Coupling Designs for Randomized Experiments with Complex Treatments*

Max Cytrynbaum[†] Fredrik Sävje[‡]

April 10, 2026

Abstract

We describe a new family of *coupling designs*, extending the basic principle of stratified randomization to experiments with continuous, constrained multivariate, text/image and other irregular treatment spaces. Our approach is to first match units into homogeneous groups, then use Monte Carlo coupling techniques to assign within-group treatments that are highly dispersed over the treatment space. We show that ensuring similar experimental units receive highly dissimilar treatments generically improves estimation efficiency. In particular, the efficiency gains from a coupling design are proportional to the product of dispersion and match quality, where dispersion measures how spread out the treatment assignments are under a given coupling relative to independent randomization. We develop a new spectral analysis, revealing how efficiency depends on a match between the smoothness and shape of the estimator's influence function and the principal directions of a given coupling. We illustrate how coupling designs work in practice using a cash transfer experiment in development economics and a discrete-choice experiment in two-sided marketplaces.

*We are grateful to seminar and conference participants at Boston College, Emory, Columbia, Harvard, Northwestern, NYU, University of Pittsburgh, Princeton, and Stanford for comments that greatly improved the paper.

[†]Department of Economics, Yale University.

[‡]Department of Economics, Uppsala University.

1 Introduction

Consider assigning treatments D_i from a distribution F to units $i \in [n]$ in a randomized experiment. It is widely recognized that for simple discrete distributions F, experimenters can improve estimation efficiency by stratifying on covariates at design time. For example, if $D_i \in \{0, 1\}$ and $F = \text{Bernoulli}(1/2)$, a *matched pairs* design would match units with similar baseline covariates $X_i \approx X_j$ into pairs, then set $D_i = 1$ and $D_j = 0$ or vice-versa, with equal probability. Doing so balances covariates between treatment and control groups, improving precision for generic parametric causal estimators (Bruhn and McKenzie, 2009; Imai et al., 2009; Bai et al., 2023). Similarly, if there are k treatments and $F = \text{Unif}([k])$, one can improve efficiency by *matched k-tuples* randomization: match units into groups of k with similar baseline covariates, then randomly assign one unit to each treatment (Cochran and Cox, 1957).

Moving beyond these well understood situations, consider a researcher interested in the effect of cash grants $d \in [0, u]$ on future household consumption $Y_i(d)$. They want to estimate an approximation of the average dose-response curve $d \mapsto n^{-1} \sum_i Y_i(d)$ for grant amounts $d \in [0, u]$. Identification of the full curve requires continuous randomization of the treatment, such as $F = \text{Unif}[0, u]$. However, this makes stratification impossible since there are infinitely many treatment levels. One possible solution is to just assign treatments independently between units, forgoing any efficiency improvements.

An alternative solution that potentially improves efficiency is to first discretize the treatment, say into $k = 20$ treatment levels, then apply matched k-tuples randomization. However, such discretization generally requires changing the causal estimand, since the dose-response is non-identified at intermediate points. Even leaving such identification issues aside, match quality deteriorates rapidly as k increases, reducing the efficiency gains from stratified randomization. Indeed, in moderately sized experiments, it can be challenging to find well-matched pairs of $k = 2$ units even for relatively low-dimensional covariates, and will be much more difficult to find well-matched groups of $k = 20$ units.

The challenge of improving efficiency in experiments with complex treatments is not limited to continuous treatments. The purpose of this paper is to develop a new family of *coupling designs* that extends the basic principle of stratification to allow for efficient randomization in experiments with continuous, constrained multivariate, text/image and other irregular treatment spaces. We make three main contributions in this paper:

1. In Sections 2 and 3, we introduce a new family of coupling designs that enables efficient randomization in experiments with complex treatments and illustrate how they can be used in practice. The key idea is to first match units into homogeneous groups, then use coupling techniques to assign treatments within these groups to be highly dispersed over the treatment space. We construct such couplings for treatment spaces $\mathcal{D} \subseteq \mathbb{R}^m$ by combining techniques from the Monte Carlo integration literature with tools from optimal transport.

2. In Sections 4 and 5, we introduce the concepts of dispersion and match quality, showing that the efficiency gain from a coupling design is proportional to the product of these two key forces. Section 5 contains our main theoretical results.

We develop a novel spectral analysis that shows how efficiency depends on both the smoothness and shape of the estimator's influence function as well as the principal directions of the implemented coupling. Section 6 applies this analysis to compare various choices of coupling designs.

3. In Section 7, we develop asymptotic theory for the family of coupling designs. We show asymptotic normality of parametric estimators under coupling design randomization, and develop consistent variance estimators enabling valid inference on general causal parameters.

1.1 Related Literature

Experimental design for causal inference problems goes back to at least Fisher (1926), and has been an active area of research in statistics and econometrics for decades. Recent surveys include Athey and Imbens (2017) and Bai et al. (2025). Several approaches to improving efficiency in randomized experiments have been developed in this literature, including rerandomization (Morgan and Rubin, 2012; Li et al., 2018), pure optimization (Kasy, 2016; Kallus, 2017), as well as approaches based on discrepancy minimization (Harshaw et al., 2024). However, these approaches are not designed to handle complex treatment spaces. For example, none of these approaches can accommodate continuous treatment distributions.

There is a large literature on improving efficiency by stratified randomization, which is closely related to our work. For example, matched pairs designs assign opposite treatments to pairs of similar units (Fisher, 1935; Greevy et al., 2004; Bruhn and McKenzie, 2009). The efficiency properties of matched pairs have been studied by Imai (2008), Fogarty (2018), Bai (2022), Bai et al. (2021), Pashley and Miratrix (2021), among others. Generalizations to matched k-tuples and other stratified designs have been considered by several authors (Cochran and Cox, 1957; Higgins et al., 2015; Bugni et al., 2018, 2019; Cytrynbaum, 2023; Bai et al., 2024, 2023). Perhaps the closest paper to our work is Koo and Pashley (2026), which provides a modern, potential outcomes based analysis of incomplete block designs (Yates, 1936; Kempthorne, 1956), which classically were studied using a restrictive, model-based approach. Incomplete block designs refer to stratified randomization where the total number of distinct treatments is larger than the block (stratum) size. Similarly, in our setting it will either be highly inefficient or outright impossible to implement every distinct treatment within each matched group.

The coupling designs we introduce in this paper extend the basic mechanism of stratified randomization to complex treatment spaces by replacing treatment permutations within strata with general negatively dependent couplings that disperse treatment assignments across the treatment space $\mathcal{D} \subseteq \mathbb{R}^m$. In addition to enabling covariate-balancing randomization in a much broader class of useful experiments, coupling designs also provide new insights into conventional stratified designs, and can even improve on such designs in classical settings by more efficiently trading off between the key forces of dispersion and match quality.

To generate highly dispersed treatments within matched groups, we combine coupling techniques from the Monte Carlo integration literature (Robert and Casella, 2004; Owen, 2013) with geometry-preserving maps from optimal transport theory (Brenier, 1991; Mérigot, 2011; Carlier et al., 2010).

2 Overview and Illustrative Applications

2.1 Generalizing Matched Pairs

Coupling designs extend the basic mechanism of stratification to allow efficient randomization from any distribution F within tightly matched groups of units. To illustrate this idea, consider extending matched pairs designs to allow randomization of continuous, univariate treatments $D_i \in \mathbb{R}$. The conventional matched pairs design can be understood as drawing $(D_i, D_j) \sim G$ from a coupling G with fixed marginals $G_i = G_j = \text{Bernoulli}(1/2)$ and $D_i = 1 - D_j$, which achieves maximal negative correlation: $\text{Corr}_G(D_i, D_j) = -1$. For more general distributions F, this perspective suggests first matching pairs of similar units i and j, then drawing $(D_i, D_j) \sim G$ from a coupling with fixed marginals $G_i = G_j = F$ and strong pairwise negative correlation: $\text{Corr}_G(D_i, D_j) \ll 0$. One way to construct such a coupling for any F is via a classic idea from Monte Carlo integration theory known as *antithetic variates* sampling (Hammersley and Morton, 1956).

Antithetic Matched Pairs. Antithetic variates sampling generates assignments $D_i^* = F^{-1}(U)$ and $D_j^* = F^{-1}(1-U)$ using a common uniform variate $U \sim \text{Unif}[0, 1]$. Since both U and $1 - U$ are uniform, the quantile transform produces $D_i^*, D_j^* \sim F$ marginally. By drawing opposing quantiles U and $1-U$, the coupling $(D_i^*, D_j^*) \sim G$ induces strong negative correlation between the treatments. In fact, the results of Hoeffding (1940) show that for any monotone function $y(\cdot)$ of the treatment, antithetic variates achieves minimal correlation:

$$\text{Corr}(y(D_i^*), y(D_j^*)) = \min_{G_i=G_j=F} \text{Corr}_G(y(D_i), y(D_j)). \tag{2.1}$$

We can construct an *antithetic matched pairs* design by assigning treatments $(D_i^*, D_j^*) \sim G$ within tightly matched pairs, inducing strong negative correlation while also implementing the chosen marginal $D_i^* \sim F$ for each unit $i \in [n]$. When $F = \text{Bernoulli}(1/2)$, the design yields $D_i^* = 1 - D_j^*$, recovering the conventional matched pairs design. However, this construction can be used more generally for any univariate distribution F, for example $F = \text{Unif}[0, u]$.

Antithetic variates were developed to improve efficiency in Monte Carlo integration problems, such as estimating $\theta_0 = E_F[y(D)]$ using $n^{-1} \sum_i y(D_i)$. In contrast to classic Monte Carlo integration, in causal inference units generally have heterogeneous responses to the treatment with $Y_i(\cdot) \neq Y_j(\cdot)$. We can use matching to enforce approximate homogeneity at the group level. After successful matching, $Y_i(\cdot) \approx Y_j(\cdot)$ within matched pairs, allowing the design to leverage the efficiency improvements from antithetic variates as if the responses were homogeneous.

Simplifying for illustration, consider an estimator $\widehat{\theta} = (1/2)(Y_i(D_i) + Y_j(D_j))$ of the average outcome $\theta_0 = (1/2)(E_F[Y_i(D)] + E_F[Y_j(D)])$. If units are perfectly matched $Y_i(\cdot) = Y_j(\cdot) = y(\cdot)$, the variance relative to independent assignment is

$$\frac{\text{Var}_G(\widehat{\theta})}{\text{Var}_{G_{iid}}(\widehat{\theta})} - 1 = \text{Corr}_G(Y_i(D_i), Y_j(D_j)) = \text{Corr}_G(y(D_i), y(D_j)). \tag{2.2}$$

If $y(\cdot)$ is monotone, then by Equation 2.1 this antithetic pairs design can significantly improve efficiency. Indeed, it minimizes variance of the estimator $\widehat{\theta}$ among all pairwise couplings G.

2.2 Coupling Designs

The simple example above illustrates how matched pairs randomization can be extended to randomize efficiently from any univariate distribution F when the response function is monotone. In what follows, we generalize this construction to provide coupling-based randomization methods for general treatment spaces $\mathcal{D} \subseteq \mathbb{R}^m$. We show that these methods improve efficiency under weak smoothness conditions on the potential outcomes $Y_i(\cdot)$.

We construct general coupling designs by first matching the experimental units into homogeneous groups of size $k \geq 2$ using covariates. Next, treatments are drawn within each group from a coupling $(D_i)_{i=1}^k \sim G$ with marginals $G_i = F$ for all $i \in [k]$, for a fixed distribution F over the treatment space $\mathcal{D} = \mathrm{Supp}(F)$. We can view this as a matched k-tuples design with coupling-based randomization within each group. The family of coupling designs can accommodate very general marginal distributions F and spaces $\mathcal{D}$. When F is discrete, we also recover conventional stratified randomization for appropriate choices of k and G. There are many possible couplings that can be used to randomize within groups. We discuss several examples and provide a general construction in Section 3 below.

Dispersed Treatments. As in Equation 2.2, we would like to produce negatively correlated treatments $(D_i)_{i=1}^k$ within matched groups of k. For multivariate $D \in \mathbb{R}^m$, this can be achieved by sampling $(D_i)_{i=1}^k$ to be highly dispersed or "spread out" over the treatment space $\mathcal{D}$. Making treatments dispersed in this way implies that for sufficiently smooth functions $\phi : \mathcal{D} \to \mathbb{R}$,

$$\mathrm{Corr}_G(\phi(D_i), \phi(D_j)) \ll 0, \qquad i \neq j. \tag{2.3}$$

When tuple size k is large, it is possible to achieve higher dispersion, since this allows us to coordinate randomization of $(D_i)_{i=1}^k$ to cover more of the treatment space. However, as tuple size k increases, match quality becomes worse and the response functions $Y_i(\cdot)$ and $Y_j(\cdot)$ of matched units i and j are less similar on average. Using a formalization of these concepts introduced in Section 4, we show that the efficiency gain from a coupling design relative to independent assignment is

$$\text{Efficiency Gain} = \text{Dispersion} \times \text{Match Quality}.$$

Intuitively, by assigning matched k-tuples of similar units to highly dissimilar treatments $(D_i)_{i=1}^k$, we prevent spurious *in-sample* correlations from arising between the treatment assignments and unit-specific heterogeneity. We formalize how this generalizes classical notions of covariate balance to much more complex treatment spaces and distributions.

To show our main result in full generality, Section 5 defines a coupling operator U_G whose eigenspaces in $L^2(F)$ can be viewed as the principal directions of the coupling G with respect to random sampling. The dispersion $\mathrm{Disp}_G(\phi)$ of any $\phi \in L^2(F)$ decomposes orthogonally over these eigenspaces, so the overall efficiency gain depends on how well the influence functions $s_i(\cdot)$ align with the high-dispersion directions of G. This yields a general decomposition of the variance reduction from coupling designs into a weighted sum of eigenspace-specific dispersion $\times$ match quality terms, where the weights reflect the approximation quality of $s_i(\cdot)$ on each eigenspace.

2.3 Illustrative Application: Discrete Choice

To make the ideas of the paper concrete, we describe an application of an experiment with complex treatments where a coupling design can be used but conventional stratification will either be impossible or have poor performance.

Consider estimating the probability that a product d is purchased, say $Y_i(d) = 1$, when it is shown to unit i. For example, the product could be a restaurant on a food delivery platform that provides individual restaurant promotions to its users. While all products are unique, they can typically be compared. For example, we could featurize each restaurant $d_r \in \mathbb{R}^m$ using a vector of attributes like cuisine type, average price, rating, and so on. Then the treatment space $\mathcal{D} = \{d_1, \ldots, d_R\}$ is an irregular, discrete subset of $\mathbb{R}^m$. Suppose the platform uses a discrete choice model to estimate user preferences. They assign $D_i \sim F$ over restaurants $\mathcal{D}$, observe choices $Y_i(D_i)$, then estimate a Logit discrete choice model with

$$P(Y_i = 1|D_i) = L(\theta' D_i). \tag{2.4}$$

In Section 3.1 below, we show the Logit MLE consistently estimates an approximation to the dose-response $n^{-1}\sum_i Y_i(d)$, the proportion of units who would choose to purchase when shown product d. We can use coupling designs to improve precision for estimating such causal parameters. To do so, we construct couplings G that produce high dispersion over the irregular point cloud of possible product types, then sample $(D_i)_{i=1}^k \sim G$ within matched k-tuples of similar users. Our construction combines tools from Monte Carlo integration theory and optimal transport (OT), see Section 3.2 for details and Figure 1 for a visualization of the treatments $(D_i)_{i=1}^k$.

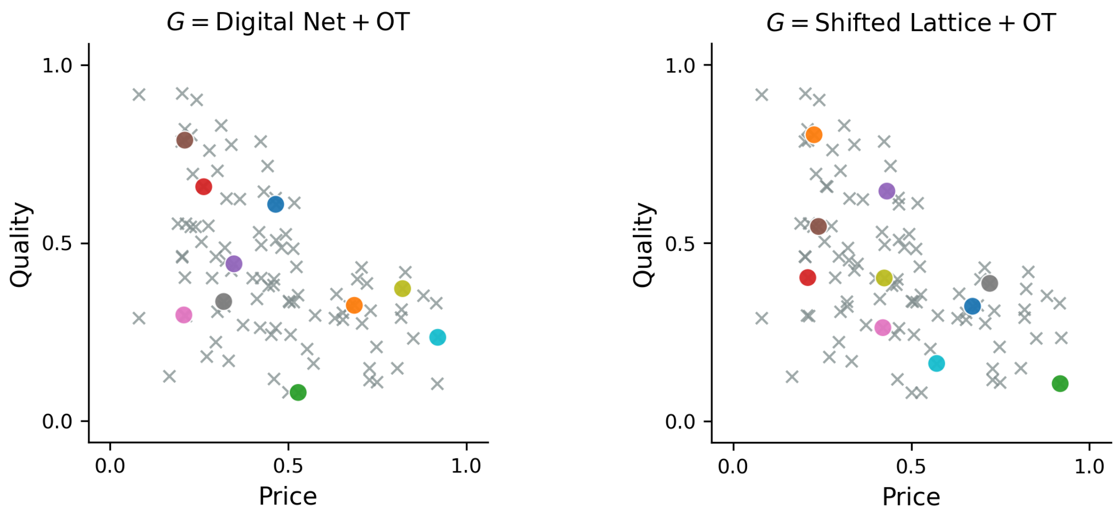


Figure 1: Dispersed treatment assignments $(D_i)_{i=1}^k \sim G$ for $k = 10$ over the irregular discrete space of restaurant types $\mathcal{D} \subseteq \mathbb{R}^2$, represented as gray x's.

For comparison, stratified randomization would randomly assign each of R restaurants without replacement to groups of $k = R$ users, uniformly at random. If the experiment size $n < R$, this is not possible. More generally, if R is large, stratified randomization will behave similarly to iid randomization, due to poor match quality within groups of $k \geq R$ units. By contrast, coupling designs allow experimenters to draw treatments $(D_i)_{i=1}^k$ within matched k-tuples for any $k \geq 2$, preserving match quality while also dispersing treatments through the space $\mathcal{D}$.

Exploiting Smoothness. This example also shows how coupling designs can naturally exploit the smoothness of the outcome functions $Y_i(d)$ over $d \in \mathcal{D}$. To see this, note that it is natural to assume that units have similar preferences $Y_i(d_r) \approx Y_i(d_l)$ over restaurants with $|d_r - d_l| \approx 0$, a form of smoothness. Because of this, assigning similar pairs of units $Y_i(\cdot) \approx Y_j(\cdot)$ to restaurants d_r and d_l effectively wastes a sample: we learn the same thing from observing $Y_i(d_r)$ and $Y_j(d_l)$. Coupling designs prevent this by assigning matched groups of similar units $Y_i(\cdot) \approx Y_j(\cdot)$ to highly dissimilar restaurants. Our core theory in Sections 5 and 6 explicitly connects the efficiency gain from coupling designs to the smoothness of the responses $Y_i(\cdot)$.

2.4 Additional Illustrative Applications

Example 2.1 (Complex Factorial Treatments)**.** Consider a cash transfer program in which each unit i both receives a grant amount $D_{i1} \in [0, u]$ and is assigned to one of t job training frequencies $D_{i2} \in \{1, \ldots, t\}$. This is a mixed continuous-discrete factorial experiment with $D_i = (D_{i1}, D_{i2}) \in \mathcal{D}$ for treatment space $\mathcal{D} = [0, u] \times \{1, \ldots, t\}$. A natural treatment distribution is $F = \text{Unif}(\mathcal{D})$. The researcher may wish to estimate a regression model with main effects and interactions:

$$Y_i \sim 1 + D_{i1} + D_{i2} + D_{i1}D_{i2}. \tag{2.5}$$

In Section 3.1 below, we show how this regression identifies a best approximation to the dose-response function $(d_1, d_2) \mapsto n^{-1} \sum_i Y_i(d_1, d_2)$.

To improve the precision of such regression estimators, we construct couplings G that spread out the treatments $(D_i)_{i=1}^k$ through the space $\mathcal{D}$, randomly within matched k-tuples of participants. More generally, our method enables efficient randomization in factorial designs with treatment space of the form $\mathcal{D} = \times_{j=1}^m \mathcal{D}_j$, where each $(\mathcal{D}_j, F_j)$ may be either continuous or discrete.

Example 2.2 (Constrained Multivariate Treatments)**.** In the previous example, the treatment space took the simple form of a product set $\mathcal{D} = [0, u] \times \{1, \ldots, t\}$. In some experiments, there may be restrictions on which treatments can be implemented due to logistical, fairness, budgetary or other considerations. For example, consider randomizing assistance $D =$ (fertilizer, seeds, loans, technical training) to farmers, in various amounts. The goal is to approximate the production function $d \mapsto n^{-1} \sum_i Y_i(d)$, where $Y_i(d)$ is farmer revenue. However, fairness considerations may restrict the variation in total treatment value $v = D'p$ across farmers for a vector of prices p of the different types of assistance. Then the feasible treatment space $\mathcal{D} = \{d \in \mathcal{D}_{pre} : l \leq d'p \leq u\}$, where $\mathcal{D}_{pre}$ is a product as in Example 2.1.

We can use coupling designs to improve efficiency by dispersing the treatments $(D_i)_{i=1}^k$ over the constrained space $\mathcal{D}$, within matched k-tuples of similar farmers. More generally, our method enables efficient randomization over constrained spaces $\mathcal{D} = \{d \in \mathcal{D}_{pre} : C(d) \leq B\}$ for a product set $\mathcal{D}_{pre}$ and constraint function $C(\cdot)$.

Example 2.3 (Text and Image Treatments)**.** Consider the correspondence study of Bertrand and Mullainathan (2004) that assigns fictitious job applicants with different racial cues and qualifications to real job postings to measure the effect of race perceptions on interview callback probability. We can view the treatment as $D = (R, Z)$, where $R \in \{0, 1\}$ is race and Z are high-dimensional application features such as the text on the resume, aspects of the photo of the applicant and so on. Suppose we let $D \sim F$ with $R \perp\!\!\!\perp Z$. We can regress $Y_i \sim 1 + R_i$ to estimate the average treatment effect

$$\theta_n = \frac{1}{n} \sum_{i=1}^{n} E_Z\big[Y_i(1, Z) - Y_i(0, Z)\big]. \tag{2.6}$$

To improve precision, coupling designs would assign $(D_i)_{i=1}^k \sim G$ so that both race and high-dimensional applicant features Z_i are dispersed within matched k-tuples of similar job postings. Alternatively, this design can be used to send multiple highly dispersed applications $(D_i)_{i=1}^k \sim G$ to the same job posting, resulting in perfect match quality.

3 Coupling Designs

3.1 Causal Estimands and Estimators

We consider estimators $\widehat{\theta} = n^{-1} \sum_i s_i(D_i)$, where the functions $s_i(\cdot)$ are non-random in a design-based framework. For example, we could have $s_i(d) = s(Y_i(d), d, X_i)$ for a fixed function $s(\cdot)$ of the data. The corresponding finite-population estimand is

$$\theta_n \equiv n^{-1} \sum_i E_F[s_i(D)]. \tag{3.1}$$

We can view θ_n as a "fully heterogeneous" version of the classic Monte Carlo estimand $\theta_0 = E_F[s(D)]$. In fact, these types of estimators and estimands are ubiquitous in causal inference problems (Harshaw et al., 2025). Thus, experimental design in causal inference can be viewed as a heterogeneous Monte Carlo integration problem. For brevity, in what follows we will denote $E_n[a_i] \equiv n^{-1} \sum_i a_i$ for any array $(a_i)_{i=1}^n$.

Example 3.1 (Dose-Response BLP)**.** Define the average dose-response function $\bar{Y}_n(d) = E_n[Y_i(d)]$ for $d \in \mathbb{R}^m$. The best linear approximation (BLP) coefficient is

$$\theta_{\text{BLP}} = \operatorname*{argmin}_{\theta} \min_{\alpha} E_F[(\bar{Y}_n(D) - \alpha - \theta' D)^2]. \tag{3.2}$$

For binary treatments, this is just the sample average treatment effect: $\theta_{\text{BLP}} = E_n[Y_i(1) - Y_i(0)]$. More generally, see Yitzhaki (1996) for an interpretation of θ_{BLP} as a weighted average of the marginal effects $\partial \bar{Y}_n(d)/\partial d$ for $d \in \mathbb{R}$. For estimation, let $H(d) \equiv \text{Var}_F(D)^{-1}(d - E_F[D])$ and $\widehat{\theta} = E_n[Y_i(D_i)H(D_i)]$, which is in the form above for $s_i(d) = Y_i(d)H(d)$. This can be understood as a generalization of the classic Horvitz-Thompson estimator. See Harshaw et al. (2023) for a related construction in the context of bipartite experiments. We have $\theta_n = n^{-1} \sum_i E_F[s_i(D)] = \theta_{\text{BLP}}$, so the estimator recovers the BLP coefficient.

Influence Functions. A variety of estimators admit the design-based asymptotic linearization $\widehat{\beta} - \beta_n = E_n[s_i(D_i)] + o_p(n^{-1/2})$. Because of this, our efficiency analysis of the quantity $E_n[s_i(D_i)]$ under coupling designs also characterizes the first-order efficiency of significantly more general parametric estimators. For example, let $\widehat{\beta}$ be the coefficient from the OLS regression $Y_i \sim 1 + D_i$. Then, under weak conditions on the design,

$$\widehat{\beta} - \theta_{\mathrm{BLP}} = E_n[s_i(D_i)] + O_p(n^{-1}). \tag{3.3}$$

We have $s_i(d) = e_i(d)H(d)$ for residual $e_i(d) = Y_i(d) - E_F[\bar{Y}_n(D)] - \theta'_{\mathrm{BLP}}(d - E_F[D])$. We provide more details in Appendix B.2. In a slight abuse of terminology, we refer to $s_i(\cdot)$ as unit i's *influence function.*

Example 3.2 (Discrete Choice)**.** Recall the binary choice setting in Section 2.3. The dose-response function $\bar{Y}_n(d) \equiv E_n[Y_i(d)]$ for potential outcomes $Y_i(d) \in \{0, 1\}$ is the proportion of units in the finite population that choose $Y_i(d) = 1$ when given treatment $d \in \mathbb{R}^m$. Suppose that we estimate the logit model $\bar{Y}_n(d) = L(\beta' d)$ using maximum likelihood, with $L(x) = 1/(1 + e^{-x})$. Let $\widehat{\beta}$ be the MLE estimator of the coefficients of the logit model. We show $\widehat{\beta} \xrightarrow{p} \beta_n$, where β_n is the best logistic approximation to the dose-response $\bar{Y}_n(\cdot)$ in terms of lowest expected KL-divergence. In particular, let $\mathrm{KL}(p \,\|\, q)$ be the KL-divergence between two Bernoulli distributions with probability parameters p and q, then

$$\beta_n = \underset{\beta}{\mathrm{argmin}} \; E_F\big[\mathrm{KL}\big(\bar{Y}_n(D) \;\|\; L(\beta' D)\big)\big]. \tag{3.4}$$

Define Jacobian $J_n = E_F[L(\beta_n' D)(1 - L(\beta_n' D))DD']$ and prediction residual $e_i(d) = Y_i(d) - L(\beta_n' d)$. We show in Appendix B.2 that $\widehat{\beta} - \beta_n = E_n[s_i(D_i)] + O_p(n^{-1})$ for influence functions $s_i(d) = e_i(d) \cdot J_n^{-1} d$. Thus, our efficiency theory for simple estimators $\widehat{\theta} = E_n[s_i(D_i)]$ also characterizes the first-order efficiency of the logit MLE $\widehat{\beta}$ and other parametric M-estimators, for example.

Examples 3.1 and 3.2 can be extended to accommodate regressions $Y_i \sim 1 + t(D_i)$ or logit with a set of basis functions $t(d)$. For example, if $D \in \mathbb{R}^2$, we could choose $t(d) = (d_1, d_2, d_1 d_2)$ for the regression $Y_i \sim 1 + D_{i1} + D_{i2} + D_{i1} D_{i2}$, as in Example 2.1 above. Similar to the previous results, this provides the best approximation of $\bar{Y}_n(d)$ among all linear functions with both main effects and two-way interactions.

3.2 Design Construction

A coupling design is implemented in three steps:

(1) **Match.** Match units into homogeneous groups g of size $|g| = k \geq 2$ using covariates. We refer to the groups as matched k-tuples, assuming k divides n for simplicity. The matching is a bijection $\tau : [n] \to [k] \times [n/k]$ that assigns each unit i to $\tau(i) = (j, g)$, a unique position $j \in [k]$ in group $g \in [n/k]$.

(2) **Disperse.** Independently for each group g, draw samples $(U_{ig})_{i=1}^k \sim G_U$ from a coupling G_U with each marginal $U_{ig} \sim \mathrm{Unif}[0, 1]^m$, such that the collection $(U_{ig})_{i=1}^k$ is highly dispersed over the unit cube $[0, 1]^m$.

(3) **Transport.** Set treatments $D_{ig} = T(U_{ig})$ for all units $i \in [k]$ in group g, where $T : [0, 1]^m \to \mathcal{D}$ is a geometry-preserving transport map such that $T(U) \sim F$ for $U \sim \mathrm{Unif}[0, 1]^m$.

The first step of matching is common to all stratified designs, while the second and third steps are unique to the coupling designs we introduce in this paper. Our theory is largely agnostic about the specific details of the matching algorithm. The ultimate aim is to find homogeneous groups of units with similar response functions $Y_i(\cdot)$. One possible proxy for this objective is to minimize a covariate discrepancy:

$$\sum_g \sum_{i,j \in [k]} |X_{ig} - X_{jg}|^2. \tag{3.5}$$

There are several algorithms for this problem (Greevy et al., 2004; Bai et al., 2021; Cytrynbaum, 2023). Due to the curse of dimensionality in matching, this should be done using a small set of covariates expected to be highly predictive for endline outcomes Y_i.

A coupling is a joint distribution G over $\mathcal{D}^k$ with fixed marginals. For tractability and expositional clarity, we consider couplings G that are exchangeable and have identical marginals $G_i = F$ for $i \in [k]$. A coupling is exchangeable if the joint distribution of $(D_i)_{i=1}^k$ is invariant to permutations of the indices $i \in [k]$. We define the set of feasible couplings $\Pi_k(F)$ to be the exchangeable joint distributions G over $\mathcal{D}^k$ with fixed marginals $G_i = F$.

3.3 Uniform Couplings

There are many strategies for sampling highly dispersed uniform random variables $(U_i)_{i=1}^k \sim G_U$. Inspired by the Monte Carlo integration literature, we consider three canonical examples based on the Gaussian copula, stratification, and randomly shifted lattices.

Example 3.3 (Gaussian Copula)**.** For univariate treatments $m = 1$, draw $Z \sim \mathcal{N}(0, \Sigma)$ with correlation matrix $\Sigma_{ii} = 1$ and maximal negative correlation $\Sigma_{ij} = -(k-1)^{-1}$ for $i \neq j$. Let Φ be the standard normal CDF and rank transform $U_i = \Phi(Z_i)$ for $i \in [k]$ so that each $U_i \sim \text{Unif}[0, 1]$. For general $m \geq 1$, independently draw $(U_{ij})_{i=1}^k$ for each dimension $j \in [m]$ using this procedure.

Example 3.4 (Latin Hypercube)**.** For $m = 1$, partition $[0, 1]$ into k bins of width $1/k$, with bin $J_r = [(r-1)/k, r/k]$ for $r \in [k]$. Draw a random permutation π of $[k]$ uniformly and assign unit i to bin $J_{\pi(i)}$. Conditional on π, draw $U_i \sim \text{Unif}(J_{\pi(i)})$, independently for $i \in [k]$. For general $m \geq 1$, sample $(U_{ij})_{i=1}^k$ as above independently for each dimension $j \in [m]$ (McKay et al., 1979). This is sometimes referred to as a k-*rooks* design. To see why, note if $m = 2$, each point U_i occupies a unique row and column of the $k \times k$ grid of bins on $[0, 1]^2$, so the k points can be viewed as non-attacking rooks on a chessboard (Shirley, 1991).

Latin hypercube produces samples $(U_i)_{i=1}^k \sim G_U$ such that the coordinate projections $(U_{ij})_{i=1}^k$ are spread out over the interval $[0, 1]$ for each dimension $j \in [m]$. However, even if the one-dimensional projections $(U_{ij})_{i=1}^k$ are highly dispersed, the samples $(U_i)_{i=1}^k$ may not be jointly well-dispersed through the hypercube $[0, 1]^m$ if $m > 1$. Because of this, Latin hypercube produces strong negative correlations $\text{Corr}_G(\phi(D_i), \phi(D_j)) \ll 0$ only for univariate functions like $\phi(d_1, \ldots, d_m) = d_1$, but tends to have weaker effects for jointly varying functions like $\phi(d) = d_1 d_2$.

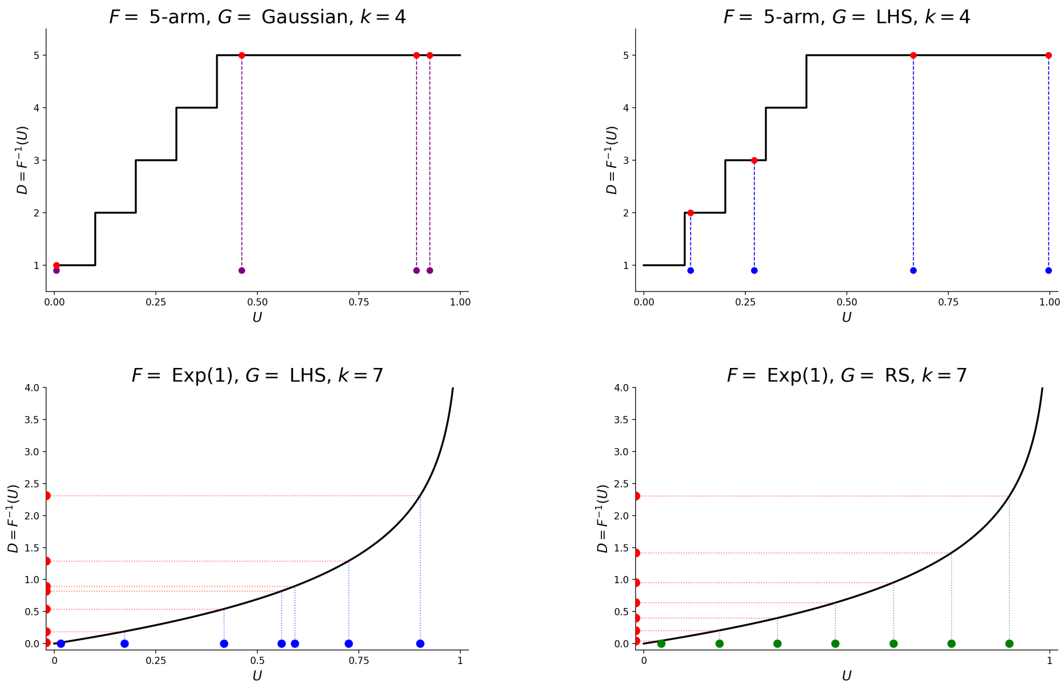


Figure 2: Uniform samples $(U_i)_{i=1}^k \sim G_U$ for couplings G_U = Gaussian copula, Latin hypercube (LHS), rotation sampling (RS). Transformed to $(D_i)_{i=1}^k \sim G$ via $D_i = F^{-1}(U_i)$. Top row: F = 5-arm distribution with tuple size $k = 4$. Bottom row: F = Exp(1) distribution with tuple size $k = 7$.

This problem can be solved with more advanced coupling constructions. Some examples include orthogonal array Latin hypercube sampling (Owen, 1992; Tang, 1993), scrambled digital nets (Owen, 1995), and shifted rank-1 lattice rules (Cranley and Patterson, 1976; Sloan and Joe, 1994). For brevity, we only formally describe shifted lattice rules. In the univariate case, such couplings draw $(U_i)_{i=1}^k \sim G_U$ by adding a uniform random shift to a regular grid $(l/k)_{l=0}^{k-1} \subseteq [0, 1]$. For more general $m \geq 1$, they apply a random shift $S \sim \text{Unif}[0, 1]^m$ to a dispersed lattice of points determined by a number-theoretic construction.

Example 3.5 (Exchangeable Shifted Lattice)**.** Pick integers z_j with $\gcd(z_j, k) = 1$ for each $j \in [m]$ and let $z = (z_1, \ldots, z_m)$. Draw a random permutation π and a shift $S \sim \text{Unif}[0, 1]^m$ with $\pi \perp\!\!\!\perp S$. Then, for each $i \in [k]$ set

$$U_i = \left(\frac{\pi(i)}{k} z + S\right) \pmod 1. \tag{3.6}$$

For $m = 1$ and $z = 1$, we have $(U_i)_{i=1}^k = (\pi(i)/k + S)_{i=1}^k \pmod 1$, a random shift and permutation of the regular grid $(l/k)_{l=0}^{k-1}$, which is known as a rotation sampling (Fishman and Huang, 1983). For $m \geq 2$, the condition $\gcd(z_j, k) = 1$ implies that the projections $(U_{ij})_{i=1}^k$ onto each coordinate $j \in [m]$ are themselves rotation samples. In addition to being marginally dispersed on each dimension, a well-chosen generating vector z ensures that the $(U_i)_{i=1}^k$ are also jointly well-dispersed through $[0, 1]^m$. For a theoretical analysis of the choice of generating vector z, see Kuo (2003). We slightly modify the standard construction of Sloan and Joe (1994), adding a random permutation π for exchangeability due to heterogeneity of the units in causal inference problems.

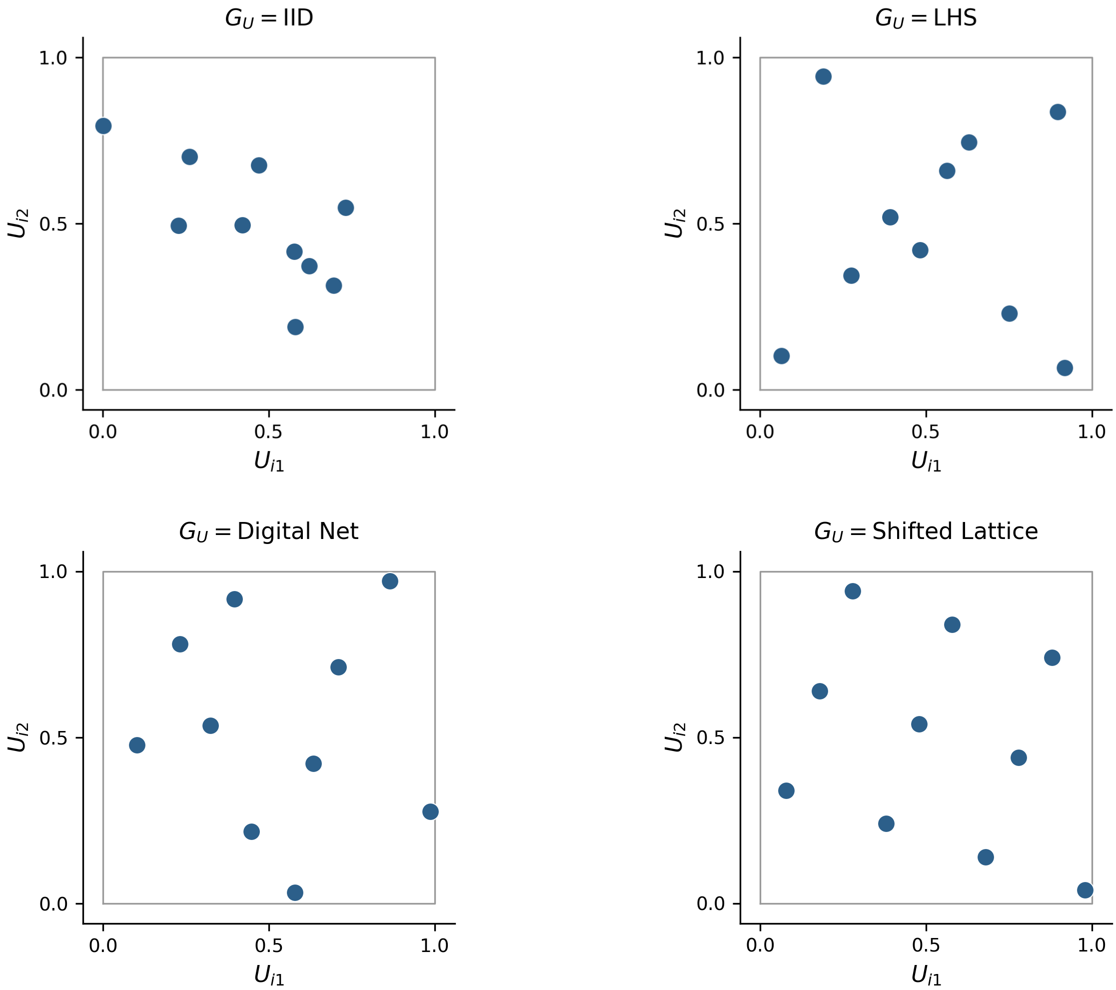


Figure 3: Multivariate uniform draws $(U_i)_{i=1}^k$ in $[0, 1]^2$ under four coupling designs with tuple size $k = 10$. The IID coupling produces clustered draws, while the Latin hypercube, scrambled digital net, and shifted lattice couplings produce increasingly dispersed samples.

3.4 Transport Maps

In general, we must map the uniform samples $(U_i)_{i=1}^k$ to treatments $D_i = T(U_i)$ that are both dispersed over $\mathcal{D}$ and have the correct marginal distribution F.

For univariate treatments $m = 1$, we can use the transport map $T(u) = F^{-1}(u)$, setting $D_i = F^{-1}(U_i)$, so that $D_i \sim F$ by the properties of the quantile function. In the multivariate case, if $F = \otimes_{j=1}^m F_j$ has independent components $D_{ij} \perp\!\!\!\perp D_{il}$ for $j \neq l \in [m]$, then we can similarly enforce $T(U_i) \sim F$ by applying the quantile transform componentwise:

$$T(U_i) = \big(F_1^{-1}(U_{i1}), \ldots, F_m^{-1}(U_{im})\big). \tag{3.7}$$

However, this will not work for general multivariate distributions F with dependent components, since $D_i = T(U_i)$ would not have the correct joint distribution. Thus, the componentwise quantile transform can only be used when the treatment space is a product set $\mathcal{D} = \times_{j=1}^m \mathcal{D}_j$ and the desired marginal distribution F has independent components. Note that several of the applications described in Section 2 do not satisfy this structure.

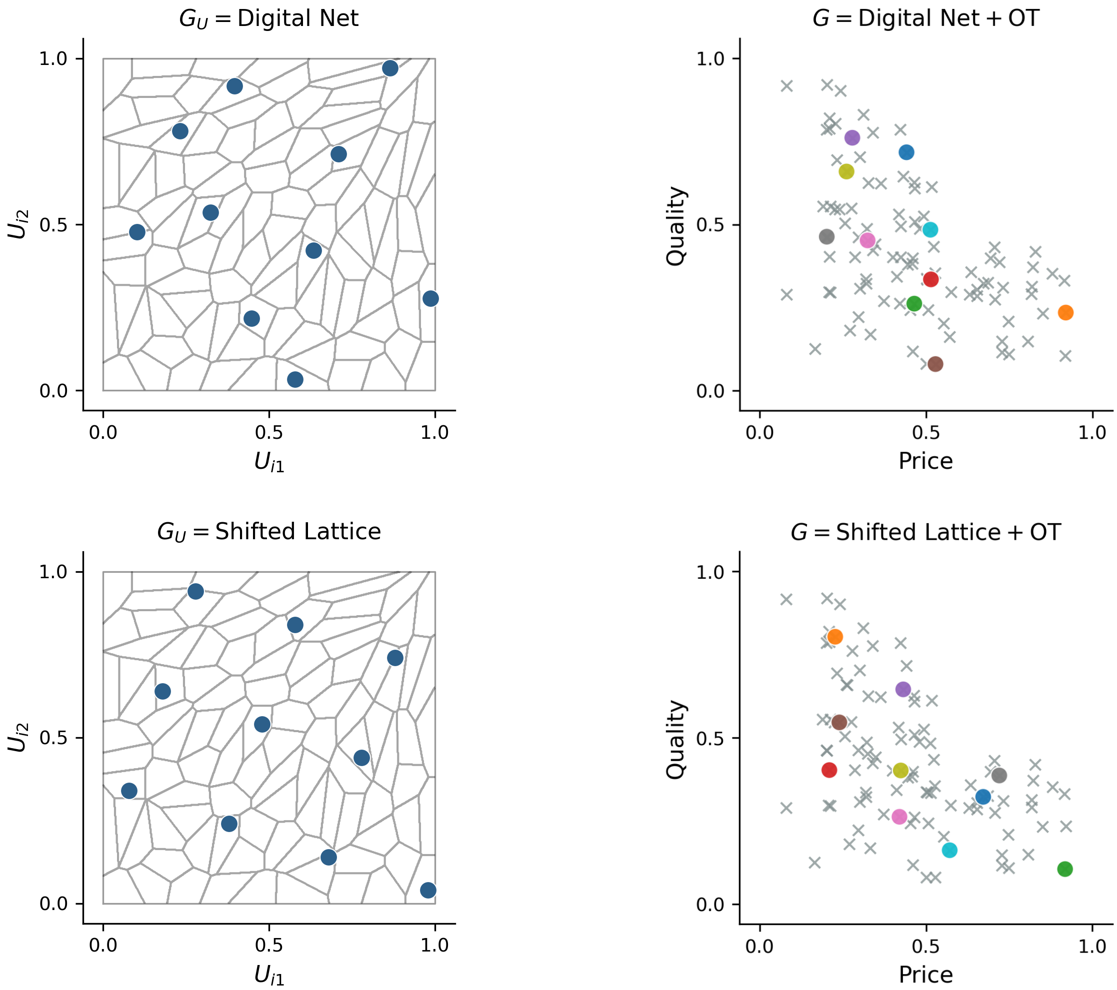


Figure 4: Coupled uniform draws (left) and transported assignments $D_i = T^*(U_i)$ (right), for $G =$ scrambled digital net (top) and randomly shifted lattice (bottom). We display the Laguerre tessellation $[0,1]^m = \cup_j C_j$ induced by the optimal transport map, where each convex cell C_j maps to a unique restaurant $T^*(C_j) = d_j$.

The quantile transform preserves the geometry of the samples $(U_i)_{i=1}^k$ in the sense that if U_i and U_j are far apart in $[0,1]$, then $F^{-1}(U_i)$ and $F^{-1}(U_j)$ will also be far apart in $\mathcal{D}$. To accommodate complex treatment spaces and dependent components, we must construct general geometry-preserving maps $T : [0,1]^m \to \mathcal{D}$ with $T(U) \sim F$. The geometry-preserving condition naturally leads us to *Brenier maps* from optimal transport (Brenier, 1991):

$$T^* = \underset{T:\, T(U)\sim F}{\operatorname{argmin}} \int_{[0,1]^m} |u - T(u)|_2^2 \, du. \tag{3.8}$$

In the univariate case and when $F = \otimes_{j=1}^m F_j$, the Brenier map recovers the quantile transform in Equation 3.7 above. However, optimal transport can be used to construct geometry-preserving maps $T^*(U) \sim F$ for much more general spaces $\mathcal{D}$. For discrete spaces $\mathcal{D} \subseteq \mathbb{R}^m$, such maps can also be computed efficiently by semi-discrete optimal transport (Mérigot, 2011). We discuss the geometric condition motivating this definition and further computational details in Appendix A.1.

4 Dispersion and Match Quality

This section describes the key mathematical objects for quantifying the relative efficiency of coupling design randomization, defining appropriate measures of *match quality* and *sample dispersion*. This allows us to show a simple fundamental relation between these objects and the efficiency gain from a coupling design:

$$\text{Efficiency Gain} = \text{Dispersion} \times \text{Match Quality}.$$

4.1 Sample Dispersion

Above, we constructed couplings that "spread out" treatments over the treatment space $\mathcal{D}$. For intuition about how this improves precision, consider the simple problem of estimating $\theta_0 = E_F[\phi(D)]$ with $\widehat{\theta} = k^{-1} \sum_{i=1}^k \phi(D_i)$. This is a homogeneous version of the more general heterogeneous estimation problems in Section 3.1.

If by random chance we sample treatments $D_i \approx D_j$ close together in the space $\mathcal{D}$, then for smooth enough functions $\phi : \mathcal{D} \to \mathbb{R}$, the samples $\phi(D_i) \approx \phi(D_j)$ will be quite similar, which effectively wastes an experimental sample. By contrast, if $(D_i)_{i=1}^k$ are dispersed over $\mathcal{D}$, then we learn more about the function $\phi(\cdot)$ from a given fixed sample size k.

To formalize this intuition, let sample variance $\mathrm{Var}_k(a_i) \equiv (k-1)^{-1} \sum_{i=1}^k (a_i - \bar{a})^2$ for any $(a_i)_{i=1}^k$ and define the sample dispersion as a normalized measure of how spread out the samples $(\phi(D_i))_{i=1}^k$ are in expectation over G.

Definition 4.1 (Dispersion)**.** For coupling G with margins F, if $\mathrm{Var}_F(\phi) > 0$, define

$$\mathrm{Disp}_G(\phi) \equiv (k-1) \left(\frac{E_G \, \mathrm{Var}_k(\phi(D_i))}{\mathrm{Var}_F(\phi)} - 1 \right). \tag{4.1}$$

If $\mathrm{Var}_F(\phi) = 0$, define $\mathrm{Disp}_G(\phi) \equiv 0$.

For the iid design $G_{iid} = \otimes_{i=1}^k F$, we have $E_{G_{iid}} \mathrm{Var}_k(\phi(D_i)) = \mathrm{Var}_F(\phi)$ by unbiasedness of the sample variance, so $\mathrm{Disp}_{G_{iid}}(\phi) = 0$. If $\mathrm{Disp}_G(\phi) > 0$, then the samples $(\phi(D_i))_{i=1}^k$ are more spread out in expectation under G than under iid randomization. For homogeneous estimation problems and smooth $\phi(\cdot)$, this improves efficiency by the mechanism described above. Indeed, the relative efficiency for the homogeneous problem above is

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} = \mathrm{Disp}_G(\phi). \tag{4.2}$$

Because $\mathrm{Var}_G(\widehat{\theta}) \geq 0$, we have $\mathrm{Disp}_G(\phi) \leq 1$ for all $\phi(\cdot)$ and $G \in \Pi_k(F)$. We also have the lower bound $\mathrm{Disp}_G(\phi) \geq -(k-1)$, which is attained for any $\phi(\cdot)$ under a clustered coupling with $D_i = D_j$ for all $i, j \in [k]$. Under exchangeable couplings, the dispersion can be interpreted as a normalized measure of negative correlation between the samples $\phi(D_i)$ and $\phi(D_j)$ for $i \neq j$.

Proposition 4.2. *Let* $0 < \mathrm{Var}_F(\phi) < \infty$ *and* $G \in \Pi_k(F)$*. Then we have*

$$\mathrm{Disp}_G(\phi) = -(k-1) \, \mathrm{Corr}_G(\phi(D_i), \phi(D_j)), \qquad i \neq j. \tag{4.3}$$

Negatively correlated samples tend to "repel" each other, making them more spread out. We work primarily with this formulation in what follows.

Role of Smoothness. Recall the discrete choice example in Section 2.3, where we argued that coupling designs can exploit smoothness of the purchase decision $y(d)$ in restaurant features $d \in \mathbb{R}^m$. Let $n = k$ and suppose preferences are homogeneous with $Y_i(d) = y(d)$. We can use the Horvitz-Thompson estimator to estimate the best linear approximation of $y(\cdot)$ as in Example 3.1. Then $\widehat{\theta} = k^{-1} \sum_{i=1}^k \phi(D_i)$ with influence function $\phi(d) = y(d)H(d)$ for weights $H(d) = \mathrm{Var}_F(D)^{-1}(d - E_F[D])$. When $y(\cdot)$ is smooth, so is $\phi(\cdot)$, and the samples $\phi(D_i) \approx \phi(D_j)$ whenever $D_i \approx D_j$. By spreading $(D_i)_{i=1}^k$ across the treatment space $\mathcal{D}$, the couplings above produce more dispersed samples $(\phi(D_i))_{i=1}^k$, increasing $\mathrm{Disp}_G(\phi)$ and improving efficiency (Equation 4.2). In Section 5, we develop the technical machinery to describe exactly how $\mathrm{Disp}_G(\phi)$ is determined by the smoothness and shape of $\phi(\cdot)$.

4.2 Match Quality

The direct connection between dispersion and efficiency in Equation 4.2 above only holds for homogeneous problems with $\widehat{\theta} = k^{-1} \sum_{i=1}^k \phi(D_i)$. For realistic causal estimation problems, we also need to account for heterogeneity of the functions $s_i(\cdot)$ within matched k-tuples of units, due to imperfect matching on only partially predictive covariates. To do so, next we define an appropriate measure of within-group match quality.

Definition 4.3 (Match Quality)**.** Let $v_{iid}(s) \equiv E_n \mathrm{Var}_F(s_i(D))$ denote the average marginal variance of the influence functions $s_i(\cdot)$. The match quality coefficient is defined as $Q_k(s) \equiv 1 - v_\Delta(s)/v_{iid}(s)$ for discrepancy

$$v_\Delta(s) \equiv \frac{1}{2n(k-1)} \sum_g \sum_{i \neq j \in [k]} \mathrm{Var}_F(s_{ig}(D) - s_{jg}(D)). \tag{4.4}$$

The term $v_\Delta(s)$ is a design-based matching discrepancy, with $v_\Delta(s) = 0$ under perfect matching, $s_{ig} = s_{jg}$. The match coefficient $Q_k(s)$ measures how homogeneous the functions $s_{ig}(\cdot)$ are within each matched k-tuple, with $Q_k(s) = 1$ under perfect matching. More generally, we have lower and upper bounds $-(k-1)^{-1} \leq Q_k(s) \leq 1$ for all populations $s = (s_i)_{i=1}^n$ and matching procedures. If units are matched at random, the expected match quality is $E_\tau[Q_k(s)] \geq -(n-1)^{-1}$ in the worst case. In theory, it is possible to approach the lower bound by purposefully matching units into k-tuples to be as dissimilar as possible, but we do not expect this to arise in practice. See Appendix B.3 for further details.

Covariate Power. Recall units are matched into k-tuples using observed baseline covariates $(X_i)_{i=1}^n$. The match quality coefficient $Q_k(s)$ will be large if both:

(a) Covariates X_i are highly predictive of heterogeneity in $s_i(\cdot)$.

(b) Matching discrepancy on X_i within k-tuples (Equation 3.5) is small.

We formalize this observation in Appendix A.2, providing conditions under which $Q_k(s) = R^2_{s|X} \cdot Q_k(\mu) + o_p(1)$ as $n \to \infty$. Here $Q_k(\mu)$ is the matching discrepancy on features of the covariates alone, and $R^2_{s|X}$ measures covariate predictive power for heterogeneity in $s_i(\cdot)$.

4.3 Efficiency from Dispersion and Match Quality

Our main theoretical result is that the efficiency gain from coupling designs is proportional to the product of dispersion $\mathrm{Disp}_G(\phi)$ and match quality $Q_k(s)$. To state this result in full generality requires additional technical machinery, which we develop in Section 5 below. To build intuition, we first state the result in a simple univariate parametric model.

Theorem 4.4 (Relative Efficiency). *Let $s_i(d) = c_i + a_i\phi(d)$ for $i \in [n]$. Then the variance of estimator $\widehat{\theta} = E_n[s_i(D_i)]$ relative to the iid design is given by*

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} = \mathrm{Disp}_G(\phi) \times Q_k(s). \tag{4.5}$$

For intuition, note that in the homogeneous special case $s_i(d) = c + a\phi(d)$ for all units $i \in [n]$, relative efficiency is exactly given by $\mathrm{Disp}_G(\phi)$. In general heterogeneous problems, the relative efficiency is dampened by imperfect matching, captured by the match quality coefficient $Q_k(s) < 1$.

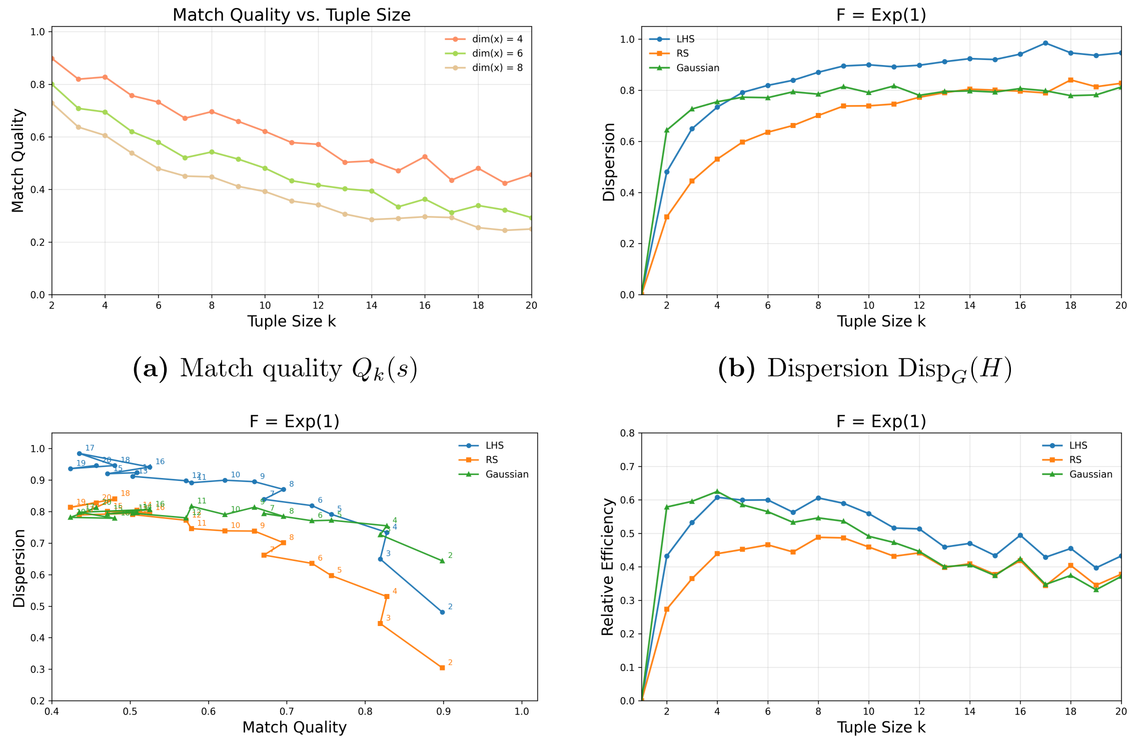
Another interesting special case occurs when $\mathrm{Disp}_G(\phi) = 1$, so that relative efficiency is exactly equal to match quality $Q_k(s) = 1 - v_\Delta(s)/v_{iid}(s)$. Rearranging, we find $n\,\mathrm{Var}_G(\widehat{\theta}) = v_\Delta(s)$, the matching discrepancy from Definition 4.3. This shows how $v_\Delta(s)$ can be interpreted as the ideal variance under perfect dispersion, where efficiency is only limited by imperfect matching.

Tuple Size Tradeoff. The quantity $\mathrm{Disp}_G(\phi)$ is generally increasing in tuple size k, since for larger k it becomes easier to jointly correlate the treatments $(D_i)_{i=1}^k$ to be spread out over $\mathcal{D}$. Our analysis in Sections 5 and 6 formalizes this effect. By contrast, match quality $Q_k(s)$ is generally decreasing in k, since it becomes harder to find many similar units to match together.

Consider the two extremes of this trade-off. At one extreme, we can set $k = 2$, sampling $(D_1, D_2) \sim G$ within matched pairs. This makes matches as tight as possible, maximizing $Q_k(s)$, then uses the coupling G to increase dispersion subject to the constraint $k = 2$.

At the opposite extreme, we can maximize dispersion by setting k very large and sampling $(D_1, \ldots, D_k) \sim G$ to be as dispersed as possible over $\mathcal{D}$. However, match quality will suffer due to the large tuple size. For general F, perfect dispersion is only possible in the limit as $k \to \infty$, but $\mathrm{Disp}_G(\phi) = 1$ is possible for some special distributions F with small support.

Illustration of Tradeoff. Consider a researcher estimating the dose-response of welfare outcomes to cash grants, randomizing from $F = \mathrm{Exp}(1)$ to provide many small grants while also trying some larger amounts. Suppose the treatment has no effect on potential outcomes, so $Y_i(d) = \mathbb{1}'X_i$ and $s_i(d) = (\mathbb{1}'X_i)H(d)$. Theorem 4.4 shows that the efficiency under a coupling G is the product of $\mathrm{Disp}_G(H)$ and match quality $Q_k(s)$. The feasible match quality vs. dispersion frontier and efficiency gain are shown in Figure 5 as functions of k. Perfect treatment dispersion is impossible here, requiring $k \to \infty$. However, efficiency is actually maximized at moderate k, e.g. around $k = 4$, accepting reduced dispersion in exchange for efficiency gains from higher match quality.



(a) Match quality $Q_k(s)$ **(b)** Dispersion $\mathrm{Disp}_G(H)$

(c) Feasible Frontier $(Q_k(s), \mathrm{Disp}_G(H))$ **(d)** Efficiency $= Q_k(s) \times \mathrm{Disp}_G(H)$

Figure 5: Tuple size tradeoff with $F = \mathrm{Exp}(1)$, $n = 100$, and $\dim(X) = 4$. Match quality (a) decreases in tuple size k, while dispersion (b) increases in k. The frontier in (c) shows the feasible (match quality, dispersion) pairs for each coupling, with labels indicating tuple size k. Relative efficiency (d) is maximized at moderate k, balancing dispersion and match quality.

Example 4.5 (Stratified Randomization)**.** For discrete treatments, conventional stratified randomization is an example of a design that lexicographically emphasizes dispersion over match quality. By completely randomizing within each stratum, stratified randomization ensures treatments are perfectly dispersed over the discrete space $\mathcal{D}$, achieving $\mathrm{Disp}_G(\phi) = 1$ for all $\phi(\cdot)$. However, this is only possible for large enough stratum size k. In particular, if $P(D_i = j) = p_j$, conventional stratified randomization is possible if and only if $p_j \cdot k \in \mathbb{N}$ for all j. In cases with many treatments or irregular assignment probabilities, this can impose a severe cost in match quality. Because efficiency is proportional to the product of both dispersion and match quality, such a lexicographic preference for dispersion will often be inefficient. See Appendix A.4 for a detailed discussion.

4.4 Covariate Balance for Complex Treatments

An important motivation for conventional stratified randomization is that it reduces covariate imbalances between the different treatment groups. These imbalances are equivalent to spurious in-sample correlations between the treatment D_i and covariates X_i. Here, we show that by assigning similar groups of units to highly dispersed treatments, coupling designs prevent such spurious correlations from arising. Because of this, such designs enable covariate-balancing randomization over complex treatment spaces.

For binary $D_i \in \{0,1\}$, covariate balance is commonly assessed using the t-statistics from a regression $X_i \sim 1 + D_i$. Up to a normalization, this is equivalent to checking the magnitude of the sample covariance $\mathrm{Cov}_n(D_i, X_i)$. If this covariance is small, then treatments are approximately independent of unit-specific heterogeneity *in-sample*. Thus, ensuring covariate balance is equivalent to randomizing in a way that enforces $E_G[\mathrm{Cov}_n(D_i, X_i)^2] \approx 0$.

To extend this balance measure to complex treatment spaces $\mathcal{D}$, let $\phi : \mathcal{D} \to \mathbb{R}$ and $b : \mathbb{R}^p \to \mathbb{R}$ be basis functions. Write $\mathrm{Var}_n(b_i) = n^{-1} \sum_{i=1}^n (b_i - \bar{b})^2$ for $b_i = b(X_i)$ and define the *imbalance* under a coupling G as the mean-squared sample covariance

$$\mathcal{I}_G(\phi, b) \equiv E_G[\,\mathrm{Cov}_n(\phi(D_i), b(X_i))^2\,]. \tag{4.6}$$

We can view D_i as approximately independent of covariates X_i if $\mathcal{I}_G(\phi, b) \approx 0$ for a rich set of basis functions $\phi(\cdot)$ and $b(\cdot)$. Intuitively, coupling designs prevent such in-sample correlations from arising between X_i and D_i by ensuring that similar units are assigned to highly dissimilar treatments. In particular, the next corollary shows that the imbalance measure $\mathcal{I}_G(\phi, b)$ is decreasing in the product $\mathrm{Disp}_G(\phi) \times Q_k(b)$ of dispersion and match quality.

Corollary 4.6 (Covariate Imbalance)**.** *If* $0 < \mathrm{Var}_F(\phi) < \infty$ *and* $\mathrm{Var}_n(b_i) > 0$*, then the covariate imbalance under a coupling design* G *relative to the iid design is*

$$\frac{\mathcal{I}_G(\phi, b)}{\mathcal{I}_{G_{iid}}(\phi, b)} = 1 - \mathrm{Disp}_G(\phi) \times Q_k(b). \tag{4.7}$$

Here, the term $Q_k(b) \equiv 1 - (n/k)^{-1} \sum_g \mathrm{Var}_k(b_{ig}) / \mathrm{Var}_n(b_i)$ *denotes within-group match quality on* $b_i = b(X_i)$*.*

5 Efficiency Theory

This section proves our efficiency result in full generality, without the simplifying parametric assumption we imposed when previewing the results in Section 4. To do so, we decompose the space of influence functions into a basis of orthogonal subspaces on which $\mathrm{Disp}_G(\cdot)$ is constant, which we view as the principal directions of the coupling G.

5.1 Dispersion Basis

Define the square-integrable functions $L^2(F) = \{\phi : E_F[\phi(D)^2] < \infty\}$. We assume $s_i(\cdot) \in L^2(F)$ for $i \in [n]$ throughout. Also denote the mean zero subspace $L^2_0(F) \equiv \{\phi \in L^2(F) : E_F[\phi(D)] = 0\}$. We define the principal directions of a coupling G to be the eigenspaces of the following linear operator, which captures the pairwise dependence structure of the design.

Definition 5.1 (Coupling Operator)**.** For $G \in \Pi_k(F)$, let $U_G : L^2(F) \to L^2(F)$

$$(U_G\phi)(d) = E_G[\phi(D_i) \mid D_j = d], \quad i \neq j. \tag{5.1}$$

This operator is well-defined for any choice of $i \neq j$ due to exchangeability of G. Since dispersion $\mathrm{Disp}_G(\phi)$ is invariant to constant shifts, it is convenient to work with the mean zero subspace $L_0^2(F)$ in what follows. Note that if $\phi \in L_0^2(F)$, then by tower law $E_F[(U_G\phi)(D)] = 0$, so U_G also maps $L_0^2(F)$ to itself. The coupling operator is self-adjoint and linear on the Hilbert space $L_0^2(F)$, so it has a real spectrum. We additionally require the following condition:

Assumption 5.2 (Discrete Spectrum)**.** *There exist eigenspaces $(E_m)_{m\geq 1}$ of U_G such that $L_0^2(F) = \oplus_{m\geq 1} E_m$.*

Assumption 5.2 holds for any coupling $G \in \Pi_k(F)$ if the marginal F is discrete. For continuous F, Lemma B.8 in the appendix shows that it holds for all of the univariate couplings described in Section 3.3. More generally, by the spectral theorem Assumption 5.2 holds whenever the operator U_G is compact.

A key insight for our analysis is that $\mathrm{Disp}_G(\phi)$ decomposes orthogonally over the eigenspaces of U_G for any $\phi \in L^2(F)$.

Theorem 5.3 (Dispersion Basis)**.** *Let $G \in \Pi_k(F)$.*

(a) If $E \subseteq L_0^2(F)$ is an eigenspace with $U_G\phi = \lambda\phi$ for $\phi \in E$, then the dispersion $\mathrm{Disp}_G(\phi) = -(k-1)\lambda$. In particular, $\mathrm{Disp}_G(\phi) = \mathrm{Disp}_G(\psi)$ for all $\phi, \psi \in E$.

(b) Impose Assumption 5.2. For $\phi \in L^2(F)$, let $P_m\phi = \mathrm{argmin}_{f\in E_m} \mathrm{Var}_F(\phi - f)$ be the orthogonal projection onto eigenspace E_m. Then

$$\mathrm{Disp}_G(\phi) = \sum_{m\geq 1} \frac{\mathrm{Var}_F(P_m\phi)}{\mathrm{Var}_F(\phi)} \mathrm{Disp}_G(E_m). \tag{5.2}$$

We view the eigenspaces of U_G as the principal directions of the coupling G in $L^2(F)$ with respect to sampling, since they control how much dispersion is produced when sampling from G. In particular, $\mathrm{Disp}_G(\phi)$ will be large if $\phi(\cdot)$ is well approximated on the high dispersion eigenspaces, e.g. with $\mathrm{Disp}_G(E_m) \approx 1$. The following corollary is immediate.

Corollary 5.4 (Principal Directions)**.** *Suppose $L_0^2(F) = E \oplus E^\perp$ are eigenspaces of the operator U_G with $\mathrm{Disp}_G(E) > \mathrm{Disp}_G(E^\perp)$. Then*

$$E = \underset{\substack{\phi\in L_0^2(F)\\ \phi\neq 0}}{\mathrm{argmax}}\ \mathrm{Disp}_G(\phi), \qquad E^\perp = \underset{\substack{\phi\in L_0^2(F)\\ \phi\neq 0}}{\mathrm{argmin}}\ \mathrm{Disp}_G(\phi). \tag{5.3}$$

More generally, if $L_0^2(F) = \oplus_{m=1}^M E_m$, then each eigenspace can be obtained by maximizing $\mathrm{Disp}_G(\phi)$ subject to orthogonality to previously found eigenspaces. This is analogous to principal components analysis (PCA), where each principal component can be found by maximizing data variance subject to orthogonality to previous components.

Coupling Analysis. Theorem 5.3 also provides a simple recipe for computing $\mathrm{Disp}_G(\phi)$ by analyzing the eigenspaces and eigenvalues of the operator U_G. We illustrate this by computing the exact dispersion for the univariate Latin hypercube coupling. Our analysis highlights the role played by smoothness of the function $\phi(\cdot)$ in guaranteeing high dispersion $\mathrm{Disp}_G(\phi)$ under this coupling. We provide a detailed comparison with other couplings in Section 6 below.

Example 5.5 (Latin Hypercube, Dispersion). Let $F = \mathrm{Unif}[0,1]$ and define the histogram space $E_{hist} \equiv \{\phi \in L_0^2(F) : \phi(d) = \sum_l \alpha_l \cdot \mathbb{1}(d \in J(l))\}$ on bins $J(l) = [(l-1)/k, l/k)$ for $l \in [k]$. We show that $L_0^2(F) = E_{hist} \oplus E_{hist}^{\perp}$, eigenspaces of U_G with dispersions 1 and 0 respectively (Lemma B.6). See Figure 6 for a graphical illustration of these spaces. Theorem 5.3 implies that for $G =$ Latin hypercube sampling (LHS), we have:

$$\mathrm{Disp}_G(\phi) = \frac{\mathrm{Var}_F(P_{hist}\phi)}{\mathrm{Var}_F(\phi)}. \tag{5.4}$$

Here, $P_{hist} : L^2(F) \to E_{hist}$ is the orthogonal projection onto E_{hist}, which is the de-meaned best histogram approximation of $\phi(\cdot)$:

$$(P_{hist}\phi)(d) = \sum_l E_F[\phi(D)|D \in J(l)] \cdot \mathbb{1}(d \in J(l)) - E_F[\phi(D)]. \tag{5.5}$$

Dispersion is large under $G =$ LHS to the extent that $\phi(\cdot)$ is well approximated by histograms on the partition $(J(l))_{l=1}^k$. For fixed tuple size k, well-approximation on the space $E_{hist} = E_{hist}(k)$ requires the function $\phi(\cdot)$ to be relatively smooth. As k increases, the partition becomes finer and the histogram space is more expressive, so we obtain high dispersion even for "rougher" functions ϕ. In particular, consider tuple sizes $k \le r$ and suppose k divides r, i.e. $k \,|\, r$. Then $E_{hist}(k) \subseteq E_{hist}(r)$, so the projection weight on $E_{hist}(r)$ is larger. It follows that dispersion is increasing in tuple size:

$$\mathrm{Disp}_{G_k}(\phi) \le \mathrm{Disp}_{G_r}(\phi) \qquad k \le r \text{ with } k \,|\, r. \tag{5.6}$$

Remark 5.6 (Canonical Marginals). It may appear that we need to separately solve for eigenspace decomposition $L_0^2(F) = \oplus_{m \ge 1} E_m$ for each choice of marginal F and coupling operator U_G. In fact, it generally suffices to solve for the eigenspaces once for a canonical choice of marginal, e.g. $F = \mathrm{Unif}[0,1]^m$. To see this, recall we constructed general $G \in \Pi_k(F)$ by sampling high dispersion $(U_i)_{i=1}^k \sim G_U$ with each $U_i \sim \mathrm{Unif}[0,1]^m$, then setting $D_i = T(U_i)$ for a transport map $T : [0,1]^m \to \mathcal{D}$. Then $s_i(D_i) = s_i(T(U_i)) \equiv \tilde{s}_i(U_i)$ for the *effective* influence function $\tilde{s}_i = s_i \circ T$. Because of this, we can compute eigenspaces E_m under the canonical marginal $F = \mathrm{Unif}[0,1]^m$. In particular, our analysis of LHS in Example 5.5 with $F = \mathrm{Unif}[0,1]$ is without loss of generality.

5.2 Efficiency

The dispersion basis in Theorem 5.3 allows us to generalize the simple relationship between efficiency and the product of dispersion $\times$ match quality from Theorem 4.4 to general influence functions $s_i(\cdot) \in L^2(F)$. Impose Assumption 5.2 and let $s_i^m(d) = (P_m s_i)(d)$ be the projection of $s_i(\cdot)$ onto E_m. Define approximation weights

$$w_m(s) \equiv \frac{n^{-1}\sum_i \mathrm{Var}_F(P_m s_i)}{n^{-1}\sum_i \mathrm{Var}_F(s_i)}. \tag{5.7}$$

The weights $w_m(s)$ quantify how well the influence functions can be approximated using functions in eigenspace E_m, on average over the experimental units. In what follows, denote $s = (s_i)_{i=1}^n$ and $s^m = (s_i^m)_{i=1}^n$. We also write $\mathrm{Disp}_G(m)$ as the common dispersion on E_m.

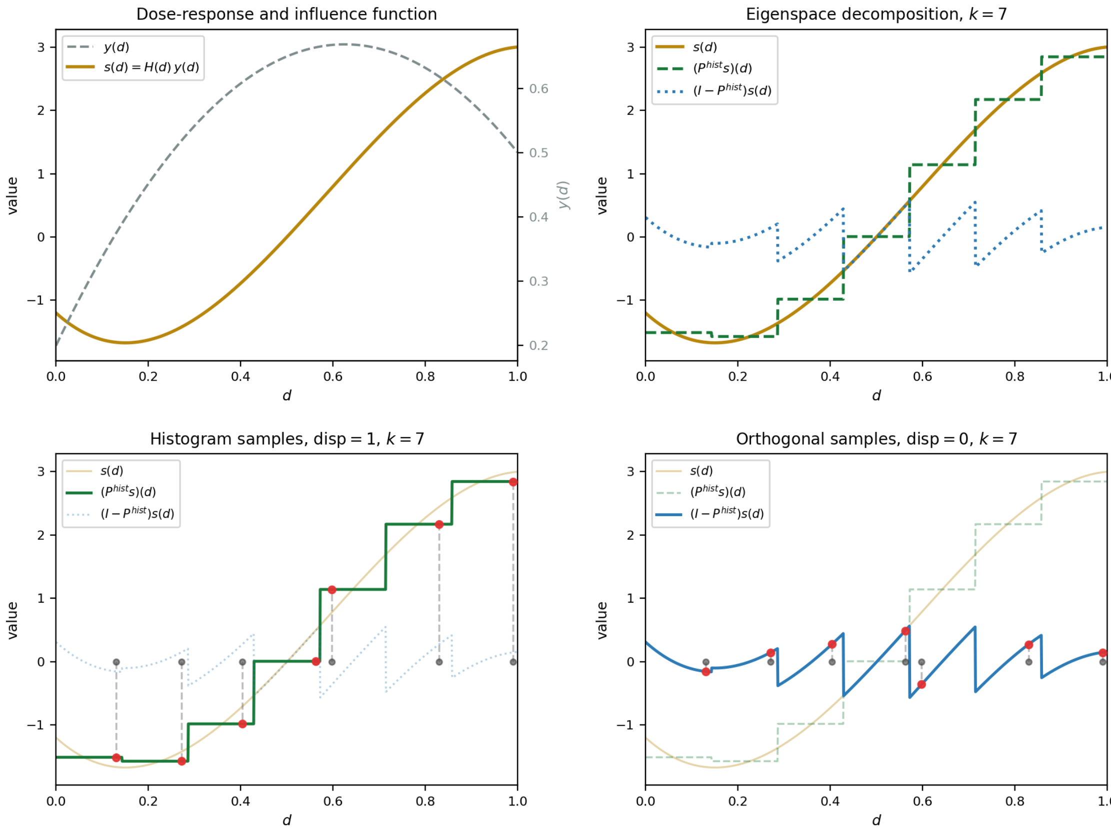


Figure 6: Eigenspaces for $G = \text{LHS}$. Top left: Individual dose-response $Y_i(d)$ and influence function $s_i(d) = H(d)\,Y_i(d)$ for estimating θ_{BLP} (Example 3.1) with $F = \text{Unif}[0,1]$. Top right: Decomposition $s_i = P_{hist}s_i + (I - P_{hist})s_i$ into histogram and orthogonal projections for $k = 7$. Bottom left: LHS samples from histogram projection $P_{hist}s_i$ with $\text{Disp}_G(E_{hist}) = 1$, highly dispersed. Bottom right: LHS samples from orthocomplement $(I - P_{hist})s_i$ with $\text{Disp}_G(E_{hist}^{\perp}) = 0$, resembling iid draws.

Theorem 5.7 (Efficiency)**.** *Impose Assumption 5.2. Then for $G \in \Pi_k(F)$*

$$1 - \frac{\text{Var}_G(\widehat{\theta})}{\text{Var}_{G_{iid}}(\widehat{\theta})} = \sum_{m \geq 1} w_m(s) \cdot \text{Disp}_G(m) Q_k(s^m). \tag{5.8}$$

The weights $w_m(s)$ are non-negative and sum to one. Thus, the theorem shows that the efficiency gain from coupling design randomization is a convex combination of products of dispersion $\text{Disp}_G(m)$ and match quality $Q_k(s^m)$ across eigenspaces E_m. The efficiency gain is therefore large when the influence functions $s_i(\cdot)$ align well with eigenspaces with high dispersion $\text{Disp}_G(m) \approx 1$ and good match quality $Q_k(s^m) \approx 1$. We show that many of the couplings introduced above achieve high dispersion generically for smooth functions $s_i(\cdot)$.

Example 5.8 (Latin Hypercube, Efficiency)**.** As in Example 5.5, for $G = \text{LHS}$, we have $L_0^2(F) = E_{hist} \oplus E_{hist}^{\perp}$ with dispersions 1 and 0. Then by Theorem 5.7,

$$1 - \frac{\text{Var}_G(\widehat{\theta})}{\text{Var}_{G_{iid}}(\widehat{\theta})} = w_{hist}(s) \cdot Q_k(s^{hist}). \tag{5.9}$$

In particular, LHS is efficient when the influence functions $s_i(\cdot)$ are smooth enough to be well-approximated on the histogram space E_{hist}, and when units are well-matched on the projection s^{hist} of the influence functions onto this space. Equation (5.9) also shows the dispersion versus match quality tradeoff discussed in Section 4.3. As tuple size k grows, the space $E_{hist}(k)$ becomes more expressive and $w_{hist}(s) = w^k_{hist}(s)$ increases. By contrast, match quality $Q_k(s^{hist})$ is generally decreasing in k.

Approximate Stratification. Recall the definition of the match quality coefficient $Q_k(s) = 1 - v_\Delta(s)/v_{iid}(s)$ from Definition 4.3, where $v_{iid}(s)$ is the iid variance and $v_\Delta(s)$ is the average within-group variance of the influence functions. When units are well-matched, we have $Q_k(s) \approx 1$ so that $v_\Delta(s) \ll v_{iid}(s)$.

Corollary 5.9 (Nominal Variance). *Under the conditions of Theorem 5.7,*

$$n \operatorname{Var}_G(\widehat{\theta}) = \sum_{m \geq 1} \operatorname{Disp}_G(m) \cdot v_\Delta(s^m) + [1 - \operatorname{Disp}_G(m)] \cdot v_{iid}(s^m). \tag{5.10}$$

If $\operatorname{Disp}_G(m) = 1$, we obtain the variance $v_\Delta(s^m)$. For simple discrete treatments, $\operatorname{Disp}_G(\cdot) = 1$ under classic stratified randomization (Example 4.5), so we can regard $v_\Delta(s)$ as the *perfectly stratified* variance. For general complex treatment spaces $\mathcal{D} \subseteq \mathbb{R}^m$, perfect stratification is impossible, but the corollary shows a sense in which we can approximate it by using high dispersion couplings.

Example 5.10 (Latin Hypercube, Approximate Stratification). Applying Corollary 5.9 to $G = \text{LHS}$, we obtain

$$n \operatorname{Var}_G(\widehat{\theta}) = v_\Delta(s^{hist}) + v_{iid}(s - s^{hist}). \tag{5.11}$$

The LHS coupling behaves like perfect stratification $v_\Delta(s^{hist})$ on the histogram projection $s_i^{hist}(\cdot)$ of each influence function, but behaves like iid randomization on the residual $(s_i - s_i^{hist})(\cdot)$. See Figure 6 for a visualization of this effect.

6 Coupling Analysis and Comparisons

Next, we illustrate how the theory developed in the previous section can be applied to compare the efficiency and robustness of designs based on LHS with designs based on rotation sampling (RS) and the Gaussian copula. The analysis shows that RS also generically produces high dispersion for large enough k, but is less robust to adversarial influence function shapes $s_i(\cdot)$ than LHS. By contrast, for moderate k the Gaussian copula only produces high dispersion for approximately linear functions, a strong parametric restriction.

6.1 Rotation Sampling

Recall from Example 3.5 that a rotation sample $(U_i)_{i=1}^k \sim G_U$ lies on a randomly shifted equispaced grid, so that $U_i = U_j \oplus l/k$ for some $l \in [k]$ and $i, j \in [k]$, where $a \oplus b \equiv a + b \pmod 1$ for $a, b \in \mathbb{R}$. The canonical marginal for univariate rotation sampling is thus $F = \text{Unif}[0, 1]$. We show that this coupling is efficient if influence functions $s_i(\cdot)$ are smooth, in a sense defined below.

Suppose a function $\phi(\cdot)$ is perfectly $1/k$-cyclic, with $\phi(x) = \phi(x \oplus 1/k)$ for all $x \in [0, 1]$. Then our samples $\phi(U_i) = \phi(U_j)$ for $i, j \in [k]$ are perfectly correlated

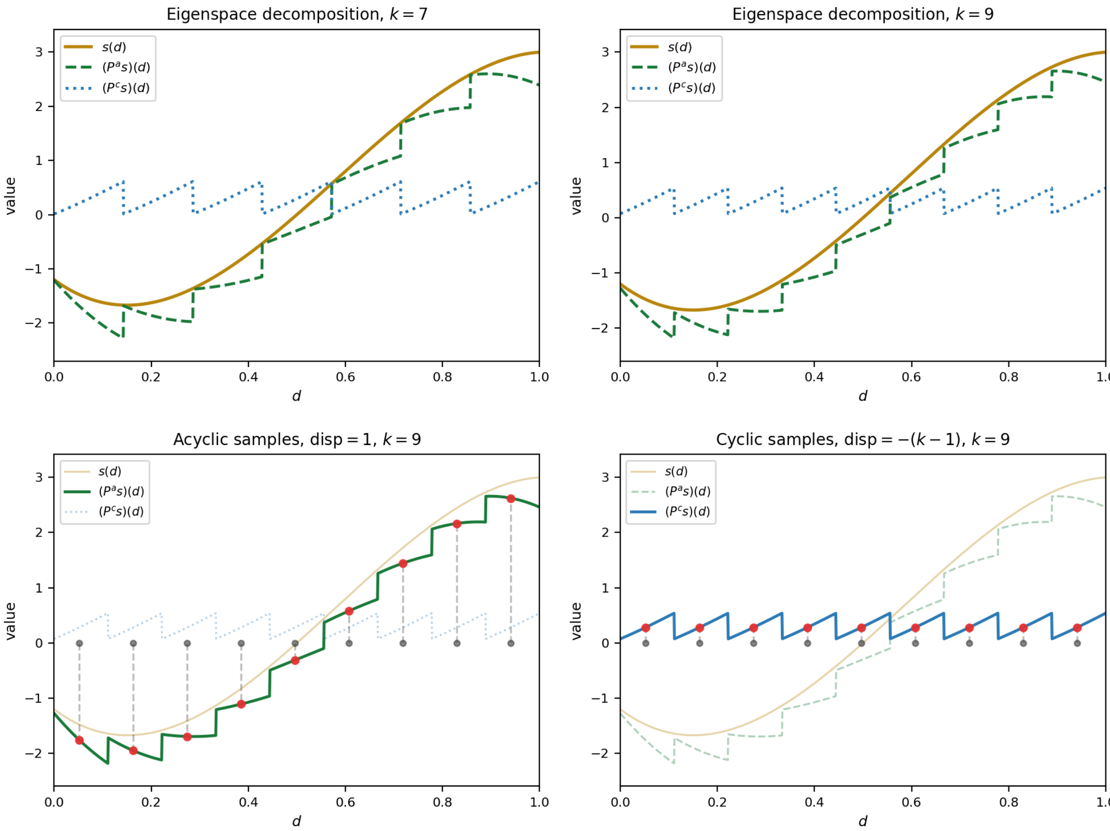


Figure 7: Eigenspace decomposition for G = rotation sampling (RS). Top left: Single influence function $s_i(d)$ decomposed into acyclic projection $P_a s_i$ and cyclic projection $P_c s_i$, $k = 7$. Top right: decomposition for $k = 9$, showing how E_c shrinks as k increases. Bottom left: Samples from acyclic projection $P_a s_i$ with $\text{Disp}_G(E_a) = 1$. Bottom right: Samples from cyclic projection $P_c s_i$ with $\text{Disp}_G(E_c) = -(k-1)$, note clustering behavior.

under rotation sampling, effectively yielding a clustered sample. The low dispersion eigenspace turns out to be exactly this space of cyclic functions:

$$E_c \equiv \{\phi \in L_0^2(F) : \phi(x) = \phi(x \oplus 1/k), \forall x \in [0,1]\}. \tag{6.1}$$

The acyclic subspace is defined as the orthogonal complement $E_a \equiv E_c^{\perp}$ in $L_0^2(F)$. Our analysis shows that $L_0^2(F) = E_c \oplus E_a$, which are eigenspaces of the coupling operator U_G with dispersions $\text{Disp}_G(E_a) = 1$ and $\text{Disp}_G(E_c) = -(k-1)$. Projections on E_c and E_a have closed forms, with $P_c\phi(d) = k^{-1}\sum_{l=1}^{k}\phi(d \oplus l/k) - E_F[\phi(D)]$ the de-meaned cyclic average and residual $P_a\phi(d) = \phi(d) - k^{-1}\sum_{l=1}^{k}\phi(d \oplus l/k)$. Then by Theorem 5.3, for any $\phi \in L^2(F)$,

$$\text{Disp}_G(\phi) = \frac{\text{Var}_F(P_a\phi)}{\text{Var}_F(\phi)} - (k-1)\frac{\text{Var}_F(P_c\phi)}{\text{Var}_F(\phi)}. \tag{6.2}$$

Remark 6.1 (Smoothness and Shape)**.** Equation 6.2 shows that $\text{Disp}_G(\phi)$ is large under the RS coupling if $\phi(\cdot)$ is well-approximated on the acyclic space E_a. This can be understood as smoothness and shape restrictions on $\phi(\cdot)$. To see this, expand E_c in a Fourier basis, noting $E_c = \text{span}\{\sin(2\pi lkx), \cos(2\pi lkx) : l \geq 1\}$ consists of frequencies that are exact multiples of k, while E_a contains all remaining frequencies. Since E_a includes all frequencies lower than k, smooth functions are well-approximated on E_a for moderate k. As k increases, the cyclic space E_c shrinks, reducing the degree of smoothness needed for high dispersion. See also L'Ecuyer and Lemieux (2000) for a related Fourier analysis perspective on the variance of (non-exchangeable) shifted lattice designs.

Let $w_a(s)$ and $w_c(s)$ denote the approximation weights (Equation 5.7) on E_a and E_c respectively, so that $w_c(s)$ quantifies how cyclic the influence functions $s_i(\cdot)$ are, on average over $i \in [n]$. We apply Theorem 5.7 to $G = \text{RS}$, using the facts above.

Theorem 6.2 (Rotation Sampling)**.** *Let G be the RS coupling. Then*

$$1 - \frac{\text{Var}_G(\widehat{\theta})}{\text{Var}_{G_{iid}}(\widehat{\theta})} = w_a(s) \cdot Q_k(s^a) - w_c(s) \cdot (k-1) \cdot Q_k(s^c). \tag{6.3}$$

The low dispersion cyclic space $E_c = E_c(k)$, where the design performs poorly, shrinks as k increases. In particular, we have $E_c(r) \subseteq E_c(k)$ for $k \mid r$, so the weights satisfy $w_c^r(s) \leq w_c^k(s)$. This shows a sense in which dispersion is monotonically increasing in tuple size k for rotation sampling.

Robustness Comparison. Recall that for LHS, the low dispersion subspace $E_{hist}^{\perp}$ behaves like an iid design with $\text{Disp}_G(E_{hist}^{\perp}) = 0$ (Example 5.8). By contrast, the low dispersion space under RS has $\text{Disp}_G(E_c) = -(k-1)$, actually harming relative efficiency through the negative term in Equation (6.3). This effect can also be seen in the nominal variance. Let group mean $\bar{s}_g(\cdot) = k^{-1} \sum_{i \in [k]} s_{ig}(\cdot)$ and define the *group variance* $v_g(s) \equiv (n/k)^{-1} \sum_g \text{Var}_F(\bar{s}_g)$. By Corollary 5.9, we have

$$n \, \text{Var}_G(\widehat{\theta}) = v_\Delta(s^a) + k \cdot v_g(s^c). \tag{6.4}$$

This shows RS achieves the ideal perfectly stratified variance $v_\Delta(s^a)$ on E_a, but has variance $k \cdot v_g(s^c)$ on E_c, reducing the effective sample size by a factor of k. Thus, the RS coupling is less robust to adversarial influence function shapes than LHS. However, for this worst case to arise in practice, the influence functions $s_i(\cdot)$ must be strongly cyclic with high frequency, which is rare in typical social science applications.

6.2 Gaussian Copula

We show that when k is moderate, the Gaussian copula produces high dispersion only for approximately linear influence functions $s_i(\cdot)$, a restrictive condition relative to the general smooth functions that achieve high dispersion under the LHS and RS couplings. This suggests caution when using the Gaussian copula to generate dispersion in experimental design.

For the Gaussian copula, it is convenient to use canonical measure $F = \mathcal{N}(0, 1)$. In this case, the coupling operator U_G coincides with the Mehler kernel operator (Mehler, 1866). We use the eigenbasis expansion $L_0^2(F) = \oplus_{m \geq 1} \text{span}(h_m)$ for U_G, where $(h_m)_{m \geq 1}$ are the normalized probabilist's Hermite polynomials (Thangavelu, 1993). Each $h_m(x)$ is a polynomial of order m, for example, $h_1(x) = x$ and $h_2(x) =$

$(x^2-1)/\sqrt{2}$. As shown in Theorem 5.3, the dispersion of a polynomial h_m can be obtained from its eigenvalue λ_m under U_G, with $\mathrm{Disp}_G(h_m) = -(k-1)\lambda_m$. The projections onto $E_m = \mathrm{span}(h_m)$ are $s_i^m(\cdot) = \mathrm{Cov}_F(s_i, h_m) \cdot h_m(\cdot)$. An application of Theorem 5.7 yields the following result.

Theorem 6.3 (Gaussian)**.** *Let G be the Gaussian coupling. For $k \geq 3$*

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} = \sum_{m \geq 1} w_m(s) \cdot [-1/(k-1)]^{m-1} \cdot Q_k(s^m). \tag{6.5}$$

Define the space of linear functions $E_L = \{\phi : \phi(d) = a + bd\}$. For any non-constant $\phi \in E_L$, we have $\mathrm{Disp}_G(\phi) = \mathrm{Disp}_G(h_1) = 1$, so this is a high dispersion subspace. By contrast, $|\,\mathrm{Disp}_G(h_m)| \leq (k-1)^{-(m-1)}$ for any Hermite polynomial of order $m \geq 2$, which is rapidly decreasing as tuple size k increases. For larger k, the Gaussian copula produces high dispersion only for the linear component of $s_i(\cdot)$, performing no better than iid randomization on $E_L^{\perp}$.

Corollary 6.4. *Let $s_i^L(\cdot)$ denote the orthogonal projection of $s_i(\cdot)$ onto E_L in $L^2(F)$. The variance $n\,\mathrm{Var}_G(\widehat{\theta}) = v_\Delta(s^L) + v_{iid}(s - s^L) + O(k^{-1})$ and*

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} = w_L \cdot Q_k(s^L) + O(k^{-1}). \tag{6.6}$$

Corollary 6.4 shows that, for moderate k, the Gaussian copula is efficient only if the influence functions $s_i(\cdot)$ are approximately linear. Note that $s_i(\cdot)$ may not be linear even if the potential outcomes $Y_i(\cdot)$ are linear in the treatment, since each $s_i(\cdot)$ also depends on the estimand and the design. For example, in the cash transfer experiment of Example 2.1, if the responses are linear in the grant amount, $Y_i(d) = a_i + b_i d$, the influence function for the best linear approximation coefficient (Example 3.1) is $s_i(d) = Y_i(d)H(d)$ for $H(d) = (d - E_F[D])/\mathrm{Var}_F(D)$, which is quadratic in d.

Remark 6.5 (Antithetic Variates)**.** The results above characterized the performance of the Gaussian copula for moderate to large k. At the opposite extreme $k = 2$, the Gaussian copula is equivalent to antithetic variates. In Appendix A.3, we apply our core efficiency results to study antithetic variates in more detail, extending the discussion in Section 2.1.

6.3 Parametric vs. Nonparametric Couplings

Our analysis of the various couplings above suggests that they can be sorted into two categories: non-parametric couplings like LHS and RS, which produce high dispersion under weak smoothness conditions on $s_i(\cdot)$, and parametric couplings like the Gaussian copula, which produce high dispersion only for a restricted class of influence functions with specific shapes. Here we formalize this claim.

For any $\phi : [0,1] \to \mathbb{R}$, define *total variation* $V_{[0,1]}(\phi) = \sup_\Pi \sum_{j=1}^r |\phi(t_j) - \phi(t_{j-1})|$, where the supremum is over all finite partitions $(t_j)_{j=0}^r$ of $[0,1]$. Define the class of *bounded variation* functions $\mathcal{H}(b) \equiv \{\phi : V_{[0,1]}(\phi) \leq b\}$. This provides a weak notion of smoothness for functions on $[0,1]$, allowing for discontinuous functions (e.g.

histograms) provided they don't oscillate too much over the interval $[0,1]$. This condition is meaningful for both discrete and continuous treatments. For example, if F is discrete, the effective influence function $\tilde{s}_i(\cdot)$ on $[0,1]$ defined by $s_i(D) = s_i(F^{-1}(U)) = \tilde{s}_i(U)$ will be discontinuous, but may have small total variation.

Theorem 6.6 (Dispersion Limits). *Let $\phi \in L^2(F)$. As $k \to \infty$:*

(a) If $G =$ Gaussian, then $\mathrm{Disp}_G(\phi) \to \mathrm{Var}_F(P_L\phi)/\mathrm{Var}_F(\phi)$.

(b) If $G \in \{LHS, RS\}$, then for any $\epsilon > 0$,

$$\inf_{\substack{\phi \in \mathcal{H}(b) \\ \mathrm{Var}_F(\phi) > \epsilon}} \mathrm{Disp}_G(\phi) = 1 + o(1). \tag{6.7}$$

The theorem shows that the Gaussian copula produces high dispersion only for functions $\phi(\cdot)$ that are well-approximated by linear functions, while the non-parametric LHS and RS couplings produce high dispersion for any function $\phi(\cdot)$ with bounded total variation.

To relate this to the variance of the estimator $\widehat{\theta}$, define a smoothness coefficient $\eta_{\mathrm{TV}}(s) \equiv E_n[V_{[0,1]}(s_i)^2]/v_{iid}(s)$, which is a normalized measure of average total variation of the $s_i(\cdot)$.

Theorem 6.7 (Efficiency from Smoothness). *Let $G \in \{LHS, RS\}$. Then*

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} \geq Q_k(s) - \frac{\eta_{\mathrm{TV}}(s)}{k}. \tag{6.8}$$

In particular, the theorem guarantees that LHS and RS improve precision relative to iid randomization when $Q_k(s) \geq \eta_{\mathrm{TV}}(s)/k$, so that the match quality is sufficiently high relative to the roughness of the influence functions $s_i(\cdot)$. This lower bound on efficiency can be improved by using stronger notions of smoothness. For example, by imposing uniform Lipschitz continuity on the $s_i(\cdot)$, the lower bound for LHS can be improved to $Q_k(s) - O(1/k^2)$. However, such smoothness conditions typically rule out discrete treatments, motivating the weaker total variation notion of smoothness that we use here.

7 Asymptotics and Inference

7.1 Consistency

We consider an asymptotic regime with a sequence of experimental populations and coupling designs indexed by n with $n \to \infty$. Thus, in this section all variables are implicitly indexed by n, denoting their place in the sequence. For example, influence functions $s_i(\cdot) = s_i^n(\cdot)$ for $i \in [n]$, and similarly the design parameters $G = G(n)$ and $k = k(n)$. We often suppress the indexing for brevity.

The experimental design literature has noted an important tradeoff between efficiency and robustness, showing how balancing covariates to improve efficiency can entail a loss of robustness in adverse experimental settings (Efron, 1971; Harshaw et al., 2024). Motivated by this, we study a robust notion of consistency, requiring that coupling designs perform well uniformly over a range of possible empirical settings (Harshaw et al., 2025). Given a family of influence functions $\mathcal{S}$, the uniform mean square error $R_n(G, \mathcal{S})$ for a coupling G is

$$R_n(G, \mathcal{S}) \equiv \sup_{s \in \mathcal{S}} n E_G\big[(\widehat{\theta} - \theta_n)^2\big]. \tag{7.1}$$

If $R_n(G, \mathcal{S}) = O(1)$, we say that $\widehat{\theta}$ is *uniformly* $\sqrt{n}$*-consistent* under the family $\mathcal{S}$ and the coupling sequence $G(n)$.

We can study the worst-case performance of our designs by requiring uniform consistency under weak restrictions on $\mathcal{S}$. For example, we can set $\mathcal{S}_{\mathrm{m}}$ to be the set of all $s_i(\cdot)$ with bounded average moments: $\mathcal{S}_{\mathrm{m}} = \{s_1, \dots, s_n \in L^2(F) : E_n[\mathrm{Var}_F(s_i)] \leq 1\}$. This allows for settings with highly non-smooth influence functions and negative match quality, requiring good performance even if an adversary matched units into groups that are maximally dissimilar. We show that coupling designs are still uniformly $\sqrt{n}$-consistent over $\mathcal{S}_{\mathrm{m}}$ under weak conditions on the design parameters, providing a strong robustness guarantee.

Theorem 7.1. *Let* $\mathrm{ID}_G = \inf_{\phi \neq c} \mathrm{Disp}_G(\phi)$ *and* $\mathrm{SD}_G = \sup_{\phi \neq c} \mathrm{Disp}_G(\phi)$ *be the extremal dispersions of G over non-constant $\phi \in L^2(F)$. Impose Assumption 5.2. For any $G \in \Pi_k(F)$,*

$$R_n(G, \mathcal{S}_{\mathrm{m}}) = 1 + \max\left(-\mathrm{ID}_G, \frac{\mathrm{SD}_G}{k-1}\right). \tag{7.2}$$

In particular, $1 \leq R_n(G, \mathcal{S}_{\mathrm{m}}) \leq k$. Thus, if $k = O(1)$, then uniform $\sqrt{n}$-consistency is attained for any coupling sequence $G(n)$.

Remark 7.2 (Minimaxity)**.** We have $R_n(G_{iid}, \mathcal{S}_{\mathrm{m}}) = 1$, so the iid design is minimax optimal over $\mathcal{S}_{\mathrm{m}}$. If $G \in \{\mathrm{LHS}, \mathrm{Gaussian}\}$, then $R_n(G, \mathcal{S}_{\mathrm{m}}) = k/(k-1) \to 1$ as $k \to \infty$, so these couplings are asymptotically minimax optimal over $\mathcal{S}_{\mathrm{m}}$. Note also that if $G = \mathrm{RS}$ then $R_n(G, \mathcal{S}_{\mathrm{m}}) = k$, attaining the worst case rate. This reflects the fact that rotation sampling can be exploited by highly non-smooth, perfectly cyclic influence functions $s_i(\cdot)$.

The influence function configurations that attain the worst case rate in Theorem 7.1 are generally pathological, with either highly non-smooth influences $s_i(\cdot)$ or negative match quality, which might not be relevant for empirical practice. For example, for $G = \mathrm{LHS}$, $R_n(G, \mathcal{S}_{\mathrm{m}})$ is attained by placing perfectly negatively correlated influence functions $s_i(\cdot)$ within each group, which assumes that our matching was not only ineffectual, but actually much worse than random. We can place mild regularity conditions on the set $\mathcal{S}_{\mathrm{m}}$ to rule out such pathological settings, obtaining bounds that are more informative about the performance of coupling designs in typical applications.

To illustrate this, we consider the nonparametric couplings $G \in \{\mathrm{LHS}, \mathrm{RS}\}$ under a restricted family of influence functions $\mathcal{S}_{\mathrm{r}}$ that have reasonable match quality and finite total variation. The next result shows that under such conditions, these couplings uniformly dominate the iid design.

Theorem 7.3. *Let $\mathcal{S}_{\mathrm{r}} = \{s \in \mathcal{S}_{\mathrm{m}} : Q_k(s) \geq q_0 > 0 \;\text{ and }\; \eta_{\mathrm{TV}}(s) \leq \bar{\eta}\}$ be the family of influence functions with match quality at least q_0 and smoothness coefficient at most $\bar{\eta}$. Impose Assumption 5.2. Then for $G \in \{LHS, RS\}$,*

$$R_n(G, \mathcal{S}_{\mathrm{r}}) \leq 1 - q_0 + \bar{\eta}/k. \tag{7.3}$$

We have $R_n(G_{iid}, \mathcal{S}_\mathrm{r}) = 1$, so the LHS and RS coupling designs dominate the iid design in a minimax sense over $\mathcal{S}_\mathrm{r}$ when $\bar{\eta}/k < q_0$, which holds for large enough k. Intuitively, this is because iid randomization has no way of exploiting the structure imposed on the influence functions by $\mathcal{S}_\mathrm{r}$. Moreover, after ruling out arbitrarily non-smooth functions $s_i(\cdot)$, rotation sampling $G = \mathrm{RS}$ is uniformly $\sqrt{n}$-consistent over $\mathcal{S}_\mathrm{r}$ even as $k \to \infty$.

7.2 Asymptotic Normality

We provide conditions under which the estimator $\widehat{\theta}$ is asymptotically normal.

Assumption 7.4 (CLT)**.** *The following hold:*

(1). (Bounded Fourth Moments) $M_{4,n} = E_n\big[E_F[s_i(D)^4]\big] = O(1)$.

(2). (Not Superefficient) $E_G[(\widehat{\theta} - \theta_n)^2] = \Omega(n^{-1})$.

(3). (Group Size) $k = o(n^{1/3})$.

Part (2) of Assumption 7.4 is a high level condition ruling out certain degenerate estimation problems. For example, under perfect homogeneity $s_i(d) = s(d)$ with $s(\cdot)$ Lipschitz continuous, one can show that super-efficient estimation is possible using the Latin hypercube coupling with $k = n$. Such perfectly homogeneous settings are not empirically relevant.

Proposition 7.5 (CLT)**.** *Impose Assumption 7.4. Then for* $\sigma_n^2 = \mathrm{Var}_G(\widehat{\theta})$,

$$(\widehat{\theta} - \theta_n)/\sigma_n \Rightarrow \mathcal{N}(0, 1), \tag{7.4}$$

Since each of the n/k groups has independent treatments, and each group's contribution to $\widehat{\theta}$ is bounded, the CLT follows from standard Lindeberg-Feller arguments provided the number of groups n/k grows sufficiently fast. Assumption 7.4 ensures this by requiring $k = o(n^{1/3})$.

7.3 Variance Estimation and Inference

We construct a variance estimator for $\mathrm{Var}_G(\widehat{\theta})$ using a collapsed strata approach (Hansen et al., 1953). Let $\pi : [n/k] \to [n/k]$ be a permutation with no fixed points, $\pi(g) \neq g$ for all $g \in [n/k]$, which associates each group g with a paired group $\pi(g)$. Typically, this is a matching $\pi(\pi(g)) = g$, but we allow for non-matching permutations to accommodate, for example, an odd number of groups.

Let $\widehat{\theta}_g = k^{-1}\sum_{i=1}^k s_{ig}(D_i)$ and $\theta_g = E_F[\widehat{\theta}_g] = k^{-1}\sum_{i=1}^k \theta_{ig}$ for $\theta_{ig} = E_F[s_{ig}(D)]$. The variance estimator is the scaled squared difference between the mean outcomes in the paired groups:

$$\widehat{\sigma}_n^2 = \frac{k^2}{2n^2}\sum_{g=1}^{n/k}\big(\widehat{\theta}_g - \widehat{\theta}_{\pi(g)}\big)^2. \tag{7.5}$$

Let Δ_n^2 be the average squared difference in group effects across the paired groups:

$$\Delta_n^2 = \frac{1}{n/k}\sum_{g=1}^{n/k}(\theta_g - \theta_{\pi(g)})^2. \tag{7.6}$$

Proposition 7.6. *The variance estimator has expectation and variance given by*

$$E_G\left[\frac{\widehat{\sigma}_n^2}{\sigma_n^2}\right] = 1 + \frac{k\Delta_n^2}{2n\sigma_n^2}, \qquad \mathrm{Var}_G\left(\frac{\widehat{\sigma}_n^2}{\sigma_n^2}\right) \le \frac{12k^3}{n^3\sigma_n^4} M_{4,n}. \tag{7.7}$$

Therefore,

(a) The variance estimator is conservative: $\inf_n E_G[\widehat{\sigma}_n^2/\sigma_n^2] \ge 1$.

(b) The normalized variance estimator is convergent in mean square given Assumption 7.4: $\mathrm{Var}_G(\widehat{\sigma}_n^2/\sigma_n^2) = o(1)$.

The magnitude of the bias depends on the tuple size k and the heterogeneity in group effects. Under $\sqrt{n}$-consistency, $\sigma_n^2 \asymp n^{-1}$, the normalized bias is $O(k\Delta_n^2)$. We generally expect Δ_n^2 to be asymptotically bounded but not to vanish, so the variance estimator is typically biased upwards also asymptotically. The normalized bias is asymptotically bounded only if $k = O(1)$. However, the unnormalized bias is $O(k\Delta_n^2/n)$, so it vanishes provided that $k = o(n)$. Because the variance estimator is conservative, confidence intervals constructed using $\widehat{\sigma}_n^2$ are asymptotically valid.

Proposition 7.7. *Let* $\widehat{\sigma}_n = \sqrt{\widehat{\sigma}_n^2}$ *be the standard error estimator, and impose Assumption 7.4. Then, for any* $\alpha \in (0, 1)$*, the confidence interval*

$$\mathrm{CI}_{1-\alpha} = \left[\widehat{\theta} - z_{1-\alpha/2}\,\widehat{\sigma}_n,\ \ \widehat{\theta} + z_{1-\alpha/2}\,\widehat{\sigma}_n\right],$$

where $z_{1-\alpha/2}$ *is the* $(1-\alpha/2)$*-quantile of the standard normal distribution, satisfies*

$$\liminf_{n\to\infty} \Pr_G(\theta_n \in \mathrm{CI}_{1-\alpha}) \ge 1 - \alpha.$$

8 Concluding Remarks

Coupling designs provide a powerful approach for improving the efficiency of experimental designs in complex treatment spaces. The core insight is that the mechanism underlying conventional stratification can be extended to complex treatment spaces by matching units into similar groups, then assigning within-group treatments that are highly dispersed over the space $\mathcal{D}$. The efficiency gain from coupling designs is proportional to the product of sample dispersion and match quality. The attained dispersion depends on the shape of the influence functions $s_i(\cdot)$, where high dispersion is achieved when the $s_i(\cdot)$ are well-approximated on the high-dispersion eigenspaces of the coupling operator U_G associated with the relevant coupling design.

Several directions for future work remain. First, we focused on exchangeable couplings with fixed marginals for tractability and exposition. Alternative designs that use non-exchangeable couplings or allow for flexible marginal distributions may offer further efficiency improvements, but require new tools to analyze their properties. Second, the key insight of producing high sample dispersion among similar experimental units applies very broadly, and the stratified structure of coupling designs is not essential to leverage this insight. Alternative designs that do not partition units into groups, but still assign highly dispersed treatments to similar units, may be possible and could offer further efficiency improvements. Finally, integrating coupling designs with response-adaptive methods that update the treatment distribution over successive experimental waves is a promising avenue for combining the efficiency gains from both approaches.

A Appendix

A.1 Multivariate Transport Maps

We want to find a map $T : [0,1]^m \to \mathcal{D}$ with $T(U) \sim F$ for $U \sim \mathrm{Unif}[0,1]^m$, allowing for cases without independence $F = \otimes_{j=1}^m F_j$ or a simple product structure $\mathcal{D} = \times_{j=1}^m \mathcal{D}_j$. This map should have geometry-preservation properties, so that highly dispersed samples $(U_i)_{i=1}^k \sim G_U$ are mapped to treatments $(D_i)_{i=1}^k \sim G$ that are also jointly dispersed over $\mathcal{D}$.

Rank Preservation. For example, in the univariate case $m = 1$ we might hope that such a map not only has $T(U) \sim F$, but also satisfies a *rank-preservation* property, so that $u \le v$ if and only if $T(u) \le T(v)$ for any $u, v \in [0,1]$. In fact, it's well known that the quantile transform $T = F^{-1}$ is the unique map satisfying both rank preservation and the marginal transport condition. This is one justification for using the quantile transform for the univariate couplings in Section 3.4, showing how $T = F^{-1}$ has a strong geometry-preservation guarantee.

We would like to generalize this rank preservation condition to multivariate treatments $m \ge 1$. To do so, first note that rank preservation $u \le v$ iff $T(u) \le T(v)$ is equivalent to $(u - v)(T(u) - T(v)) \ge 0$ for all $u, v \in [0,1]$. For $m \ge 1$, we can generalize this by requiring that T satisifies a *pairwise monotonicity* property: $(u - v)'(T(u) - T(v)) \ge 0$ for all $u, v \in [0,1]^m$. Intuitively, this is a geometric constraint requiring $T(\cdot)$ to preserve the relative orientation of input points. A slightly stronger condition, which implies pairwise monotonicity, is the *cyclic monotonicity* condition. For any finite set $\{u_1, \dots, u_L\} \subseteq [0,1]^m$ and permutation σ of $[L]$, we can require that

$$\sum_{l \in [L]} u_l' T(u_l) \ge \sum_{l \in [L]} u_l' T(u_{\sigma(l)}). \tag{A.1}$$

This requires that each output $T(u)$ is the "most aligned" with u in the sense that no reassignment among finitely many points can increase the total inner product.

Optimal Transport. There exists a unique T^* satisfying both the marginal condition $T^*(U) \sim F$ and the cyclic monotonicity constraint above (Brenier, 1991). This unique function T^* is known as the *Brenier map* and is given by the solution to the optimal transport problem:

$$T^* = \underset{T:\, T(U) \sim F}{\operatorname{argmin}} \int_{[0,1]^m} |u - T(u)|_2^2 \, du. \tag{A.2}$$

The Brenier map can be viewed as a multivariate generalization of the quantile transform F^{-1}, see e.g. Chernozhukov et al. (2017). In particular, in the univariate case $T^* = F^{-1}$. More generally, if $F = \otimes_{j=1}^m F_j$ then $T^* = (F_1^{-1}, \dots, F_m^{-1})$ as in Equation 3.7. Thus, optimal transport extends the basic quantile transform approach advocated above, see Appendix B.2 for a proof. However, T^* can also be used to construct high dispersion couplings $(D_i)_{i=1}^k \sim G$ for more complex treatment spaces $\mathcal{D}$ and distributions F, as we now discuss.

Applications. Recall in Example 2.3 we considered binary choice experiments where the treatment space is a non-manipulable, irregular set $\mathcal{D} = \{d_1, \ldots, d_N\} \subseteq \mathbb{R}^m$. Similarly, for the constrained treatment space in Example 2.2, if $\mathcal{D}_{pre} = \Pi_{j=1}^m [N_j]$, then $\mathcal{D} = \{d \in \mathcal{D}_{pre} : C(d) \leq B\} \subseteq \mathbb{R}^m$ is an irregular discrete point set. In both cases, finding $T^* : [0,1]^m \to \mathcal{D}$ in Equation 3.8 becomes a *semi-discrete* optimal transport problem, which is efficiently solvable by convex optimization (Mérigot, 2011). In particular, the map T^* has a simple form in this case, partitioning $[0,1]^m$ into disjoint convex regions $[0,1]^m = \cup_j C_j$, where each region is mapped to a unique point $T^*(C_j) = d_j$ in $\mathcal{D}$. This partition is known as a Laguerre tessellation.

Solving for T^* can be more difficult if $\mathcal{D}$ is fully continuous. For example, if the ambient space $\mathcal{D}_{pre} = [0,1]^2$, then $\mathcal{D} = \{d \in \mathcal{D}_{pre} : C(d) \leq B\}$ is a constrained continuous space. One natural approach is to use rejection sampling to construct a fine discrete approximation $\mathcal{D}_{approx} = \{d_1, \ldots, d_N\}$ of $\mathcal{D}$. This can be done by proposing candidate points $d \sim \text{Unif}(\mathcal{D}_{pre})$ and accepting if $C(d) \leq B$. Then one can compute the Brenier map T^* to this discretized approximation using semi-discrete optimal transport as above. Importantly, since discretization preserves the underlying geometry of $\mathcal{D}$, we will still obtain $\text{Corr}_G(\phi(D_i), \phi(D_j)) \ll 0$ for smooth functions $\phi : \mathcal{D} \to \mathbb{R}$. For truly continuous randomization, if the constraint functions $C(d) \leq B$ are simple enough, it is sometimes possible to analytically construct a transport map $T : [0,1]^m \to \mathcal{D}$ using Knothe-Rosenblatt transport (Carlier et al., 2010) instead of optimal transport.

A.2 Covariate Power

To formalize the relationship between match quality and covariate predictive power, suppose just for illustration in the remainder of this subsection that $(X_i, s_i(\cdot)) \sim P$ iid for some fixed measure P. This guarantees that the units in our experiment are "typical" relative to a fixed relationship P between X_i and $s_i(\cdot)$. Let $\mu(d, X_i) \equiv E_P[s_i(d)|X_i]$ and the predictive power of covariates by

$$R^2_{s|X} \equiv \frac{E_P \operatorname{Var}_F(\mu(D, X_i))}{E_P \operatorname{Var}_F(s_i(D))}. \tag{A.3}$$

This measures how well X_i predicts heterogeneity in influence functions $s_i(d)$ across treatment levels, the predictive power for "design heterogeneity." Note how this differs from the typical superpopulation R^2 coefficient, say in a linear regression. For example, if $\mu(d, X_i) = \mu(X_i)$, then $R^2_{s|X} = 0$, even if $\mu(X_i)$ is highly predictive of $s_i(\cdot)$ over the distribution P.

Proposition A.1 (Covariate Power). *Let $(X_i, s_i(\cdot)) \sim P$ iid with $E_P[\operatorname{Var}_F(s_i)] > 0$ and $E_P[E_F[s_i(D)^4]] < \infty$. Suppose the matching τ is $\sigma(X_{1:n})$-measurable. Then if $k = O(1)$ as $n \to \infty$*

$$Q_k(s) = R^2_{s|X} \cdot Q_k(\mu) + o_p(1). \tag{A.4}$$

The term $Q_k(\mu)$ measures within-group match quality on the covariate features $\mu_i(\cdot) = \mu(\cdot, X_i)$, on average over $D \sim F$. This proposition formalizes the intuition above, showing that, under slightly stronger assumptions, $Q_k(s)$ is increasing in the product of both $R^2_{s|X}$ and $Q_k(\mu)$.

Practical Recommendations. Proposition A.1 suggests that we want to measure baseline covariates X_i with large predictive power $R^2_{s|X}$. In particular, we want to match tightly on the features of those covariates $\mu(\cdot, X_i)$ most predictive of $s_i(\cdot)$. Of course, both of these objects are unknown at design time before the experimenter has access to data. One could try to estimate them with a pilot, but the literature has raised concerns about the finite sample properties of pilot-based designs that try to estimate related quantities (Cai and Rafi, 2024; Cytrynbaum, 2021).

A.3 Antithetic Variates

In this section, we show how our framework characterizes the exact efficiency of the antithetic variates coupling for $k = 2$, generalizing the discussion in Section 2.1 beyond monotone influence functions. Define the even and odd functions in $L_0^2(F)$ by $E_{even} = \{\phi \in L_0^2(F) : \phi(d) = \phi(1-d)\}$ and $E_{odd} = E_{even}^{\perp}$ in $L_0^2(F)$ to be functions with $\phi(d) = -\phi(1-d)$. We show the decomposition $L_0^2(F) = E_{even} \oplus E_{odd}$, which are eigenspaces of U_G with dispersions $\mathrm{Disp}_G(E_{odd}) = 1$ and $\mathrm{Disp}_G(E_{even}) = -1$. The projections have closed forms, with $\phi^{even}(d) = (1/2)(\phi(d) + \phi(1-d)) - E_F[\phi(D)]$ and $\phi^{odd}(d) = (1/2)(\phi(d) - \phi(1-d))$. Applying Theorem 5.7:

Theorem A.2 (AV). *The variance* $n\,\mathrm{Var}_G(\widehat{\theta}) = v_\Delta(s^{odd}) + 2 \cdot v_g(s^{even})$ *and*

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} = w_{odd} \cdot Q_k(s^{odd}) - w_{even} \cdot Q_k(s^{even}). \tag{A.5}$$

Antithetic variates performs well if w_{odd} is close to 1 and units are well-matched. Similar to rotation sampling, efficiency is penalized if the $s_i(\cdot)$ have large weight on the negative dispersion eigenspace. To see why $\mathrm{Disp}_G(E_{even}) = -1$, note that $D_1 = U$ and $D_2 = 1 - U$, even functions produce perfectly correlated samples $\phi(D_1) = \phi(D_2)$. The term $2 \cdot v_g(s^{even})$ in the nominal variance reflects this clustering effect, which reduces effective sample size by a factor of 2.

Remark A.3 (Gaussian Equivalence)**.** For $k = 2$, the Gaussian coupling is equivalent to antithetic variates randomization. To see this, note if $(Z_1, Z_2) \sim G$ with $\mathrm{Corr}_G(Z_1, Z_2) = -1$ then $Z_1 = -Z_2$. The rank transform satisfies $\Phi(z) = 1 - \Phi(-z)$ for $z \in \mathbb{R}$, so for $U_i = \Phi(Z_i)$ we have $U_1 = 1 - U_2$.

A.4 Dispersion of Stratified Randomization

Stratified randomization *completely randomizes* treatment assignments within strata of similar units. This is a particular example of a coupling design, where the coupling is given by complete randomization. By definition, complete randomization assigns equal probability to all treatment allocations $d_{1:k} \in \mathcal{D}^k$ that exactly reproduce the target distribution F. For example, if $F = \mathrm{Unif}(\{1, 2, 3\})$, complete randomization with $k = 3$ randomizes uniformly over allocations with exactly one unit assigned to each treatment. We can formalize this as follows:

Definition A.4 (Complete Randomization)**.** Let $\mathcal{D} = [m]$ and $f_j = F(D = j)$. For an allocation $d_{1:k} = (d_1, \ldots, d_k)$, define the realized proportion of units $i \in [k]$ assigned to $d = j$ by $\widehat{f}_j(d_{1:k}) = k^{-1} \sum_{i=1}^{k} \mathbb{1}(d_i = j)$. Then define

$$G^c \equiv \mathrm{Unif}\{d_{1:k} : \widehat{f}(d_{1:k}) = f\}. \tag{A.6}$$

As the distribution F becomes more complex, we may need very large group sizes k to satisfy this condition, as the next result shows:

Theorem A.5 (Stratified Randomization)**.** *The coupling $G^c \in \Pi_k(F)$ exists if and only if $k \cdot f_j \in \mathbb{N}$ for all $j \in [m]$. If G^c exists, then for any $0 < \mathrm{Var}_F(\phi) < \infty$*

$$\mathrm{Disp}_{G^c}(\phi) = 1. \tag{A.7}$$

The theorem shows that for discrete F, complete randomization is a simple heuristic to attain perfect dispersion $\mathrm{Disp}_G(\cdot) = 1$ within groups. However, in settings with many treatments or irregular treatment probabilities, this may only exist for very large k-tuples, potentially destroying match quality. By Theorem 4.4, stratified randomization will generally be suboptimal in such settings since efficiency increases with the product dispersion $\times$ match quality, not dispersion alone.

B Proofs

This section contains proofs of our main results, as well as unlabeled minor claims and equations throughout the text.

B.1 Proofs for Section 1

Proof of Equation 2.2. Under perfect matching $Y_i(\cdot) = Y_j(\cdot) = y(\cdot)$, the estimator is $\widehat{\theta} = (1/2)(y(D_i) + y(D_j))$ with $\theta_0 = E_F[y(D)]$. Since $G_i = G_j = F$, both $y(D_i)$ and $y(D_j)$ have variance $\mathrm{Var}_F(y(D))$, so $\mathrm{Var}_G(\widehat{\theta})$ is

$$
\begin{aligned}
&= (1/4)\,\mathrm{Var}_G(y(D_i) + y(D_j)) = (1/4)[2\,\mathrm{Var}_F(y(D)) + 2\,\mathrm{Cov}_G(y(D_i), y(D_j))] \\
&= (1/2)\,\mathrm{Var}_F(y(D))[1 + \mathrm{Corr}_G(y(D_i), y(D_j))].
\end{aligned}
$$

Under iid randomization, $\mathrm{Cov}(y(D_i), y(D_j)) = 0$, so $\mathrm{Var}_{G_{iid}}(\widehat{\theta}) = (1/2)\,\mathrm{Var}_F(y(D))$. Dividing through gives the result. $\square$

Proof of Matched Pairs Equivalence. Let $F = \mathrm{Bernoulli}(1/2)$, so $F^{-1}(u) = \mathbb{1}(u \geq 1/2)$. Under the antithetic variates coupling, $U \sim \mathrm{Unif}[0,1]$ and $D_i^* = F^{-1}(U)$, $D_j^* = F^{-1}(1-U)$. If $U < 1/2$, then $D_i^* = 0$ and $1 - U > 1/2$ so $D_j^* = 1$. If $U \geq 1/2$, then $D_i^* = 1$ and $1 - U \leq 1/2$ so $D_j^* = 0$. In both cases $D_i^* = 1 - D_j^*$, so (D_i^*, D_j^*) is supported on $\{(0,1), (1,0)\}$ each with probability $1/2$. This is exactly the matched pairs distribution. $\square$

B.2 Proofs for Section 3

Proof of Identification in Example 3.1. First, consider θ_n. We have

$$
\begin{aligned}
\theta_n &= E_n E_F[s_i(D)] = E_F[\bar{Y}_n(D) H(D)] = \mathrm{Var}_F(D)^{-1} E_F[\bar{Y}_n(D)(D - E_F[D])] \\
&= \mathrm{Var}_F(D)^{-1}\,\mathrm{Cov}_F(\bar{Y}_n(D), D).
\end{aligned}
$$

Then by concentrating out, $\theta_{\mathrm{BLP}} = \mathrm{argmin}_{\theta \in \mathbb{R}^m} \min_\alpha E_F[(\bar{Y}_n(D) - \alpha - \theta' D)^2] = \mathrm{argmin}_{\theta \in \mathbb{R}^m} \mathrm{Var}_F(\bar{Y}_n(D) - \theta' D)$. The FOC gives $\mathrm{Var}_F(D)\theta_{\mathrm{BLP}} = \mathrm{Cov}_F(\bar{Y}_n(D), D)$, so $\theta_{\mathrm{BLP}} = \mathrm{Var}_F(D)^{-1}\,\mathrm{Cov}_F(\bar{Y}_n(D), D) = \theta_n$. $\square$

Assumption B.1 (OLS)**.** *Assume $k = O(1)$, $E_F[|D|_2^4] < \infty$, $E_n E_F[Y_i(D)^4] < \infty$, and $\mathrm{Var}_F(D) \succ 0$.*

Proof of OLS Linearization in Equation 3.3. By Frisch-Waugh, the OLS coefficient $\widehat{\beta} = \mathrm{Var}_n(D_i)^{-1}\,\mathrm{Cov}_n(D_i, Y_i)$. Define $y_n = E_n E_F[Y_i(D)]$. We have

$$
\begin{aligned}
\mathrm{Cov}_n(D_i, Y_i) &= E_n[(D_i - E_n[D_i])Y_i] = E_n[(D_i - E_n[D_i])(Y_i - y_n)] \\
&= E_n[(D_i - E_F[D] + E_F[D] - E_n[D_i])(Y_i - y_n)] \\
&= E_n[(D_i - E_F[D])(Y_i - y_n)] + (E_F[D] - E_n[D_i])(E_n[Y_i] - y_n)
\end{aligned}
$$

Write this as $\mathrm{Cov}_n(D_i, Y_i) = A_n + B_n$. By Lemma B.15 and Assumption B.1, we have $E_n[D_i] = E_F[D] + O_p(n^{-1/2})$ and $E_n[Y_i(D_i)] = y_n + O_p(n^{-1/2})$. Then $B_n = O_p(n^{-1})$.

$$
\begin{aligned}
A_n &= E_n[(D_i - E_F[D])(Y_i - y_n - \theta_n'(D_i - E_F[D]))] \\
&\quad + E_n[(D_i - E_F[D])(D_i - E_F[D])']\theta_n = A_n^1 + A_n^2
\end{aligned}
$$

We claim $A_n^1 = O_p(n^{-1/2})$ and $A_n^2 = \mathrm{Var}_n(D_i)\theta_n + O_p(n^{-1})$. Note that for A_n^1,

$$E_n E_F[(D - E_F[D])e_i(D)] = E_n \,\mathrm{Cov}_F(D, Y_i(D)) - \mathrm{Var}_F(D)\theta_n = 0.$$

To apply Lemma B.15 we show $E_n E_F[|r_i(D)|_2^2] = O(1)$. This is $E_F[|r_i(D)|_2^2] = E_F[|D - E_F[D]|_2^2 e_i(D)^2] \le (E_F[|D - E_F[D]|_2^4])^{1/2}(E_F[e_i(D)^4])^{1/2}$. By the c_r inequality, $E_F[e_i(D)^4] \lesssim E_F[Y_i(D)^4] + y_n^4 + |\theta_n|_2^4 E_F[|D - E_F[D]|_2^4]$, so $E_n E_F[r_i(D)^2] = O(1)$ under Assumption B.1. Then $A_n^1 = O_p(n^{-1/2})$. The proof for A_n^2 is similar. Then overall $\mathrm{Var}_n(D_i)^{-1}[A_n^1 + A_n^2] = \theta_n + \mathrm{Var}_F(D)^{-1}A_n^1 + O_p(n^{-1})$. This proves the claim with influence function $s_i(d) = \mathrm{Var}_F(D)^{-1}(d - E_F[D])e_i(d)$. □

Assumption B.2 (Logit). *Assume $k = O(1)$, $E_F[|D|_2^4] < \infty$, $E_n E_F[Y_i(D)^4] < \infty$, $\sup_n \|\theta_n\| < \infty$, $\mathrm{Var}_F(D) \succ 0$.*

Theorem B.3 (Logit Identification and Influence Function). *Impose Assumption B.2. Then Equation (3.4) and the claim $\widehat{\theta} = E_n[s_i(D_i)] + O_p(n^{-1})$ hold.*

Proof of Theorem B.3. Step 1: First we, show consistency. Define objectives

$$\begin{aligned}
\widehat{Q}_n(\theta) &= E_n[Y_i \log(L(\theta' D_i)) + (1 - Y_i)\log(1 - L(\theta' D_i))] \\
Q_n(\theta) &= E_F[\bar{Y}_n(D)\log(L(\theta' D)) + (1 - \bar{Y}_n(D))\log(1 - L(\theta' D))].
\end{aligned}$$

We claim $\widehat{Q}_n(\theta) = Q_n(\theta) + O_p(n^{-1/2})$ for each $\theta \in \Theta$. To see this, fix θ and define $s_i(d) = Y_i(d)\log L(\theta' d) + (1 - Y_i(d))\log(1 - L(\theta' d))$, so that $\widehat{Q}_n(\theta) = E_n[s_i(D_i)]$ and $Q_n(\theta) = E_n E_F[s_i(D)]$. By Lemma B.15 it suffices to check $v_{iid}(s) = O(1)$. Note $|\log L(v)| \le |v| + \log 2$ and $|\log(1 - L(v))| \le |v| + \log 2$, so that influence $|s_i(d)| \le (|Y_i(d)| + 1)(|\theta' d| + \log 2)$. By Cauchy-Schwarz, Young's, and c_r inequality

$$\begin{aligned}
E_F[s_i(D)^2] &\lesssim (E_F[(1 + |Y_i(D)|)^4])^{1/2}(E_F[(|\theta' D| + \log 2)^4])^{1/2} \\
&\lesssim E_F[(1 + |Y_i(D)|)^4] + E_F[(|\theta' D| + \log 2)^4] \\
&\lesssim E_F[|Y_i(D)|^4] + E_F[|\theta' D|^4] \lesssim E_F[|Y_i(D)|^4] + |\theta|_2^4 E_F[|D|_2^4]
\end{aligned}$$

Then $E_n E_F[s_i(D)^2] \lesssim E_n E_F[|Y_i(D)|^4] + |\theta|_2^4 E_F[|D|_2^4] = O(1)$ under Assumption B.2. This proves the claim. Note that $\widehat{Q}_n(\cdot)$ is concave. Also observe that $|\nabla_\theta Q_n(\theta)|_2 = |E_F[(\bar{Y}_n(D) - L(\theta' D))D]|_2 \le E_F[|D|_2] < \infty$ by Assumption B.2. Then $Q_n(\theta)$ is uniformly Lipschitz in n, hence equicontinuous. A suitably modified version of the convexity lemma in Pollard (1991) then implies $\sup_{\theta \in K} |\widehat{Q}_n(\theta) - Q_n(\theta)| = o_p(1)$ for any compact $K \subseteq \mathbb{R}^m$. We now prove $\widehat{\theta} \xrightarrow{p} \theta_n$. Since $\sup_n \|\theta_n\| < \infty$ by assumption, choose R large enough that $B(\theta_n, 1) \subseteq \mathrm{int}(B_R)$ for all n, where $B_R = \{\theta : \|\theta\|_2 \le R\}$. Let $\widehat{\theta}_R = \mathrm{argmax}_{\theta \in B_R} \widehat{Q}_n(\theta)$. The ULLN on B_R gives $\sup_{\theta \in B_R} |\widehat{Q}_n(\theta) - Q_n(\theta)| = o_p(1)$. Since $\widehat{\theta}_R$ maximizes $\widehat{Q}_n$ over B_R and $\theta_n \in B_R$,

$$\begin{aligned}
Q_n(\theta_n) - Q_n(\widehat{\theta}_R) &\le [Q_n(\theta_n) - \widehat{Q}_n(\theta_n)] + [\widehat{Q}_n(\widehat{\theta}_R) - Q_n(\widehat{\theta}_R)] \\
&\le 2 \sup_{\theta \in B_R} |\widehat{Q}_n(\theta) - Q_n(\theta)| = o_p(1).
\end{aligned}$$

We claim for any $\epsilon > 0$, there exists $\eta > 0$ such that $Q_n(\theta_n) - Q_n(\theta) \ge \eta$ whenever $\|\theta - \theta_n\| \ge \epsilon$ and $\theta \in B_R$. To see this, $Q_n(\theta_n) - Q_n(\theta) = -\frac{1}{2}(\theta - \theta_n)'\nabla^2 Q_n(\bar{\theta})(\theta - \theta_n)$ for some $\bar{\theta} \in [\theta_n, \theta]$ by Taylor's theorem, using $\nabla Q_n(\theta_n) = 0$. Since $\bar{\theta} \in B_R$, we have $|\bar{\theta}' D| \le R|D|_2$, so $L'(\bar{\theta}' D) = L(\bar{\theta}' D)(1 - L(\bar{\theta}' D)) \ge L(R|D|_2)(1 - L(R|D|_2)) \equiv$

$w_R(D) > 0$. Then $-\nabla^2 Q_n(\bar\theta) = E_F[L'(\bar\theta' D)DD'] \succeq E_F[w_R(D)DD'] \equiv M_R$. Note that $M_R \succ 0$, since if $M_R v = 0$ for some $v \neq 0$, then $w_R(D)(v'D)^2 = 0$. Since $w_R(D) > 0$, we have $v'D = 0$, contradicting $\mathrm{Var}_F(D) \succ 0$. Also, M_R does not depend on n. Then $Q_n(\theta_n) - Q_n(\theta) \geq \frac{1}{2}\lambda_{\min}(M_R)\|\theta - \theta_n\|^2 \geq \frac{1}{2}\lambda_{\min}(M_R)\epsilon^2 \equiv \eta > 0$. Then $P(|\widehat\theta_R - \theta_n| > \epsilon) \leq P(Q_n(\theta_n) - Q_n(\widehat\theta_R) > \eta) = o(1)$, so that $\widehat\theta_R \xrightarrow{p} \theta_n$. Then $P(\widehat\theta = \widehat\theta_R) \geq P(|\widehat\theta_R - \theta_n| < 1) \to 1$, so $\widehat\theta \xrightarrow{p} \theta_n$.

Step 2: Next we derive the influence function. Define $r_i(d,\theta) = (Y_i(d) - L(\theta' d))d$, so the MLE solves $E_n[r_i(D_i, \widehat\theta)] = 0$. Observe that $\nabla_\theta r_i(d,\theta) = -L'(\theta' d)dd'$ where $L'(v) = L(v)(1 - L(v))$. Then by the mean value theorem, there exists $\bar\theta_n \in [\theta_n, \widehat\theta]$ such that $0 = E_n[r_i(D_i, \theta_n)] - E_n[L'(\bar\theta_n' D_i)D_i D_i'](\widehat\theta - \theta_n)$, By Step 1, $\widehat\theta \xrightarrow{p} \theta_n$, so $\bar\theta_n \xrightarrow{p} \theta_n$. We claim $E_n[L'(\bar\theta_n' D_i)D_i D_i'] = J_n + o_p(1)$. Since $|L''(v)| \leq 1/4$, the mean value theorem again gives $|L'(\bar\theta_n' D_i) - L'(\theta_n' D_i)| \leq \frac{1}{4}|\bar\theta_n - \theta_n|_2 |D_i|_2$, so

$$|E_n[(L'(\bar\theta_n' D_i) - L'(\theta_n' D_i))D_i D_i']|_{op} \leq \tfrac{1}{4}|\bar\theta_n - \theta_n|_2 \cdot E_n[|D_i|_2^3].$$

By Jensen, $E_n[|D_i|_2^3] \leq (E_n[|D_i|_2^4])^{3/4}$. The latter is $O_p(1)$ by Lemma B.15 since $E_F[|D|_2^4] < \infty$. Since $|\bar\theta_n - \theta_n|_2 = o_p(1)$, we have shown that $E_n[L'(\bar\theta_n' D_i)D_i D_i'] = E_n[L'(\theta_n' D_i)D_i D_i'] + o_p(1)$. Working entry-wise with $z_i(d) \equiv L'(\theta_n' d)d_j d_l$, we have $v_{iid}(z) \leq E_F[D_j^2 D_l^2] \leq E_F[|D|_2^4] = O(1)$ since $L' \leq 1$. Since $k = O(1)$, this implies $E_n[L'(\theta_n' D_i)D_i D_i'] = J_n + O_p(n^{-1/2})$ by Lemma B.15, proving the claim. Next we show $E_n[r_i(D_i, \theta_n)] = O_p(n^{-1/2})$. Since $E_n E_F[r_i(D, \theta_n)] = 0$ by the first-order condition $\nabla Q_n(\theta_n) = 0$, it suffices by Lemma B.15 to check $E_n E_F[r_i^j(D,\theta_n)^2] = O(1)$ for each component $r_i^j(d, \theta_n) = (Y_i(d) - L(\theta_n' d))d_j$. Since $|Y_i(d) - L(\theta_n' d)| \leq 1$, we have $E_F[r_i^j(D,\theta_n)^2] \leq E_F[D_j^2] = O(1)$. Since $k = O(1)$, the claim follows. Then

$$\widehat\theta - \theta_n = E_n[L'(\bar\theta_n' D_i)D_i D_i']^{-1} E_n[r_i(D_i, \theta_n)] = (J_n + o_p(1))^{-1} O_p(n^{-1/2}).$$

We claim $(J_n + o_p(1))^{-1} = J_n^{-1} + o_p(1)$. Write $E_n[L'(\bar\theta_n' D_i)D_i D_i'] = J_n + R_n$ where $R_n = o_p(1)$. Then $(J_n + R_n)^{-1} = J_n^{-1} - J_n^{-1}R_n(J_n + R_n)^{-1}$, so it suffices to show $|J_n^{-1}R_n(J_n + R_n)^{-1}|_{op} = o_p(1)$. Since $\sup_n |\theta_n|_2 < \infty$, there exists R such that $|\theta_n' D| \leq R|D|_2$ for all n, so $L'(\theta_n' D) \geq w_R(D) \equiv L(R|D|_2)(1 - L(R|D|_2)) > 0$. Then $J_n \succeq M_R \equiv E_F[w_R(D)DD'] \succ 0$ for all n, where $M_R \succ 0$ by $\mathrm{Var}_F(D) \succ 0$ as in Step 1. In particular $|J_n^{-1}|_{op} \leq \lambda_{\min}(M_R)^{-1} = O(1)$. Similarly, $|(J_n + R_n)^{-1}|_{op} = O_p(1)$. Since $R_n = o_p(1)$, the claim follows. Therefore

$$\widehat\theta - \theta_n = J_n^{-1}E_n[r_i(D_i, \theta_n)] + o_p(n^{-1/2}) = E_n[s_i(D_i)] + o_p(n^{-1/2}),$$

where $s_i(d) = J_n^{-1}e_i(d) \cdot d$ for residual $e_i(d) = Y_i(d) - L(\theta_n' d)$. *Step 3:* Finally, we show identification. Note that $\mathrm{KL}(p\|q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$, so we have $p\log q + (1-p)\log(1-q) = -\mathrm{KL}(p\|q) + h(p)$ where $h(p) = p\log p + (1-p)\log(1-p)$ does not depend on q. Then

$$\begin{aligned}\theta_n &= \underset{\theta}{\operatorname{argmax}}\ E_F[\bar Y_n(D)\log L(\theta' D) + (1 - \bar Y_n(D))\log(1 - L(\theta' D))] \\ &= \underset{\theta}{\operatorname{argmin}}\ E_F[\mathrm{KL}(\bar Y_n(D)\|L(\theta' D))],\end{aligned}$$

This finishes the proof. ☐

B.3 Proofs for Section 4

Proof of Proposition 4.2. By Lemma C.3, we have

$$\operatorname{Var}_k(\phi(D_i)) = k^{-1}\sum_{i\in[k]}\phi(D_i)^2 - (k(k-1))^{-1}\sum_{i\neq j\in[k]}\phi(D_i)\phi(D_j).$$

Since $G \in \Pi_k(F)$, by fixed marginals and exchangeability,

$$\begin{aligned}
E_G \operatorname{Var}_k(\phi(D_i)) &= E_F[\phi(D)^2] - E_G[\phi(D_1)\phi(D_2)] \\
&= E_F[\phi(D)^2] - E_F[\phi(D)]^2 + E_F[\phi(D)]^2 - E_G[\phi(D_1)\phi(D_2)] \\
&= \operatorname{Var}_F(\phi) - \operatorname{Cov}_G(\phi(D_1), \phi(D_2)).
\end{aligned}$$

Then to finish the proof, observe

$$(k-1)\left(\frac{E_G \operatorname{Var}_k(\phi(D_i))}{\operatorname{Var}_F(\phi)} - 1\right) = -(k-1)\operatorname{Corr}_G(\phi(D_1), \phi(D_2)) = \operatorname{Disp}_G(\phi).$$

□

Proof of Definition 4.1 Bounds. Let $Z = \sum_{i=1}^k \phi(D_i)$. Since $G \in \Pi_k(F)$, we have $\operatorname{Var}_G(\phi(D_i)) = \operatorname{Var}_F(\phi)$ for all $i \in [k]$, and $\operatorname{Cov}_G(\phi(D_i), \phi(D_j)) = \rho \operatorname{Var}_F(\phi)$ for $\rho = \operatorname{Corr}_G(\phi(D_1), \phi(D_2))$ for all $i \neq j$ by exchangeability. Then

$$\begin{aligned}
0 \leq \operatorname{Var}_G(Z) &= \sum_{i\in[k]}\operatorname{Var}_G(\phi(D_i)) + \sum_{i\neq j\in[k]}\operatorname{Cov}_G(\phi(D_i), \phi(D_j)) \\
&= k\operatorname{Var}_F(\phi) + k(k-1)\rho\operatorname{Var}_F(\phi) = k\operatorname{Var}_F(\phi)(1+(k-1)\rho).
\end{aligned}$$

Since $\operatorname{Var}_F(\phi) > 0$, we must have $1 + (k-1)\rho \geq 0$, so that $\rho \geq -(k-1)^{-1}$. Then we have $\operatorname{Disp}_G(\phi) = -(k-1)\rho \leq 1$, finishing the proof. □

Proof of Equation 4.2. Fix $G \in \Pi_k(F)$ and let $\rho = \operatorname{Corr}_G(\phi(D_i), \phi(D_j))$ for $i \neq j$.

$$\begin{aligned}
\operatorname{Var}_G(\widehat{\theta}) &= k^{-2}\sum_{i,j\in[k]}\operatorname{Cov}_G(\phi(D_i), \phi(D_j)) = k^{-2}\big(k\operatorname{Var}_F(\phi) + k(k-1)\rho\operatorname{Var}_F(\phi)\big) \\
&= k^{-1}\operatorname{Var}_F(\phi)(1+(k-1)\rho) = k^{-1}\operatorname{Var}_F(\phi)(1-\operatorname{Disp}_G(\phi))).
\end{aligned}$$

Under G_{iid}, $\rho = 0$, so $\operatorname{Var}_{G_{iid}}(\widehat{\theta}) = k^{-1}\operatorname{Var}_F(\phi)$. Then $\operatorname{Var}_G(\widehat{\theta})/\operatorname{Var}_{G_{iid}}(\widehat{\theta}) = 1 - \operatorname{Disp}_G(\phi)$, and the claim follows. □

Proposition B.4 (Random Matching)**.** *Suppose* $\operatorname{Var}_F(s_i) < \infty$ *for* $i \in [n]$*. Let* τ *be a uniform random partition of* $[n]$ *into* n/k *groups. Define* $\bar{s}_n(d) \equiv n^{-1}\sum_{i=1}^n s_i(d)$*.*

$$E_\tau[Q_k(s)] = \frac{n}{n-1}\cdot\frac{\operatorname{Var}_F(\bar{s}_n)}{v_{iid}(s)} - \frac{1}{n-1}. \tag{B.1}$$

Proof. Define $c(s) \equiv n^{-1}\sum_g\sum_{i\neq j\in[k]}\operatorname{Cov}_F(s_{ig}, s_{jg})$. By Lemma C.1, the match coefficient $Q_k(s) = (k-1)^{-1}c(s)/v_{iid}(s)$. Then we calculate

$$\begin{aligned}
E_\tau[c(s)] &= n^{-1}E_\tau\sum_{i\neq j\in[n]}\mathbb{1}(g(i)=g(j))\operatorname{Cov}_F(s_i, s_j) \\
&= n^{-1}\sum_{i\neq j\in[n]}P_\tau(g(i)=g(j))\operatorname{Cov}_F(s_i, s_j) = \frac{k-1}{n(n-1)}\sum_{i\neq j\in[n]}\operatorname{Cov}_F(s_i, s_j).
\end{aligned}$$

The third equality since $P_\tau(g(i) = g(j)) = (k-1)/(n-1)$ by a counting argument. Then $E_\tau[Q_k(s)] = n^{-1}(n-1)^{-1}\sum_{i\neq j\in[n]} \mathrm{Cov}_F(s_i, s_j)/v_{iid}(s)$.

$$\mathrm{Var}_F(\bar{s}_n) = n^{-2}\sum_{i,j}\mathrm{Cov}_F(s_i, s_j) = n^{-2}\sum_i \mathrm{Var}_F(s_i) + n^{-2}\sum_{i\neq j}\mathrm{Cov}_F(s_i, s_j).$$

Rearranging gives $\sum_{i\neq j}\mathrm{Cov}_F(s_i, s_j) = n^2\,\mathrm{Var}_F(\bar{s}_n) - nv_{iid}(s)$. Then

$$\begin{aligned} E_\tau[Q_k(s)] &= \frac{1}{n(n-1)v_{iid}(s)}(n^2\,\mathrm{Var}_F(\bar{s}_n) - nv_{iid}(s)) \\ &= \frac{n}{n-1}\frac{\mathrm{Var}_F(\bar{s}_n)}{v_{iid}(s)} - \frac{1}{n-1}. \end{aligned}$$

This completes the proof. □

Proof of Proposition A.1. We claim that (1) $c(s) = c(\mu) + o_p(1)$ for covariance $c(s) = n^{-1}\sum_g\sum_{i\neq j\in[k]}\mathrm{Cov}_F(s_{ig}, s_{jg})$ and (2) $v_{iid}(\mu) = E_P[\mathrm{Var}_F(\mu_i)] + o_p(1)$ and $v_{iid}(s) = E_P[\mathrm{Var}_F(s_i)] + o_p(1)$. Under these assumptions, by Lemma C.1, $Q_k(\mu) = (k-1)^{-1}c(\mu)/v_{iid}(\mu)$, so

$$\begin{aligned} Q_k(s) &= \frac{(k-1)^{-1}c(s)}{v_{iid}(s)} = \frac{(k-1)^{-1}[c(\mu) + o_p(1)]}{v_{iid}(s)} \\ &= \frac{(k-1)^{-1}c(\mu)}{v_{iid}(s)} + o_p(1) = \frac{v_{iid}(\mu)}{v_{iid}(s)}\cdot Q_k(\mu) + o_p(1). \end{aligned}$$

The third equality since $v_{iid}(s) \xrightarrow{p} E_P[\mathrm{Var}_F(s_i)] > 0$ by assumption. Since $Q_k(\mu) \in [-(k-1)^{-1}, 1]$ is bounded and $v_{iid}(\mu)/v_{iid}(s) \xrightarrow{p} R^2_{s|X}$ by the claims above and continuous mapping theorem, we conclude $Q_k(s) = R^2_{s|X}\cdot Q_k(\mu) + o_p(1)$.

Consider claim (1). We expand $\mathrm{Cov}_F(s_{ig}, s_{jg}) = \mathrm{Cov}_F(\mu_{ig}, \mu_{jg}) + \mathrm{Cov}_F(\mu_{ig}, e_{jg}) + \mathrm{Cov}_F(e_{ig}, \mu_{jg}) + \mathrm{Cov}_F(e_{ig}, e_{jg})$ for $i \neq j$. We can decompose $c(s) = c(\mu) + A_n + B_n$, where A_n collects the cross covariance terms $\mathrm{Cov}_F(\mu_{ig}, e_{jg})$ and $\mathrm{Cov}_F(e_{ig}, \mu_{jg})$, and B_n collects the $\mathrm{Cov}_F(e_{ig}, e_{jg})$ terms. We show $A_n, B_n = o_p(1)$. To see this, note

$$E[A_n|X_{1:n}] = n^{-1}\sum_g\sum_{i\neq j\in[k]}\big[E[\mathrm{Cov}_F(\mu_{ig}, e_{jg})|X_{1:n}] + E[\mathrm{Cov}_F(e_{ig}, \mu_{jg})|X_{1:n}]\big].$$

We have $\mathrm{Cov}_F(\mu_{ig}, e_{jg}) = E_F[\mu_{ig}(D)e_{jg}(D)] - E_F[\mu_{ig}(D)]E_F[e_{jg}(D)]$. Recall matching $\tau : [n] \to [k] \times [n/k]$ with $\tau \in \sigma(X_{1:n})$. Then $E_F[\mu_{ig}(D)e_{jg}(D)] = \sum_{r,l}\mathbb{1}(\tau(r) = ig,\, \tau(l) = jg)\, E_F[\mu(X_r, D)\, e_l(D)]$. Note $\mathbb{1}(\tau(r) = ig,\, \tau(l) = jg)$ and $\mu(X_r, d)$ are $\sigma(X_{1:n})$-measurable for each d. Moreover, we claim $E_P[E_F[\mu(X_r, D)\, e_l(D)]|X_{1:n}] = 0$ for each $r \neq l$. It suffices to show $E_P[E_F[\mu(X_r, D)\, e_l(D)]\cdot h(X_{1:n})] = 0$ for any bounded measurable $h(\cdot)$. To see this, note

$$\begin{aligned} &E_P\big[E_F[|\mu(X_r, D)\, e_l(D)\, h(X_{1:n})|]\big] \lesssim E_P\big[E_F[|\mu(X_r, D)\, e_l(D)|]\big] \\ &= E_F\big[E_P[|\mu(X_r, D)\, e_l(D)|]\big] \le E_F\big[E_P[\mu(X_r, D)^2]^{1/2}\, E_P[e_l(D)^2]^{1/2}\big] \\ &\le E_F\big[E_P[\mu(X_r, D)^2] + E_P[e_l(D)^2]\big] \le E_F\big[E_P[s_i(D)^2]\big] = E_P\big[E_F[s_i(D)^2]\big] < \infty. \end{aligned}$$

The first inequality since $|h| \le M$ for some $M < \infty$. The first equality is Tonelli's theorem. The second and third inequalities are Cauchy-Schwarz and Young's. The fourth since $E_P[\mu(X_r,d)^2] = E_P[(E_P[s_i(d)|X_r])^2] \le E_P[s_i(d)^2]$ by Jensen's inequality, and $E_P[e_l(d)^2] \le E_P[s_i(d)^2]$. The last equality is Tonelli's theorem, and finiteness holds by assumption. Then by Fubini's theorem,

$$E_P\big[E_F[\mu(X_r,D)\,e_l(D)]\cdot h(X_{1:n})\big] = E_F\big[E_P[\mu(X_r,D)\,e_l(D)\cdot h(X_{1:n})]\big].$$

For each fixed d, since $\mu(X_r,d)$ and $h(X_{1:n})$ are $\sigma(X_{1:n})$-measurable, the tower property gives $E_P[\mu(X_r,d)\,e_l(d)\cdot h(X_{1:n})] = E_P\big[\mu(X_r,d)\,h(X_{1:n})\,E_P[e_l(d)|X_{1:n}]\big]$. Since $(X_i, s_i(\cdot)) \sim P$ iid, the residual $e_l(\cdot)$ satisfies $E_P[e_l(d)|X_{1:n}] = E_P[e_l(d)|X_l] = 0$ for all d. Hence the integrand is zero for all d, so $E_P[E_F[\mu(X_r,D)\,e_l(D)]\cdot h(X_{1:n})] = 0$. Since $h(\cdot)$ was arbitrary, we conclude $E_P[E_F[\mu(X_r,D)\,e_l(D)]|X_{1:n}] = 0$. Then

$$\begin{aligned}
&E_P[E_F[\mu_{ig}(D)e_{jg}(D)]|X_{1:n}] \\
&= \sum_{r,l} \mathbb{1}(\tau(r)=ig,\,\tau(l)=jg)\,E_P[E_F[\mu(X_r,D)\,e_l(D)]|X_{1:n}] = 0.
\end{aligned}$$

By the same argument, $E_P[E_F[\mu_{ig}(D)]E_F[e_{jg}(D)]|X_{1:n}] = 0$, since $E_F[e_{jg}(D)] = \sum_l \mathbb{1}(\tau(l) = jg)\,E_F[e_l(D)]$ and $E_P[E_F[e_l(D)]|X_{1:n}] = 0$. Then we have shown $E_P[\mathrm{Cov}_F(\mu_{ig}, e_{jg})|X_{1:n}] = 0$ for each $i \neq j \in [k]$. By symmetry, $E_P[A_n|X_{1:n}] = 0$. Similarly, one can show $\mathrm{Var}_P(A_n|X_{1:n}) = O_p(n^{-1})$. Then by conditional Chebyshev, $A_n = O_p(n^{-1/2})$. The proof that $B_n = O_p(n^{-1/2})$ is similar.

Next, we show claim (2). Consider $v_{iid}(\mu) = E_n[\mathrm{Var}_F(\mu_i)]$. Note $E_P[\mathrm{Var}_F(\mu_i)] \le E_P[E_F[\mu_i(D)^2]] = E_F[E_P[\mu_i(D)^2]] \le E_F[E_P[s_i(D)^2]] = E_P[E_F[s_i(D)^2]] < \infty$. The equalities are by Tonelli's theorem to exchange E_P and E_F. The second inequality since $E_P[\mu_i(d)^2] = E_P[(E_P[s_i(d)|X_i])^2] \le E_P[s_i(d)^2]$ by Jensen's inequality. Then by WLLN, $v_{iid}(\mu) = E_P[\mathrm{Var}_F(\mu_i)] + o_p(1)$. Similarly, $v_{iid}(s) = E_n[\mathrm{Var}_F(s_i)] = E_P[\mathrm{Var}_F(s_i)] + o_p(1)$. This proves the claim. □

Proof of Theorem 4.4. For $\widehat{\theta} = E_n[s_i(D_i)]$ with $s_i(d) = c_i + a_i\phi(d)$, we have

$$\begin{aligned}
n\,\mathrm{Var}_G(\widehat{\theta}) &= v_{iid}(s) + n^{-1}\sum_g \sum_{i\neq j\in[k]} \mathrm{Cov}_G(s_{ig}(D_{ig}), s_{jg}(D_{jg})) \\
&= v_{iid}(s) + n^{-1}\sum_g \sum_{i\neq j\in[k]} a_{ig}a_{jg}\,\mathrm{Cov}_G(\phi(D_{ig}), \phi(D_{jg})) \\
&= v_{iid}(s) - \mathrm{Disp}_G(\phi)\cdot(k-1)^{-1}n^{-1}\sum_g \sum_{i\neq j\in[k]} a_{ig}a_{jg}\cdot \mathrm{Var}_F(\phi).
\end{aligned}$$

The first equality is by definition of coupling designs, with $\widehat{\theta} = n^{-1}\sum_g\sum_i s_{ig}(D_{ig})$. The second equality is by our parametric assumption. The third equality is by Definition 4.1. By Lemma C.1, match quality $Q_k(s) = (k-1)^{-1}c(s)/v_{iid}(s)$ for the coefficient $c(s) = n^{-1}\sum_g\sum_{i\neq j\in[k]}\mathrm{Cov}_F(s_{ig}, s_{jg}) = \mathrm{Var}_F(\phi)\cdot n^{-1}\sum_g\sum_{i\neq j\in[k]} a_{ig}a_{jg}$. Then continuing the calculation above, this is

$$\begin{aligned}
&= v_{iid}(s) - \mathrm{Disp}_G(\phi)\cdot(k-1)^{-1}c(s) = v_{iid}(s) - v_{iid}(s)\,\mathrm{Disp}_G(\phi)Q_k(s) \\
&= v_{iid}(s)(1 - \mathrm{Disp}_G(\phi)Q_k(s)).
\end{aligned}$$

This finishes the proof. □

Proof of Theorem A.5. First, suppose such G^c exists. Then there exists $z \in [m]^k$ with $\widehat{f}(z) = f$. In particular, $\widehat{f}_j(z) = k^{-1}\sum_{i\in[k]} \mathbb{1}(z_i = j) = f_j$, so $k \cdot f_j = \sum_{i\in[k]} \mathbb{1}(z_i = j) \in \mathbb{N}$. Conversely, if $n_j \equiv k \cdot f_j \in \mathbb{N}$ for all j, we have $\sum_j n_j = k$. Then we can construct an allocation $z \in [m]^k$ with $z_i = j$ for exactly n_j indices i. This allocation has $\widehat{f}_j(z) = n_j/k = f_j$ for all $j \in [m]$, so $z \in \{z : \widehat{f}(z) = f\}$. Then the set $\{z : \widehat{f}(z) = f\}$ is non-empty and $G^c = \mathrm{Unif}\{z : \widehat{f}(z) = f\}$ exists.

Next, we show $G^c \in \Pi_k(F)$. Claim G^c is exchangeable. We must show $P_{G^c}(Z_\sigma = z) = P_{G^c}(Z = z)$ for all $z \in [m]^k$ and any permutation $\sigma \in S_k$. Note $P_{G^c}(Z_\sigma = z) = P_{G^c}(Z = z_{\sigma^{-1}})$. Since G^c is uniform on $S \equiv \{z : \widehat{f}(z) = f\}$, we have $P_{G^c}(Z = z) = |S|^{-1}\mathbb{1}(z \in S)$. Then it suffices to show $z \in S \iff z_\sigma \in S$ for all $z \in [m]^k$ and $\sigma \in S_k$, noting $\{\sigma : \sigma \in S_k\} = \{\sigma^{-1} : \sigma \in S_k\}$. For $z \in [m]^k$ and $\sigma \in S_k$,

$$\widehat{f}_j(z_\sigma) = k^{-1}\sum_{i\in[k]} \mathbb{1}(z_{\sigma(i)} = j) = k^{-1}\sum_{l\in[k]} \mathbb{1}(z_l = j) = \widehat{f}_j(z)$$

where the second equality is by the substitution $\sigma^{-1}(l) = i$. Then $\widehat{f}(z_\sigma) = f$ if and only if $\widehat{f}(z) = f$, so $z \in S \iff z_\sigma \in S$. This proves the claim. Next, we show the marginal $G^c_l = F$. Fix $j \in [m]$ and note that

$$f_j = E_{G^c}\big[\widehat{f}_j(Z)\big] = k^{-1}\sum_{i=1}^{k} P_{G^c}(Z_i = j) = P_{G^c}(Z_l = j) \quad \text{for any } l \in [k].$$

The first equality holds by definition of G^c, the final equality by exchangeability. This finishes the proof of the first result. Next consider the claim about dispersion. Let $(D_i)_{i\in[k]} \sim G^c$. By definition,

$$k^{-1}\sum_{i\in[k]} \phi(D_i) = \sum_{j\in[m]} \phi(j)\cdot \widehat{f}_j(D) = \sum_{j\in[m]} \phi(j)\cdot f_j = E_F[\phi(D)].$$

Then by Equation 4.2, we have $0 = k\,\mathrm{Var}_{G^c}(\widehat{\theta}) = \mathrm{Var}_F(\phi)(1 - \mathrm{Disp}_{G^c}(\phi))$, so $\mathrm{Disp}_{G^c}(\phi) = 1$. This finishes the proof. □

Proof of Corollary 4.6. Define $\psi(d) = \phi(d) - E_F[\phi(D)]$ and $z_i = b_i - \bar{b}$. Set $s_i(d) = z_i \cdot \psi(d)$, so that $\mathrm{Cov}_n(\phi(D_i), b_i) = E_n[\psi(D_i)\cdot z_i] = E_n[s_i(D_i)]$. This has the form studied in Theorem 4.4. Since $E_F[\psi(D)] = 0$, we have $E_F[s_i(D_i)] = 0$ for $i \in [n]$, so $E_G[\mathrm{Cov}_n(\phi(D_i), b_i)^2] = E_G[(E_n[s_i(D_i)])^2] = \mathrm{Var}_G(E_n[s_i(D_i)])$. By Theorem 4.4,

$$n \cdot E_G[\mathrm{Cov}_n(\phi(D_i), b_i)^2] = v_{iid}(s)\big(1 - \mathrm{Disp}_G(\phi)\cdot Q_k(s)\big). \tag{B.2}$$

The iid variance is $v_{iid}(s) = E_n[z_i^2\,\mathrm{Var}_F(\psi)] = \mathrm{Var}_F(\phi)\cdot\mathrm{Var}_n(b_i)$. For match quality, by the last identity in Lemma C.1, $v_\Delta(s) = (k-1)^{-1}(n/k)^{-1}\sum_g\sum_{i\in[k]}\mathrm{Var}_F(s_{ig} - \bar{s}_g)$. Since $s_{ig}(d) - \bar{s}_g(d) = (z_{ig} - \bar{z}_g)\psi(d)$ and $z_{ig} - \bar{z}_g = b_{ig} - \bar{b}_g$, we have $\mathrm{Var}_F(s_{ig} - \bar{s}_g) = (b_{ig} - \bar{b}_g)^2\,\mathrm{Var}_F(\phi)$. Then by equality above $v_\Delta(s) = \mathrm{Var}_F(\phi)\cdot(n/k)^{-1}\sum_g \mathrm{Var}_k(b_{ig})$. Then

$$Q_k(s) = 1 - \frac{v_\Delta(s)}{v_{iid}(s)} = 1 - \frac{(n/k)^{-1}\sum_g \mathrm{Var}_k(b_{ig})}{\mathrm{Var}_n(b_i)} = Q_k(b).$$

By definition, $\mathcal{I}_G(\phi, b) = E_G[\mathrm{Cov}_n(\phi(D_i), b_i)^2]$, so Equation B.2 and $Q_k(s) = Q_k(b)$ give $n\cdot\mathcal{I}_G(\phi, b) = v_{iid}(s)(1 - \mathrm{Disp}_G(\phi)\cdot Q_k(b))$. Setting $G = G_{iid}$ and using $\mathrm{Disp}_{G_{iid}}(\phi) = 0$, we obtain $n\cdot\mathcal{I}_{G_{iid}}(\phi, b) = v_{iid}(s)$. Dividing, $\mathcal{I}_G(\phi, b)/\mathcal{I}_{G_{iid}}(\phi, b) = 1 - \mathrm{Disp}_G(\phi)\cdot Q_k(b)$, which finishes the proof. □

B.4 Proofs for Section 5

Lemma B.5 (Operator Properties)**.** *For $G \in \Pi_k(F)$, the coupling operator U_G is a self-adjoint contraction on $L^2(F)$. If E is an eigenspace of U_G, then E is closed in $L^2(F)$. If E, E' are eigenspaces of U_G with $\lambda_E \neq \lambda_{E'}$, then $E \perp E'$.*

Proof of Lemma B.5. Write $U = U_G$. First, we show U is a contraction on $L^2(F)$. For any $\phi \in L^2(F)$, by conditional Jensen's inequality and tower law

$$\begin{aligned} E_F[(U\phi)(D)^2] &= E_F[E_G[\phi(D_1)|D_2]^2] \leq E_F[E_G[\phi(D_1)^2|D_2]] \\ &= E_G[\phi(D_1)^2] = E_F[\phi(D)^2]. \end{aligned}$$

Next we show self-adjointness. For $\phi, \psi \in L^2(F)$,

$$\begin{aligned} \langle U\phi, \psi\rangle_F &= E_F[(U\phi)(D)\psi(D)] = E_F[E_G[\phi(D_1)|D_2]\psi(D_2)] = E_G[\phi(D_1)\psi(D_2)] \\ &= E_G[\phi(D_2)\psi(D_1)] = \langle U\psi, \phi\rangle_F. \end{aligned}$$

The second equality follows since $D_2 \sim F$. The fourth equality since by exchangeability, $(D_1, D_2) \sim (D_2, D_1)$, so $(\phi(D_1), \psi(D_2)) \sim (\phi(D_2), \psi(D_1))$. For closedness, since U is a contraction it is bounded, and the eigenspace $E = \ker(U - \lambda I)$ is the kernel of a bounded linear operator, hence closed. For orthogonality, let $\phi \in E$ and $\psi \in E'$ with $\lambda_E \neq \lambda_{E'}$. Then by self-adjointness, $\lambda_E\langle\phi, \psi\rangle_F = \langle U\phi, \psi\rangle_F = \langle\phi, U\psi\rangle_F = \lambda_{E'}\langle\phi, \psi\rangle_F$. Since $\lambda_E \neq \lambda_{E'}$, we have $\langle\phi, \psi\rangle_F = 0$. ☐

Proof of Theorem 5.3. First consider (a). We have $\mathrm{Var}_F(\phi) > 0$, so

$$\begin{aligned} \mathrm{Disp}_G(\phi) &= -(k-1)\,\mathrm{Corr}_G(\phi(D_1), \phi(D_2)) = -(k-1)\frac{E_G[\phi(D_1)\phi(D_2)]}{\mathrm{Var}_F(\phi)} \\ &= -(k-1)\frac{\langle U_G\phi, \phi\rangle_F}{\mathrm{Var}_F(\phi)} = -(k-1)\frac{\mathrm{Var}_F(\phi)}{\mathrm{Var}_F(\phi)} \cdot \lambda = -(k-1)\cdot\lambda. \end{aligned}$$

The third equality by calculations in Lemma B.5. The fourth equality since $\phi \in E \subseteq L^2_0(F)$. This proves the claim. Next consider (b). First, we show the projection P_m exists and has the required form. By Lemma B.5, each eigenspace E_m is closed in $L^2(F)$, so the projection $P_m\phi = \mathrm{argmin}_{f \in E_m} E_F[(\phi(D) - f(D))^2]$ exists and is unique. Moreover, recall $E_F[f(D)] = 0$ for any $f \in E_m \subset L^2_0(F)$. Let $\phi' = \phi - E_F[\phi]$. Then

$$\begin{aligned} E_F[(\phi(D) - f(D))^2] &= E_F\left[(\phi'(D) + E_F[\phi] - f(D))^2\right] \\ &= E_F[\phi]^2 + E_F\left[(\phi'(D) - f(D))^2\right] = \mathrm{Var}_F(\phi - f) + E_F[\phi]^2. \end{aligned}$$

The second equality since $E_F[\phi' - f] = 0$, so the cross-term cancels. This shows that $P_m\phi = \mathrm{argmin}_{f \in E_m} \mathrm{Var}_F(\phi - f)$. For the main result, let $n = k$ and define $s_i = \phi$ for each $i \in [k]$. Then combining Equation 4.2 and Theorem 5.7,

$$\mathrm{Disp}_G(\phi) = 1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} = \sum_{m \geq 1} w_m \cdot \mathrm{Disp}_G(E_m) Q_k(\phi^m). \tag{B.3}$$

Since $s_i = \phi$ for all i, the match quality $Q_k(\phi^m) = 1$. The projection weight $w_m = E_n \mathrm{Var}_F(P_m s_i)/E_n \mathrm{Var}_F(s_i) = \mathrm{Var}_F(P_m\phi)/\mathrm{Var}_F(\phi)$. This finishes the proof. ☐

Proof of Corollary 5.4. Let $L_0^2(F) = E \oplus E^\perp$ be eigenspaces of U_G with dispersions $\text{Disp}_G(E)$ and $\text{Disp}_G(E^\perp)$. Without loss, suppose $\text{Disp}_G(E) > \text{Disp}_G(E^\perp)$. Let $\phi \in L^2(F)$ with $\text{Var}_F(\phi) > 0$. By Theorem 5.3, we have $\text{Disp}_G(\phi) = w \cdot \text{Disp}_G(E) + (1-w) \cdot \text{Disp}_G(E^\perp)$, for $w = \text{Var}_F(P_E\phi)/\text{Var}_F(\phi) \in [0,1]$. Then apparently $\text{Disp}_G(E^\perp) \leq \text{Disp}_G(\phi) \leq \text{Disp}_G(E)$. To show the upper bound is achieved, take $\phi \in E$ with $\text{Var}_F(\phi) > 0$. Then $P_E\phi = \phi$ and $P_{E^\perp}\phi = 0$, so $w = \text{Var}_F(\phi)/\text{Var}_F(\phi) = 1$, so $\text{Disp}_G(\phi) = \text{Disp}_G(E)$ is achieved. Conversely, suppose $\text{Disp}_G(\phi) = \text{Disp}_G(E)$ for some $\phi \neq c$. Then $w \cdot \text{Disp}_G(E) + (1-w) \cdot \text{Disp}_G(E^\perp) = \text{Disp}_G(E)$, which requires $w = 1$ since $\text{Disp}_G(E) \neq \text{Disp}_G(E^\perp)$. Then $P_{E^\perp}\phi = 0$ and $\phi \in E$. This completes the proof. □

Proof of Theorem 5.7, Corollary 5.9. Define $r_i \equiv s_i - E_F[s_i] \in L_0^2(F)$ and let $r_i^m \equiv P_m r_i$ and $s_i^m = P_m s_i$. We have $P_m E_F[s_i] = 0$ since $E_F[s_i] \perp E_m$. Then $P_m r_i = P_m s_i$ by linearity. Since $r_i^m \in E_m$, by definition $U r_i^m = \lambda_m r_i^m$ for all m, and (b) $r_i^m \perp r_i^l$ for $m \neq l$, by orthogonality of eigenspaces. By Assumption 5.2, $r_i = \sum_m P_m r_i = \sum_m r_i^m$. Note the key fact $\text{Cov}_G(\phi(D_1), \psi(D_2)) = E_G[\phi(D_1)\psi(D_2)] = E_G[E_G[\phi(D_1)|D_2]\psi(D_2)] = E_F[(U\phi)(D)\psi(D)] = \langle U\phi, \psi\rangle_F$ for any $\phi, \psi \in L_0^2(F)$. Denote $U_G = U$ and calculate

$$
\begin{aligned}
n \,\text{Var}_G(\widehat{\theta}) &= n^{-1}\sum_i \text{Var}_F(s_i(D)) + n^{-1}\sum_g \sum_{i\neq j\in[k]} \text{Cov}_G(s_{ig}(D_1), s_{jg}(D_2)) \\
&= n^{-1}\sum_i \text{Var}_F(r_i(D)) + n^{-1}\sum_g \sum_{i\neq j\in[k]} \text{Cov}_G(r_{ig}(D_1), r_{jg}(D_2)) \\
&= n^{-1}\sum_i |\sum_m r_i^m|_F^2 + n^{-1}\sum_g \sum_{i\neq j\in[k]} \text{Cov}_G\big(\sum_m r_{ig}^m(D_1), \sum_l r_{jg}^l(D_2)\big) \\
&= n^{-1}\sum_i |\sum_m r_i^m|_F^2 + n^{-1}\sum_g \sum_{i\neq j\in[k]} \langle U\sum_m r_{ig}^m, \sum_l r_{jg}^l\rangle_F
\end{aligned}
$$

The first equality by definition of the design. The second equality since $G \in \Pi_k(F)$ so $s_{ig}(D_1) - E_G[s_{ig}(D_1)] = s_{ig}(D_1) - E_F[s_{ig}(D)] = r_{ig}(D_1)$. The third equality since $r_i = \sum_m r_i^m$ by Assumption 5.2. The fourth equality by the key fact. Continuing,

$$
\begin{aligned}
&= n^{-1}\sum_i |\sum_m r_i^m|_F^2 + n^{-1}\sum_g \sum_{i\neq j\in[k]} \langle \sum_m \lambda_m r_{ig}^m, \sum_l r_{jg}^l\rangle_F \\
&= n^{-1}\sum_i \sum_m |r_i^m|_F^2 + n^{-1}\sum_g \sum_{i\neq j\in[k]} \sum_m \lambda_m \langle r_{ig}^m, r_{jg}^m\rangle_F \\
&= n^{-1}\sum_m \sum_i |r_i^m|_F^2 + n^{-1}\sum_m \lambda_m \sum_g \sum_{i\neq j\in[k]} \langle r_{ig}^m, r_{jg}^m\rangle_F.
\end{aligned}
$$

The first equality since $U\sum_m r_{ig}^m = \sum_m U r_{ig}^m = \sum_m \lambda_m r_{ig}^m$ by continuity of U, since U is a linear contraction by Lemma B.5. The first term in the second equality follows by Parseval's theorem. The second term follows by continuity of the inner product map $(\phi, \psi) \to \langle \phi, \psi\rangle_F$ and orthogonality. To see this, note $\langle \sum_m \lambda_m r_{ig}^m, \sum_l r_{jg}^l\rangle_F = \lim_{K\to\infty}\langle \sum_{m=1}^K \lambda_m r_{ig}^m, \sum_{l=1}^K r_{jg}^l\rangle_F = \lim_{K\to\infty}\sum_{m=1}^K \lambda_m \langle r_{ig}^m, r_{jg}^m\rangle_F$. For the third equality above, we need to rigorously justify exchange of sums. By Tonelli's Theorem, we can exchange the first sums $\sum_i \sum_m |r_i^m|_F^2 = \sum_m \sum_i |r_i^m|_F^2$. For exchanging sums in the second term,

$$\begin{aligned}
n^{-1}\sum_m\sum_g\sum_{i\neq j\in[k]}|\lambda_m\langle r_{ig}^m, r_{jg}^m\rangle_F| &\le n^{-1}\sum_m\sum_g\sum_{i\neq j\in[k]}|\langle r_{ig}^m, r_{jg}^m\rangle_F| \\
\le n^{-1}\sum_m\sum_g\sum_{i\neq j\in[k]}|r_{ig}^m|_F|r_{jg}^m|_F &\le n^{-1}\sum_m\sum_g\Big(\sum_{i\in[k]}|r_{ig}^m|_F\Big)^2 \\
\le kn^{-1}\sum_m\sum_g\sum_{i\in[k]}|r_{ig}^m|_F^2 = k\sum_m E_n|r_i^m|_F^2 &= kE_n\sum_m|r_i^m|_F^2.
\end{aligned}$$

The first inequality since U is a contraction by Lemma B.5. The second is Cauchy-Schwarz. The third is Jensen's inequality. The second equality by Tonelli's Theorem. Finally, note that $E_n\sum_m|r_i^m|_F^2 = E_n|\sum_m r_i^m|_F^2 = E_n\operatorname{Var}_F(s_i) < \infty$ by assumption. Then by Fubini's Theorem the interchange of sums is justified. This finishes the justification of the second align environment above. Continuing from above, expand $\langle r_{ig}^m, r_{jg}^m\rangle_F = -(1/2)(|r_{ig}^m - r_{jg}^m|_F^2 - |r_{ig}^m|_F^2 - |r_{jg}^m|_F^2)$. Then the term $n^{-1}\sum_m\lambda_m\sum_g\sum_{i\neq j\in[k]}\langle r_{ig}^m, r_{jg}^m\rangle_F$ above can be expanded as

$$\begin{aligned}
&-(2n)^{-1}\sum_m\lambda_m\sum_g\sum_{i\neq j\in[k]}(|r_{ig}^m - r_{jg}^m|_F^2 - |r_{ig}^m|_F^2 - |r_{jg}^m|_F^2) \\
&=\sum_m\lambda_m\big[(k-1)n^{-1}\sum_i|r_i^m|_F^2 - (2n)^{-1}\sum_g\sum_{i\neq j\in[k]}|r_{ig}^m - r_{jg}^m|_F^2\big] \\
&=\sum_m\operatorname{Disp}_G(m)\big[(2n(k-1))^{-1}\sum_g\sum_{i\neq j\in[k]}|r_{ig}^m - r_{jg}^m|_F^2 - E_n|r_i^m|_F^2\big].
\end{aligned}$$

Note that $|r_i^m|_F^2 = \operatorname{Var}_F(s_i^m)$. Then putting both terms together, we get

$$\begin{aligned}
n\operatorname{Var}_G(\widehat{\theta}) = &\sum_m(1-\operatorname{Disp}_G(m))\cdot E_n\operatorname{Var}_F(s_i^m) \\
&+\operatorname{Disp}_G(m)\cdot(2n(k-1))^{-1}\sum_g\sum_{i\neq j\in[k]}\operatorname{Var}_F(s_{ig}^m - s_{jg}^m) \\
&=\sum_m(1-\operatorname{Disp}_G(m))\cdot v_{iid}(s^m) + \operatorname{Disp}_G(m)\cdot v_\Delta(s^m).
\end{aligned}$$

This proves Corollary 5.9. Then $\operatorname{Efficiency}(G) = 1 - \operatorname{Var}_G(\widehat{\theta})/\operatorname{Var}_{G_{iid}}(\widehat{\theta})$ has

$$\begin{aligned}
\operatorname{Efficiency}(G) &= 1 - v_{iid}(s)^{-1}\sum_m[v_{iid}(s^m)(1-\operatorname{Disp}_G(m)) + v_\Delta(s^m)\operatorname{Disp}_G(m)] \\
&= v_{iid}(s)^{-1}\sum_m\operatorname{Disp}_G(m)(v_{iid}(s^m) - v_\Delta(s^m)) \\
&= \sum_m\frac{v_{iid}(s^m)}{v_{iid}(s)}\operatorname{Disp}_G(m)\left(1-\frac{v_\Delta(s^m)}{v_{iid}(s^m)}\right) = \sum_m w_m\operatorname{Disp}_G(m)Q_k(s^m).
\end{aligned}$$

This proves Theorem 5.7. Finally, we prove an extra decomposition needed for Theorem 6.3. By Lemma C.1, $v_\Delta(s^m) = v_{iid}(s^m) - (k-1)^{-1}c(s^m)$ and by Theorem

5.3 we have $\mathrm{Disp}_G(m) = -(k-1)\lambda_m$, so we can write

$$\begin{aligned} n\,\mathrm{Var}_G(\widehat{\theta}) &= \sum_m v_{iid}(s^m) + \mathrm{Disp}_G(m) \cdot (v_\Delta(s^m) - v_{iid}(s^m)) \\ &= \sum_m v_{iid}(s^m) + \mathrm{Disp}_G(m) \cdot (-(k-1)^{-1}c(s^m)) \\ &= \sum_m v_{iid}(s^m) + \lambda_m c(s^m) = v_{iid}(s) + \sum_m \lambda_m \cdot c(s^m). \end{aligned}$$

This finishes the proof. □

LHS Coupling. Next, we consider Examples 5.5, 5.8, and 5.10. Let G be the LHS coupling for canonical marginal $F = \mathrm{Unif}[0,1]$. Define $J(l) = [(l-1)/k, l/k)$ for $l \in [k]$ and $I(d) = \sum_l J(l)\mathbb{1}(d \in J(l))$. Define the demeaned histogram space $E_{hist} = \{\phi \in L_0^2(F) : \phi(x) = \sum_l a_l \mathbb{1}(x \in J(l))\}$, constants $E_1 = \{\phi : \phi(x) = c\}$, and remainder $E_c = E_{hist}^{\perp}$ (orthogonal complement in $L_0^2(F)$). First, we show that Assumption 5.2 holds and characterize the projection operators.

Lemma B.6 (LHS)**.** *The operator $(U_G\phi)(d) = E[\phi(D)|D \notin I(d)]$. The direct sum $L_0^2(F) = E_{hist} \oplus E_{hist}^{\perp}$ holds, eigenspaces of U_G with eigenvalues $\lambda = -(k-1)^{-1}$, and 0, verifying Assumption 5.2. The projection is given by*

$$P_{hist}\phi(d) = E[\phi(D)|D \in I(d)] - E_F[\phi(D)].$$

Proof. First, we characterize the operator U_G. Recall that $D_i = k^{-1}(\pi_i - 1 + U_i)$, where π is a random permutation of $[k]$ and $U_i \sim \mathrm{Unif}[0,1]$, with $\pi \perp\!\!\!\perp U$. By tower law $E[\phi(D_2)|D_1] = E[E[\phi(D_2)|D_1, \pi]|D_1]$. Since (U_1, U_2, π) are jointly independent, we have $U_2 \perp\!\!\!\perp U_1|\pi$. Moreover, $D_1 \in \sigma(U_1, \pi)$ and $D_2 \in \sigma(U_2, \pi_2)$, so $D_1 \perp\!\!\!\perp D_2|\pi$. Then we conclude $E[\phi(D_2)|D_1, \pi] = E[\phi(D_2)|\pi]$. Next, we analyze $E[\phi(D_2)|\pi]$. The distribution $D_2|\pi \sim k^{-1}(\pi_2 - 1) + \mathrm{Unif}[0, k^{-1}]$. Then $E[\phi(D_2)|\pi] = m(\pi_2)$ for discrete function $m(l) \equiv k\int_{(l-1)/k}^{l/k} \phi(x)dx$. Then we have

$$\begin{aligned} E[\phi(D_2)|D_1] &= E[E[\phi(D_2)|\pi]|D_1] = E[m(\pi_2)|D_1] = E[E[m(\pi_2)|U_1, \pi_1]|D_1] \\ &= E[E[m(\pi_2)|\pi_1]|D_1] = E[m(\pi_2)|\pi_1]. \end{aligned}$$

The third equality by tower law, since $D_1 \in \sigma(U_1, \pi_1)$. The fourth equality since $\pi \perp\!\!\!\perp U$. The last equality since $\pi_1 = f(D_1)$ for $f(D_1) = \lfloor kD_1 \rfloor + 1$, so $\sigma(\pi_1) \subseteq \sigma(D_1)$. By calculations in Lemma B.9, $\pi_2|\pi_1$ is uniform on $[k] \setminus \{\pi_1\}$. Then the expectation of $m(\pi_2)$ given π_1 is

$$E[m(\pi_2)|\pi_1] = (k-1)^{-1} \sum_{l \neq \pi_1} m(l) = (k-1)^{-1} \sum_{l \neq \pi_1} k \int_{(l-1)/k}^{l/k} \phi(x)\,dx.$$

Note $\pi_1 = \lfloor kD_1 \rfloor + 1$. Set $I(d) = J(l)$ if $d \in J(l)$. So $(U_G\phi)(d) = E[\phi(D_2)|D_1 = d]$:

$$\begin{aligned} (U_G\phi)(d) &= (k-1)^{-1} \sum_{l \neq \lfloor kd \rfloor + 1} k \int_{(l-1)/k}^{l/k} \phi(t)dt = \frac{k}{k-1}\int_{I(d)^c} \phi(t)dt \\ &= E[\phi(D)|D \notin I(d)] = (k-1)^{-1}(k \cdot E_F[\phi] - E_F[\phi(D)|D \in I(d)]). \end{aligned}$$

The final equality follows by tower law. This finishes the characterization of U_G.

Next, we show the claimed decomposition $L_0^2(F) = E_{hist} \oplus E_{hist}^\perp$. For $\phi \in E_{hist}$ we have $E_F[\phi] = 0$ by definition and clearly $E_F[\phi(D)|D \in I(d)] = \phi(d)$ so $(U_G\phi)(d) = -(k-1)^{-1}\phi(d)$ by the calculation above. Then E_{hist} is an eigenspace with $\lambda = -(k-1)^{-1}$. Finally, note that we can write $(U_G\phi)(d) = k(k-1)^{-1}\langle\phi, t_d\rangle_F$ for $t_d(x) = \mathbb{1}(x \notin I(d))$. Since $t_d \in \text{span}(1) \oplus E_{hist}$, we have $\langle\phi, t_d\rangle_F = 0$ for $\phi \in E_{hist}^\perp$, noting the orthocomplement is taken within $L_0^2(F)$. Then $E_{hist}^\perp$ is an eigenspace with $\lambda = 0$. Note E_{hist} is closed as a finite-dimensional linear subspace, so $L_0^2(F) = E_{hist} \oplus E_{hist}^\perp$, verifying Assumption 5.2.

Finally, let $P\phi(x) = E_F[\phi(D)|D \in I(x)] - E_F[\phi(D)]$. We claim $P = P_{hist}$. Let $\phi \in E_{hist}$, then since ϕ is piecewise constant on $J(l)$, $E_F[\phi(D)|D \in I(x)] - E_F[\phi(D)] = \phi(x) - 0 = \phi(x)$, so $P\phi = \phi$. Also if $\phi = c$ then $P\phi = 0$. Third, note if $\phi \perp E_{hist}$ and $\phi \in L_0^2(F)$, we have $\phi \perp \mathbb{1}(D \in J(l))$ for each l, so $E_F[\phi(D)|D \in I(x)] = 0$, and $E_F[\phi(D)] = 0$, giving $P\phi = 0$. For any $\phi \in L^2(F) = \text{span}(1) \oplus E_{hist} \oplus E_{hist}^\perp$, we have $\phi = E_F[\phi] + P_{hist}\phi + (1 - P_{hist})\phi$. We have shown P is a linear operator that fixes E_{hist} and annihilates $E_{hist}^\perp \oplus \text{span}(1)$, so applying P to equation above gives $P\phi = PP_{hist}\phi = P_{hist}\phi$ for any $\phi \in L^2(F)$. This finishes the proof. □

Canonical Spaces. Next, we develop the machinery about canonical spaces needed for Lemma B.8. Suppose for some Q the space $L_0^2(Q) = \oplus_{m\geq 1} W_m$, where W_m are eigenspaces of the operator $(T_H\phi)(u) = E_H[\phi(U_i)|U_j = u]$ for $i \neq j$ with $H \in \Pi_k(Q)$, as in Assumption 5.2. Define the coupling $G \in \Pi_k(F)$ by $(D_i)_{i=1}^k \sim G$ for $D_i = v(U_i)$ for $(U_i)_{i=1}^k \sim H$ and some $v : \text{Supp}(Q) \to \text{Supp}(F)$. In particular, this implies $D = v(U) \sim F$ for $U \sim Q$. Define the linear pushforward operator $R : L^2(F) \to L^2(Q)$ by $(R\phi)(u) = (\phi \circ v)(u)$.

Lemma B.7 (Canonical Spaces)**.** *The coupling operator has $U_G = R^*T_HR$. If the map $v(\cdot)$ is injective with measurable inverse v^{-1} on $v(\text{Supp}(Q))$, then $R^* = R^{-1}$ and $L_0^2(F) = \oplus_{m\geq 1} E_m$ for eigenspaces $E_m = R^{-1}(W_m)$ of U_G. In particular, Assumption 5.2 is satisfied for G. Also the eigenvalues $\lambda(E_m) = \lambda(W_m)$ and projection operators $P_m^E = R^{-1}P_m^W R$.*

Proof of Lemma B.7. First, we show R is well-defined. To see this, $E_Q[(R\phi)(U)^2] = E_Q[\phi(v(U))^2] = E_F[\phi(D)^2] < \infty$ for any $\phi \in L^2(F)$. Next, claim $R : L^2(F) \to L^2(Q)$ is a linear isometry. Linearity is clear. For the isometry property,

$$\begin{aligned}\langle\phi, \psi\rangle_F &= E_F[\phi(D)\psi(D)] = E_Q[\phi(v(U))\psi(v(U))] \\ &= E_Q[(R\phi)(U)(R\psi)(U)] = \langle R\phi, R\psi\rangle_Q.\end{aligned}$$

In particular, $\|R\| = 1$, so R is bounded and the adjoint R^* exists by Riesz representation. Note R maps $L_0^2(F)$ into $L_0^2(Q)$, since $E_Q[(R\phi)(U)] = E_F[\phi(D)] = 0$ for $\phi \in L_0^2(F)$. Next, we show $\langle R^*T_HR\phi, \psi\rangle_F = \langle U_G\phi, \psi\rangle_F$ for $\phi, \psi \in L^2(F)$.

$$\begin{aligned}\langle U_G\phi, \psi\rangle &= E_F[U_G\phi(D)\psi(D)] = E_G[\phi(D_1)\psi(D_2)] = E_h[(R\phi)(U_1)(R\psi)(U_2)] \\ &= E_H[E[(R\phi)(U_1)|U_2](R\psi)(U_2)] = E_Q[T_H(R\phi)(U)(R\psi)(U)] \\ &= \langle T_HR\phi, R\psi\rangle_Q = \langle R^*T_HR\phi, \psi\rangle_F.\end{aligned}$$

Observe that if $\langle a - b, \psi\rangle = 0$ for any ψ, then $a = b$ by taking $\psi = a - b$. By this observation, we have shown $R^*T_HR\phi = U_G\phi$ for each $\phi \in L^2(F)$, so $R^*T_HR = U_G$, proving the first claim.

Next, impose injectivity of $v(\cdot)$. Note that $D = v(U)$, so $F(D \in \text{Image}(v)) = 1$. Define $S : L^2(Q) \to L^2(F)$ by $(S\phi)(d) = \phi(v^{-1}(d))$ for $d \in \text{Image}(v)$ and $(S\phi)(d) = 0$ for $d \notin \text{Image}(v)$. Note that $\text{Image}(v) = (v^{-1})^{-1}(\mathbb{R})$, which is measurable since v^{-1} is measurable by assumption. Then $S\phi$ is measurable. Moreover, $E_F[(S\phi)(D)^2] = E_F[\phi(v^{-1}(D))^2] = E_Q[\phi(U)^2] < \infty$ for any $\phi \in L^2(Q)$, showing S is well-defined. Then we have $RS\phi(u) = (S\phi)(v(u)) = \phi(v^{-1}(v(u))) = \phi(u)$. Then $RS = I$ on $L^2(Q)$. In particular, R is surjective and S is injective. Moreover, $R^*R = I$ on $L^2(F)$, so R is injective. Then $R^*RS = R^*$, so $S = R^*$. Then $SR = RS = I$, so $R^* = S = R^{-1}$. Clearly R^{-1} is also an isometry, hence continuous.

We will show $L^2_0(F) = \oplus_{m\geq 1} E_m$. Let $\phi \in E_m$. Then $T_H R\phi = \lambda_m R\phi$, so $R^{-1}T_H R\phi = \lambda_m R^{-1}R\phi = \lambda_m \phi$, so E_m is an eigenspace with eigenvalue λ_m. Note R^{-1} is continuous, since R^{-1} is a linear isometry, hence a bounded linear map. Each W_m is closed by assumption, so $E_m = R^{-1}(W_m)$ is closed by continuity of R^{-1}. Let $\phi \in L^2_0(F)$. Then by assumption $R\phi = \sum_{m\geq 1} P^W_m R\phi = \lim_{l\to\infty} \sum_{m\geq 1}^{l} P^W_m R\phi$. By linearity and continuity of R^{-1}, $\phi = \sum_{m\geq 1} R^{-1}P^W_m R\phi$. Since $P^W_m R\phi \in W_m$, then $R^{-1}P^W_m R\phi \in R^{-1}(W_m) = E_m$. Let $\phi \in E_m$ and $\psi \in E_l$ for $l \neq m$. Then $R\phi \in W_m$ and $R\psi \in W_l$, so by isometry $\langle \phi, \psi\rangle_F = \langle R\phi, R\psi\rangle_Q = 0$ since $W_m \perp W_l$ by assumption. This shows $E_m \perp E_l$ for $l \neq m$. Putting this all together, we have shown $L^2_0(F) = \oplus_{m\geq 1} E_m$ with the required properties. □

Lemma B.8. *Assumption 5.2 is satisfied for any $G \in \Pi_k(F)$ if $|\text{Supp}(F)| < \infty$. If F is continuous, then the assumption holds for the antithetic variates, Latin hypercube, rotation sampling, and Gaussian couplings in Section 3.3.*

Proof of Lemma B.8. Consider the first statement. We claim $U_G|_{L^2_0(F)}$ is compact. To see this, let $B \subseteq L^2_0(F)$ be a bounded set. We need to show $U_G B$ is relatively compact in $L^2_0(F)$. Let $\text{Supp}(F) = \{d_1, \ldots, d_M\}$. Then $L^2(F) = \text{span}\{\mathbb{1}(D = d_l) : l \in [M]\}$ is finite dimensional, so $L^2_0(F) \subseteq L^2(F)$ is also finite dimensional. Then the Heine-Borel theorem applies, so the closure $\bar{B}$ is compact. By Lemma B.5, U_G is a linear contraction, hence bounded and continuous. Then $U_G(B) \subseteq U_G(\bar{B})$ compact, so $U_G(B)$ is contained in a compact set, so it is relatively compact. Also by Lemma B.5, $U_G|_{L^2_0(F)}$ is self-adjoint on $L^2_0(F)$. By the spectral theorem for compact self-adjoint operators, $L^2_0(F) = \oplus_{m\geq 1} E_m$, verifying Assumption 5.2.

Next, consider the second statement. Lemmas B.12, B.10, B.6, and B.11 show that $L^2_0(Q) = \oplus_{m\geq 1} W_m$ for eigenspaces W_m of the operator T_H, where $H =$ antithetic variates, rotation sampling, Latin hypercube, and Gaussian respectively with canonical marginals $Q = \mathcal{N}(0, 1)$ in the Gaussian case and $Q = \text{Unif}[0, 1]$ otherwise. The distributions $(D_i)_{i=1}^k \sim G \in \Pi_k(F)$ for all of these couplings satisfy the conditions of Lemma B.7, with $D_i = v(U_i)$ for $(U_i)_{i=1}^k \sim H$ with $v = F^{-1}$ for the uniform cases and $v = F^{-1} \circ \Phi$ for the Gaussian coupling. We claim that if F is continuous, then in each case v is injective with measurable inverse on its image. For injectivity, note that $F \circ F^{-1} = I$ on $(0, 1)$ by Lemma C.6, so F^{-1} is injective with left-inverse F on its range. By continuity, the left-inverses $v^{-1} = F$ and $v^{-1} = \Phi^{-1} \circ F$ are both measurable. Then by Lemma B.7, Assumption 5.2 is satisfied for each coupling G. This finishes the proof. □

B.5 Proofs for Section 6

First, consider rotation sampling (RS). Recall that for $U \sim \text{Unif}[0,1]$ and π a random permutation of $\{1,\dots,k\}$ with $U \perp\!\!\!\perp \pi$, we set $D_i = U \oplus k^{-1}\pi_i$. Then $(D_i)_{i=1}^k \sim G_R$ is the RS coupling. We begin with a lemma characterizing the bivariate marginal (D_i, D_j) for $i \neq j$. Recall $a \oplus b \equiv a + b \pmod 1$ for $a, b \in \mathbb{R}$.

Lemma B.9. *Let $(D_i)_{i=1}^k \sim G_R$. Then $(D_1, D_2) \sim (V, V \oplus R)$, where $V \sim \text{Unif}[0,1]$ and $R \sim \text{Unif}\{1/k, \dots, (k-1)/k\}$ with $V \perp\!\!\!\perp R$.*

Proof. Observe that $D_2 = U \oplus k^{-1}\pi_2 = U \oplus k^{-1}\pi_1 \oplus k^{-1}(\pi_2 - \pi_1) = D_1 \oplus R$ with $R \equiv k^{-1}(\pi_2 - \pi_1) \pmod 1$. We claim that $D_1 \perp\!\!\!\perp R$ and $D_1 \sim \text{Unif}[0,1]$. Note that $U + a \pmod 1 \sim \text{Unif}[0,1]$ for any fixed $a \in \mathbb{R}$. Then $D_1|\pi = U \oplus k^{-1}\pi_1|\pi \sim \text{Unif}[0,1]$ since $U \perp\!\!\!\perp \pi$. Then $D_1 \sim \text{Unif}[0,1]$. This also shows $D_1 \perp\!\!\!\perp \pi$, so $D_1 \perp\!\!\!\perp R$ since $R \in \sigma(\pi)$, proving the claim. Then $(D_1, D_2) = (D_1, D_1 \oplus R) \sim (V, V \oplus R)$ with $V \perp\!\!\!\perp R$ and $V \sim \text{Unif}[0,1]$. Finally, we show the distribution of R. A calculation shows $\pi_2|\pi_1$ is uniform on $[k] \setminus \{\pi_1\}$. Define $f(\pi_2) = \pi_2 - \pi_1 \pmod k$. This is a bijection from $[k] \setminus \{\pi_1\}$ to $\{1, \dots, k-1\}$. Then $\pi_2 - \pi_1 \pmod k | \pi_1 \sim \text{Unif}\{1, \dots, k-1\}$. Then this also holds marginally, completing the proof. $\square$

Lemma B.10 (RS)**.** *For the RS coupling, $(U_G\phi)(d) = (k-1)^{-1}\sum_{l=1}^{k-1}\phi(d \oplus lk^{-1})$. Also, $L_0^2(F) = E_c \oplus E_a$ is a direct sum of eigenspaces of U_G with eigenvalues 1 on E_c and $-(k-1)^{-1}$ on E_a. In particular, Assumption 5.2 holds. Finally, the projection operators P_c and P_a on $L^2(F)$ have $(P_c\phi)(d) \equiv k^{-1}\sum_{l=1}^k \phi(d \oplus lk^{-1}) - E_F[\phi(D)]$ and $(P_a\phi)(d) = \phi(d) - k^{-1}\sum_{l=1}^k \phi(d \oplus lk^{-1})$.*

Proof. First, we establish the direct sum and projection formulas. For $\phi \in L^2(F)$, define $(P\phi)(d) \equiv k^{-1}\sum_{l=1}^k \phi(d \oplus lk^{-1}) - E_F[\phi(D)]$. We claim that $P_c = P$. Write $(S\phi)(d) \equiv k^{-1}\sum_{l=1}^k \phi(d \oplus lk^{-1})$, so that $P\phi = S\phi - E_F[\phi(D)]$. First, we will show that $P\phi$ is cyclic. For any $m \geq 1$, note

$$(S\phi)(d \oplus mk^{-1}) = k^{-1}\sum_{l\in[k]}\phi(d \oplus lk^{-1} \oplus mk^{-1}) = k^{-1}\sum_{l\in[k]}\phi(d \oplus lk^{-1}) = (S\phi)(d).$$

Then $S\phi$ is cyclic, so $P\phi = S\phi - E_F[\phi(D)]$ is as well. Next, we show $P\phi$ is mean-zero. Note $E_F[(S\phi)(D)] = k^{-1}\sum_{l=1}^k E_F[\phi(D \oplus lk^{-1})] = E_F[\phi(D)]$, since $D \oplus lk^{-1} \sim \text{Unif}[0,1]$ whenever $D \sim \text{Unif}[0,1]$. Then $E_F[(P\phi)(D)] = E_F[(S\phi)(D)] - E_F[\phi(D)] = 0$. Then $P\phi \in E_c$ for any $\phi \in L^2(F)$. Next, we show idempotency, $P^2 = P$. Note if $\phi \in E_c$, then $\phi(d \oplus lk^{-1}) = \phi(d)$ for all l, so $(S\phi)(d) = \phi(d)$. Also $E_F[\phi(D)] = 0$ since $\phi \in E_c \subseteq L_0^2(F)$. Then $(P\phi)(d) = \phi(d) - 0 = \phi(d)$. Then for any $\phi \in L^2(F)$, $P\phi \in E_c$, so $P(P\phi) = P\phi$, showing $P^2 = P$. This also implies $\text{Im}(P) = E_c$. Next, we show P is self-adjoint. For any $\phi, \psi \in L^2(F)$, we have $\langle P\phi, \psi\rangle_F = \langle S\phi, \psi\rangle_F - E_F[\phi(D)]\langle 1, \psi\rangle_F = \langle S\phi, \psi\rangle_F - E_F[\phi(D)]E_F[\psi(D)]$. Similarly, $\langle \phi, P\psi\rangle_F = \langle \phi, S\psi\rangle_F - E_F[\psi(D)]E_F[\phi(D)]$, so it suffices to show S is self-adjoint.

This follows since for any $\phi, \psi \in L^2(F)$, writing $a \ominus b = a - b \pmod 1$,

$$\begin{aligned}
\langle S\phi, \psi \rangle_F &= \int_0^1 (S\phi)(x)\psi(x)dx = \int_0^1 k^{-1} \sum_{l=1}^k \phi(x \oplus lk^{-1})\psi(x)dx \\
&= k^{-1} \sum_{l=1}^k \int_0^1 \phi(u)\psi(u \ominus lk^{-1})du = \int_0^1 \phi(u) k^{-1} \sum_{l=1}^k \psi(u \ominus lk^{-1})du \\
&= \int_0^1 \phi(u)(S\psi)(u)du = \langle \phi, S\psi \rangle_F.
\end{aligned}$$

Then $\langle P\phi, \psi \rangle_F = \langle \phi, P\psi \rangle_F$, so P is self-adjoint on $L^2(F)$. Since P is a self-adjoint idempotent on $L^2(F)$ with image E_c, it is the orthogonal projection onto E_c in $L^2(F)$. Then E_c is closed in $L^2(F)$, hence also closed in $L_0^2(F)$. Then the projection theorem for Hilbert spaces gives $L_0^2(F) = E_c \oplus E_a$, where $E_a = E_c^\perp$ is the orthocomplement within $L_0^2(F)$. The projection formula $P_c = P$ holds on $L^2(F)$. Note also that P annihilates E_a, since for $\phi \in E_a \subseteq L_0^2(F)$, $P\phi = S\phi - 0$, and $S\phi \in E_c$ (since $P\phi \in E_c$ and $E_F[\phi(D)] = 0$), so $\langle S\phi, S\phi \rangle_F = \langle S^2\phi, \phi \rangle_F = \langle S\phi, \phi \rangle_F = 0$ since $\phi \in E_a \perp S\phi$. Then $S\phi = P\phi = 0$ on E_a. We claim $I - S = P_a$. To see this, note $\phi = E_F[\phi] + P_c\phi + P_a\phi$. It's clear $I - S$ annihilates $\mathrm{span}(1) \oplus E_c$. Then it suffices to show $I - S$ fixes E_a. For $\phi \in E_a$, we showed $S\phi = 0$ above, so $(I - S)\phi = \phi$. This finishes our characterization of the projections.

Eigenspaces. Next, we characterize the eigenspaces of U_G. Denote $U_G = U$. Lemma B.9 showed that bivariate marginal $(D_1, D_2) \sim (V, V \oplus R)$, with $V \sim \mathrm{Unif}[0,1]$, $R \sim \mathrm{Unif}\{1/k, \ldots, (k-1)/k\}$ and $V \perp\!\!\!\perp R$. Then

$$(U\phi)(d) = E[\phi(V \oplus R) | V = d] = E[\phi(d \oplus R)] = (k-1)^{-1} \sum_{l \in [k-1]} \phi(d \oplus lk^{-1}).$$

The second equality since $V \perp\!\!\!\perp R$. This proves the first claim. Since $\phi(d \oplus mk^{-1}) = \phi(d)$ for any $m \geq 1$ and $\phi \in E_c$, then $U\phi(d) = (k-1)^{-1} \sum_{l=1}^{k-1} \phi(d) = \phi(d)$ and E_c is an eigenspace of U with eigenvalue 1. Next, note the identity

$$\begin{aligned}
(U\phi)(d) &= (k-1)^{-1} \sum_{l \in [k-1]} \phi(d \oplus lk^{-1}) = (k-1)^{-1} \big(\sum_{l \in [k]} \phi(d \oplus lk^{-1}) - \phi(d) \big) \\
&= (k-1)^{-1} (k(S\phi)(d) - \phi(d)).
\end{aligned}$$

If $\phi \in E_a$, we showed $S\phi = 0$ above, so $U\phi = -(k-1)^{-1}\phi$, showing that E_a is an eigenspace with $\lambda = -(k-1)^{-1}$. This completes the proof. $\square$

Proof of Theorem 6.2, Equation 6.4. Relative efficiency follows from Lemma B.10 and Theorem 5.7. For the marginal variance, by Corollary 5.9 and Lemma C.1,

$$\begin{aligned}
n \operatorname{Var}_G(\widehat{\theta}) &= (1 - \mathrm{Disp}_G(s^c)) v_{iid}(s^c) + \mathrm{Disp}_G(s^c) v_\Delta(s^c) \\
&\quad + (1 - \mathrm{Disp}_G(s^a)) v_{iid}(s^a) + \mathrm{Disp}_G(s^a) v_\Delta(s^a) \\
&= k \cdot v_{iid}(s^c) - (k-1) v_\Delta(s^c) + v_\Delta(s^a) \\
&= k \cdot v_{iid}(s^c) - (k-1)(v_{iid}(s^c) - (k-1)^{-1} c(s^c)) + v_\Delta(s^a) \\
&= v_{iid}(s^c) + c(s^c) + v_\Delta(s^a) = k \cdot v_g(s^c) + v_\Delta(s^a).
\end{aligned}$$

$\square$

Lemma B.11 (Gaussian)**.** *Let $F = \mathcal{N}(0,1)$ and G the Gaussian coupling with correlation $\rho = -(k-1)^{-1}$ and $k \geq 3$. Write $U_G = U$ Let $(h_m)_{m\geq 0}$ be the normalized probabilist's Hermite polynomials. Then $L^2(F) = \oplus_{m\geq 0} \operatorname{span}(h_m)$ with $Uh_m = \lambda_m$ for $\lambda_m = (-1)^m(k-1)^{-m}$. Restricting to $L_0^2(F)$ gives $L_0^2(F) = \oplus_{m\geq 1} \operatorname{span}(h_m)$, verifying Assumption 5.2.*

Proof. Denote $\rho = -(k-1)^{-1}$. Then $D_2|D_1 = x \sim \mathcal{N}(\rho x, (1-\rho^2))$, so the operator

$$
\begin{aligned}
U\phi(x) = E[\phi(D_2)|D_1 = x] &= \int_{\mathbb{R}} \phi(y) \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) dy \\
&= \int_{\mathbb{R}} \phi(y) \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2(x^2+y^2) - 2\rho xy}{2(1-\rho^2)}\right) dF(y) \equiv \int_{\mathbb{R}} \phi(y) K(y,x) dF(y).
\end{aligned}
$$

The third equality follows by algebra, using $dF(y) = (2\pi)^{-1/2}e^{-y^2/2}$. The function $K(x,y)$ is the Mehler kernel, with $K(x,y) = \sum_{m=0}^{\infty} \rho^m h_m(x)h_m(y)$. For example, see Thangavelu (1993). Define the finite sum $K_M = \sum_{m=0}^{M} \rho^m h_m(x)h_m(y)$ and $U_M\phi(x) = E_F[K_M(x,Y)\phi(Y)]$. For $\phi \in L_2(F)$, we have $|U_M\phi - U\phi| \leq |U_M - U|_{op}|\phi|_F$. Moreover, since $U_M - U$ is a Hilbert-Schmidt operator with kernel $K_M - K$, we have $|U_M - U|_{op} \leq |K_M - K|_{L_2(F\otimes F)}$, where $(K - K_M)(x,y) = \sum_{m=M+1}^{\infty} \rho^m h_m(x)h_m(y)$. Note that $e_m(x,y) = h_m(x)h_m(y)$ is an ON collection in $L_2(F \otimes F)$. Moreover, $K = \sum_m \rho^m e_m$ with $\sum_m (\rho^m)^2 < \infty$ since $\rho < 1$ for $k \geq 3$. Then by Riesz-Fischer, $K_M \to K$ in $L_2(F \otimes F)$, so that $|U_M - U|_{op} \to 0$. Clearly $U_M h_m = \rho^m h_m$ for $m \leq M$ by orthonormality, so $Uh_m = \lim_M U_M h_m = \lim_M \rho^m h_m = \rho^m h_m$. Then we have shown $L_2(F) = \oplus_{m\geq 0} E_m$ for $E_m = \operatorname{span}(h_m)$ with eigenvalue $\lambda_m = \rho^m = (-1)^m(k-1)^{-m}$. Restricting to $L_0^2(F)$ gives $L_0^2(F) = \oplus_{m\geq 1} E_m$, since $E_0 = \operatorname{span}(h_0) = \{\text{constants}\}$. This completes the proof □

Proof of Theorem 6.3. From the proof of Theorem 5.7, we have the variance identity $n \operatorname{Var}_G(\widehat{\theta}) = v_{iid}(s) + \sum_{m\geq 1} \lambda_m c(s^m)$. By Lemma B.11, the eigenspaces correspond to Hermite polynomials with eigenvalues $\lambda_m = (-1)^m(k-1)^{-m}$. By Lemma C.1, $c(s^m) = (k-1)v_{iid}(s^m)Q_k(s^m)$. Substituting these into the variance expression yields

$$
n \operatorname{Var}_G(\widehat{\theta}) = v_{iid}(s) + \sum_{m\geq 1} (-1)^m (k-1)^{-(m-1)} v_{iid}(s^m) Q_k(s^m).
$$

Then dividing by $n \operatorname{Var}_{G_{iid}}(\widehat{\theta}) = v_{iid}(s)$ and using weights $w_m = v_{iid}(s^m)/v_{iid}(s)$, we have

$$
\frac{\operatorname{Var}_G(\widehat{\theta})}{\operatorname{Var}_{G_{iid}}(\widehat{\theta})} = 1 + \sum_{m\geq 1} (-1)^m (k-1)^{-(m-1)} w_m Q_k(s^m).
$$

Rearranging for relative efficiency $\operatorname{Efficiency}(G) = 1 - \operatorname{Var}_G(\widehat{\theta}) / \operatorname{Var}_{G_{iid}}(\widehat{\theta})$ gives the result, noting $-(-1)^m = (-1)^{m-1}$. □

Proof of Corollary 6.4. We show $n \operatorname{Var}_G(\widehat{\theta}) = v_\Delta(s^L) + v_{iid}(s - s^L) + R_k(s)$ with remainder $|R_k(s)| \leq (k-1)^{-1} v_{iid}(s - s^L)$. Write $s_i^m = \langle s_i, h_m \rangle_F \cdot h_m$ for the projection onto h_m. By the proof of Theorem 5.7, $n \operatorname{Var}_G(\widehat{\theta}) = \sum_m v_{iid}(s^m) + \lambda_m c(s^m)$. Then by Lemma C.1, since $v_\Delta(s) = v_{iid}(s) - (k-1)^{-1}c(s)$ we calculate

$$
\begin{aligned}
n \operatorname{Var}_G(\widehat{\theta}) &= v_\Delta(s^1) + \sum_{m\geq 2} (v_{iid}(s^m) + \lambda_m c(s^m)) \\
&= v_\Delta(s^1) + v_{iid}(s - s^1) + \sum_{m\geq 2} (-1)^m (k-1)^{-m} c(s^m).
\end{aligned}
$$

The second equality is from Lemma C.2. Define $R_k(s) \equiv \sum_{m\geq 2}(-1)^m(k-1)^{-m}c(s^m)$. For the first two terms, note that $s_i^1(D) = \langle h_1, s_i\rangle_F \cdot h_1(D) = \langle D, s_i\rangle_F \cdot D = \mathrm{Cov}_F(D, s_i(D)) \cdot D$ since $E_F[D] = 0$. We have $s_i^L(D) = a_i^* + b_i^* D$, where

$$b_i^* = \operatorname*{argmin}_b \min_a \mathrm{Var}_F(s_i(D) - a - bD) = \operatorname*{argmin}_b \mathrm{Var}_F(s_i(D) - bD).$$

The solution is $b_i^* = \mathrm{Cov}_F(s_i(D), D) = \langle h_1, s_i\rangle_F$. Then we have shown $s_i^L = a_i^* + s_i^1$, so $v_\Delta(s^1) = v_\Delta(s^L)$ and $v_{iid}(s - s^1) = v_{iid}(s - s^L)$. Summarizing, we showed that $n \mathrm{Var}(\widehat{\theta}) = v_\Delta(s^L) + v_{iid}(s - s^L) + R_k(s)$. Next, we show the bound on $R_k(s)$. We have $|c(s^j)| \leq (k-1)v_{iid}(s^j)$. To see this, recall $c(s^m) = n^{-1}\sum_g \sum_{i\neq j} \mathrm{Cov}_F(s_{ig}^m, s_{jg}^m)$. By the triangle inequality, Cauchy-Schwarz and Young's inequality

$$\begin{aligned}|c(s^m)| &\leq n^{-1}\sum_g\sum_{i\neq j}|\mathrm{Cov}_F(s_{ig}^m, s_{jg}^m)| \leq n^{-1}\sum_g\sum_{i\neq j}(1/2)(\mathrm{Var}_F(s_{ig}^m) + \mathrm{Var}_F(s_{jg}^m))\\ &= n^{-1}(k-1)\sum_g\sum_{i\in[k]}\mathrm{Var}_F(s_{ig}^m) = (k-1)E_n \mathrm{Var}_F(s_i^m) = (k-1)\cdot v_{iid}(s^m).\end{aligned}$$

Then by the triangle inequality, we have

$$\begin{aligned}|R_k(s)| &\leq \sum_{m=2}^{\infty}(k-1)^{-m}|c(s^m)| \leq \sum_{m=2}^{\infty}(k-1)^{-m}(k-1)\cdot v_{iid}(s^m)\\ &= (k-1)^{-1}\sum_{m=2}^{\infty}(k-1)^{-(m-2)}v_{iid}(s^m) \leq (k-1)^{-1}\sum_{m=2}^{\infty}v_{iid}(s^m)\\ &= (k-1)^{-1}v_{iid}(s - s^L).\end{aligned}$$

The last equality $v_{iid}(s - s^L) = \sum_{m=2}^{\infty} v_{iid}(s^m)$ follows since

$$\begin{aligned}v_{iid}(s - s^L) &= E_n \mathrm{Var}_F\left(\sum_{m=2}^{\infty} s_i^m\right) = E_n\left|\sum_{m=2}^{\infty} s_i^m\right|_F^2\\ &= E_n\sum_{m=2}^{\infty}|s_i^m|_F^2 = \sum_{m=2}^{\infty}E_n|s_i^m|_F^2 = \sum_{m=2}^{\infty}v_{iid}(s^m).\end{aligned}$$

The third equality is Parseval's theorem. The second to last equality is Tonelli's Theorem. This shows the claimed bound. □

Lemma B.12 (Antithetic Variates). *Let G be the antithetic variates coupling. The coupling operator $U_G\phi(x) = E_G[\phi(D_1)|D_2 = x] = \phi(1-x)$. Also $L_0^2(F) = E_{even} \oplus E_{odd}$, which are eigenspaces of U_G with eigenvalues 1 on E_{even} and -1 on E_{odd}. In particular, Assumption 5.2 holds. Define $S_e\phi \equiv (1/2)(\phi(x) + \phi(1-x)) - E_F[\phi(D)]$ and $S_o\phi \equiv (1/2)(\phi(x) - \phi(1-x))$. Then $P_{even}\phi = S_e\phi$ and $P_{odd}\phi = S_o\phi$.*

Proof. Let $\phi \in L^2(F)$, then clearly $\phi = S_e\phi + S_o\phi + E_F[\phi(D)]$ with $S_e\phi \in E_{even}$ and $S_o\phi \in E_{odd}$ and $E_F[\phi(D)] \in \mathrm{span}(1)$. If $\phi \in E_{even}$ then $S_e\phi = \phi$. If $\phi \in E_{odd}$ then $E_F[\phi(D)] = -E_F[\phi(1-D)]$, so $E_F[(S_e\phi)(D)] = 0$. Then $S_e\phi = (1/2)(\phi(x) + \phi(1-x)) = (1/2)(-\phi(1-x) + \phi(1-x)) = 0$. Then for any $\phi \in L^2(F)$ and $\psi \in E_{even}$,

$$\begin{aligned}\langle \phi - S_e\phi, \psi\rangle_F &= \langle S_o\phi + E_F[\phi(D)], \psi\rangle_F = E_F[(S_o\phi)(D)\psi(D)]\\ &= -E_F[(S_o\phi)(1-D)\psi(1-D)] = -\langle S_o\phi, \psi\rangle_F\end{aligned}$$

Then we must have $\langle S_o\phi, \psi\rangle_F = 0$. The 2nd equality above follows since $\psi \in E_{even} \perp \text{span}(1)$. The 3rd equality using $S_o\phi \in E_{odd}$ and $\psi \in E_{even}$. The final equality since $D \sim 1 - D$. This shows $\phi - S_e\phi \perp E_{even}$, so $S_e\phi = P_{even}\phi$. A similar argument shows $S_o\phi = P_{odd}\phi$. This finishes our proof of the projection operators. Next we show the direct sum. By the above work, if $\phi \in L_0^2(F)$, then $\phi = S_e\phi + S_o\phi$, so $L_0^2(F) = E_{even} + E_{odd}$. Orthogonality follows by the same argument in the display above, so $L_0^2(F) = E_{even} \oplus E_{odd}$. Finally, we characterize the operator $U_G = U$. We have $(D_1, D_2) = (Z, 1 - Z)$ for $Z \sim \text{Unif}[0, 1]$. Then $U\phi(x) = E_G[\phi(D_1)|D_2 = x] = E[\phi(Z)|1-Z = x] = \phi(1-x)$. For any $\phi \in E_{even}$, we have $U\phi(x) = \phi(1-x) = \phi(x)$, so $\lambda = 1$. For any $\phi \in E_{odd}$, we have $U\phi(x) = \phi(1 - x) = -\phi(x)$, so $\lambda = -1$. This finishes the proof. $\square$

Proof of Theorem A.2. By Lemma B.12, $L_0^2(F) = E_{even} \oplus E_{odd}$, a direct sum of eigenspaces of U_G with dispersions $\text{Disp}_G(\phi) = 1$ and $\text{Disp}_G(\phi) = -1$ respectively. Then the claim in Equation A.5 follows directly from Theorem 5.7. Moreover, by Corollary 5.9 and the variance identities in Lemma C.1, we have

$$\begin{aligned} n\,\text{Var}_G(\widehat{\theta}) &= 2v_{iid}(s^e) - v_\Delta(s^e) + v_\Delta(s^o) = 2v_{iid}(s^e) - [v_{iid}(s^e) - c(s^e)] + v_\Delta(s^o) \\ &= v_{iid}(s^e) + c(s^e) + v_\Delta(s^o) = 2v_g(s^e) + v_\Delta(s^o). \end{aligned}$$

This finishes the proof. $\square$

Lemma B.13 (RS Uniform Bound)**.** *Let $F = \text{Unif}[0, 1]$ and E_c be the $1/k$-cyclic subspace. Then for any $(s_i(\cdot))_{i=1}^n \subseteq L^2(F)$ we have $k \cdot v_g(s^c) \le k^{-1}E_nV_{[0,1]}(s_i)^2$.*

Proof. Our strategy will be to show $s_i^c(x) = k^{-1}\sum_{l\in[k]} s_i(x\oplus l/k) \approx E_F[s_i]+O(k^{-1})$. Define sets $E_l = [(l - 1)/k, l/k) \subseteq [0, 1]$ and $\mathcal{S}(x) = \{x \oplus l/k : l \in [k]\}$. Clearly $|\mathcal{S}(x) \cap E_l| = 1$ for $l \in [k]$. Let $t_l(x)$ denote this point and define the histogram approximation $f_i(x, t) = \sum_{l\in[k]} \mathbb{1}(t \in E_l) \cdot s_i(t_l(x))$. Then we have

$$\int_0^1 f_i(x, t)dt = k^{-1} \sum_{l\in[k]} s_i(t_l(x)) = k^{-1} \sum_{l\in[k]} s_i(x \oplus l/k)$$

The discrepancy $\delta s_i(x) \equiv s_i^c(x) - E_F[s_i]$ may be written

$$\delta s_i(x) = s_i^c(x) - E_F[s_i] = k^{-1} \sum_{l\in[k]} s_i(x \oplus l/k) - E_F[s_i] = \int_0^1 [f_i(x, t) - s_i(t)]dt$$

Note $|f_i(x, t) - s_i(t)| \le \sum_{l\in[k]} \mathbb{1}(t \in E_l)|s_i(t_l(x)) - s_i(t)| \le \sum_{l\in[k]} \mathbb{1}(t \in E_l)V_{E_l}(s_i)$. The last inequality follows since $t_l(x), t \in E_l$ on if $\mathbb{1}(t \in E_l) = 1$. Then above

$$\begin{aligned} |\delta s_i(x)| &\le \int_0^1 |f_i(x, t) - s_i(t)|dt \le \int_0^1 \sum_{l\in[k]} \mathbb{1}(t \in E_l)V_{E_l}(s_i)dt \\ &= k^{-1} \sum_{l\in[k]} V_{E_l}(s_i) = k^{-1}V_{[0,1]}(s_i). \end{aligned}$$

In particular, $\sup_{x\in[0,1]} |\delta s_i(x)| \le k^{-1}V_{[0,1]}(s_i)$. Then the variance

$$\text{Var}_F(s_i^c) = E_F[\delta s_i(D)^2] \le \sup_x |\delta s_i(x)|^2 \le k^{-2}V(s_i)^2.$$

By Lemma C.1, $k \cdot v_g(s^c) \le k \cdot v_{iid}(s^c)$. Then $kv_g(s^c) \le kv_{iid}(s^c) = kE_n\,\text{Var}_F(s_i^c) \le k^{-1}E_nV_{[0,1]}(s_i)^2$. This finishes the proof. $\square$

Lemma B.14 (LHS Uniform Bound). *Let $F = \mathrm{Unif}[0,1]$ and let E_{hist} be the k-histogram subspace on bins $E_l = [(l-1)/k, l/k)$ for $l \in [k]$. Let $s_i^{hist} \equiv P_{hist} s_i$ be the $L^2(F)$ projection in Equation (5.5). Then for any $(s_i(\cdot))_{i=1}^n \subseteq L^2(F)$,*

$$v_{iid}(s - s^{hist}) \le k^{-1} E_n V_{[0,1]}(s_i)^2.$$

Proof. Fix a unit i and write $s_i = s_i(\cdot)$. For each bin E_l, define the within-bin mean $m_l \equiv E_F[s_i(D) \mid D \in E_l] = k \int_{E_l} s_i(t)dt$. By Equation (5.5), we have $s_i^{hist}(t) = \sum_{l \in [k]} m_l \mathbb{1}(t \in E_l)$. Define residual $r_i(t) \equiv s_i(t) - s_i^{hist}(t)$. Clearly $E_F[r_i(D)] = 0$ so $\mathrm{Var}_F(r_i) = E_F[r_i(D)^2] = \int_0^1 r_i(t)^2 dt$. Next, we calculate for $t \in E_l$

$$|r_i(t)| = |s_i(t) - m_l| \le k \int_{E_l} |s_i(t) - s_i(t')|dt' \le \sup_{u,v \in E_l} |s_i(u) - s_i(v)| \le V_{E_l}(s_i).$$

Then, $\int_{E_l} r_i(t)^2 dt \le \int_{E_l} V_{E_l}(s_i)^2 dt = k^{-1} V_{E_l}(s_i)^2$. Summing over bins gives

$$\mathrm{Var}_F(s_i - s_i^{hist}) = \int_0^1 r_i(t)^2 dt \le k^{-1} \sum_{l \in [k]} V_{E_l}(s_i)^2 \le k^{-1} \Big(\sum_{l \in [k]} V_{E_l}(s_i) \Big)^2 = k^{-1} V_{[0,1]}(s_i)^2.$$

The second inequality since $(\sum_l a_l)^2 \ge \sum_l a_l^2$ for $a_l \ge 0$, and the final equality using $\sum_l V_{E_l}(s_i) = V_{[0,1]}(s_i)$. Finally, $v_{iid}(s - s^{hist}) = E_n \mathrm{Var}_F(s_i - s_i^{hist}) \le k^{-1} E_n V_{[0,1]}(s_i)^2$. This finishes the proof. □

Proof of Theorem 6.6. For the first claim, consider $G = \mathrm{RS}$. By the dispersion formula in Section 6.1, letting $w_c = \mathrm{Var}_F(P_c\phi)/\mathrm{Var}_F(\phi)$, we have

$$\mathrm{Disp}_G(\phi) = w_a - (k-1)w_c = (1 - w_c) - (k-1)w_c = 1 - kw_c.$$

By Lemma B.13, we have $k \mathrm{Var}_F(P_c\phi) = ks_g(P_c\phi) \le k^{-1} V_{[0,1]}(\phi)^2$, where $V_{[0,1]}(\phi)$ is the total variation of ϕ on $[0,1]$. Then $\sup_{\phi \in \mathcal{H}(b,\epsilon)} kw_c(\phi) \le b^2 k/\epsilon = o(1)$, so that $\inf_{\phi \in \mathcal{H}(b,\epsilon)} \mathrm{Disp}_G(\phi) = 1 + o(1)$, as claimed. Next, consider $G = \mathrm{LHS}$. By Equation 5.4, we have $\mathrm{Disp}_G(\phi) = w_h = \mathrm{Var}_F(P_{hist}\phi)/\mathrm{Var}_F(\phi) = 1 - \mathrm{Var}_F(\phi - P_{hist}\phi)/\mathrm{Var}_F(\phi)$. By Lemma B.14, $\mathrm{Var}_F(\phi - P_{hist}\phi) \le k^{-1} V_{[0,1]}(\phi)^2$. The conclusion follows. Finally, consider the statement about the Gaussian coupling. Let $n = k$ and $s_i = \phi$ for all $i \in [k]$. Then match quality $Q_k(\phi^L) = 1$ and weights $w_L = \mathrm{Var}_F(P_L\phi)/\mathrm{Var}_F(\phi)$. By Equation 4.2 and Corollary 6.4, we have

$$\mathrm{Disp}_G(\phi) = 1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} = w_L \cdot 1 + O(k^{-1}) = \frac{\mathrm{Var}_F(P_L\phi)}{\mathrm{Var}_F(\phi)} + o(1).$$

This proves the second claim. □

Proof of Theorem 6.7. First consider $G = \mathrm{LHS}$. By Equation 5.11, $n \mathrm{Var}_G(\widehat{\theta}) = v_\Delta(s^{hist}) + v_{iid}(s - s^{hist})$. Using the orthogonal decomposition $L_0^2(F) = E_{hist} \oplus E_{hist}^\perp$ and Lemma C.2, we have $v_\Delta(s^{hist}) = v_\Delta(s) - v_\Delta(s - s^{hist})$, so

$$n \mathrm{Var}_G(\widehat{\theta}) = v_\Delta(s) + v_{iid}(s - s^{hist}) - v_\Delta(s - s^{hist}).$$

Since $v_\Delta(\cdot) \ge 0$ by Lemma C.1, we obtain $n \mathrm{Var}_G(\widehat{\theta}) \le v_\Delta(s) + v_{iid}(s - s^{hist})$. By Lemma B.14, $v_{iid}(s - s^{hist}) \le k^{-1} E_n[V_{[0,1]}(s_i)^2]$. Dividing by $n \mathrm{Var}_{G_{iid}}(\widehat{\theta}) = v_{iid}(s)$ and using $Q_k(s) = 1 - v_\Delta(s)/v_{iid}(s)$:

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} \ge Q_k(s) - \frac{E_n[V_{[0,1]}(s_i)^2]}{k \cdot v_{iid}(s)} = Q_k(s) - \frac{\eta_{\mathrm{TV}}(s)}{k}.$$

Next consider $G = \text{RS}$. By Equation 6.4, $n \operatorname{Var}_G(\widehat{\theta}) = v_\Delta(s^a) + k \cdot v_g(s^c)$. Using the decomposition $L_0^2(F) = E_c \oplus E_a$ and Lemma C.2, $v_\Delta(s^a) = v_\Delta(s) - v_\Delta(s^c)$, so

$$n \operatorname{Var}_G(\widehat{\theta}) = v_\Delta(s) + k \cdot v_g(s^c) - v_\Delta(s^c).$$

Since $v_\Delta(s^c) \geq 0$, we obtain $n \operatorname{Var}_G(\widehat{\theta}) \leq v_\Delta(s) + k \cdot v_g(s^c)$. By Lemma C.1, $k \cdot v_g(s^c) \leq k \cdot v_{iid}(s^c)$, and by Lemma B.13, $k \cdot v_{iid}(s^c) \leq k^{-1} E_n[V_{[0,1]}(s_i)^2]$. Dividing by $v_{iid}(s)$ as before gives $1 - \operatorname{Var}_G(\widehat{\theta}) / \operatorname{Var}_{G_{iid}}(\widehat{\theta}) \geq Q_k(s) - \eta_{\text{TV}}(s)/k$. □

B.6 Proofs of Uniform Consistency Results

Proof of Theorem 7.1. First, we show the main result in Equation 7.2. We begin with an upper bound on the worst case variance, then show it is attained for some functions s_i. For eigenspace E_m of U_G, let $s^m = (P_m s_i)_{i=1}^n$ be the orthogonal projections and λ_m the eigenvalue. From the proof of Theorem 5.7,

$$n \operatorname{Var}_G(\widehat{\theta}) = v_{iid}(s) + \sum_{m \geq 1} \lambda_m \cdot c(s^m). \tag{B.4}$$

By Lemma C.1, $-v_{iid}(s^m) \leq c(s^m) \leq (k-1) v_{iid}(s^m)$ for each m. If $\lambda_m \geq 0$, then $\lambda_m c(s^m) \leq (k-1)\lambda_m v_{iid}(s^m)$. If $\lambda_m < 0$, then $\lambda_m c(s^m) \leq -\lambda_m v_{iid}(s^m)$. Recall from Theorem 5.3 that $\operatorname{Disp}_G(m) = -(k-1)\lambda_m$. Putting this together, we have

$$\lambda_m \cdot c(s^m) \leq \max(-\operatorname{Disp}_G(m), \operatorname{Disp}_G(m)/(k-1)) \cdot v_{iid}(s^m).$$

Then from Equation (B.4) and using $v_{iid}(s) = \sum_m v_{iid}(s^m)$ from Lemma C.2:

$$n \operatorname{Var}_G(\widehat{\theta}) \leq \sum_m [1 + \max(-\operatorname{Disp}_G(m), \operatorname{Disp}_G(m)/(k-1))] \cdot v_{iid}(s^m).$$

Next, we claim that $\inf_{m \geq 1} \operatorname{Disp}_G(m) = \text{ID}_G$ and $\sup_{m \geq 1} \operatorname{Disp}_G(m) = \text{SD}_G$. Write $\phi \neq c$ for ϕ non-constant. By Theorem 5.3, for $\sum_{m \geq 1} w_m = 1$, we have $\operatorname{Disp}_G(\phi) = \sum_{m \geq 1} w_m \operatorname{Disp}_G(m) \geq \inf_m \operatorname{Disp}_G(m)$. Then $\inf_{\phi \neq c} \operatorname{Disp}_G(\phi) \geq \inf_m \operatorname{Disp}_G(m)$. Conversely, picking $\phi_m \in E_m$ non-constant for $m \geq 1$, we have $\inf_{\phi \neq c} \operatorname{Disp}_G(\phi) \leq \inf_m \operatorname{Disp}_G(\phi_m) = \inf_m \operatorname{Disp}_G(m)$. This proves the claim for the inf, and the sup follows by symmetry. Using this, we have $-\operatorname{Disp}_G(m) \leq -\inf_m \operatorname{Disp}_G(m) = -\text{ID}_G$ and $\operatorname{Disp}_G(m) \leq \sup_m \operatorname{Disp}_G(m) = \text{SD}_G$, so the above is

$$\begin{aligned} n \operatorname{Var}_G(\widehat{\theta}) &\leq \sum_m [1 + \max(-\text{ID}_G, \text{SD}_G/(k-1))] \cdot v_{iid}(s^m) \\ &= [1 + \max(-\text{ID}_G, \text{SD}_G/(k-1))] \cdot v_{iid}(s). \end{aligned}$$

This establishes the upper bound.

Next we show that the bound is achieved. *Case 1:* Suppose $-\text{ID}_G \geq \text{SD}_G/(k-1)$. Choose l with $\operatorname{Disp}_G(l) < \text{ID}_G + \epsilon$. Pick $\phi \in E_l$ with $\operatorname{Var}_F(\phi) = 1$ and set $s_i = \phi$ for all $i \in [n]$. Then $v_{iid}(s) = 1$ and, since all influence functions are identical, we have $c(s^l) = n^{-1} \sum_g \sum_{i \neq j} \operatorname{Cov}_F(\phi, \phi) = (k-1) \operatorname{Var}_F(\phi) = (k-1)$ and $c(s^m) = 0$ for $m \neq l$ by Lemma C.2. Since only s^l is nonzero, (B.4) gives

$$n \operatorname{Var}_G(\widehat{\theta}) = 1 + (k-1)\lambda_l = 1 - \operatorname{Disp}_G(l) > 1 - \text{ID}_G - \epsilon.$$

Case 2: Suppose $\mathrm{SD}_G/(k-1) > -\mathrm{ID}_G$. Choose l with $\mathrm{Disp}_G(l) > \mathrm{SD}_G - \epsilon$. Pick $\phi \in E_l$ with $\mathrm{Var}_F(\phi) = 1$ and choose coefficients $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ satisfying $\sum_{i=1}^k \alpha_i = 0$ and $k^{-1}\sum_{i=1}^k \alpha_i^2 = 1$. Within each group g, set $s_{ig}(d) = \alpha_i \phi(d)$ for the ith unit. Then $v_{iid}(s) = k^{-1}\sum_i \alpha_i^2 \mathrm{Var}_F(\phi) = 1$, and

$$c(s) = k^{-1}\sum_{i\neq j} \alpha_i\alpha_j \mathrm{Var}_F(\phi) = k^{-1}\left[\left(\sum_i \alpha_i\right)^2 - \sum_i \alpha_i^2\right] = -1.$$

Thus $n\,\mathrm{Var}_G(\widehat{\theta}) = 1 - \lambda_l = 1 + \mathrm{Disp}_G(l)/(k-1) > 1 + \mathrm{SD}_G/(k-1) - \epsilon/(k-1)$. Since $\epsilon > 0$ is arbitrary, the supremum equals $1 + \max(-\mathrm{ID}_G,\ \mathrm{SD}_G/(k-1))$.

Consider part (a). Recall from Section 4.1 the dispersion bounds $-(k-1) \leq \mathrm{Disp}_G(\phi) \leq 1$. For the upper bound, by (7.2), $R_n(G) = 1 + \max(-\mathrm{ID}_G,\ \mathrm{SD}_G/(k-1)) \leq 1 + \max(k-1,\ 1/(k-1)) = 1 + (k-1) = k$. For the lower bound, we show $\max(-\mathrm{ID}_G,\ \mathrm{SD}_G/(k-1)) \geq 0$. If $\mathrm{ID}_G \leq 0$, then $-\mathrm{ID}_G \geq 0$ and the conclusion is immediate. If $\mathrm{ID}_G > 0$, then $\mathrm{SD}_G \geq \mathrm{ID}_G > 0$, so $\mathrm{SD}_G/(k-1) > 0$. In either case $R_n(G) \geq 1$. Minimaxity of G_{iid} is immediate since $R_n(G_{iid}) = 1 + \max(0, 0) = 1$. For $G = \mathrm{RS}$, Lemma B.10 gives eigenvalues $\lambda = 1$ on E_c and $\lambda = -(k-1)^{-1}$ on E_a, so by Theorem 5.3 the dispersions are $\mathrm{Disp}_G(E_c) = -(k-1)$ and $\mathrm{Disp}_G(E_a) = 1$. Hence $\mathrm{ID}_G = -(k-1)$ and $\mathrm{SD}_G = 1$, giving $R_n(G) = 1 + \max(k-1,\ 1/(k-1)) = k$. Finally, if $k = O(1)$, then $R_n(G) \leq k = O(1)$ for any coupling sequence, so $\widehat{\theta}$ is uniformly $\sqrt{n}$-consistent.

Next, consider part (b). The argument above showed $\mathrm{ID}_G = \inf_m \mathrm{Disp}_G(m)$ and $\mathrm{SD}_G = \sup_m \mathrm{Disp}_G(m)$ for $L_0^2(F) = \oplus_{m\geq 1} E_m$ an eigenspace decomposition for U_G. For $G = \mathrm{LHS}$, Lemma B.6 gives $L_0^2(F) = E_{hist} \oplus E_{hist}^\perp$ with eigenvalues $\lambda = -(k-1)^{-1}$ on E_{hist} and $\lambda = 0$ on $E_{hist}^\perp$. By Theorem 5.3, $\mathrm{Disp}_G(E_{hist}) = -(k-1)(-(k-1)^{-1}) = 1$ and $\mathrm{Disp}_G(E_{hist}^\perp) = -(k-1)\cdot 0 = 0$. Hence $\mathrm{ID}_G = 0$ and $\mathrm{SD}_G = 1$, so $R_n(G) = 1 + \max(0,\ 1/(k-1)) = k/(k-1)$. For $G = $ Gaussian, Lemma B.11 gives $L_0^2(F) = \oplus_{m\geq 1}\mathrm{span}(h_m)$ with eigenvalues $\lambda_m = (-1)^m(k-1)^{-m}$. By Theorem 5.3, $\mathrm{Disp}_G(h_m) = -(k-1)\cdot(-1)^m(k-1)^{-m} = (-1)^{m+1}(k-1)^{1-m}$. For odd m, $\mathrm{Disp}_G(h_m) = (k-1)^{1-m} > 0$, which is maximized at $m = 1$ giving $\mathrm{Disp}_G(h_1) = 1$. For even m, $\mathrm{Disp}_G(h_m) = -(k-1)^{1-m} < 0$, which is minimized at $m = 2$ giving $\mathrm{Disp}_G(h_2) = -1/(k-1)$. Hence $\mathrm{SD}_G = 1$ and $\mathrm{ID}_G = -1/(k-1)$, so $R_n(G) = 1 + \max(1/(k-1),\ 1/(k-1)) = k/(k-1)$. In both cases $R_n(G) = k/(k-1) = O(1)$, so uniform $\sqrt{n}$-consistency holds for any sequence $k(n)$. If $k \to \infty$, then $R_n(G) = k/(k-1) \to 1$, and since $R_n(G) \geq 1$ by part (a), both designs are asymptotically minimax optimal. $\square$

Lemma B.15 (Pointwise Consistency)**.** *Let $G \in \Pi_k(F)$ with $k = O(1)$. If $v_{iid}(s) = O(1)$, then $E_n[s_i(D_i)] = \theta_n + O_p(n^{-1/2})$.*

Proof. By the proof of Theorem 7.1, $n\,\mathrm{Var}_G(\widehat{\theta}) \leq k v_{iid}(s)$ for any $G \in \Pi_k(F)$. Since $k = O(1)$ and $v_{iid}(s) = O(1)$, we have $\mathrm{Var}_G(E_n[s_i(D_i)]) = O(n^{-1})$. By Chebyshev's inequality, $E_n[s_i(D_i)] - \theta_n = O_p(n^{-1/2})$. $\square$

Proof of Theorem 7.3. By Theorem 6.7, for $G \in \{\mathrm{LHS}, \mathrm{RS}\}$,

$$1 - \frac{\mathrm{Var}_G(\widehat{\theta})}{\mathrm{Var}_{G_{iid}}(\widehat{\theta})} \ \geq\ Q_k(s) - \frac{\eta_{\mathrm{TV}}(s)}{k}.$$

Rearranging and using $n \operatorname{Var}_{G_{iid}}(\widehat{\theta}) = v_{iid}(s)$,

$$n \operatorname{Var}_G(\widehat{\theta}) \;\le\; v_{iid}(s) \left(1 - Q_k(s) + \frac{\eta_{\mathrm{TV}}(s)}{k}\right).$$

For $s \in \mathcal{S}_{\mathrm{r}}$, we have $v_{iid}(s) \le 1$, $Q_k(s) \ge q_0$, and $\eta_{\mathrm{TV}}(s) \le \bar{\eta}$, so $n \operatorname{Var}_G(\widehat{\theta}) \le 1 - q_0 + \bar{\eta}/k$. Taking the supremum over $s \in \mathcal{S}_{\mathrm{r}}$ completes the proof. ☐

C Asymptotics Proofs

C.1 Proof of Proposition 7.5

Proof of Proposition 7.5. For each matched group $g \in [n/k]$, define the centered scaled group sum

$$S_g = n^{-1} \sum_{i=1}^{k} [s_{ig}(D_i) - E_F[s_{ig}(D)]].$$

Since treatments are independent across matched groups, $\{S_g\}_{g=1}^{n/k}$ are independent with

$$\widehat{\theta} - \theta_n = \sum_{g=1}^{n/k} S_g \qquad \text{and} \qquad \sigma_n^2 = \operatorname{Var}_G(\widehat{\theta}) = \sum_{g=1}^{n/k} \operatorname{Var}_G(S_g).$$

We verify the Lindeberg condition: for every $\varepsilon > 0$,

$$\frac{1}{\sigma_n^2} \sum_{g=1}^{n/k} E_G\left[S_g^2 \mathbb{1}(|S_g| > \varepsilon \sigma_n)\right] \to 0.$$

By Markov's inequality,

$$E_G\left[S_g^2 \mathbb{1}(|S_g| > \varepsilon \sigma_n)\right] \le \frac{E_G[S_g^4]}{\varepsilon^2 \sigma_n^2}.$$

By the power mean inequality applied pointwise,

$$\begin{aligned} E_G[S_g^4] &= \frac{1}{n^4} E_G\left[\left(\sum_{i=1}^{k} [s_{ig}(D_i) - E_F[s_{ig}(D)]]\right)^4\right] \\ &\le \frac{k^3}{n^4} \sum_{i=1}^{k} E_F\left[(s_{ig}(D) - E_F[s_{ig}(D)])^4\right] \le \frac{3k^3}{n^4} \sum_{i=1}^{k} E_F[s_i(D)^4] \end{aligned}$$

where the second to last step uses the fact that $D_i \sim F$ marginally under G, and the last step uses $E_F[(s_{ig}(D) - E_F[s_{ig}(D)])^4] \le 3E_F[s_i(D)^4]$. Summing over all groups,

$$\frac{1}{\sigma_n^2} \sum_{g=1}^{n/k} E_G\left[S_g^2 \mathbb{1}(|S_g| > \varepsilon \sigma_n)\right] \le \frac{3k^3}{\varepsilon^2 n^4 \sigma_n^4} \sum_{i=1}^{n} E_F[s_i(D)^4] = \frac{3k^3 M_{4,n}}{\varepsilon^2 n^3 \sigma_n^4},$$

where $M_{4,n} = E_n\left[E_F[s_i(D)^4]\right]$. By Assumption 7.4 part 1, $M_{4,n} = O(1)$. By Assumption 7.4 part 2, $\sigma_n^2 = \Omega(n^{-1})$, so $1/\sigma_n^4 = O(n^2)$. Therefore,

$$\frac{3k^3 M_4}{\varepsilon^2 n^3 \sigma_n^4} = O(k^3/n) \to 0,$$

since $k^3/n \to 0$ by Assumption 7.4 part 3. The Lindeberg–Feller CLT can therefore be applied using the group sums S_g as the independent variates. ☐

C.2 Proof of Proposition 7.6

Proof of Proposition 7.6. Expectation. Because groups g and $\pi(g)$ are independent (since π has no fixed points), we have $E_G[\widehat{\theta}_g\widehat{\theta}_{\pi(g)}] = E_G[\widehat{\theta}_g]E_G[\widehat{\theta}_{\pi(g)}] = \theta_g\theta_{\pi(g)}$. Therefore,

$$\begin{aligned} E_G\big[(\widehat{\theta}_g - \widehat{\theta}_{\pi(g)})^2\big] &= E_G[\widehat{\theta}_g^2] + E_G[\widehat{\theta}_{\pi(g)}^2] - 2\theta_g\theta_{\pi(g)} \\ &= \mathrm{Var}_G(\widehat{\theta}_g) + \mathrm{Var}_G(\widehat{\theta}_{\pi(g)}) + (\theta_g - \theta_{\pi(g)})^2. \end{aligned}$$

Summing over g and using that π is a permutation,

$$\sum_{g=1}^{n/k} E_G\big[(\widehat{\theta}_g - \widehat{\theta}_{\pi(g)})^2\big] = 2\sum_{g=1}^{n/k} \mathrm{Var}_G(\widehat{\theta}_g) + \sum_{g=1}^{n/k}(\theta_g - \theta_{\pi(g)})^2.$$

Since $\widehat{\theta} = (n/k)^{-1}\sum_g \widehat{\theta}_g$ and the groups are independent,

$$\sigma_n^2 = \mathrm{Var}_G(\widehat{\theta}) = (n/k)^{-2}\sum_{g=1}^{n/k} \mathrm{Var}_G(\widehat{\theta}_g),$$

so $\sum_g \mathrm{Var}_G(\widehat{\theta}_g) = (n/k)^2\sigma_n^2$. Substituting and using $\sum_g(\theta_g - \theta_{\pi(g)})^2 = (n/k)\Delta_n^2$,

$$E_G[\widehat{\sigma}_n^2] = \frac{k^2}{2n^2}\left[2(n/k)^2\sigma_n^2 + (n/k)\Delta_n^2\right] = \sigma_n^2 + \frac{k\Delta_n^2}{2n}.$$

Dividing by σ_n^2 gives the expectation result.

Variance. Let $Q_g = (\widehat{\theta}_g - \widehat{\theta}_{\pi(g)})^2$. Since Q_g depends only on groups g and $\pi(g)$, the terms Q_g and Q_h are independent unless $\{g, \pi(g)\} \cap \{h, \pi(h)\} \neq \emptyset$. For each g, the dependent indices are $h \in \{g, \pi(g), \pi^{-1}(g)\}$, so

$$\mathrm{Var}_G(\widehat{\sigma}_n^2) = \frac{k^4}{4n^4}\sum_{g=1}^{n/k}\sum_{h\in\{g,\pi(g),\pi^{-1}(g)\}} \mathrm{Cov}_G(Q_g, Q_h).$$

By the Cauchy–Schwarz and AM–GM inequalities, $\mathrm{Cov}_G(Q_g, Q_h) \leq [\mathrm{Var}_G(Q_g) + \mathrm{Var}_G(Q_h)]/2$. Each inner sum has at most three terms, and reindexing via the permutation π preserves the sum, giving

$$\mathrm{Var}_G(\widehat{\sigma}_n^2) \leq \frac{3k^4}{4n^4}\sum_{g=1}^{n/k} \mathrm{Var}_G(Q_g).$$

Next, $\mathrm{Var}_G(Q_g) \leq E_G[Q_g^2] = E_G[(\widehat{\theta}_g - \widehat{\theta}_{\pi(g)})^4] \leq 8(E_G[\widehat{\theta}_g^4] + E_G[\widehat{\theta}_{\pi(g)}^4])$, where the last step uses the convexity of $x \mapsto |x|^4$. Because π is a permutation, $\sum_g(E_G[\widehat{\theta}_g^4] + E_G[\widehat{\theta}_{\pi(g)}^4]) = 2\sum_g E_G[\widehat{\theta}_g^4]$, so

$$\mathrm{Var}_G(\widehat{\sigma}_n^2) \leq \frac{12k^4}{n^4}\sum_{g=1}^{n/k} E_G[\widehat{\theta}_g^4].$$

By Jensen's inequality applied to $x \mapsto |x|^4$,

$$E_G[\widehat{\theta}_g^4] = E_G\left[\left(k^{-1}\sum_{i=1}^{k} s_{ig}(D_i)\right)^4\right] \le k^{-1}\sum_{i=1}^{k} E_G[s_{ig}(D_i)^4].$$

Therefore,

$$\mathrm{Var}_G(\widehat{\sigma}_n^2) \le \frac{12k^3}{n^4}\sum_{i=1}^{n} E_F[s_i(D)^4] = \frac{12k^3}{n^3} M_{4,n}.$$

Dividing by σ_n^4 gives the variance result. □

C.3 Proof of Proposition 7.7

Proof of Proposition 7.7. Let $T_n = (\widehat{\theta} - \theta_n)/\sigma_n$ and $R_n = \widehat{\sigma}_n/\sigma_n$. By Proposition 7.5, $T_n \Rightarrow \mathcal{N}(0,1)$. By Proposition 7.6, $E_G[R_n^2] \ge 1$ and $\mathrm{Var}_G(R_n^2) = o(1)$. Note that $\theta_n \in \mathrm{CI}_{1-\alpha}$ if and only if $|T_n| \le z_{1-\alpha/2}\, R_n$. For any $\delta > 0$,

$$\Pr_G(|T_n| \le z_{1-\alpha/2}\, R_n) \ge \Pr_G(|T_n| \le z_{1-\alpha/2}\sqrt{1-\delta},\ R_n^2 \ge 1-\delta).$$

Since $E_G[R_n^2] \ge 1$, Chebyshev's inequality gives $\Pr_G(R_n^2 < 1-\delta) \le \mathrm{Var}_G(R_n^2)/\delta^2 = o(1)$. Therefore,

$$\Pr_G(|T_n| \le z_{1-\alpha/2}\, R_n) \ge \Pr_G(|T_n| \le z_{1-\alpha/2}\sqrt{1-\delta}) - o(1).$$

Taking limits and using $T_n \Rightarrow \mathcal{N}(0,1)$,

$$\liminf_{n\to\infty} \Pr_G(\theta_n \in \mathrm{CI}_{1-\alpha}) \ge \Pr_G(|Z| \le z_{1-\alpha/2}\sqrt{1-\delta}),$$

where $Z \sim \mathcal{N}(0,1)$. Since this holds for all $\delta > 0$, taking $\delta \to 0$ gives $\liminf_n \Pr_G(\theta_n \in \mathrm{CI}_{1-\alpha}) \ge 1-\alpha$. □

C.4 Lemmas

Define $v_{iid}(s) \equiv E_n \mathrm{Var}_F(s_i)$ and $v_\Delta(s) \equiv (2n(k-1))^{-1}\sum_g \sum_{i\ne j} \mathrm{Var}_F(s_{ig} - s_{jg})$. Also let $\bar{s}_g \equiv k^{-1}\sum_{i\in[k]} s_{ig}$ and define $v_g(s) \equiv (n/k)^{-1}\sum_g \mathrm{Var}_F(\bar{s}_g)$. Also define $c(s) \equiv n^{-1}\sum_g \sum_{i\ne j} \mathrm{Cov}_F(s_{ig}, s_{jg})$.

Lemma C.1 (Variance Identities). *We have $v_\Delta(s) = v_{iid}(s) - (k-1)^{-1}c(s)$ and $k \cdot v_g(s) = v_{iid}(s) + c(s)$. Moreover, $kv_g(s) \le kv_{iid}(s)$ and $v_\Delta(s) \le 2v_{iid}(s)$. In particular, $-v_{iid}(s) \le c(s) \le (k-1)v_{iid}(s)$. The match quality coefficient $Q_k(s) = (k-1)^{-1}c(s)/v_{iid}(s)$. Consequently, $Q_k(s) \in [-1/(k-1),\, 1]$. Finally, for $k \ge 2$, we have*

$$v_\Delta(s) = (k-1)^{-1}(n/k)^{-1}\sum_g \sum_{i\in[k]} \mathrm{Var}_F(s_{ig} - \bar{s}_g)$$

Proof. Consider the first identity. We again expand the variance of the difference:

$$\begin{aligned}
v_\Delta(s) &= (2n(k-1))^{-1} \sum_g \sum_{i \neq j \in [k]} \big(\operatorname{Var}_F(s_{ig}) + \operatorname{Var}_F(s_{jg}) - 2\operatorname{Cov}_F(s_{ig}, s_{jg})\big) \\
&= (n(k-1))^{-1} \sum_g \Big((k-1) \sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) - \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg})\Big) \\
&= n^{-1} \sum_g \sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) - (n(k-1))^{-1} \sum_g \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg}) \\
&= v_{iid}(s) - (n(k-1))^{-1} \sum_g \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg}).
\end{aligned}$$

Consider the second identity. We start from the variance of the group mean $\bar{s}_g$:

$$k^2 \operatorname{Var}_F(\bar{s}_g) = \operatorname{Var}_F \sum_{i \in [k]} s_{ig} = \sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) + \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg}).$$

Summing over all groups g and multiplying by n^{-1} gives

$$\begin{aligned}
n^{-1} k^2 \sum_g \operatorname{Var}_F(\bar{s}_g) &= n^{-1} \sum_g \sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) + n^{-1} \sum_g \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg}) \\
&= v_{iid}(s) + n^{-1} \sum_g \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg}).
\end{aligned}$$

For the next inequality, note that

$$\begin{aligned}
|c(s)| &= \left| n^{-1} \sum_g \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg}) \right| \leq n^{-1} \sum_g \sum_{i \neq j \in [k]} |\operatorname{Cov}_F(s_{ig}, s_{jg})| \\
&\leq n^{-1} \sum_g \sum_{i \neq j \in [k]} (1/2)(\operatorname{Var}_F(s_{ig}) + \operatorname{Var}_F(s_{jg})) \\
&= n^{-1} \sum_g (k-1) \sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) = (k-1) E_n \operatorname{Var}_F(s_i) = (k-1) v_{iid}(s).
\end{aligned}$$

The second inequality is by Cauchy-Schwarz and Young's inequality. Then $kv_g(s) = v_{iid}(s) + c(s) \leq kv_{iid}(s)$ and $v_\Delta(s) = v_{iid}(s) - (k-1)^{-1}c(s) \leq 2v_{iid}(s)$. For the bounds on $c(s)$: the lower bound $c(s) \geq -v_{iid}(s)$ follows from $kv_g(s) = v_{iid}(s) + c(s) \geq 0$, and the upper bound $c(s) \leq (k-1)v_{iid}(s)$ from $v_\Delta(s) = v_{iid}(s) - (k-1)^{-1}c(s) \geq 0$. The statement about match quality follows by rearranging the identity $v_\Delta(s) = v_{iid}(s) - (k-1)^{-1}c(s)$, noting $Q_k(s) = 1 - v_\Delta(s)/v_{iid}(s)$. Finally consider the last statement. Expand

$$\begin{aligned}
\sum_{i \neq j \in [k]} \operatorname{Var}_F(s_{ig} - s_{jg}) &= \sum_{i \neq j \in [k]} \big(\operatorname{Var}_F(s_{ig}) + \operatorname{Var}_F(s_{jg}) - 2\operatorname{Cov}_F(s_{ig}, s_{jg})\big) \\
&= 2(k-1) \sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) - 2 \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg}).
\end{aligned}$$

Next, we use $\sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) = \sum_{i \in [k]} \operatorname{Var}_F(s_{ig} - \bar{s}_g) + k \operatorname{Var}_F(\bar{s}_g)$. Also, from the definition of $\bar{s}_g$, we have $k^2 \operatorname{Var}_F(\bar{s}_g) = \sum_{i \in [k]} \operatorname{Var}_F(s_{ig}) + \sum_{i \neq j \in [k]} \operatorname{Cov}_F(s_{ig}, s_{jg})$.

Combining these gives

$$\begin{aligned}\sum_{i\neq j\in[k]} \operatorname{Var}_F(s_{ig}-s_{jg}) &= 2(k-1)\sum_{i\in[k]}\operatorname{Var}_F(s_{ig}) - 2\big(k^2\operatorname{Var}_F(\bar{s}_g) - \sum_{i\in[k]}\operatorname{Var}_F(s_{ig})\big)\\ &= 2k\sum_{i\in[k]}\operatorname{Var}_F(s_{ig}) - 2k^2\operatorname{Var}_F(\bar{s}_g)\\ &= 2k\big(\sum_{i\in[k]}\operatorname{Var}_F(s_{ig}) - k\operatorname{Var}_F(\bar{s}_g)\big) = 2k\sum_{i\in[k]}\operatorname{Var}_F(s_{ig}-\bar{s}_g).\end{aligned}$$

Summing over g and multiplying by $(2n(k-1))^{-1}$ yields

$$\begin{aligned}v_\Delta(s) &= (2n(k-1))^{-1}\sum_g 2k\sum_{i\in[k]}\operatorname{Var}_F(s_{ig}-\bar{s}_g)\\ &= (k-1)^{-1}(n/k)^{-1}\sum_g\sum_{i\in[k]}\operatorname{Var}_F(s_{ig}-\bar{s}_g).\end{aligned}$$

This proves the first identity. □

Lemma C.2 (Orthogonal Decomposition). *Let $L_2(F) = H\oplus H^\perp$ where H is a closed linear subspace with $1\in H$. Let $s_i^H = P_H s_i$ be the orthogonal projection onto H. Then the following hold:*

1. $v_{iid}(s) = v_{iid}(s^H) + v_{iid}(s-s^H)$
2. $c(s) = c(s^H) + c(s-s^H)$
3. $v_\Delta(s) = v_\Delta(s^H) + v_\Delta(s-s^H)$
4. $Q_k(s) = w_H\cdot Q_k(s^H) + (1-w_H)\cdot Q_k(s-s^H)$, *where* $w_H = v_{iid}(s^H)/v_{iid}(s)$.

Proof. For (1), note that $s_i^H\in H$ and $s_i - s_i^H\in H^\perp$ by definition of projection. Since $1\in H$, we have $\langle s_i - s_i^H, 1\rangle_F = E_F[s_i - s_i^H] = 0$. Then $\operatorname{Var}_F(s_i) = \operatorname{Var}_F(s_i - s_i^H + s_i^H) = \operatorname{Var}_F(s_i - s_i^H) + 2\operatorname{Cov}_F(s_i - s_i^H, s_i^H) + \operatorname{Var}_F(s_i^H)$. Then the middle term $\operatorname{Cov}_F(s_i - s_i^H, s_i^H) = E_F[(s_i - s_i^H)s_i^H] = \langle s_i - s_i^H, s_i^H\rangle_F = 0$ by orthgonality. The conclusion follows since $v_{iid}(s) = E_n\operatorname{Var}_F(s_i)$. Next consider the second claim. We expand the covariance term:

$$\begin{aligned}\operatorname{Cov}_F(s_{ig}, s_{jg}) &= \operatorname{Cov}_F\left(s_{ig} - s_{ig}^H + s_{ig}^H, s_{jg} - s_{jg}^H + s_{jg}^H\right)\\ &= \operatorname{Cov}_F(s_{ig}^H, s_{jg}^H) + \operatorname{Cov}_F(s_{ig} - s_{ig}^H, s_{jg} - s_{jg}^H)\\ &\quad + \operatorname{Cov}_F(s_{ig}^H, s_{jg} - s_{jg}^H) + \operatorname{Cov}_F(s_{ig} - s_{ig}^H, s_{jg}^H).\end{aligned}$$

The cross term is $\operatorname{Cov}_F(s_{ig}^H, s_{jg} - s_{jg}^H) = \langle s_{ig}^H, s_{jg} - s_{jg}^H\rangle_F = 0$ since $s_{ig}^H\in H$ and $s_{jg} - s_{jg}^H\in H^\perp$ with $H^\perp\perp 1$, and similarly for the second term $\operatorname{Cov}_F(s_{ig} - s_{ig}^H, s_{jg}^H)$. The claim now follows since $c(s) = n^{-1}\sum_g\sum_{i\neq j}\operatorname{Cov}_F(s_{ig}, s_{jg})$. Finally, the third claim follows from the identity $v_\Delta(s) = v_{iid}(s) - (k-1)^{-1}c(s)$ in Lemma C.1 and the additivity of $v_{iid}(\cdot)$ and $c(\cdot)$ shown above. For (4), write $w_\perp = v_{iid}(s-s^H)/v_{iid}(s) = 1-w_H$ by (1). If $v_{iid}(s^H) = 0$ or $v_{iid}(s-s^H) = 0$, then $w_H = 0$ or $w_\perp = 0$ respectively, and the corresponding v_Δ term also vanishes since $v_\Delta\le 2v_{iid}$ by Lemma C.1, so the claim holds trivially. Otherwise, dividing (3) by $v_{iid}(s)$ and using (1):

$$\frac{v_\Delta(s)}{v_{iid}(s)} = w_H\frac{v_\Delta(s^H)}{v_{iid}(s^H)} + w_\perp\frac{v_\Delta(s-s^H)}{v_{iid}(s-s^H)} = w_H(1-Q_k(s^H)) + w_\perp(1-Q_k(s-s^H)).$$

Then $Q_k(s) = 1 - v_\Delta(s)/v_{iid}(s) = 1 - w_H - w_\perp + w_H Q_k(s^H) + w_\perp Q_k(s - s^H) = w_H Q_k(s^H) + w_\perp Q_k(s - s^H)$, using $w_H + w_\perp = 1$. $\square$

Define $\mathrm{Var}_g(a_{ig}) = (k-1)^{-1}\sum_{i=1}^k (a_{ig} - \bar{a}_g)(a_{ig} - \bar{a}_g)'$.

Lemma C.3. *The following equalities hold*

$$
\begin{aligned}
(k-1)^{-1} \cdot n^{-1} \sum_g \sum_{i \neq j \in [k]} a_{ig} a'_{jg} &= E_n[a_i a'_i] - (n/k)^{-1} \sum_g \mathrm{Var}_g(a_{ig}) \\
&= E_n[a_i a'_i] - (2n(k-1))^{-1} \sum_g \sum_{i \neq j \in [k]} (a_{ig} - a_{jg})(a_{ig} - a_{jg})'
\end{aligned}
$$

In particular, $(2n(k-1))^{-1}\sum_g \sum_{i\neq j \in [k]} (a_{ig} - a_{jg})(a_{ig} - a_{jg})' = (n/k)^{-1}\sum_g \mathrm{Var}_g(a_{ig})$.

Proof. For the first equality, we calculate

$$
\begin{aligned}
S(a) = (k-1)^{-1} n^{-1} \sum_g \sum_{i \neq j \in [k]} a_{ig} a'_{jg} &= (k-1)^{-1}(2n)^{-1} \sum_g \sum_{i \neq j \in [k]} (a_{ig} a'_{jg} + a_{jg} a'_{ig}) \\
&= (2n(k-1))^{-1} \sum_g \sum_{i \neq j \in [k]} a_{ig} a'_{ig} + a_{jg} a'_{jg} - (a_{ig} - a_{jg})(a_{ig} - a_{jg})' \\
&= E_n[a_i a'_i] - (2n(k-1))^{-1} \sum_g \sum_{i \neq j \in [k]} (a_{ig} - a_{jg})(a_{ig} - a_{jg})'.
\end{aligned}
$$

Next, for each group g define $\Delta_g = \sum_{i \neq j \in g} (a_{ig} - a_{jg})(a_{ig} - a_{jg})'$, which is

$$
\Delta_g = \sum_{i,j \in g} (a_{ig} - a_{jg})(a_{ig} - a_{jg})' = \sum_{i,j \in g} \big(a_{ig} a'_{ig} + a_{jg} a'_{jg} - a_{ig} a'_{jg} - a_{jg} a'_{ig}\big).
$$

Since $\sum_{i,j \in g} a_{ig} a'_{ig} = k \sum_{i \in g} a_{ig} a'_{ig}$ and $\sum_{i,j \in g} a_{ig} = \sum_{j \in g} a_{jg} = k\bar{a}_g$, this becomes

$$
\begin{aligned}
\Delta_g &= k \sum_{i \in g} a_{ig} a'_{ig} + k \sum_{j \in g} a_{jg} a'_{jg} - k^2 \bar{a}_g \bar{a}'_g - k^2 \bar{a}_g \bar{a}'_g \\
&= 2k \sum_{i \in g} a_{ig} a'_{ig} - 2k^2 \bar{a}_g \bar{a}'_g = 2k^2 (k^{-1} \sum_{i \in g} a_{ig} a'_{ig} - \bar{a}_g \bar{a}'_g) \\
&= 2k^2 k^{-1} \sum_{i \in g} (a_{ig} - \bar{a}_g)(a_{ig} - \bar{a}_g)' = 2k(k-1)\,\mathrm{Var}_g(a_{ig}).
\end{aligned}
$$

Then $\Delta = \frac{1}{2n(k-1)} \sum_g \Delta_g = n^{-1} k \sum_g \mathrm{Var}_g(a_{ig})$, which finishes the proof. $\square$

We define functions of bounded variation.

Definition C.4 (Bounded variation)**.** A function $h : [a,b] \to \mathbb{R}$ is said to be of bounded variation on $[a,b]$ if $V_{[a,b]}(h) \equiv \sup_\Pi \sum_{j=1}^m \big|h(t_j) - h(t_{j-1})\big| < \infty$, where the sup is over all finite partitions $\Pi : a = t_0 < t_1 < \cdots < t_m = b$ of $[a,b]$.

Lemma C.5 (Composition)**.** *Let $h : [A,B] \to \mathbb{R}$ of bounded variation and $r : [a,b] \to [A,B]$ weakly monotone. Then $h \circ r : [a,b] \to \mathbb{R}$ has $V_{[a,b]}(h \circ r) \le V_{[A,B]}(h)$.*

Proof. Suppose r is nondecreasing. For any partition $\Pi : a = t_0 < \cdots < t_m = b$, set $y_j = r(t_j)$. Then $(y_j)_{j=0}^m$ is nondecreasing in $[A, B]$, so it partitions $r([a, b]) = [r(a), r(b)] \subseteq [A, B]$. Then we have

$$\sum_{j=1}^{m} \big|(h \circ r)(t_j) - (h \circ r)(t_{j-1})\big| = \sum_{j=1}^{m} \big|h(y_j) - h(y_{j-1})\big| \leq V_{r([a,b])}(h) \leq V_{[A,B]}(h).$$

Taking the sup over all partitions Π gives $V_{[a,b]}(h \circ r) \leq V_{r([a,b])}(h) \leq V_{[A,B]}(h)$. The proof for the monotone decreasing case is similar. □

Lemma C.6. *Let F be a continuous CDF. Then $F(F^{-1}(u)) = u$ for all $u \in (0, 1)$.*

Proof. Recall $F^{-1}(u) = \inf\{x : F(x) \geq u\}$. If $F(F^{-1}(u)) > u$, then by continuity $F(F^{-1}(u) - \epsilon) > u$ for some $\epsilon > 0$, contradicting the definition of F^{-1}, so $F(F^{-1}(u)) \leq u$. Similarly, we must have $F(F^{-1}(u)) \geq u$ by definition of the infimum, so $F(F^{-1}(u)) = u$. □

Lemma C.7. *Let $T : H \to H$ be a bounded linear operator on a Hilbert space H. Then for any $\lambda \in \mathbb{C}$, the eigenspace $E_\lambda = \ker(T - \lambda I)$ is a closed subspace of H.*

Proof. Let $S = T - \lambda I$, which is bounded and linear since T and I are. Bounded linear operators are continuous. The eigenspace $E_\lambda = S^{-1}\{0\}$. Since $\{0\}$ is a closed set in H and S is continuous, the preimage E_λ must be closed. □

References

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. *Handbook of Economic Field Experiments.*

Bai, Y. (2022). Optimality of matched-pair designs in randomized controlled trials. *American Economic Review.*

Bai, Y., Liu, J., Shaikh, A. M., and Tabord-Meehan, M. (2023). On the efficiency of finely stratified experiments. arXiv preprint arXiv:2307.15181.

Bai, Y., Liu, J., and Tabord-Meehan, M. (2024). Inference for matched tuples and fully blocked factorial designs. *Quantitative Economics*, 15(2):279–330.

Bai, Y., Romano, J. P., and Shaikh, A. M. (2021). Inference in experiments with matched pairs. *Journal of the American Statistical Association.*

Bai, Y., Shaikh, A. M., and Tabord-Meehan, M. (2025). A primer on the analysis of randomized experiments and a survey of some recent advances. arXiv:2405.03910.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, pages 200–232.

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association.*

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics.*

Cai, Y. and Rafi, A. (2024). On the performance of the neyman allocation with small pilots. *Journal of Econometrics*, 242(1).

Carlier, G., Galichon, A., and Santambrogio, F. (2010). From Knothe's transport to Brenier's map and a continuation method for optimal transport. *SIAM Journal on Mathematical Analysis*, 41(6):2554–2576.

Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223–256.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*. John Wiley & Sons, New York, 2nd edition.

Cranley, R. and Patterson, T. N. L. (1976). Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914.

Cytrynbaum, M. (2021). The optimal design for a fixed estimator. Working Paper.

Cytrynbaum, M. (2023). Optimal stratification of survey experiments. arXiv:2111.08157.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, 58(3):403–417.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, 33:503–513.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Fishman, G. S. and Huang, B. D. (1983). Antithetic variates revisited. *Communications of the ACM*, 26(11):964–971.

Fogarty, C. B. (2018). On mitigating the analytical limitations of finely stratified experiments. *Journal of the Royal Statistical Society: Series B*, 80(5).

Greevy, R., Lu, B., Silber, J. H., and Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics*, 5(2):263–275.

Hammersley, J. M. and Morton, K. W. (1956). A new Monte Carlo technique: Antithetic variates. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52(3):449–475.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample Survey Methods and Theory*. Wiley.

Harshaw, C., Sävje, F., Eisenstat, D., Mirrokni, V., and Pouget-Abadie, J. (2023). Design and analysis of bipartite experiments under a linear exposure-response model. *Electronic Journal of Statistics*, 17(2):4528–4572.

Harshaw, C., Sävje, F., Spielman, D. A., and Zhang, P. (2024). Balancing covariates in randomized experiments with the Gram–Schmidt walk design. *Journal of the American Statistical Association*, 119(548):2934–2946.

Harshaw, C., Sävje, F., and Wang, Y. (2025). A general design-based framework and estimator for randomized experiments. arXiv:2210.08698.

Higgins, M. J., Sävje, F., and Sekhon, J. S. (2015). Improving massive experiments with threshold blocking. *Proceedings of the National Academy of Sciences*.

Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5:179–233.

Imai, K. (2008). Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in Medicine*, 27:4857–4873.

Imai, K., King, G., and Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation. *Statistical Science*.

Kallus, N. (2017). Optimal a priori balance in the design of controlled experiments. *Journal of the Royal Statistical Society: Series B*.

Kasy, M. (2016). Why experimenters might not always want to randomize, and what they could do instead. *Political Analysis*, pages 1–15.

Kempthorne, O. (1956). The efficiency factor of an incomplete block design. *Annals of Mathematical Statistics*, 27:846–849.

Koo, T. and Pashley, N. E. (2026). Design-based causal inference for incomplete block designs. *Biometrika*.

Kuo, F. Y. (2003). Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces. *Journal of Complexity*, 19(3):301–320.

L'Ecuyer, P. and Lemieux, C. (2000). Variance reduction via lattice rules. *Management Science*, 46(9):1214–1235.

Li, X., Ding, P., and Rubin, D. B. (2018). Asymptotic theory of rerandomization in treatment–control experiments. *Proceedings of the National Academy of Sciences*.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

Mehler, F. G. (1866). Ueber die Entwicklung einer Function von beliebig vielen Variablen nach Laplaceschen Functionen höherer Ordnung. *Journal für die reine und angewandte Mathematik*, 66:161–176.

Mérigot, Q. (2011). A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5):1583–1592.

Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Annals of Statistics*, 40(2).

Owen, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, 2(2):439–452.

Owen, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. In Niederreiter, H. and Shiue, P. J.-S., editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*, pages 299–317. Springer, New York.

Owen, A. B. (2013). *Monte Carlo theory, methods and examples*. https://artowen.su.domains/mc/.

Pashley, N. E. and Miratrix, L. W. (2021). Insights on variance estimation for blocked and matched pairs designs. *Journal of Educational and Behavioral Statistics*.

Pollard, D. (1991). Asymptotics for least absolute deviation estimators. *Econometric Theory*, 7(2):186–199.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag.

Shirley, P. (1991). Discrepancy as a quality measure for sample distributions. In *Proceedings of Eurographics*, pages 183–193.

Sloan, I. H. and Joe, S. (1994). *Lattice Methods for Multiple Integration*. Clarendon Press, Oxford.

Tang, B. (1993). Orthogonal array-based Latin hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397.

Thangavelu, S. (1993). *Lectures on Hermite and Laguerre Expansions*, volume 42 of *Mathematical Notes*. Princeton University Press.

Yates, F. (1936). Incomplete randomized blocks. *Annals of Eugenics*, 7(2):121–140.

Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, 14(4):478–486. Also available via JSTOR stable 1392256.