A Design-Based Riesz Representation Framework For Randomized Experiments

Christopher Harshaw
Department of Computer Science
UC Berkeley
charshaw@berkeley.edu

Yitan Wang
Department of Statistics & Data Science
Yale University
yitan.wang@yale.edu

Fredrik Sävje
Department of Political Science
Department of Statistics & Data Science
Yale University
fredrik.savje@yale.edu

Abstract

We describe a new design-based framework for drawing causal inference in randomized experiments. Estimands in the framework are defined as arbitrary linear functionals of the potential outcome functions, which are posited to live in an experimenter-specified function class. This makes the framework expressive, allowing experimenters to formulate and investigate a wide range of causal questions. We describe a class of estimators for estimands defined using the framework and investigate their properties. The construction of the estimators is based on insights from the Riesz representation theorem. We provide necessary and sufficient conditions for unbiasedness and consistency. Finally, we provide conditions under which the estimators are asymptotically normal, and describe a conservative variance estimator to facilitate inference about the estimands.

1 Introduction

Randomized experiments have been widely adopted in a diverse set of fields, including clinical trials, policy evaluation, and social science research. However, the use of experiments is limited, partly because conventional experimental methodologies cannot accommodate the range of settings empirical researchers are interested in. An important such limitation is the assumption of no interference, stating that a unit's outcome or behavior is unaffected by the treatment assigned to other units. Another limitation is the widespread focus on binary treatments, as opposed to, for example, continuous or set-valued treatments.

A growing body of work seeks to address these limitations, but both conventional and recently developed frameworks lack sufficient expressiveness to formulate the range of questions empirical researchers are interested in, and the frameworks generally still require strong assumptions. Furthermore, most work in this literature is justified by sampling from a super-population rather than using a design-based justification in which the randomness under consideration stems from the experiment itself. This restricts the applicability and interpretability of prior frameworks and methods.

NeurIPS 2022Workshop on Causal Machine Learning for Real-World Impact (CML4Impact 2022).

In this paper, we describe a new design-based experimental framework for causal estimation under interference to address these limitations [1]. The purpose of the framework is to be sufficiently expressive to allow experimenters to define and investigate a wide range of causal questions involving continuous or discrete treatments under rich and complex interference structures. At the same time, the framework is constructed to be sufficiently tractable to admit precise estimation and inference of the estimands defined with it. The framework unifies and generalizes all previously developed design-based frameworks that we are aware of. There are three main contributions of this paper.

1. We describe a new design-based causal inference framework, in which causal effects are defined using linear functionals on potential outcome functions. This gives empirical researchers an expressive tool to formulate new types of estimands that are relevant for policy and substantive theory.
2. We describe a new treatment effect estimator, which we call the Riesz estimator. We provide a large sample analysis of the estimator, including conditions for consistency and asymptotic normality.
3. We describe a new conservative variance estimator for the Riesz estimator, and provide conditions under which the variance estimator is consistent. This facilitates the construction of asymptotically valid confidence intervals.

The statistical methodology developed in this paper relies on insights from functional analysis. We believe these insights shed light on the underlying principles that facilitate inference of causal effects more generally, both with and without interference, and we believe these insights will be of independent interest to many researchers working in causal inference. Similar insights have recently been used in the semi-parametric observational causal inference literature. While the application of these insights in this paper is different from the application in the semi-parametric literature, we hope that this work will serve as starting point for building bridges between the design-based and semi-parametric perspectives on causal inference.

This is a short version of a longer paper, which contains more technical rigor as well as motivating examples. We refer the interested reader to our longer preprint, which is presently uncited to maintain anonymity. Because the framework and methods we describe in this paper generalize and unify approaches previously described in the causal inference literature, we find it advantageous to review and highlight these connections throughout the paper, rather than describing the connections to prior work in a dedicated section.

2 Illustration: Cash Transfer Programs

The purpose of this section is to illustrate policy-relevant questions that cannot be investigated using existing design-based frameworks and methods, but which can be investigated using the framework described in this paper. A type of policy that has received attention in the social sciences in general, and development economics in particular, is conditional or unconditional cash transfer programs, the latter of which is sometimes referred to as universal basic income (UBI) programs. An example is the study by [HS16], who investigate how different amounts of cash in an unconditional cash transfer program targeting poor households in Kenya affected consumption and well-being. Here, the experimental units are people who each receive a continuous individual treatment $Z_i \in [0, 1]$, which the the cash transfer where $Z_i = 0$ denotes no transfer and $Z_i = 1$ denotes the maximum possible transfer.

The typical estimand in a cash transfer experiment is a contrast between two levels of the transfer, resulting in a version of the all-or-nothing effect. A possible alternative estimand is the effect of a marginal increase of the transfer, which is policy relevant because it is informative of whether changes to existing programs would make sense. We can use the derivative of the potential outcome function to capture such a marginal effect. For example, the overall effect of a marginal increase the

[1] Here, the term "design-based" refers to the setting where randomization of treatment assignment is the sole basis of inference. This is in contrast to "sampling-based" inference, where experimental units are assumed to be drawn i.i.d. from a larger population, and inference is valid only when this strong assumption holds.

transfer is captured by

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y_i}{\partial z_j}(\boldsymbol{s}),$$

where $\boldsymbol{s} \in [0,1]^n$ is the current transfer level and $y_i : [0,1]^n \to \mathbb{R}$ is a unit's potential outcome function which maps *all treatment assignments* to the observed outcome. Note that this marginal effect takes into account all spillover effects of the increase; we could imagine that increasing the transfer could have a positive effect for the person who receives the transfer but be detrimental to other people because of crowding-out or inflationary effects. As highlighted by [EHM$^+$22], among others, general equilibrium effects, which is a type of interference, is an important consideration when evaluating cash transfer programs.

3 The Framework

There are n units in the experiment, indexed by integers $i \in [n]$, which may be for example participants in the experiment.

Treatments Let $\mathcal{Z}$ denote the set of all possible treatments, upon which we place no restrictions. The experimenter selects a random treatment Z from $\mathcal{Z}$ to apply in the experiment. For conventional experiments where each unit receives an individual treatment, the set of treatments will have a product structure, i.e. $\mathcal{Z}$ is $[0,1]^n$ or $\{0,1\}^n$. In these cases, the treatment is considered to be the collection of all individual treatments, i.e. $Z = (Z_1, \dots Z_n)$. We refer to the distribution of Z as the *experimental design*, which we take to be known to the experimenter. The only randomness under consideration in this paper is that which is induced by the experimental design.

Outcome Model Spaces Each unit $i \in [n]$ has a potential outcome function $y_i : \mathcal{Z} \to \mathbb{R}$ that maps each treatment to a potential observation. While this potential outcome function is unknown to experimenters, they will use their domain knowledge to impose structure on the potential outcome functions, which facilitates estimation and inference. Formally, we suppose the experimenter provides a *model space* $\mathcal{M}_i$ which is a subspace of real-valued measurable functions on $\mathcal{Z}$ with bounded variation under the design. The model spaces are flexible constructions which shall encode the interference structure. In this paper, we focus on finite-dimensional model spaces for technical brevity, though infinite-dimensional model spaces are addressed in the full preprint version of the paper. Throughout the paper, we implicitly assume that the model space is correctly specified in the sense that $y_i \in \mathcal{M}_i$ for all units $i \in [n]$.

Effect Functionals Each individual effect is defined by a linear functional evaluated at the corresponding unit's potential outcome function. Formally, for each unit $i \in [n]$, the experimenter specifies a linear *effect functional* $\theta_i : \mathcal{M}_i \to \mathbb{R}$ and the *individual effect* is defined as $\tau_i = \theta_i(y_i)$. The use of effect functionals is considerably more expressive than conventional approaches, and it accommodates a wide range of types of treatments. Our estimand of interest is the *aggregated treatment effect*, which is the average of the individual effects:

$$\tau = \frac{1}{n}\sum_{i=1}^{n}\tau_i = \frac{1}{n}\sum_{i=1}^{n}\theta_i(y_i) \quad .$$

Depending on the choice of effect functionals and model spaces, the aggregated treatment effect can express conventionally studied global treatment effects, spillover effects, and direct effects as well as new types of effects including marginal effects of continuous treatment.

3.1 Examples of the Framework

The classical no-interference setting with binary treatments, often referred to as SUTVA [Rub80, Ney23], is recovered when $\mathcal{Z} = \{0,1\}^n$ and the model space $\mathcal{M}_i$ is the span of indicators $\mathbf{1}[Z_i = 1]$ and $\mathbf{1}[Z_i = 0]$. The ATE is expressed by the effect functional $\theta_i(y_i) = y_i(\mathbf{1}) - y_i(\mathbf{0})$.

The exposure mapping framework [AS17] is one of most commonly used frameworks for experiments under interference, where outcomes are considered to be functions of low-dimensional representation of the treatment variable. Formally, let Δ be a finite set of exposure and for each unit $i \in [n]$, let $d_i : \mathcal{Z} \to \Delta$ be the exposure mapping of unit i. Expressed in our notation, the model space $\mathcal{M}_i$ contains all functions of the form $y_i(z) = \sum_{e \in \Delta} \alpha_{i,e} \mathbf{1}[d_i(z) = e]$ for real coefficients $\alpha_{i,e}$. Commonly considered effects such as global treatment effect, indirect effect, and direct effect can be written using effect functionals $\theta_i(y_i) = y_i(z_{a,i}) - y_i(z_{b,i})$ where $z_{a,i}, z_{b,i}$ are two pre-specified unit-specific treatments whose outcomes we wish to contrast.

As an alternative to the exposure mapping approach, consider a design-based version of the linear-in-means model, which is used extensively in applied work [see, e.g., CJS15, Dup14, OT12]. Methodological investigations of the model are provided by [Chi19, Leu20, HLW22], among others. Here, the outcome functions are assumed to take the following form:

$$y_i(z) = \alpha_{i,0} + \alpha_{i,1} z_i + \alpha_{i,2} \sum_{j \sim i} z_j \quad ,$$

where $i \sim j$ if units i and j are neighbors in a graph. Typically, the literature has focused on binary treatments $\mathcal{Z} = \{0,1\}^n$, but within our framework there is no conceptual difficulty in extending this to continuous treatments $\mathcal{Z} = [0,1]^n$. While the contrastive effects discussed above are most often used in this setting, our framework of effect functionals allows for marginal spill-over effects based on derivatives (Section 2) as well as integral-based functions which capture types of accumulation.

In the full pre-print version of the paper, we highlight how the framework can formulate investigations into dose-response, treatment timing, and spatial interference experiments [ASW21, PILL20, Pol20].

4 The Riesz Estimator

In this section, we present the Riesz estimator, which is a general method for estimating an aggregated treatment effect, as defined in Section 3. All proofs in this section appear in Sections D and A.

We begin by presenting a positivity assumption which is both necessary and sufficient for unbiased estimation of the aggregated treatment effect. We say that two potential outcome functions u, v in a model space $\mathcal{M}_i$ are *observationally equivalent* if $u(Z) = u(Z)$ almost surely under the experimental design.

Assumption 1 (Positivity)**.** For each unit $i \in [n]$, all pairs of observationally equivalent functions $u, v \in \mathcal{M}_i$ produce the same effect: $\theta_i(u) = \theta_i(v)$.

Proposition 4.1. *There exists an unbiased estimator of the aggregated treatment effect if and only if Assumption 1 holds.*

Assumption 1 and Proposition 4.1 formalize the way in which the experimental design must provide the experimenter sufficient information of the potential outcome functions in order to estimate the aggregated effect. Assumption 1 may be seen as a direct generalization of the more familiar positivity assumption in the standard SUTVA no-interference setting: $\Pr(Z_i = 1), \Pr(Z_i = 0) > 0$.

We are now ready to define the Riesz estimator. We begin by formally defining the Riesz representors in this context. Additionally, we show that the Riesz representors exist when positivity holds.

Definition 4.1 (Riesz Representor)**.** Given an effect functional $\theta_i : \mathcal{M}_i \to \mathbb{R}$, a function $\psi_i \in \mathcal{M}_i$ is a *Riesz representor* if $\theta_i(u) = \mathrm{E}[u(Z)\psi_i(Z)]$ for all $u \in \mathcal{M}_i$.

Proposition 4.2. *Under Assumption 1, a Riesz representor ψ_i exists for each effect functional θ_i.*

The usefulness of this representation is that it essentially gives us a direct, albeit stochastic, observation of the effect functional. A priori, it is unclear how to relate the observed outcome $y_i(Z)$ to the individual effect $\theta_i(y_i)$, which is the reason for the conventional ad hoc approach to estimation in interference settings. The usefulness of the Riesz representor is that it tells us exactly how to relate the observed outcome $y_i(Z)$ to to the effect functional $\theta_i(y_i)$.

By the definition of the Riesz representors, we may view the product $\widehat{\tau}_i = \psi_i(Z) y_i(Z)$ as an unbiased estimator for the individual treatment effect. Each of these individual-level estimators will be terribly imprecise, but they are unbiased by construction, so precision may be achieved by averaging. This motivates the Riesz estimator $\widehat{\tau}$, which is defined as

$$\widehat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau}_i = \frac{1}{n} \sum_{i=1}^{n} \psi_i(Z) y_i(Z) \quad .$$

In some ways, the Riesz estimator is similar to other design-based, linear estimators based on reweighting; indeed, in SUTVA and exposure mapping frameworks, the Riesz estimator recovers the Horvitz–Thompson estimator. However, whereas the Horvitz–Thompson estimator generally requires a small number of discrete treatments, the Riesz estimator does not have any such limitations. The following theorem shows that the Riesz estimator is unbiased under the positivity assumption.

Theorem 4.1 (Unbiasedness)**.** *Under Assumption 1, the Riesz estimator is unbiased:* $\mathrm{E}[\widehat{\tau}] = \tau$.

When the model spaces are finite dimensional, the Riesz representors may be explicitly constructed and evaluated. Formally, we require that each $\mathcal{M}_i$ is represented by an explicit set of basis functions $g_{i,1}, \dots, g_{i,d}$ on which the experimenter can compute cross-moments $\mathrm{E}[g_{i,\ell}(Z) g_{i,k}(Z)]$ and evaluate the effect functional $\theta_i(g_{i,\ell})$. In this case, a Gram–Schmidt type procedure may be used to construct the Riesz representor, and the details appear in Section D of the appendix.

5 Large Sample Analysis

In this section, we provide a large sample analysis of the Riesz estimator. In particular, we provide sufficient conditions under which the estimator is consistent and asymptotically normal. In the longer preprint version of this paper, we provide necessary and sufficient conditions for consistency by investigating the operator norm of a variance characterizing linear operator; however, this has been removed from this short version for space considerations. All proofs in this section appear in Sections B and C.

We consider an asymptotic regime that consists of a sequence of experiments with a growing number of units. Each experiment in the sequence is associated with its own experimental design, set of model spaces, linear functionals, and so on. Therefore, nearly all variables and parameters are implicitly subscripted by n to denote which experiment in the sequence they are associated with, but we omit this subscript for notational clarity. We investigate the asymptotic properties of the Riesz estimator subject to conditions on the sequence. This asymptotic regime, or minor variations of it, is commonly used in the design-based causal inference literature [see, e.g., Fre08, Lin13, SAH21, Leu22a].

We measure regularity of the collection of potential outcome functions and the Riesz representors using moment conditions. Let $\mathcal{M}^n = \mathcal{M}_1 \times \dots \times \mathcal{M}_n$ be the product of the individual model spaces. An element $\boldsymbol{u} \in \mathcal{M}^n$ is the vector-valued function $\boldsymbol{u}(Z) = (u_1(Z) \dots u_n(Z))$. We write $\boldsymbol{y}$ and $\boldsymbol{\psi}$ for the collected potential outcome function and the collected Riesz representor function, respectively. An example of a norm that has been widely used in the design-based causal inference, albeit implicitly so, is what we refer to as the max-p norm:

$$\|\boldsymbol{u}\|_{\max,p} = \max_{i \in [n]} \mathrm{E}\big[|u_i(Z)|^p\big]^{1/p}.$$

For example, it is common to assume that the fourth moment of the observed outcome exists for all units, and this is equivalent to assuming that the max-p norm of potential outcome functions $\boldsymbol{y}$ is asymptotically bounded for $p = 4$.

We measure the amount of dependence in the randomized experiment via the use of dependency neighborhoods. Two collections of functions $\mathcal{U}$ and $\mathcal{V}$ are said to be *independent with respect to the experimental design* if for every positive integer k and any functions $u_1, \dots u_k \in \mathcal{U}$ and $v_1 \dots v_k \in \mathcal{V}$, we have that the collection of random variables $(u_1(Z), \dots, u_k(Z))$ are jointly independent from the collection of random variables $(v_1(Z), \dots, v_k(Z))$ under the experimental design.

Definition 5.1. The *model dependency neighborhood* for unit i is the smallest $\mathcal{D}_i \subset [n]$ such that $\mathcal{M}_i$ and $\bigcup_{j \in [n] \setminus \mathcal{D}_i} \mathcal{M}_j$ are independent with respect to the experimental design. Let $d_{\mathrm{avg}} = n^{-1} \sum_{i=1}^{n} |\mathcal{D}_i|$ denote the average size of the dependence neighborhoods, and let $d_{\max} = \max_i |\mathcal{D}_i|$ denote the size of the largest dependence neighborhood.

Intuitively, functions associated with unit i will be independent of all functions associated with unit j if $j \notin \mathcal{D}_i$. The size of a dependency neighborhoods is one way to characterize the amount of dependence for a given unit, and the average d_{avg} and maximum $d_{\max}$ characterize the overall amount of dependence in the experiment.

5.1 Consistency

Proposition 5.1. *Let p and q be values satisfying $1/p + 1/q = 1/2$. The variance of the Riesz estimator is upper bounded by*

$$\sqrt{n \operatorname{Var}(\widehat{\tau})} \leq d_{\text{avg}}^{1/2} \|\boldsymbol{y}\|_{\max,p} \|\boldsymbol{\psi}\|_{\max,q} \ .$$

The magnitude of $\boldsymbol{\psi}$ captures the degree to which the outcomes are weighted in the Riesz estimator, and it can be interpreted as how difficult it is to estimate the treatment effect. If $\|\boldsymbol{\psi}\|_{\max,q}$ is large, the outcomes are heavily weighted and the effect is difficult to estimate. Positivity, as defined in Assumption 1, implies that $\|\boldsymbol{\psi}\|_{\max,q} < \infty$ for any fixed n, but it is possible that $\|\boldsymbol{\psi}\|_{\max,q} \to \infty$ as n grows even when positivity holds. Experimenters should try to ensure that $\|\boldsymbol{\psi}\|_{\max,q}$ is asymptotically bounded, which essentially is a slight strengthing of the positivity assumption. In a setting with discrete treatments, such as when using exposure mappings, this corresponds to ensuring that the assignment probabilities of the relevant exposures are bounded away from zero by a constant. That is, if $d_i(z) = a$ is an exposure of interest for unit i, then positivity stipulates that $\Pr(d_i(Z) = a) > 0$, while $\|\boldsymbol{\psi}\|_{\max,q} = \mathcal{O}(1)$ stipulates that $\Pr(d_i(Z) = a) \geq c$ for some constant $c > 0$.

The following corollary of Proposition 5.1 shows that consistency is ensured for the Riesz estimator provided that the potential outcomes and Riesz representors are well-behaved, and that there is not too much dependence between model spaces.

Corollary 5.1. *Under Assumption 1 and supposing regularity conditions $\|\boldsymbol{y}\|_{\max,p} = \mathcal{O}(1)$ and $\|\boldsymbol{\psi}\|_{\max,q} = \mathcal{O}(1)$ for $1/p + 1/q = 1/2$, and limited average model dependence, $d_{\text{avg}} = o(n)$, the Riesz estimator is consistent in mean square. If the condition on the average model dependence is strengthened to $d_{\text{avg}} = \mathcal{O}(1)$, the Riesz estimator is root-n consistent.*

5.2 Asymptotic Normality

The conditions used to ensure asymptotic normality are stronger than the conditions for consistency. We also impose an additional regularity condition to ensure that the limiting distribution of the Riesz estimator is not degenerate, which is implemented as an assumption that the convergence rate is not faster than root-n, which is the parametric rate. In principle, there are sequence in our asymptotic regime for which the rate is faster than this, but they are knife-edge situations where, for example, the dependence structure in the design perfectly aligns with the potential outcomes. These situations are not practically relevant, so essentially nothing is lost by imposing this assumption. Versions of this assumption is common in the design-based causal inference literature, such as Condition 6 in [AS17] and Assumption 5 in [Leu22b].

Assumption 2 (Non-degeneracy)**.** $\sqrt{n \operatorname{Var}(\widehat{\tau})} \geq c$ for some $c > 0$ and all n.

Proposition 5.2. *Under Assumptions 1 and 2, regularity conditions $\|\boldsymbol{y}\|_{\max,p}$ and $\|\boldsymbol{\psi}\|_{\max,q}$ are asymptotically bounded $\mathcal{O}(1)$ for some $1/p + 1/q = 1/4$, and limited maximum model dependence, $d_{\max} = o(n^{1/4})$, the limiting distribution of the Riesz estimator is normal:*

$$\frac{\widehat{\tau} - \tau}{\sqrt{\operatorname{Var}(\widehat{\tau})}} \xrightarrow{d} \mathcal{N}(0, 1).$$

6 Valid Inference and Variance Estimation

The asymptotic results in Section 5 suggest that the normal-based confidence intervals will be asymptotically valid for the aggregated treatment effect under certain regularity conditions. The one remaining step is to construct a variance estimator which is consistent in large samples. Full details of the variance estimator appear in the larger preprint version of the paper, though we sketch the main ideas here.

First, let us decompose the variance of the Riesz estimator into individual covariance terms:

$$\operatorname{Var}(\widehat{\tau}) = \operatorname{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} y_i(Z)\psi_i(Z)\Big) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{Cov}(y_i(Z)\psi_i(Z), y_j(Z)\psi_j(Z)) \ .$$

This decomposition demonstrates that the variance of the Riesz estimator is the average of bilinear functionals on the product of model spaces. This mirrors the structure of the aggregated treatment effect, which was the average of linear functionals on the model spaces. Thus, our approach to variance estimation is again based on Riesz representation. There are however, two main differences that must be addressed: first, the bilinear functionals must be identified with linear functionals on the tensor product $\mathcal{M}_i \times \mathcal{M}_j$ in order for us to apply a Riesz representation theorem. Second, positivity will typically not hold for each of these covariance functionals, in which case we must apply upper bounds to construct a variance estimator which is conservative in expectation. This approach is following the convention in the design-based causal inference literature [see, e.g., AGL14, Fog18, IM21]. The bound we use is inspired by the so-called Aronow–Samii bound described by [AS13, AS17] for discrete treatments, and our contribution is to generalize the bound to general tensor products of model spaces. In the full preprint version of the paper, we show that a conservative variance estimator constructed in this way is (conservatively) consistent, which facilitates the construction of asymptotically valid confidence intervals.

7 Conclusion

In this paper, we have presented a new design-based causal inference framework which allows empirical researchers to formulate and investigate an exciting range of causal questions. In addition, we have provide a general methodology for point estimation and inference (confidence intervals) based on Riesz representation. Although we have demonstrated several new types of causal questions which the framework addresses, we are most excited about the formulations which are yet to come.

References

[AGL14] Peter M. Aronow, Donald P. Green, and Donald K. K. Lee. Sharp bounds on the variance in randomized experiments. *Annals of Statistics*, 42(3):850–871, 2014.

[AS13] Peter M. Aronow and Cyrus Samii. Conservative variance estimation for sampling designs with zero pairwise inclusion probabilities. *Survey Methodology*, 39(1):231–241, 2013.

[AS17] Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference. *Annals of Applied Statistics*, 11(4):1912–1947, 2017.

[ASW21] P. M. Aronow, Cyrus Samii, and Ye Wang. Design-based inference for spatial experiments with interference. arXiv:2010.13599, 2021.

[Chi19] Alex Chin. Regression adjustments for estimating the global treatment effect in experiments with interference. *Journal of Causal Inference*, 7(2), 2019.

[CJS15] Jing Cai, Alain De Janvry, and Elisabeth Sadoulet. Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108, 2015.

[Dup14] Pascaline Dupas. Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica*, 82(1):197–228, 2014.

[EHM+22] Dennis Egger, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. General equilibrium effects of unconditional cash transfers: Experimental evidence from Kenya. *Econometrica*, in press, 2022.

[Fog18] Colin B. Fogarty. On mitigating the analytical limitations of finely stratified experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):1035–1056, 2018.

[Fre08] David A. Freedman. On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40:180–193, 2008.

[HLW22] Yuchen Hu, Shuangning Li, and Stefan Wager. Average direct and indirect causal effects under interference. *Biometrika*, in print, 2022.

[HS16] Johannes Haushofer and Jeremy Shapiro. The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *The Quarterly Journal of Economics*, 131(4):1973–2042, 2016.

[IM21] Guido Imbens and Konrad Menzel. A causal bootstrap. *The Annals of Statistics*, 49(3):1460–1488, 2021.

[Leu20] Michael P. Leung. Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 102(2):368–380, 2020.

[Leu22a] Michael P. Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90(1):267–293, 2022.

[Leu22b] Michael P. Leung. Rate-optimal cluster-randomized designs for spatial interference. arXiv:2111.04219, 2022.

[Lin13] Winston Lin. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, 7(1):295–318, 2013.

[Ney23] Jerzy Neyman. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 5(4):465–472, 1923. Reprinted in 1990.

[OT12] Emily Oster and Rebecca Thornton. Determinants of technology adoption: Peer effects in menstrual cup take-up. *Journal of the European Economic Association*, 10(6):1263–1293, 2012.

[PILL20] Georgia Papadogeorgou, Kosuke Imai, Jason Lyall, and Fan Li. Causal inference with spatio-temporal data: Estimating the effects of airstrikes on insurgent violence in Iraq. arXiv:2003.13555, 2020.

[Pol20] Michael Pollmann. Causal inference for spatial treatments. arXiv:2011.00373, 2020.

[Ros11] Nathan Ross. Fundamentals of stein's method. *Probability Surveys*, 8:210–293, 2011.

[Rub80] Donald B. Rubin. Comment: Randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association*, 75(371):591, 1980.

[SAH21] Fredrik Sävje, Peter M. Aronow, and Michael G. Hudgens. Average treatment effects in the presence of unknown interference. *Annals of Statistics*, 49(2):673–701, 2021.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [No] Potential negative impacts are not explicitly discussed.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A] No simulation were run.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A] No simulation were run.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A] No simulation were run.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A] No simulation were run.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Positivity and Unbiased Estimation

We begin by proving that positivity characterizes whether unbiased estimation is possible.

Proposition 4.1. *There exists an unbiased estimator of the aggregated treatment effect if and only if Assumption 1 holds.*

Proof. When positivity holds, the Riesz estimator is an unbiased estimator of the aggregated treatment effect.

Suppose that positivity does not hold, so that for some unit $i \in [n]$, there exist two observationally equivalent functions $u \sim v$ which have different individual effects $\theta_i(u) \neq \theta_i(v)$. An estimator is a function of observed outcomes and so any estimator must produce the same estimate whether $y_i = u$ or $y_i = v$. However, under these two possible potential outcome functions, the aggregated treatment effect is different. Thus, no unbiased estimator exists. □

Next, we show that the Riesz estimator is unbiased under the positivity assumption.

Theorem 4.1. *Under Assumption 1, the Riesz estimator is unbiased:* $\mathrm{E}[\widehat{\tau}] = \tau$.

Proof. Under positivity, individual Riesz representors $\psi_1, \dots, \psi_n$ for the effect functionals $\theta_1, \dots \theta_n$ exist. Thus, using linearity of expectation and the definition of the Riesz representors, we can compute the expectation of the Riesz estimator as

$$\mathrm{E}[\widehat{\tau}] = \mathrm{E}\Big[\frac{1}{n}\sum_{i=1}^{n} y_i(Z)\psi_i(Z)\Big] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[y_i(Z)\psi_i(Z)] = \frac{1}{n}\sum_{i=1}^{n} \theta_i(y_i) = \tau \ .$$

□

B Consistency of the Riesz Estimator

In this section, we prove the sufficient conditions for consistency of the Riesz estimator, which appear in Section 5.1. We begin by deriving an upper bound on the variance of an individual Riesz estimator in terms of the moments of the potential outcome function and the individual Riesz representor.

Lemma B.1. *Let p and q be values satisfying $1/p + 1/q = 1/2$. The variance of an individual treatment estimator is bounded as*

$$\mathrm{Var}(\widehat{\tau}_i) \leq \Big(\mathrm{E}[|y_i(Z)|^p]^{1/p} \cdot \mathrm{E}[|\psi_i(Z)|^q]^{1/q}\Big)^2 \ .$$

Proof. We can upper bound the variance by the raw second moment:

$$\mathrm{Var}(\widehat{\tau}_i) = \mathrm{E}[\widehat{\tau}_i^2] - \mathrm{E}[\widehat{\tau}_i]^2 \leq \mathrm{E}[\widehat{\tau}_i^2] = \mathrm{E}[|y_i(Z)\psi_i(Z)|^2] \ .$$

Define $p' = p/2$ and $q' = q/2$ and observe that p' and q' are conjugate pairs as

$$\frac{1}{p'} + \frac{1}{q'} = \frac{2}{p} + \frac{2}{q} = 2 \cdot \Big(\frac{1}{p} + \frac{1}{q}\Big) = 2 \cdot \frac{1}{2} = 1 \ .$$

Thus, we may use Holder's inequality to obtain that

$$\begin{aligned}
\mathrm{Var}(\widehat{\tau}_i) &\leq \mathrm{E}[|y_i(Z)\psi_i(Z)|^2] \\
&\leq \mathrm{E}[|y_i(Z)|^{2p'}]^{1/p'} \cdot \mathrm{E}[|\psi_i(Z)|^{2q'}]^{1/q'} \\
&= \Big(\mathrm{E}[|y_i(Z)|^{2p'}]^{1/2p'} \cdot \mathrm{E}[|\psi_i(Z)|^{2q'}]^{1/2q'}\Big)^2 \\
&= \Big(\mathrm{E}[|y_i(Z)|^p]^{1/p} \cdot \mathrm{E}[|\psi_i(Z)|^q]^{1/q}\Big)^2 \ .
\end{aligned}$$

□

Next, we use the bounds on the variance of the individual Riesz estimator together with the dependency neighborhood conditions to obtain a bound on the variance of the Riesz estimator.

Lemma B.2. *Let p and q be values satisfying $1/p + 1/q = 1/2$. The variance of the Riesz estimator is bounded above by*

$$\mathrm{Var}(\widehat{\tau}) \leq \frac{1}{n^2}\sum_{i=1}^{n} |\mathcal{D}_i| \Big(\mathrm{E}[|y_i(Z)|^p]^{1/p} \cdot \mathrm{E}[|\psi_i(Z)|^q]^{1/q}\Big)^2 \ .$$

Proof. We begin by decomposing the variance as

$$\mathrm{Var}(\widehat{\tau}) = \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}\widehat{\tau}_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(\widehat{\tau}_i,\widehat{\tau}_j) \ ,$$

By definition of the dependency neighborhoods, if $i \notin \mathcal{D}_j$ or $j \notin \mathcal{D}_i$, then $\widehat{\tau}_i$ and $\widehat{\tau}_j$ are independent, so that $\mathrm{Cov}(\widehat{\tau}_i,\widehat{\tau}_j) = 0$. Thus, we have the identity $\mathrm{Cov}(\widehat{\tau}_i,\widehat{\tau}_j) = \mathbf{1}[i \in \mathcal{D}_j]\mathbf{1}[j \in \mathcal{D}_i]\,\mathrm{Cov}(\widehat{\tau}_i,\widehat{\tau}_j)$. Substituting this identity into the variance calculation and using Cauchy-Schwarz inequality together with Lemma B.1, we obtain that

$$\begin{aligned}
\mathrm{Var}(\widehat{\tau}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}[i \in \mathcal{D}_j]\mathbf{1}[j \in \mathcal{D}_i]\,\mathrm{Cov}(\widehat{\tau}_i,\widehat{\tau}_j) && \text{(indicator identity)}\\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}[i \in \mathcal{D}_j]\mathbf{1}[j \in \mathcal{D}_i]\sqrt{\mathrm{Var}(\widehat{\tau}_i)\,\mathrm{Var}(\widehat{\tau}_j)} && \text{(Cauchy-Schwarz)}\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{\mathbf{1}[j \in \mathcal{D}_i]\,\mathrm{Var}(\widehat{\tau}_i)\cdot\mathbf{1}[i \in \mathcal{D}_j]\,\mathrm{Var}(\widehat{\tau}_j)}\\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}\Big(\mathbf{1}[j \in \mathcal{D}_i]\,\mathrm{Var}(\widehat{\tau}_i) + \mathbf{1}[i \in \mathcal{D}_j]\,\mathrm{Var}(\widehat{\tau}_j)\Big) && \text{(AM-GM inequality)}\\
&= \frac{1}{n^2}\sum_{i=1}^{n}|\mathcal{D}_i|\,\mathrm{Var}(\widehat{\tau}_i) && \text{(collecting terms)}\\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}|\mathcal{D}_i|\Big(\mathrm{E}[|y_i(Z)|^p]^{1/p}\cdot\mathrm{E}[|\psi_i(Z)|^q]^{1/q}\Big)^2 && \text{(Lemma B.1)} \ . \square
\end{aligned}$$

Finally, we are ready to use the previous lemma to derive a finite sample bound on the variance of the Riesz estimator.

Proposition 5.1. *Let p and q be values satisfying $1/p + 1/q = 1/2$. The variance of the Riesz estimator is upper bounded by*

$$\sqrt{n\,\mathrm{Var}(\widehat{\tau})} \leq d_{\mathrm{avg}}^{1/2}\|\boldsymbol{y}\|_{\max,p}\|\boldsymbol{\psi}\|_{\max,q} \ .$$

Proof. Using Lemma B.2 and the definition of the max-p norms, we have that the variance of the Riesz estimator may be bounded as

$$\begin{aligned}
\mathrm{Var}(\widehat{\tau}) &\leq \frac{1}{n^2}\sum_{i=1}^{n}|\mathcal{D}_i|\Big(\mathrm{E}[|y_i(Z)|^p]^{1/p}\cdot\mathrm{E}[|\psi_i(Z)|^q]^{1/q}\Big)^2\\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}|\mathcal{D}_i|\big(\|\boldsymbol{y}\|_{\max,p}\|\boldsymbol{\psi}\|_{\max,q}\big)^2\\
&= \frac{1}{n}\Big(\frac{1}{n}\sum_{i=1}^{n}|\mathcal{D}_i|\Big)\Big(\|\boldsymbol{y}\|_{\max,p}\|\boldsymbol{\psi}\|_{\max,q}\Big)^2\\
&= \frac{d_{\mathrm{avg}}}{n}\Big(\|\boldsymbol{y}\|_{\max,p}\|\boldsymbol{\psi}\|_{\max,q}\Big)^2 \ .
\end{aligned}$$

The proof follows by rearranging terms. □

Finally, the following corollary re-interprets the finite sample bound above in an asymptotic setting.

Corollary 5.1. *Under Assumption 1 and supposing regularity conditions $\|\boldsymbol{y}\|_{\max,p} = \mathcal{O}(1)$ and $\|\boldsymbol{\psi}\|_{\max,q} = \mathcal{O}(1)$ for $1/p + 1/q = 1/2$, and limited average model dependence, $d_{\mathrm{avg}} = o(n)$, the Riesz estimator is consistent in mean square. If the condition on the average model dependence is strengthened to $d_{\mathrm{avg}} = \mathcal{O}(1)$, the Riesz estimator is root-n consistent.*

Proof. Observe that under Assumption 1, the Riesz estimator is unbiased so that its mean squared error is equal to its variance. In other words, we have that

$$\sqrt{\mathrm{E}[(\widehat{\tau}-\tau)^2]} = \sqrt{\mathrm{Var}(\widehat{\tau})} \leq \sqrt{\frac{d_{\mathrm{avg}}}{n}} \cdot \|\boldsymbol{y}\|_{\max,p} \|\boldsymbol{\psi}\|_{\max,q} \enspace .$$

By assumption, $\|\boldsymbol{y}\|_{\max,p}$ and $\|\boldsymbol{\psi}\|_{\max,q}$ are asymptotically bounded as $\mathcal{O}(1)$ and so the normalized mean squared error is on the order of $\sqrt{\mathrm{E}[(\widehat{\tau}-\tau)^2]} = \mathcal{O}(\sqrt{d_{\mathrm{avg}}/n})$. Thus, the estimator is consistent in mean square if $d_{\mathrm{avg}} = o(n)$.

A similar calculation shows that $d_{\mathrm{avg}} = \mathcal{O}(1)$ guarantees $\sqrt{n}$-convergence in mean squared error, i.e. $\sqrt{n \cdot \mathrm{E}[(\widehat{\tau}-\tau)^2]} = \mathcal{O}(1)$. □

C Asymptotic Normality of the Riesz Estimation

In this section, we prove that the Riesz estimator is asymptotically normal under moment conditions on the outcomes and Riesz representors and a bound on the dependency between model spaces. Our technique will be the dependency graph version of Stein's method.

We begin by re-defining the notion of dependency neighborhoods used in this literature on Stein's method [Ros11]. Let $A_1, \dots A_n$ be random variables indexed by integers $[n]$. For each index $i \in [n]$, we define the *dependency neighborhood* to be the smallest set $\mathcal{D}_i \subset [n]$ such that

$$A_i \text{ is jointly independent of the variables } \{A_j : j \in [n] \setminus \mathcal{D}_i\} \enspace .$$

In the main paper, we introduce the notion of dependency neighborhoods at the level of the model spaces, but they induce dependency neighborhoods of random variables of the form $u_i(Z)$ for $u_i \in \mathcal{M}_i$

The following lemma is a finite sample bound on the Wassterstein distance between the normalized sum of random variables $A_1, \dots A_n$ and a normal distribution.

Lemma C.1 (Lemma 3.6 of [Ros11])**.** *Let $A_1, A_2, \dots A_n$ be random variables such that $\mathrm{E}[A_i^4] < \infty$, $\mathrm{E}[A_i] = 0$. Define $S = \frac{1}{n}\sum_{i=1}^n A_i$ and define $\sigma^2 = \mathrm{Var}(S)$, and define $X = S/\sigma$. Then for a standard normal $B \sim \mathcal{N}(0,1)$, we have*

$$d_W(X, B) \leq \frac{d_{\max}^2}{\sigma^3 n^3} \sum_{i=1}^n \mathrm{E}[|A_i|^3] + \sqrt{\frac{28}{\pi}} \cdot \frac{d_{\max}^{3/2}}{n^2 \sigma^2} \sqrt{\sum_{i=1}^n \mathrm{E}[A_i^4]} \enspace ,$$

where $d_{\max} = \max_{i \in [n]} |\mathcal{D}_i|$ is the maximum dependency degree of the random variables and $d_W(\cdot,\cdot)$ is the Wasserstein distance.

We will be considering the random variables $A_i = \widehat{\tau}_i - \tau_i$, which are the errors of the individual treatment effect estimates. To this end, we need to bound the pth moments of the absolute error of the individual treatment effect estimates. We begin with the following lemma which holds for general random variables.

Lemma C.2. *For $1 \leq p \leq \infty$ and a random variable X, we have that*

$$\mathrm{E}\Big[\big|X - \mathrm{E}[X]\big|^p\Big] \leq 2^p\, \mathrm{E}\Big[|X|^p\Big] \enspace .$$

Proof. We may use Minkowski's inequality together with Jensen's inequality applied to $x \to |x|^p$, we have that

$$\begin{aligned}
\mathrm{E}\Big[\big|X - \mathrm{E}[X]\big|^p\Big]^{1/p} &\leq \mathrm{E}[|X|^p]^{1/p} + (|\mathrm{E}[X]|^p)^{1/p} && \text{(Minkowski's inequality)} \\
&\leq \mathrm{E}[|X|^p]^{1/p} + \mathrm{E}[|X|^p]^{1/p} && \text{(Jensen's inequality)} \\
&= 2\, \mathrm{E}[|X|^p]^{1/p} \enspace . && \square
\end{aligned}$$

Next, we use this lemma together with Holder's inequality to obtain a bound on the pth moments of absolute error of the individual treatment effect estimates.

Lemma C.3. *Fix $1 \le r \le \infty$ and let p, q be value satisfying $1/p + 1/q = 1/r$. The rth moment of the absolute error of the individual treatment effect estimate is*

$$\mathrm{E}\Big[|\widehat{\tau}_i - \tau_i|^r\Big] \le \Big(2\,\mathrm{E}\Big[|y_i(Z)|^p\Big]^{1/p} \cdot \mathrm{E}\Big[|\psi_i(Z)|^q\Big]^{1/q}\Big)^r \enspace .$$

Proof. Define $p' = p/r$ and $q' = q/r$ and observe that p' and q' are conjugate pairs as

$$\frac{1}{p'} + \frac{1}{q'} = \frac{r}{p} + \frac{r}{q} = r \cdot \Big(\frac{1}{p} + \frac{1}{q}\Big) = r \cdot \frac{1}{r} = 1 \enspace .$$

We have that $\mathrm{E}[\widehat{\tau}_i] = \tau_i$. Thus, we may apply Lemma C.2 and obtain

$$\begin{aligned}
\mathrm{E}\Big[|\widehat{\tau}_i - \tau_i|^r\Big] &\le 2^r\,\mathrm{E}\Big[|\widehat{\tau}_i|^r\Big] && \text{(Lemma C.2)}\\
&= 2^r\,\mathrm{E}\Big[|y_i(Z)\psi_i(Z)|^r\Big] && \text{(definition of } \widehat{\tau}_i)\\
&\le 2^r\,\mathrm{E}\Big[|y_i(Z)|^{rp'}\Big]^{1/p'} \cdot \mathrm{E}\Big[|\psi_i(Z)|^{rq'}\Big]^{1/q'} && \text{(Holder's Inequality)}\\
&= \Big(2\,\mathrm{E}\Big[|y_i(Z)|^{rp'}\Big]^{1/rp'} \cdot \mathrm{E}\Big[|\psi_i(Z)|^{rq'}\Big]^{1/rq'}\Big)^r\\
&= \Big(2\,\mathrm{E}\Big[|y_i(Z)|^{p}\Big]^{1/p} \cdot \mathrm{E}\Big[|\psi_i(Z)|^{q}\Big]^{1/q}\Big)^r \enspace . && \square
\end{aligned}$$

Finally, we are ready to prove Proposition 5.2, which establishes asymptotic normality of the Riesz estimator under moment conditions and the assumption of limited dependence. We restate the proposition below for completeness.

Proposition 5.2. *Under Assumptions 1 and 2, regularity conditions $\|\boldsymbol{y}\|_{\max,p}$ and $\|\boldsymbol{\psi}\|_{\max,q}$ are asymptotically bounded $\mathcal{O}(1)$ for some $1/p + 1/q = 1/4$, and limited maximum model dependence, $d_{\max} = o(n^{1/4})$, the limiting distribution of the Riesz estimator is normal:*

$$\frac{\widehat{\tau} - \tau}{\sqrt{\mathrm{Var}(\widehat{\tau})}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. We seek to use Lemma C.1 on the random variables $A_i = \widehat{\tau}_i - \tau_i$, which are the errors of the individual treatment effect estimates. Note that in this case, $S = (1/n)\sum_{i=1}^n A_i = \widehat{\tau} - \tau$, $\sigma = \sqrt{\mathrm{Var}(\widehat{\tau})}$, and $X = \frac{\widehat{\tau} - \tau}{\sqrt{\mathrm{Var}(\widehat{\tau})}}$.

First, let us verify that condition required for Lemma C.1 hold. By definition of the Riesz representors, the individual treatment effect estimators are unbiased so that $\mathrm{E}[A_i] = \mathrm{E}[\widehat{\tau}_i - \tau_i] = 0$. By Lemma C.3 together with asymptotic boundedness of $\|\boldsymbol{y}\|_{\max,p}$ and $\|\boldsymbol{\psi}\|_{\max,q}$ yields

$$\mathrm{E}[A_i^4]^{1/4} = \mathrm{E}[(\widehat{\tau}_i - \tau_i)^4]^{1/4} \le 2\,\mathrm{E}\Big[|y_i(Z)|^p\Big]^{1/p} \cdot \mathrm{E}\Big[|\psi_i(Z)|^q\Big]^{1/q} \le 2\|\boldsymbol{y}\|_{\max,p}\|\boldsymbol{\psi}\|_{\max,q} < \infty \enspace .$$

Using Lemma C.1, we have that for $B \sim \mathcal{N}(0, 1)$,

$$d_W(X, B) \le \frac{d_{\max}^2}{\sigma^3 n^3} \sum_{i=1}^n \mathrm{E}[|\widehat{\tau}_i - \tau_i|^3] + \sqrt{\frac{28}{\pi}} \cdot \frac{d_{\max}^{3/2}}{n^2\sigma^2} \sqrt{\sum_{i=1}^n \mathrm{E}[|\widehat{\tau}_i - \tau_i|^4]} \enspace .$$

The first sum may be bounded as

$$\sum_{i=1}^n \mathrm{E}[|\widehat{\tau}_i - \tau_i|^3] \le \sum_{i=1}^n \Big(2\,\mathrm{E}\Big[|y_i(Z)|^p\Big]^{1/p} \cdot \mathrm{E}\Big[|\psi_i(Z)|^q\Big]^{1/q}\Big)^3 \le n(2\|\boldsymbol{y}\|_{\max,p}\|\boldsymbol{\psi}\|_{\max,q})^3 \enspace .$$

Similarly, the second sum may be bounded as

$$\sum_{i=1}^n \mathrm{E}[|\widehat{\tau}_i - \tau_i|^4] \le \sum_{i=1}^n \Big(2\,\mathrm{E}\Big[|y_i(Z)|^p\Big]^{1/p} \cdot \mathrm{E}\Big[|\psi_i(Z)|^q\Big]^{1/q}\Big)^4 \le n(2\|\boldsymbol{y}\|_{\max,p}\|\boldsymbol{\psi}\|_{\max,q})^4 \enspace .$$

By assumption, the moment quantities $\|\boldsymbol{y}\|_{\max,p}$ and $\|\boldsymbol{\psi}\|_{\max,q}$ are asymptotically bounded by $\mathcal{O}(1)$ so the Wasserstein distance is bounded by

$$d_W(X, B) = \mathcal{O}\Big(\frac{d_{\max}^2}{\sigma^3 n^2} + \frac{d_{\max}^{3/2}}{\sigma^2 n^{3/2}}\Big) \ .$$

By Assumption 2, the variance is bounded below as $\mathrm{Var}(\widehat{\tau}) = \sigma^2 \geq \Omega(1/n)$ and so the Wasserstein distance is bounded by

$$d_W(X, B) = \mathcal{O}\Big(\frac{d_{\max}^2}{n^{1/2}} + \frac{d_{\max}^{3/2}}{n^{1/2}}\Big) = \mathcal{O}\Big(\frac{d_{\max}^2}{n^{1/2}}\Big) \ .$$

By assumption, $d_{\max} = o(n^{1/4})$, so that the Wasserstein distance goes to zero. □

D Construction of Riesz Representors

In this section, we provide a sketch of how to construct Riesz representors for the finite dimensional model spaces. For purely notational convenience, we assume that each model space $\mathcal{M}_i$ is d dimensional, where d is constant across units $i \in [n]$.

Throughout the section, we assume that each model space $\mathcal{M}_i$ is explicitly represented by a set of basis functions $g_{i,1} \dots g_{i,d}$. That is, each function $u \in \mathcal{M}_i$ admits a unique decomposition $u = \sum_{\ell=1}^{d} \alpha_\ell g_{i,\ell}$. We assume that the experimenter can do the following three computational primitives:

1. Evaluate the basis function $g_{i,\ell}(z)$ for all $z \in \mathcal{Z}$.
2. Compute cross moments $\mathrm{E}[g_{i,\ell}(Z) g_{i,k}(Z)]$.
3. Evaluate effect functionals $\theta_i(g_{i,\ell})$.

Using these primitives, we show that the Riesz representor may be explicitly represented as $\psi_i = \sum_{\ell=1}^{d} \beta_{i,\ell} g_{i,\ell}$ and thus evaluated.

As a warm-up, we show how to construct the Riesz representor under the assumption of strong positivity. A model space satisfies *strong positivity* under the experimental design if for all functions $u \in \mathcal{M}_i$ which are not identically zero, $\mathrm{E}[u(Z)^2] > 0$. This is a stronger version of the positivity condition in the following sense: under strong positivity, if u and v are observationally equivalent, then strong positivity implies that $u = v$, so we trivially have that $\theta_i(u) = \theta_i(v)$. To see this, suppose that u and v are observationally equivalent so that $\mathrm{E}[(u(Z) - v(Z))^2] = 0$. Strong positivity implies that only the zero function has this property, so that $u - v = 0$.

Under strong positivity assumption, the model space $\mathcal{M}_i$ together with the inner product $\langle u, v \rangle = \mathrm{E}[uZvZ]$ form a finite-dimensional Hilbert space. Note that strong positivity guarantees that $\langle u, v \rangle = \mathrm{E}[uZvZ]$ is indeed a valid inner product, as $\langle u, u \rangle = 0$ implies that $u = 0$. Thus, we may run Gram–Schmidt orthogonalization on the basis functions $g_{i,1} \dots g_{i,d}$ to obtain an orthonormal set of functions $\phi_{i,1} \dots \phi_{i,d}$ satisfying $\mathrm{E}[\phi_{i,\ell}(Z)\phi_{i,k}(Z)] = \mathbf{1}[\ell = k]$. In this case, every function $u \in \mathcal{M}_i$ admits the decomposition $u = \sum_{\ell=1}^{d} \mathrm{E}[u(Z)\phi_{i,\ell}(Z)]\phi_{i,\ell}$. The Riesz representor admits the following closed form:

$$\psi_i(Z) = \sum_{\ell=1}^{d} \theta_i(\phi_{i,\ell})\phi_{i,\ell}(Z) \ .$$

To see this, observe that

$$
\begin{aligned}
\mathrm{E}[\psi_i(Z)u(Z)] &= \mathrm{E}\Big[\Big(\sum_{k=1}^{d}\theta_i(\phi_{i,k})\phi_{i,k}(Z)\Big)\Big(\sum_{\ell=1}^{d}\mathrm{E}[u(Z)\phi_{i,\ell}(Z)]\phi_{i,\ell}(Z)\Big)\Big] \\
&= \sum_{k=1}^{d}\sum_{\ell=1}^{d}\theta_i(\phi_{i,k})\,\mathrm{E}[u(Z)\phi_{i,\ell}(Z)]\,\mathrm{E}[\phi_{i,\ell}(Z)\phi_{i,k}(Z)] \\
&= \sum_{\ell=1}^{d}\theta_i(\phi_{i,\ell})\,\mathrm{E}[u(Z)\phi_{i,\ell}(Z)] \\
&= \theta_i\Big(\sum_{\ell=1}^{d}\mathrm{E}[u(Z)\phi_{i,\ell}(Z)]\phi_{i,\ell}\Big) \\
&= \theta_i(u) \ .
\end{aligned}
$$

When strong positivity does not hold, the above approach fails because $\mathrm{E}[uZvZ]$ no longer forms a valid inner product on $\mathcal{M}_i$ and so an orthonormal basis cannot be obtained. However, by passing through a quotient space and invoking positivity (Assumption 1), a very similar construction of the Riesz representor is possible.

Recall that we write $u \sim v$ to denote observational equivalence, i.e. $u \sim v$ if $\mathrm{E}[(u(Z) - v(Z))^2] = 0$. Observe that this defines an equivalence relation and therefore partitions $\mathcal{M}_i$ into equivalence classes. We write $[u]$ to denote the equivalence class containing u. Let $S_i = \{u \in \mathcal{M}_i : u \sim \mathbf{0}\}$ be the set of functions which are observationally equivalent to the zero function, which we remark is a subspace of $\mathcal{M}_i$. Define the vector quotient space $\mathcal{M}_i' = \mathcal{M}_i/S_i$, which is the vector space induced by all equivalence classes. The bilinear form $\langle u, v\rangle = \mathrm{E}[u(Z)v(Z)]$ forms a well-defined inner product on the finite-dimensional quotient space $\mathcal{M}_i'$, and so the pair form a Hilbert space. At this point, a modified Gram–Schmidt orthogonalization process may be used to obtain an orthonormal basis for $\mathcal{M}_i'$.

The final ingredient is constructing a linear functional $\tilde{\theta}_i$ on the quotient space $\mathcal{M}_i'$ which is a contraction of the effect functional θ_i on $\mathcal{M}_i$. By positivity, this is possible using the natural contraction, $\tilde{\theta}_i([u]) = \theta_i(u)$. This contracted functional $\tilde{\theta}_i$ is well-defined exactly because positivity implies that θ_i is constant on each equivalence class.

The discussion above proves Proposition 4.2.