$\theta_0 \in \mathbb{R}^p.$$

For each coordinate $j = 1, \dots, p$, define the coordinate derivative functional

$$G_j(\eta) := \frac{\partial}{\partial \theta_j} \mu_\eta(\theta)\Big|_{\theta=\theta_0}.$$

Assume each G_j is a bounded linear functional on $\mathcal{H}$. By Riesz, there exist unique elements

$$s_1^\star, \dots, s_p^\star \in \mathcal{H}$$

such that

$$G_j(\eta) = \langle s_j^\star, \eta \rangle_\Omega \qquad \text{for all } \eta \in \mathcal{H}.$$

Define the oracle information matrix

$$\mathcal{I}^\star := \big(\langle s_j^\star, s_k^\star \rangle_\Omega\big)_{j,k=1}^p.$$

If $\mathcal{I}^\star$ is nonsingular, the vector-parameter oracle variance is

$$V^\star = (\mathcal{I}^\star)^{-1}.$$

For a finite dictionary of size M, let

$$G_M := \frac{\partial}{\partial \theta} \mathbb{E}[\eta_n^{(M)}(\theta)]\Big|_{\theta=\theta_0} \in \mathbb{R}^{M \times p},$$

and let $\Omega_M \in \mathbb{R}^{M \times M}$ denote the asymptotic covariance matrix of the stacked moments. The finite-dimensional optimal GMM variance is

$$V_M^\star = (G_M^\top \Omega_M^{-1} G_M)^{-1}.$$

Equivalently, the M-dimensional moment vector can be compressed without first-order loss to the p-dimensional optimal moment

$$G_M^\top \Omega_M^{-1} \eta_n^{(M)}(\theta).$$

The feasible two-step GMM estimator uses $\widehat{\Omega}_M^{-1}$ in place of Ω_M^{-1}. Under the density condition in Assumption C.15(S1),

$$V_M^\star \to V^\star,$$

where $V^\star = (\mathcal{I}^\star)^{-1}$. If the fixed-dimensional feasible GMM conditions hold pointwise for every M, the same diagonal argument as in Theorem C.16(ii) yields a sufficiently slow deterministic sequence $M_n \to \infty$ along which the feasible estimator attains this vector oracle variance.

Remark C.18 (Discrete or non-smooth exposure parameters). The Riesz derivative argument is a

local regularity result. It applies most directly to smooth exposure parameters, such as decay rates in smooth spatial kernels. For rings, finite grids of radii, or network K-hop exposure parameters, the relevant object is generally not the local derivative

$$\partial_\theta \mu_\eta(\theta_0).$$

Those cases should be treated as finite-grid or nonregular selection problems. The product residualized moment class remains the source of valid design-based moments, but the oracle derivative geometry in this appendix should be interpreted as applying to the regular smooth case. △

D Details and Proofs for Section 5

This appendix formalizes the two-step inference problem in Section 5. The main text states the regular continuous case. Here we also allow a finite-grid exposure component and show that, after it has been selected consistently, the remaining two-step estimator is a finite-dimensional Z-estimator based on a stacked map. The first block is the *projected* Stage 1 GMM equation.

Throughout, write the exposure-map parameter as

$$\theta = (\vartheta, \rho),$$

where ϑ is a finite-grid component, possibly absent, and $\rho \in \mathbb{R}^{p_\rho}$ is a regular continuous component, possibly absent. Under Corollary B.15, assume that the finite-grid component is selected consistently:

$$\Pr(\hat{\vartheta}_n = \vartheta_0) \to 1.$$

All statements below are conditional on the selected class ϑ_0. If there is no finite-grid component, simply drop ϑ and read $\rho = \theta$. If there is no regular continuous first-stage component, the first block below is absent.

D.1 Two-step estimator as a stacked Z-estimator

Stage 1. Let $\Psi_{1i}(\rho) \in \mathbb{R}^{q_1}$ denote the Stage 1 design-based moment vector conditional on ϑ_0, with sample average

$$\bar{\Psi}_{1,n}(\rho) := \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \Psi_{1i}(\rho).$$

Let

$$\mu_1(\rho) := \lim_{n \to \infty} \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \mathbb{E}[\Psi_{1i}(\rho)]$$

denote the deterministic limit of the Stage 1 moment map, and define

$$G_1 := \frac{\partial \mu_1(\rho_0)}{\partial \rho^\top},$$

with the derivative understood in the mean-differentiability sense of Assumption B.11. The Stage 1 GMM estimator for the regular component satisfies the projected estimating equation

$$G_1^\top \Lambda_{1,n} \bar{\Psi}_{1,n}(\hat{\rho}_n) = o_p(N_n^{-1/2}), \tag{D.1}$$

where $\Lambda_{1,n}$ is the Stage 1 weight matrix. Equation (D.1) is the same projected GMM/Z condition used in Theorem 3.5. When the Stage 1 moments are exactly identified, it reduces to the usual moment equation up to a nonsingular transformation.

Stage 2. For the linear exposure–response projection, define

$$g_i(\rho) := g(W; X_i, \vartheta_0, \rho), \qquad Z_i(\rho) := \begin{pmatrix} 1 \\ g_i(\rho) \end{pmatrix}.$$

The Stage 2 target (α_0, β_0) is defined by the population moment condition

$$\lim_{n\to\infty} \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \mathbb{E}\Big[Z_i(\rho_0)\{Y_i - \alpha_0 - \beta_0^\top g_i(\rho_0)\} \Big] = 0. \tag{D.2}$$

The linear conditional mean restriction

$$\mathbb{E}[Y_i \mid g_i(\rho_0)] = \alpha_0 + \beta_0^\top g_i(\rho_0)$$

is a sufficient condition for (D.2), but is not required for the projection interpretation. Define the Stage 2 moment vector

$$\Psi_{2i}(\rho, \alpha, \beta) := Z_i(\rho)\{Y_i - \alpha - \beta^\top g_i(\rho)\} \in \mathbb{R}^{1+k}, \tag{D.3}$$

and its sample average

$$\bar{\Psi}_{2,n}(\rho, \alpha, \beta) := \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \Psi_{2i}(\rho, \alpha, \beta).$$

The Stage 2 estimator solves

$$\bar{\Psi}_{2,n}(\hat{\rho}_n, \hat{\alpha}_n, \hat{\beta}_n) = o_p(N_n^{-1/2}). \tag{D.4}$$

Stacked map. Collect the regular parameters into

$$\zeta := (\rho^\top, \alpha, \beta^\top)^\top, \qquad \zeta_0 := (\rho_0^\top, \alpha_0, \beta_0^\top)^\top.$$

Define the stacked estimating map

$$\bar{\Phi}_n(\zeta) := \begin{pmatrix} G_1^\top \Lambda_{1,n} \bar{\Psi}_{1,n}(\rho) \\ \bar{\Psi}_{2,n}(\rho, \alpha, \beta) \end{pmatrix}. \tag{D.5}$$

Equations (D.1) and (D.4) imply

$$\bar{\Phi}_n(\hat{\zeta}_n) = o_p(N_n^{-1/2}), \qquad \hat{\zeta}_n := (\hat{\rho}_n^\top, \hat{\alpha}_n, \hat{\beta}_n^\top)^\top.$$

Thus the two-step estimator is a Z-estimator for the regular parameter vector ζ.

D.2 Joint asymptotic normality

Let $\Phi(\zeta)$ denote the deterministic limit of (D.5):

$$\Phi(\zeta) := \begin{pmatrix} G_1^\top \Lambda_1 \mu_1(\rho) \\ \mu_2(\rho, \alpha, \beta) \end{pmatrix},$$

where $\Lambda_{1,n} \xrightarrow{p} \Lambda_1$ and

$$\mu_2(\rho, \alpha, \beta) := \lim_{n\to\infty} \frac{1}{N_n} \sum_{i \in \mathcal{U}_n} \mathbb{E}[\Psi_{2i}(\rho, \alpha, \beta)].$$

By construction, $\Phi(\zeta_0) = 0$.

Assumption D.1 (Two-step high-level regularity)**.** *Conditional on the selected finite-grid exposure component ϑ_0, the following conditions hold.*

(S1) ***Consistency and approximate zero.*** *The estimator satisfies $\hat{\zeta}_n \xrightarrow{p} \zeta_0$ and*

$$\bar{\Phi}_n(\hat{\zeta}_n) = o_p(N_n^{-1/2}).$$

(S2) ***Mean differentiability and rank.*** *The map Φ is mean-differentiable at ζ_0:*

$$\Phi(\zeta) = A(\zeta - \zeta_0) + r(\zeta), \qquad \|r(\zeta)\| = o(\|\zeta - \zeta_0\|),$$

for a nonsingular matrix A.

(S3) ***Joint affinity-set CLT.*** *The stacked centered map satisfies*

$$\sqrt{N_n}\{\bar{\Phi}_n(\zeta_0) - \Phi(\zeta_0)\} \Rightarrow \mathcal{N}(0, \Omega_\Phi),$$

for a finite positive definite matrix Ω_Φ*. This condition should be verified by applying the affinity-set CLT to the stacked array underlying* (D.5)*; equivalently, the CLT must hold for every fixed linear combination of the Stage 1 projected moments and Stage 2 moments.*

(S4) ***Local stochastic expansion.*** *For every sequence* $\zeta_n \xrightarrow{p} \zeta_0$*,*

$$\sqrt{N_n}\Big[\{\bar{\Phi}_n(\zeta_n) - \Phi(\zeta_n)\} - \{\bar{\Phi}_n(\zeta_0) - \Phi(\zeta_0)\}\Big] = o_p\big(1 + \sqrt{N_n}\,\|\zeta_n - \zeta_0\|\big).$$

Assumption D.1 is the exact analogue of the high-level Z-estimator assumptions used in Theorem 3.5, applied to the stacked Stage 1/Stage 2 map. Smooth exposure-response systems can verify it using derivative-level LLNs. Non-smooth exposure maps, such as rings, can verify it through mean differentiability of Φ and stochastic equicontinuity of the centered stacked process; no sample-path differentiability of the ring exposure is required.

Theorem D.2 (Joint asymptotic normality for the two-step estimator)**.** *Suppose Assumption D.1 holds. Then, under the designs* $\mathcal{D}_n$*,*

$$\sqrt{N_n}(\hat{\zeta}_n - \zeta_0) \Rightarrow \mathcal{N}(0, V_\zeta)\,, \qquad V_\zeta := A^{-1}\Omega_\Phi A^{-\top}.$$

Proof. We apply Lemma B.13 to the stacked Stage 1/Stage 2 map, taking $M_n := \bar{\Phi}_n$, $M := \Phi$, and parameter ζ. Its hypotheses are supplied by Assumption D.1: consistency $\hat{\zeta}_n \xrightarrow{p} \zeta_0$ and the approximate-zero condition $\bar{\Phi}_n(\hat{\zeta}_n) = o_p(N_n^{-1/2})$ from (S1); $\Phi(\zeta_0) = 0$ (the population moment vanishes at the truth) with nonsingular derivative A from (S2); the joint CLT $\sqrt{N_n}\{\bar{\Phi}_n(\zeta_0) - \Phi(\zeta_0)\} \Rightarrow \mathcal{N}(0, \Omega_\Phi)$ from (S3); and the local stochastic expansion condition (Z) from (S4). Because condition (Z) normalizes the remainder by $1 + \sqrt{N_n}\|\zeta_n - \zeta_0\|$, the lemma delivers the $\sqrt{N_n}$ rate rather than presupposing it, and yields

$$\sqrt{N_n}(\hat{\zeta}_n - \zeta_0) = -A^{-1}\sqrt{N_n}\{\bar{\Phi}_n(\zeta_0) - \Phi(\zeta_0)\} + o_p(1).$$

The joint CLT in (S3) and Slutsky's theorem then give $\sqrt{N_n}(\hat{\zeta}_n - \zeta_0) \Rightarrow \mathcal{N}(0, A^{-1}\Omega_\Phi A^{-\top}) = \mathcal{N}(0, V_\zeta)$. □

Remark D.3 (Discrete exposure components). The theorem is stated after conditioning on a con-

sistently selected finite-grid exposure component. If $\theta = (\vartheta, \rho)$ and $\Pr(\hat{\vartheta}_n = \vartheta_0) \to 1$, then

$$\hat{\zeta}_n(\hat{\vartheta}_n) = \hat{\zeta}_n(\vartheta_0) \qquad \text{w.p.a.1.}$$

Consequently, finite-grid selection has no first-order contribution to the distribution in Theorem D.2. This statement relies on fixed-gap selection consistency; it is not a claim about nonregular post-selection inference under ties or local-to-zero gaps. △

D.3 Inference for the exposure–response coefficient β_0

Let K_β denote the selection matrix that extracts the β coordinates from ζ:

$$\beta = K_\beta \zeta, \qquad \beta_0 = K_\beta \zeta_0, \qquad \hat{\beta}_n = K_\beta \hat{\zeta}_n.$$

Define

$$V_\beta := K_\beta V_\zeta K_\beta^\top.$$

Corollary D.4 (Asymptotic normality of $\hat{\beta}_n$)**.** *Under the assumptions of Theorem D.2,*

$$\sqrt{N_n}(\hat{\beta}_n - \beta_0) \Rightarrow \mathcal{N}(0, V_\beta).$$

Proof. This follows immediately from Theorem D.2 and the continuous mapping theorem applied to the linear map $\zeta \mapsto K_\beta \zeta$. □

In practice, estimate A and Ω_Φ by plug-in analogues evaluated at $\hat{\zeta}_n$. The covariance Ω_Φ should be estimated using the same affinity-set or graph-HAC logic as in Section 3, applied to the stacked moment contributions underlying (D.5): the projected Stage 1 contribution $G_1^\top \Lambda_1 \Psi_{1i}(\hat{\rho}_n)$ and the Stage 2 contribution $\Psi_{2i}(\hat{\rho}_n, \hat{\alpha}_n, \hat{\beta}_n)$. The resulting sandwich estimate

$$\hat{V}_\zeta = \hat{A}^{-1} \hat{\Omega}_\Phi \hat{A}^{-\top}$$

can be projected to obtain

$$\hat{V}_\beta = K_\beta \hat{V}_\zeta K_\beta^\top.$$

This delivers design-based inference for the Stage 2 exposure–response coefficients that accounts for regular continuous first-stage exposure uncertainty and for the same spatial/network dependence structure used in Stage 1.

Remark D.5 (Conservativeness of the Stage-2 variance estimator). Write $u_i := Y_i - \alpha_0 - \beta_0^\top g_i(\rho_0)$ and

$$\mu_{2,i} := \mathbb{E}[\Psi_{2i}(\zeta_0)] = \mathbb{E}[Z_i(\rho_0)\, u_i].$$

The projection target (D.2) imposes only $\frac{1}{N_n}\sum_i \mu_{2,i} \to 0$. The individual $\mu_{2,i}$ need not vanish, and they are not identified from the realized assignment: each is a design expectation involving $\widetilde{Y}_i$ at exposure values other than the realized one. The covariance Ω_Φ in Assumption D.1 is built from contributions centered at these unit-level means, whereas the feasible $\hat{\Omega}_\Phi$ described above centers at the common sample mean, the only centering available. Evaluated at ζ_0, with A_i the affinity set of unit i (Section 3.2; not the Jacobian A),

$$\mathbb{E}[\hat{\Omega}_\Phi] = \Omega_\Phi + \underbrace{\frac{1}{N_n}\sum_i \sum_{j\in A_i} \begin{pmatrix} 0 & 0 \\ 0 & \mu_{2,i}\mu_{2,j}^\top \end{pmatrix}}_{=: B_n} + o(1);$$

the Stage-1 block and the cross block are unbiased because the Stage-1 unit means vanish identically (Corollary 2.7), and evaluating at $\hat{\zeta}_n$ instead of ζ_0 adds an $o_p(1)$ term under Assumption D.1 without affecting the decomposition.

The sign of B_n determines sharpness. The own-index terms $\frac{1}{N_n}\sum_i \mu_{2,i}\mu_{2,i}^\top$ are positive semidefinite, so if the affinity sets are singletons or block-diagonal, $B_n \succeq 0$; since $V_\zeta = A^{-1}\Omega_\Phi A^{-\top}$ and $V_\beta = K_\beta V_\zeta K_\beta^\top$ are congruence transformations, which preserve the positive semidefinite order, $\hat{V}_\beta$ then weakly overstates V_β and confidence intervals for β_0 are asymptotically valid but conservative, with no assumption of correct specification. If the exposure–response is correctly specified unit by unit, so $\mu_{2,i} = 0$ for every i, then $B_n = 0$ and inference is exact. Under general overlapping dependence the cross terms $\mu_{2,i}\mu_{2,j}^\top$ are not sign-definite; a sufficient condition restoring $B_n \succeq 0$ is local alignment of the unit-level means, $\sum_i \sum_{j\in A_i}(c^\top\mu_{2,i})(c^\top\mu_{2,j}) \geq 0$ for every direction c, as when misspecification varies smoothly relative to the affinity radius. Absent such structure we confine the conservativeness statement to the block-diagonal case.

△

E Additional Application Details

The appendix follows the order of the main applications. We first report additional implementation details for the Muralidharan et al. (2023) application, where the design-based procedure largely validates the original exposure scale. We then report additional results for the Egger et al. (2022) application, where the exposure choice is more consequential for the downstream IV estimates and multiplier accounting.

E.1 Additional Stage-1 implementation details for Muralidharan et al. (2023)

This appendix provides further details on the Stage 1 exposure-mapping exercise for Muralidharan et al. (2023) used in the main text.

Throughout, we work with the same GP-level outcomes as in the main text: NREGS earnings, total income, and wage-labor income. We use the population-weighted village-level ring exposure measure and village-consistent annulus average instruments ("bands" specification), the strict placebo design, and the two-step fixed-weight GMM criterion searched over the 0.1 km grid $\mathcal{R} = \{0.1, 0.2, \ldots, 30.0\}$ km. The baseline in the main text uses 3 km-wide bands up to $r_{\max} = 20$ km ($M = 7$ moments).

Two-step criterion and Wald interval. Let $\hat{\mu}(\theta)$ denote the stacked design-based moment vector for outcome Y at radius θ, computed using the mandal-level randomization. The Stage 1 estimator minimizes a two-step fixed-weight criterion. In the first step we form the identity-weighted pilot criterion

$$\widehat{Q}_1(\theta) \;:=\; \hat{\mu}(\theta)^\top \hat{\mu}(\theta), \qquad \hat{\theta}_1 \;:=\; \arg\min_{\theta \in \mathcal{R}} \widehat{Q}_1(\theta).$$

In the second step we estimate the weighting matrix at the pilot radius and hold it fixed: $\widehat{W}_0 := \widehat{S}(\hat{\theta}_1)^{-1}$, where $\widehat{S}(\theta)$ is the design-based covariance estimator for the moment vector at radius θ. The radius estimator is then

$$\hat{\theta} \;:=\; \arg\min_{\theta \in \mathcal{R}} \widehat{Q}_2(\theta), \qquad \widehat{Q}_2(\theta) \;:=\; \hat{\mu}(\theta)^\top \widehat{W}_0 \hat{\mu}(\theta).$$

The plotted objective in all figures is $\widehat{Q}_2(\theta)/n$ for scale comparability; this rescaling does not affect the minimizing radius.

Regarding inference: because physical interaction distances vary continuously and the 0.1 km grid step is fine, the asymptotic normality result of Section 3 applies, and we report a Wald-style standard error using a finite-difference Jacobian evaluated at the neighboring grid points,

$$\widehat{\mathrm{se}}(\hat{\theta}) \;=\; \Big[n \Big(\hat{G}^\top \widehat{W}_0 \hat{G} \Big) \Big]^{-1/2}, \qquad \hat{G} \;:=\; \frac{\hat{\mu}(\hat{\theta} + h) - \hat{\mu}(\hat{\theta} - h)}{2h},$$

where $h = 0.1$ km is the grid step. The 95% Wald interval is $\hat{\theta} \pm 1.96 \cdot \widehat{\mathrm{se}}(\hat{\theta})$. One regularity point is worth making explicit here. The *sample* criterion in the ring radius is piecewise constant, with jumps at observed inter-unit distances, so the Wald interval is justified by smoothness of the *population* (design-expectation) moment map $\theta \mapsto \mu(\theta)$ near $\hat{\theta}$, not by the fineness of the grid; a fine grid only ensures that the finite-difference Jacobian approximates the population derivative well. Where that population smoothness is in doubt (e.g., the flatter total- and wage-income objectives), the exercise is better read as finite-grid model selection over the candidate radii, and we treat the reported Wald intervals as approximate in those cases.

The fixed-radius overidentification diagnostic at the selected radius is

$$J(\hat{\theta}) \;:=\; n\, \hat{\mu}(\hat{\theta})^\top \widehat{S}(\hat{\theta})^{-1} \hat{\mu}(\hat{\theta}),$$

which, because it is evaluated at the radius $\hat{\theta}$ chosen by the criterion, is under the null asymptotically χ^2 with df $= \operatorname{rank}\big(\widehat{S}(\hat{\theta})\big) - \dim\theta$ degrees of freedom ($\dim\theta = 1$ for the scalar radius), i.e., the $M-1$ convention used in the table notes. A large p-value suggests that the selected radius is consistent with the moment restrictions; it does not imply that a wide range of radii are equally plausible.

E.1.1 Randomization design and placebo assignments

Our Stage 1 procedure relies only on the mandal-level randomization of the Smartcard rollout. Muralidharan et al. (2023) randomized 296 mandals in eight districts into multiple implementation waves, stratifying by district and by a principal component of mandal socio-economic characteristics. The randomization was conducted in two batches (six "v5" districts and two "v6" districts) using R scripts that (i) construct the first principal component of stratification variables within each district, (ii) sort mandals by district, revenue division, and this principal component, (iii) assign mandals cyclically to a set of intermediate waves (seven waves in the first batch and ten waves in the second), (iv) permute wave labels within revenue-division $\times$ wave-cell groups, and (v) collapse these intermediate waves into treatment, buffer, and control groups while enforcing district-specific targets for the number of treatment and control mandals.

To align our placebo assignments with this design as closely as possible, we obtained the original randomization scripts from the authors and implemented an analogous "strict" re-randomization procedure in our code.[23]

A subtle but important limitation is that the original experiment included a third, "buffer" wave of mandals that were scheduled to receive Smartcards after the treatment group, but no household surveys were conducted in these buffer mandals. As Muralidharan et al. (2023) emphasize, the survey data are only available for treatment and control mandals, and the timing and intensity of Smartcard rollout in buffer mandals are not fully observed. In Appendix B.1, they therefore treat buffer mandals as effectively untreated when constructing neighborhood-treatment measures, and show that re-weighting buffer mandals as partially treated (e.g., 10–50 percent of full intensity) yields qualitatively similar results and typically larger estimated total effects, suggesting that the baseline specification is conservative. Following the same spirit, our re-randomization scheme is defined over the study mandals (those with survey data and $T_m \in \{0, 1\}$) and treats the buffer group as a latent complement that is not part of the analysis sample. Thus our placebo assignments approximate the conditional randomization distribution rather than the full distribution over all 296 mandals including the buffer wave.

In addition to this strict placebo design, which mirrors the wave structure and district-level targets in the original randomization, we also consider a simpler robustness specification in which we re-draw treatment indicators within strata defined by district and discrete bins of the princi-

[23] We thank the authors for sharing the original randomization code. Our implementation follows their logic but operates on the replication files rather than the original merged census data.

pal component, holding fixed the number of treated mandals in each stratum. This "simplified" design preserves the key stratification features of the experiment but abstracts from its detailed wave structure. Our main results are based on the strict design; the simplified design yields qualitatively similar Stage 1 radius estimates and objective values, reinforcing that our conclusions are driven by the core features of the mandal-level randomization rather than by fine details of the implementation.

E.2 Additional results for Egger et al. (2022)

This appendix gives the outcome-level results behind the Egger et al. application in Section 6.2. The main text reports the three aggregates most relevant for the multiplier discussion. Here we report the same selected-support IV comparison for all outcomes and display the full set of radius-path diagnostics.

A. Linear exposure index and IV implementation

For each candidate support $R_m = 2m$ km, we use Egger et al.'s annular transfer variables to form the linear spillover exposure index

$$g_v^{(m)}(W;\theta_m) = \sum_{j=1}^{m} \theta_{m,j} T^{\neg v}_{v,2(j-1)-2j}(W), \qquad \mathbf{1}'\theta_m = 1.$$

Stage 1 estimates the index weights θ_m from the design-based exposure-sufficiency moments. Stage 2 then keeps the Egger et al. IV outcome equation and replaces the original scalar spillover variable with $g_v^{(m)}(W;\widehat{\theta}_m)$. Thus the appendix tables vary the spillover exposure map, but otherwise keep the downstream IV exercise as close as possible to the original specification.

B. Full selected-support IV results

Table 6 extends the main-text comparison to all outcomes. It reports the original Egger et al. IV estimates, the selected-support linear-index estimates, the change in point estimate and standard error, the selected support $\hat{R}$, and the p-value of the support test (the likelihood-ratio test of the full 20 km model against the restricted model that zeroes the annuli beyond $\hat{R}$) that drives the selection.

The broader pattern matches the main text: recipient expenditure and asset components are stable, while non-recipient and income-side outcomes are more sensitive and noisier. The design-based step does not move every coefficient, but it moves enough of the downstream accounting to matter.

Table 6: All outcome IV components: Egger et al. vs. selected linear exposure-index estimates

Outcome	Group	Egger IV	Linear-index IV	$\Delta\beta$, ΔSE	R	Support p-value
Household expenditure, annualized	Recipients	338.6 (100.6)	320.0 (69.4)	−18.6, −31.2	6.0	0.249
	Non-recipients	334.7 (131.1)	224.9 (85.2)	−109.8, −45.9	6.0	0.249
Non-durable expenditure, annualized	Recipients	227.2 (91.4)	224.5 (68.2)	−2.7, −23.1	6.0	0.302
	Non-recipients	317.5 (126.9)	222.1 (77.5)	−95.4, −49.4	6.0	0.302
Food expenditure, annualized	Recipients	133.8 (60.6)	85.8 (35.1)	−48.0, −25.5	4.0	0.100
	Non-recipients	133.3 (63.7)	54.6 (27.3)	−78.7, −36.3	4.0	0.100
Temptation goods expenditure, annualized	Recipients	5.9 (8.4)	7.4 (5.8)	1.5, −2.6	10.0	0.214
	Non-recipients	-0.7 (6.8)	1.1 (4.6)	1.8, −2.2	10.0	0.214
Durable expenditure, annualized	Recipients	109.0 (26.5)	108.9 (14.6)	−0.2, −11.9	4.0	0.236
	Non-recipients	8.4 (11.8)	2.0 (4.4)	−6.4, −7.4	4.0	0.236
Assets (non-land, non-house), net borrowing	Recipients	183.4 (55.2)	177.3 (31.2)	−6.1, −24.0	4.0	0.250
	Non-recipients	133.1 (101.3)	41.6 (44.8)	−91.5, −56.5	4.0	0.250
Housing value	Recipients	477.4 (56.1)	425.3 (29.3)	−52.1, −26.7	4.0	0.298
	Non-recipients	80.6 (231.6)	-32.6 (96.1)	−113.2, −135.5	4.0	0.298
Land value	Recipients	158.5 (304.3)	73.2 (156.1)	−85.3, −148.2	4.0	0.062
	Non-recipients	544.8 (416.4)	180.0 (175.3)	−364.7, −241.2	4.0	0.062
Household income, annualized	Recipients	135.7 (82.1)	97.2 (36.9)	−38.5, −45.1	4.0	0.152
	Non-recipients	225.0 (100.1)	81.1 (42.6)	−143.8, −57.4	4.0	0.152
Net value of household transfers received, annualized	Recipients	-7.4 (14.5)	-42.8 (45.3)	−35.3, 30.8	16.0	0.309
	Non-recipients	8.8 (16.8)	-93.2 (77.5)	−102.0, 60.7	16.0	0.309
Profits (ag & non-ag), annualized	Recipients	35.9 (37.7)	27.4 (21.4)	−8.5, −16.3	4.0	0.157
	Non-recipients	36.4 (46.5)	10.2 (18.6)	−26.3, −27.9	4.0	0.157
Wage earnings, annualized	Recipients	73.7 (62.4)	54.6 (30.0)	−19.0, −32.4	6.0	0.910
	Non-recipients	182.6 (90.2)	103.3 (36.8)	−79.3, −53.4	6.0	0.910
Tax paid, annualized	Recipients	-0.1 (2.2)	1.6 (1.6)	1.7, −0.6	20.0	1.000
	Non-recipients	1.7 (2.5)	0.2 (1.7)	−1.4, −0.8	20.0	1.000

Notes: Entries are IV point estimates with standard errors in parentheses. Egger IV values re-estimate the original 2 km Table 1 specification on our replication sample (cluster standard errors); they closely match the published Table I values. Recipient linear-index entries report total effects; non-recipient entries report pooled spillover effects, both with joint delta-method standard errors at each outcome's selected support $\hat{R}$. The Support p-value column reports the likelihood-ratio support-test p-value at $\hat{R}$, the diagnostic shown in the paper radius-path figures.

C. Full radius-path diagnostics

Figures 3–5 report the radius-path diagnostics for the full outcome list. The shaded region marks the selected support. The lower panel in each figure reports the support p-value path, the likelihood-ratio test of the full 20 km model against the restricted model at each candidate support. Diamonds report the sublocation-level split-sample medians with their median intervals (dotted whiskers), as in the main text.

The full figure set reinforces the main-text message. The original 2 km support is rejected for every outcome; the selected support is most often modestly larger, commonly 4–6 km, with a few outcomes selecting longer supports. At the same time, the block split checks are noisy, so we interpret the selected supports as diagnostics for the exposure map.

D. The radius path and the resolution of the annular instruments

The radius-path standard errors in Figures 3–5 are not always increasing in the support; for a few outcomes, consumption among them, the intervals widen sharply near 16 km and then tighten again. This unevenness reflects the construction of the index rather than the strength of spillovers at those distances: the annular transfer variables at neighboring radii are highly collinear when the rings are narrow, so the estimated weights are only weakly determined at particular supports and the first-step uncertainty carried into the second stage is uneven rather than smoothly increasing. The reported estimates do not depend on this. Table 7 reports the recipient total at a common 4 km support for annular instrument bands of 1, 1.5, and 2 km; the point estimates move by far less than one standard error.

Table 7: Recipient total at a common 4 km support, by instrument band width

	Annular instrument band width		
Outcome	1 km	1.5 km	2 km
Consumption	296 (59)	296 (54)	296 (62)
Assets	177 (31)	177 (29)	177 (33)
Total income	97 (37)	91 (36)	97 (40)

Notes: Recipient total effect at a fixed 4 km support, with the joint delta-method standard error in parentheses, using annular instruments of width 1, 1.5, and 2 km. The index weights and the downstream IV are otherwise as in the main specification. The design-side conditional expectations in this exhibit are estimated by unpenalized least squares on the quadratic basis rather than the main specification's cross-validated ridge fit.

E. Multiplier accounting: reparameterization and delta-method inference

This appendix details the multiplier exercise of Section 6.2. Egger et al. (2022)'s Table V maps the coefficients of several weighted IV regressions into component multipliers. For component c, role g

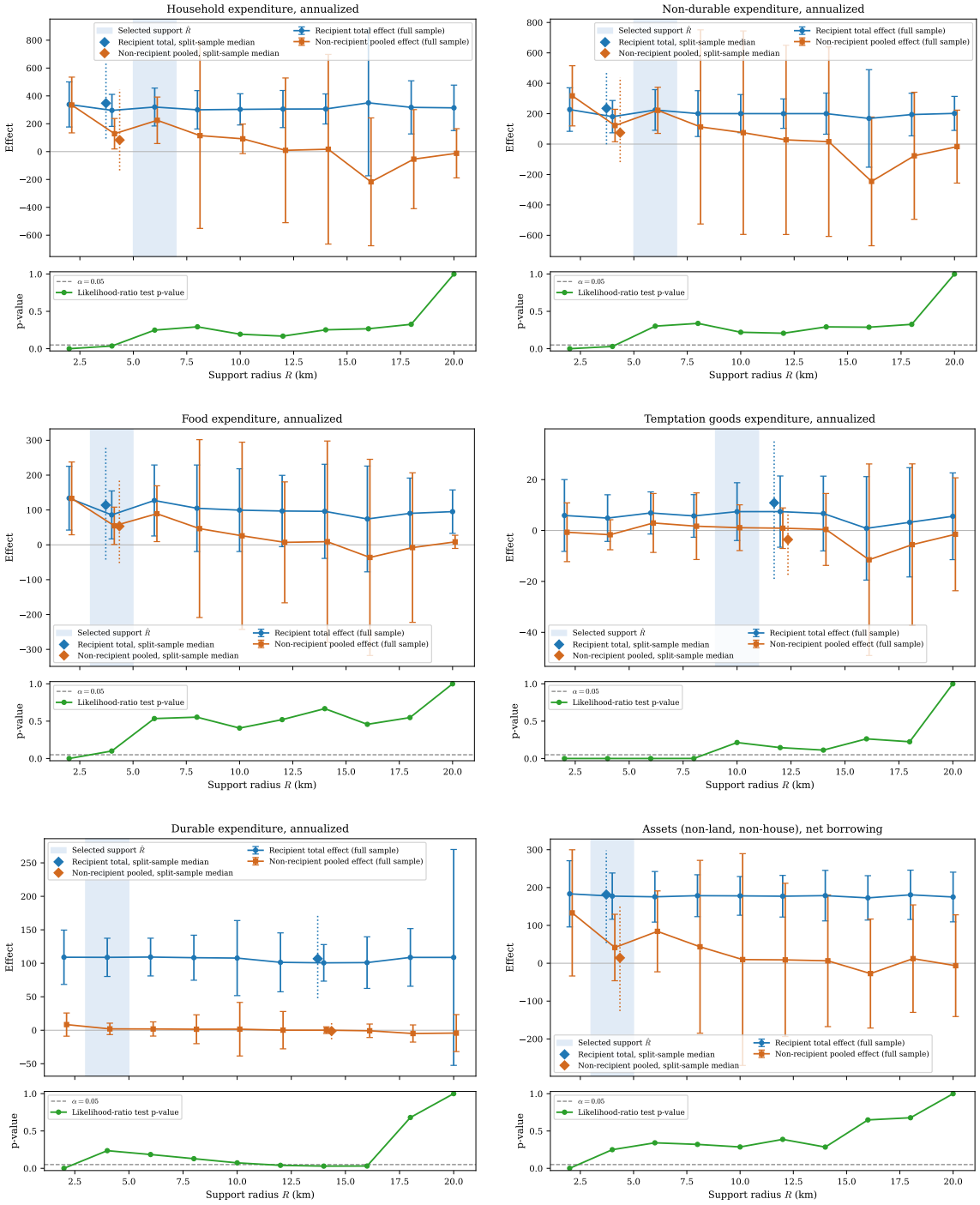


Figure 3: Radius-path diagnostics for expenditure and asset outcomes in the Egger application

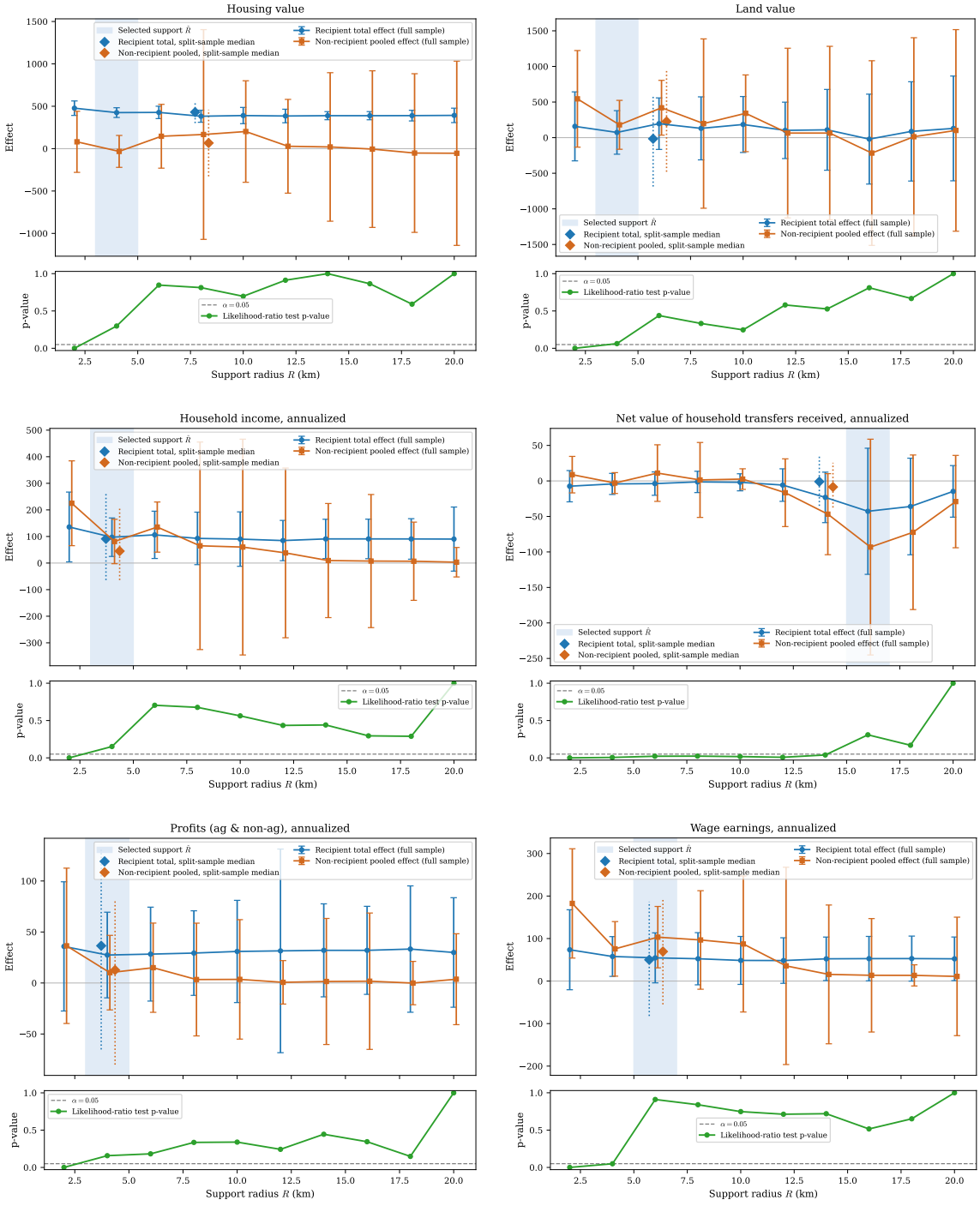


Figure 4: Radius-path diagnostics for wealth, income, and transfer outcomes in the Egger application

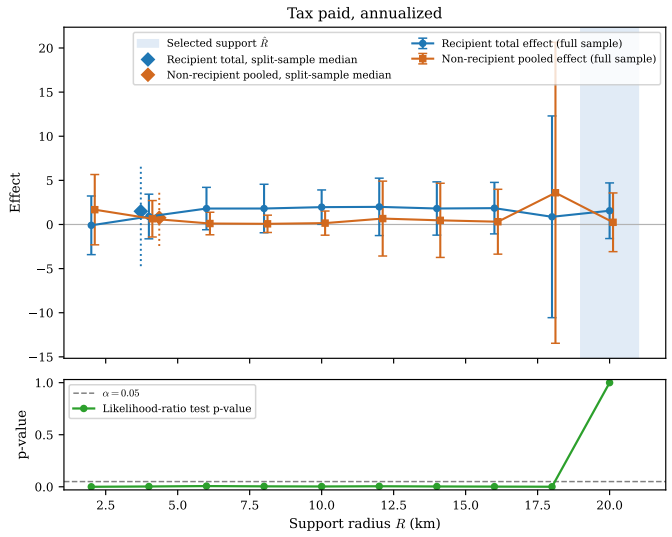


Figure 5: Radius-path diagnostics for taxes paid in the Egger application

(recipient/non-recipient), and ring k, we impose the design reparameterization

$$\beta_{cgk} = \gamma_{cg}\,\theta_{cgk}, \qquad \textstyle\sum_k \theta_{cgk} = 1,$$

so that the normalized shape θ_{cg} is exactly the Stage 1 index and γ_{cg} is a single slope on the constructed index $S_{icg}(\widehat{\theta}_{cg})$, the $\widehat{\theta}_{cg}$-weighted combination of unit i's ring exposures, which enters each Table V equation as an endogenous regressor instrumented by the corresponding index instrument. Everything else (deflators, transfer normalizations, weights, supports) is held fixed.

We propagate uncertainty to the multipliers by a cluster-robust delta-method sandwich, clustering at the sublocation level (the unit of randomization). Stacking the free shape parameters ϕ (an $m_{cg} - 1$ vector per outcome/role, with m_{cg} the number of annuli in the selected support) and the full Table V coefficient vector b into $p = (\phi', b')'$, write, for each sublocation cluster s, the stacked influence-function contribution ψ_s that combines the Stage 1 design-GMM influence, which maps the design moments to $\widehat{\phi}$, with the Table V two-stage least squares influence; the latter includes the generated-regressor adjustment by which the Stage 1 uncertainty in $\widehat{\phi}$ enters $\widehat{b}$. For the vector of reported multipliers $f(p)$ with $D = \partial f/\partial p'$, the covariance is

$$\widehat{V}_f \;=\; D\left(\tfrac{C}{C-1}\sum_{s=1}^{C}\psi_s\psi_s'\right)D', \qquad C = 84.$$

This construction never inverts the moment covariance $\widehat{\Omega} = \sum_s U_s U_s'$ (with U_s the cluster-s stacked Stage-1 and Table V moment contribution), and that is what keeps it well posed. The unit of

independence is the sublocation cluster, so $\text{rank}(\widehat{\Omega}) \leq C = 84$, while the Table V coefficient vector is high-dimensional (≈ 400); the efficient-GMM information matrix $(G'\widehat{\Omega}^{-1}G)^{-1}$, with $G = \partial\bar{g}/\partial p'$ the moment Jacobian, would invert through this rank-84 ceiling and is ill-posed. The sandwich avoids the inversion entirely: $\widehat{\Omega}$ enters only as the cluster meat, and the reported multipliers are a low-dimensional smooth function whose variance is well defined whatever the rank of $\widehat{\Omega}$. The standard errors must be read against the cluster count rather than the household count; with $C = 84$ they are comparable in magnitude to Egger et al. (2022)'s own (Table 5), and applying their wild cluster bootstrap (Rademacher signs by sublocation) to our index, holding the shape weights fixed, reproduces the Table V component to within a few percent. The resulting design-based multipliers and their benchmarks are reported in the main text (Table 5).