

Heterogeneous Treatment Effects for Networks, Panels, and other Outcome Matrices

Eric Auerbach* Yong Cai †

May 4, 2022

Abstract

We are interested in the distribution of treatment effects for an experiment where units are randomized to treatment but outcomes are measured for pairs of units. For example, we might measure risk sharing links between households enrolled in a microfinance program, employment relationships between workers and firms exposed to a trade shock, or bids from bidders to items assigned to an auction format. Such a double randomized experimental design may be appropriate when there are social interactions, market externalities, or other spillovers across units assigned to the same treatment. Or it may describe a natural or quasi experiment given to the researcher. In this paper, we propose a new empirical strategy based on comparing the eigenvalues of the outcome matrices associated with each treatment. Our proposal is based on a new matrix analog of the Fréchet-Hoeffding bounds that play a key role in the standard theory. We first use this result to bound the distribution of treatment effects. We then propose a new matrix analog of quantile treatment effects based on the difference in the eigenvalues. We call this analog spectral treatment effects.

1 Introduction

Consider a market designer who wants to learn how a change in auction format impacts the amounts bid by a set of unique bidders on a set of unique items. For example, the bidders may be logging companies and the items tracts of forest land. The designer conducts an experiment where they randomly assign bidders and items to groups, implement a different

*Department of Economics, Northwestern University. E-mail: eric.auerbach@northwestern.edu.

†Department of Economics, Northwestern University. E-mail: yongcai2023@u.northwestern.edu.

We thank Lori Beaman, Federico Bugni, Ivan Canay, Ben Golub, Bryan Graham, Joel Horowitz, Gaston Illanes, Chuck Manski, Rob Porter, and Guillaume Pouliot for helpful feedback.

auction format in each group, and observe the amount each bidder bids on each item in each group. In this scenario, the treatments are the auction formats and the object of interest is the distribution of treatment effects. That is, the law of the difference between how much a bidder would bid on an item in the first format and how much the same bidder would bid on the same item in the second format.

If each format contained one identical item then the designer’s problem reduces to that of inferring the joint distribution of two potential outcomes (what the same bidder would bid in each format) from their marginal distributions (what two unrelated bidders would bid in each format). Though not generally point identified, the distribution of treatment effects can be bounded using arguments of Fréchet (1951); Hoeffding (1940); Makarov (1982) (see for instance Manski 1997; 2003; Heckman et al. 1997; Fan and Park 2010; Abadie and Cattaneo 2018; Firpo and Ridder 2019). Under an additional assumption that the rank of an agent’s bid is the same in both formats, the distribution of treatment effects is point identified and described by the law of the difference in the quantiles of the bids made in each format. This assumption is called rank invariance and the parameter, quantile treatment effects (see for instance Abadie et al. 2002; Chernozhukov and Hansen 2005; Bitler et al. 2006; Firpo 2007).

What is an analogous way to characterize the distribution of treatment effects when both bidders and items are unique, numerous, and randomized to treatment? Our main technical contribution is to propose analogs of the Fréchet-Hoeffding bounds for this double randomized setting. A key complication is that the two dimensions of randomization make the problem quadratic rather than linear. Sharp bounds are not generally computable.

We instead consider relaxations of the quadratic problem solved by rearranging the eigenvalues of the outcome matrices associated with each treatment. We first bound the distribution of treatment effects building on work by Whitt (1976); Finke et al. (1987). We then show that under a matrix generalization of rank invariance, the distribution of treatment effects is point identified and characterized by the difference in the eigenvalues. We call this matrix analog of quantile treatment effects, spectral treatment effects.

Our setting is related to that of Bajari et al. (2021); Johari et al. (2022). Specifically, our double randomized experiment is a special case of their simple multiple randomization design, although these authors do not consider the distribution of treatment effects. Double

randomization also plays a role in the literature on social interactions and matching (see for instance Graham 2008; 2011; Graham et al. 2014).

Our setting is distinct from other literatures on heterogeneous effects for matrices, such as that on vector quantiles (see for instance Carlier et al. 2016; Galichon 2016; Arellano and Bonhomme 2021; Fan and Henry 2022). These linear tools are not generally appropriate for the kinds of quadratic problems that arise in a double randomized experiment. See Appendix Section C for a discussion.

The remainder of this paper is as follows. Section 1.1 provides some motivating examples. Section 2 reviews the single randomized experimental setting for reference. Section 3 describes our extension to the double randomized experiment. Section 4 sketches the proof behind our first result, Section 5 discusses some extensions, Section 6 provides two empirical demonstrations, and Section 7 concludes. Proof of claims and other details can be found in the appendices.

1.1 Motivating examples

We describe four motivating examples of double randomized experiments with matrix outcomes from the literature. We revisit two of these examples in more detail in Section 6.

1.1.1 Example 1: Risk sharing

Banerjee et al. (2021) study the impact of a microfinance program in a sample of villages in India. They argue that the program decreases informal risk sharing between some households. Comola and Prina (2021) study the impact of savings accounts in a sample of villages in Nepal. They argue that the program increases informal risk sharing between some households. In this example, the units are households, the treatment is program participation, and the outcomes are risk sharing links between pairs of households.

1.1.2 Example 2: Superstar extinction

Azoulay et al. (2010) study the impact of a superstar researcher's death in a sample of research groups in the life sciences. They argue that a superstars' death decreases the

quality of research conducted by other nearby researchers in the coauthorship network. In this example, the units are researchers, the treatment is the death of the superstar, and the outcomes are the amount of research conducted between coauthors.

1.1.3 Example 3: Auction format

Athey et al. (2011) study the impact of sealed versus open bid designs in a sample of US Forest Service timber auctions. They argue that a sealed bid design incentivizes some firms to participate who otherwise would not in the open bid design. In this example, the units are firms and tracts of land, the treatment is the auction format, and the outcomes are the bids made by firms on the tracts.

1.1.4 Example 4: Buyer-seller experiment

Bajari et al. (2021) model the impact of an information policy on the likelihood that a buyer buys an item from a seller. They consider an experimental design where the researcher independently randomizes buyers and sellers into groups and then assigns policies depending on the group assignments. In this example, the units are buyers and sellers, the treatment is the information policy, and the outcomes are the transactions between buyers and sellers.

2 Review of the single randomized experiment

We review the standard setting of a single randomized experiment for reference.

2.1 Model and econometric problem

2.1.1 Model

A population of agents is randomized to a binary treatment $t \in \{0, 1\}$. The population may be finite or infinite. Outcomes are measured for each agent. The potential outcomes of an agent selected uniformly at random are described by a joint distribution function H on \mathbb{R}^2 .

We define $(Y_1^*, Y_0^*) : [0, 1] \rightarrow \mathbb{R}^2$ such that $(Y_1^*(U), Y_0^*(U))$ has distribution H when U is standard uniform, see Lemma 2.7 of Whitt (1976). We interpret (Y_1^*, Y_0^*) as describing

the outcomes of a continuum of agent types indexed by $[0, 1]$. For example, $Y_t^*(u)$ might describe the income of a worker with type u that participates ($t = 1$) or does not participate ($t = 0$) in a training program.

2.1.2 Parameters of interest

We focus on the joint distribution of potential outcomes (DPO) and distribution of treatment effects (DTE). The DPO is

$$F(y_1, y_0) := P(Y_1^*(w) \leq y_1, Y_0^*(w) \leq y_0) = \int \prod_{t \in \{0,1\}} \mathbb{1}\{Y_t^*(u) \leq y_t\} du \quad (1)$$

where $y_1, y_0 \in \mathbb{R}$ are arbitrary and w is a standard uniform random variable. In words, the DPO is the mass of agent types with potential outcome less than y_1 under treatment 1 and less than y_0 under treatment 0.

The DTE is

$$\Delta(y) := P(Y_1^*(w) - Y_0^*(w) \leq y) = \int \mathbb{1}\{Y_1^*(u) - Y_0^*(u) \leq y\} du. \quad (2)$$

In words, $Y_1^*(u) - Y_0^*(u)$ is the change in outcome associated with switching the treatment status of an agent with type u from 0 to 1. The DTE is the mass of agents for which this individual treatment effect is less than y .

Our arguments naturally extend to many other parameters of potential interest, but their study is left to future work.

2.1.3 Econometric problem

The task is to identify the DPO and DTE. It is complicated by the fact that the researcher does not observe both Y_1^* and Y_0^* for the same population of agents. Agents are assigned to treatment 1 or treatment 0 but not both.

For example, the researcher may assign workers to participate or not participate in a training program. They observe the income of a participating worker under the treatment that the worker participates in the program. They do not observe the income that the

participating worker would have received had they not participated in the program. To infer this missing potential outcome, the researcher must use the incomes of the workers that did not participate in the training program.

This econometric problem is formalized by the assumption that the researcher observes the marginal distributions of the potential outcomes given by $F_1(\cdot) := H(\cdot, \infty)$ and $F_0(\cdot) := H(\infty, \cdot)$ on \mathbb{R} , but not any other feature of their joint distribution $H(\cdot, \cdot)$ on \mathbb{R}^2 .

We represent this assumption in a potentially nonstandard way. Specifically, we assume the the researcher observes not (Y_1^*, Y_0^*) but $(Y_1, Y_0) : [0, 1] \rightarrow \mathbb{R}^2$ where Y_t is equivalent to Y_t^* up to an unknown measure preserving transformation. That is,

$$Y_t(\varphi_t(u)) = Y_t^*(u) \tag{3}$$

for some unknown $\varphi_t \in \mathcal{M} := \{\phi : [0, 1] \rightarrow [0, 1] \text{ with } |\phi^{-1}(A)| = |A| \text{ for any measurable } A \subseteq [0, 1]\}$. Intuitively, Y_t is a rearranged version of Y_t^* so that the two have the same marginal distribution, but no other feature of the joint distribution of Y_0^* and Y_1^* can be learned from Y_0 and Y_1 . To be sure, (3) holds when Y_t is the inverse marginal distribution (quantile) function associated with Y_t^* (see Whitt 1976, Theorem 5.1), and so it is implied by the usual assumption. We do not formally reference the marginal distribution or quantile function of Y_t^* in our version of the econometric problem, however, because there is no natural analog in the double randomized setting. See generally Lovász (2012), Section 7.3.

2.2 Some standard results for the single randomized experiment

We first bound the DPO and DTE following Fréchet (1951); Hoeffding (1940); Makarov (1982). We then show that under a rank invariance assumption the DPO and DTE are point identified and that they can be written as functionals of the quantiles of the outcomes associated with each treatment following Doksum (1974); Lehmann (1975); Whitt (1976). These results are known to the econometrics literature. We state them as a point of reference for our extension to the double randomized experiment in Section 3.

2.2.1 Bounds on the DPO and DTE

Plugging (3) into (1) gives the sharp bounds

$$\begin{aligned} \min_{\varphi_0, \varphi_1 \in \mathcal{M}} \int \prod_{t \in \{0,1\}} \mathbb{1}\{Y_t(\varphi_t(u)) \leq y_t\} du &\leq F(y_1, y_0) \\ &\leq \max_{\varphi_0, \varphi_1 \in \mathcal{M}} \int \prod_{t \in \{0,1\}} \mathbb{1}\{Y_t(\varphi_t(u)) \leq y_t\} du. \end{aligned} \quad (4)$$

These bounds have a simple analytical solution. Let $F_t(y_t) := \int \mathbb{1}\{Y_t(u) \leq y_t\} du$. Then

Standard Result 1: For any $(y_1, y_0) \in \mathbb{R}^2$

$$\max(F_1(y_1) + F_0(y_0) - 1, 0) \leq F(y_1, y_0) \leq \min(F_1(y_1), F_0(y_0)).$$

Standard Result 1 is often attributed to Fréchet (1951); Hoeffding (1940), although our proof sketch in Section 4.2 follows Whitt (1976). The bounds are straightforward to compute or estimate (in cases of sampled, mismeasured, or missing outcomes). See Manski (1997); Heckman et al. (1997) for details and applications to program evaluation in economics.

Bounds on the DPO imply bounds on the DTE. Specifically,

Standard Result 2: For any $y \in \mathbb{R}$

$$\sup_{\substack{(y_1, y_0) \in \mathbb{R}^2: \\ y_1 - y_0 \leq y}} \max(F_1(y_1) - F_0(y_0), 0) \leq \Delta(y) \leq 1 + \inf_{\substack{(y_1, y_0) \in \mathbb{R}^2: \\ y_1 - y_0 \leq y}} \min(F_1(y_1) - F_0(y_0), 0).$$

Standard Result 2 is often attributed to Makarov (1982). These bounds are also straightforward to compute or estimate. See Fan and Park (2010); Firpo and Ridder (2019) for details and applications to program evaluation in economics.

2.2.2 Point identification of the DTE under rank invariance

The Quantile Treatment Effects parameter (QTE) refers to the difference in the quantile functions of Y_1 and Y_0 . Specifically, $QTE(u) := Q_1(u) - Q_0(u)$ where $Q_t(u) := \inf\{y \in \mathbb{R} : u \leq F_t(y)\}$ is the inverse distribution (quantile) function associated with Y_t^* , or equivalently, Y_t . It is often attributed to Doksum (1974); Lehmann (1975).

Although Q_t and Y_t^* generally have the same marginal distribution for $t \in \{0, 1\}$, $Q_1 - Q_0$ and $Y_1^* - Y_0^*$ do not. In fact, the difference in quantiles is a more conservative notion of the effect of treatment as measured, for example, by mean squared error. That is,

Standard Result 3: $\int (Q_1(u) - Q_0(u))^2 du \leq \int (Y_1^*(u) - Y_0^*(u))^2 du.$

See Corollary 2.9 of Whitt (1976).

However, under a rank invariance assumption, $Q_1 - Q_0$ and $Y_1^* - Y_0^*$ do have the same distribution. A treatment effect is rank invariant if $Y_1^* = g(Y_0^*)$ for some nondecreasing $g : \mathbb{R} \rightarrow \mathbb{R}$.

Standard Result 4: $\Delta(y) = \int \mathbb{1}\{Q_1(u) - Q_0(u) \leq y\} du$ under rank invariance.

See Theorem 2.5 of Whitt (1976). The DPO is also identified under rank invariance with $F(y_1, y_0) = \int \prod_{t \in \{0,1\}} \mathbb{1}\{Q_t(u) \leq y_t\} du.$

Intuitively, rank invariance says that the rank of an agent's outcome in the population is the same under both treatments, i.e. $\int \mathbb{1}\{Y_0^*(s) \leq Y_0^*(u)\} ds = \int \mathbb{1}\{Y_1^*(s) \leq Y_1^*(u)\} ds$ for every $u \in [0, 1]$. See for example Abadie et al. (2002); Chernozhukov and Hansen (2005); Firpo (2007) for details and applications of QTE to program evaluation in economics. We suspect that the issues of endogeneity that arise in this work can also be addressed in our setting through the use of control or instrument variables, but we leave this to future work.

3 The double randomized experiment

We extend the framework and results of Section 2 to the double randomized setting.

3.1 Model and econometric problem

3.1.1 Model

Our focus is on symmetric outcome matrices indexed by one population as in Examples 1 and 2 of Section 1.1. Asymmetric matrices or matrices indexed by two different populations as in Examples 3 and 4 are handled by symmetrization. See Section 5.1.1 for details.

A population of agents is randomized to a binary treatment $t \in \{0, 1\}$. The population may be finite or infinite. Outcomes are bounded and measured for each pair of agents.

We extend the model from Section 2.1 to the double randomized setting by defining the potential outcome functions to be indexed by pairs of agent types. Specifically, we define $(Y_1^*, Y_0^*) : [0, 1]^2 \rightarrow \mathbb{R}^2$ such that $(Y_1^*(U, V), Y_0^*(U, V))$ gives the distribution of potential outcomes between a pair of agents, each drawn independently and uniformly at random from the population, when U and V are independent standard uniform. We interpret (Y_1^*, Y_0^*) as describing the outcomes of a continuum of agent types pairs indexed by $[0, 1]^2$. For example, $Y_t^*(u, v)$ may describe the existence of a risk sharing link between households with types u and v in the case that both enroll ($t = 1$) or do not enroll ($t = 0$) in a microfinance program.

Intuitively, the double indexed function Y_t^* represents a random matrix just as its single indexed counterpart from Section 2 represents a random variable. This analogy is well developed in the random matrix literature, see generally Lovász (2012); Graham (2020). In our model, the only source of uncertainty is from the random sampling of agents. Uncertainty due to missing data, measurement error, etc. are incorporated in Section 5.3.

3.1.2 Parameters of interest

We define the DPO and DTE as in the single randomized setting. The DPO is

$$F(y_1, y_0) := P(Y_1^*(w, \tilde{w}) \leq y_1, Y_0^*(w, \tilde{w}) \leq y_0) = \int \int \prod_{t \in \{0, 1\}} \mathbb{1}\{Y_t^*(u, v) \leq y_t\} dudv \quad (5)$$

where $y_1, y_0 \in \mathbb{R}$ and w and \tilde{w} are independent standard uniform random variables. In words, the DPO is the mass of agent type pairs with potential outcome less than y_1 under treatment 1 and less than y_0 under treatment 0.

Similarly, the DTE is

$$\Delta(y) := P(Y_1^*(w, \tilde{w}) - Y_0^*(w, \tilde{w}) \leq y) = \int \int \mathbb{1}\{Y_1^*(u, v) - Y_0^*(u, v) \leq y\} dudv. \quad (6)$$

In words, $Y_1^*(u, v) - Y_0^*(u, v)$ is the change in outcome associated with switching the treatment status of a pair of agents with types u and v from 0 to 1. The DTE is the mass of pairs of

agent types for which this individual treatment effect is less than y .

3.1.3 Econometric problem

As before, the task is to identify the DPO and the DTE. It is also complicated by the fact that the researcher observes at most one potential outcome for any pair of agents.

For example, the researcher may assign households to participate or not participate in a microfinance program. They observe whether pairs of participating households form a risk sharing link under the treatment that both households participate in the program. They do not observe whether these households would have formed a link under the counterfactual treatment that neither household participates. To infer this missing potential outcome, the researcher must use the risk sharing links between the nonparticipating households.

As in Section 2.1.3, we formalize the econometric problem by assuming that the researcher observes not (Y_1^*, Y_0^*) but $(Y_1, Y_0) : [0, 1]^2 \rightarrow \mathbb{R}^2$ where Y_t is equivalent to Y_t^* up to an unknown measure preserving transformation. That is,

$$Y_t(\varphi_t(u), \varphi_t(v)) = Y_t^*(u, v) \tag{7}$$

for some unknown $\varphi_t \in \mathcal{M}$. Like the single randomized setting, (7) says that Y_t and Y_t^* represent the same random object. However, Y_0 and Y_1 do not reveal any additional information about how the entries of Y_0^* and Y_1^* are related. Unlike the single randomized setting, there is no canonical Y_t that serves the role of the inverse marginal distribution (quantile) function in the double randomized setting. The “marginal distribution of Y_t^* ” is instead an equivalence class of functions Y_t that satisfy (7). This construction is sometimes called a graphon in the random matrix theory literature, see for example Lovász (2012), Section 7.3.

3.2 Some new results for the double randomized experiment

We first bound the DPO and DTE. We then propose a new matrix generalization of rank invariance under which the DPO and DTE are point identified and can be written as functionals of the eigenvalues of the potential outcome functions associated with each treatment. Eigenvalues of functions are defined a bit differently than their matrix counterparts, see

Appendix A.2 or Lovász (2012), Section 7.5 for a review.

3.2.1 Bounds on the DPO and DTE

As in the single randomized setting, plugging (7) into (5) gives sharp bounds on the DPO

$$\begin{aligned} \min_{\varphi_0, \varphi_1 \in \mathcal{M}} \int \int \prod_{t \in \{0,1\}} \mathbb{1}\{Y_t(\varphi_t(u), \varphi_t(v)) \leq y_t\} dudv &\leq F(y_1, y_0) \\ &\leq \max_{\varphi_0, \varphi_1 \in \mathcal{M}} \int \int \prod_{t \in \{0,1\}} \mathbb{1}\{Y_t(\varphi_t(u), \varphi_t(v)) \leq y_t\} dudv. \end{aligned} \quad (8)$$

We do not consider these bounds, however, because their quadratic structure makes them analytically and computationally intractable in general. See Cela (2013), Section 1.5.

We instead propose bounds that are not generally sharp but are tractable. For any $y_t \in \mathbb{R}$ and $t \in \{0, 1\}$, let $\lambda_{1t}(y_t), \dots, \lambda_{Rt}(y_t)$ be the R largest (in absolute value) eigenvalues of $\mathbb{1}\{Y_t(\cdot, \cdot) \leq y_t\}$ ordered to be decreasing and $s_R(r) = R - r$. For any $t, t' \in \{0, 1\}$, let $\sum_r \lambda_{rt} \lambda_{rt'}$ and $\sum_r \lambda_{rt} \lambda_{s(r)t'}$ refer to $\sum_{r=1}^{\bar{R}} \lambda_{rt} \lambda_{rt'}$ and $\sum_{r=1}^{\bar{R}} \lambda_{rt} \lambda_{s_{\bar{R}}(r)t'}$ where $\bar{R} = \max\{R \in \mathbb{N} \cup \{\infty\} : \max_{t \in \{0,1\}} \min_{r \in \{1, \dots, R\}} |\lambda_{rt}| > 0\}$.

In words, \bar{R} is the larger of the number of nonzero eigenvalues of $\mathbb{1}\{Y_1 \leq y_1\}$ and $\mathbb{1}\{Y_0 \leq y_0\}$. When $\mathbb{1}\{Y_1 \leq y_1\}$ and $\mathbb{1}\{Y_0 \leq y_0\}$ have the same number of nonzero eigenvalues and it is finite, then the sum $\sum_r \lambda_{rt} \lambda_{rt'}$ is the product of the largest eigenvalues plus the product of the second largest eigenvalues, and so on. The sum $\sum_r \lambda_{rt} \lambda_{s(r)t'}$ is the product of the largest eigenvalue of $\mathbb{1}\{Y_t \leq y_t\}$ and the smallest eigenvalue of $\mathbb{1}\{Y_{t'} \leq y_{t'}\}$ plus the product of the second largest eigenvalue of $\mathbb{1}\{Y_t \leq y_t\}$ and the second smallest eigenvalue of $\mathbb{1}\{Y_{t'} \leq y_{t'}\}$, and so on. When $\mathbb{1}\{Y_t \leq y_t\}$ has more nonzero eigenvalues than $\mathbb{1}\{Y_{t'} \leq y_{t'}\}$, we add 0s to the sequence of eigenvalues for $\mathbb{1}\{Y_{t'} \leq y_{t'}\}$ until they are the same size. When one or both of the functions have an infinite number of nonzero eigenvalues, we apply the above algorithm to the R largest nonzero eigenvalues (in absolute value) of each function, and then take R to infinity.

Our first result is

Proposition 1: For any $(y_1, y_0) \in \mathbb{R}^2$

$$\begin{aligned} \max \left(\sum_r (\lambda_{r1}^2 + \lambda_{r0}^2) - 1, \sum_r \lambda_{r1} \lambda_{s(r)0}, 0 \right) &\leq F(y_1, y_0) \\ &\leq \min \left(\sum_r \lambda_{r1}^2, \sum_r \lambda_{r0}^2, \sum_r \lambda_{r1} \lambda_{r0} \right). \end{aligned} \quad (9)$$

We defer a discussion of Proposition 1 to Section 3.4, only remarking here that unlike the infeasible bounds in (8), those in (9) are straightforward to compute because they only depend on the eigenvalues of $\mathbb{1}\{Y_t^* \leq y_t\}$, or equivalently, $\mathbb{1}\{Y_t \leq y_t\}$. They can be computed or estimated (in cases of sampled, mismeasured, or missing outcomes) using standard tools, see Section 5.1.3.

Following Makarov (1982), (9) also implies bounds on the DTE. Our second result is

Proposition 2: For any $y \in \mathbb{R}$

$$\begin{aligned} \sup_{\substack{(y_1, y_0) \in \mathbb{R}^2: \\ y_1 - y_0 \leq y}} \max \left(\sum_r (\lambda_{r1}^2 - \lambda_{r0}^2), \sum_r (\lambda_{r1}^2 - \lambda_{r1} \lambda_{r0}), 0 \right) &\leq \Delta(y) \\ &\leq 1 + \inf_{\substack{(y_1, y_0) \in \mathbb{R}^2: \\ y_1 - y_0 \leq y}} \min \left(\sum_r (\lambda_{r1}^2 - \lambda_{r0}^2), \sum_r (\lambda_{r1} \lambda_{r0} - \lambda_{r0}^2), 0 \right) \end{aligned} \quad (10)$$

where $\lambda_{11}(y_1), \lambda_{21}(y_1), \dots$ and $\lambda_{10}(y_0), \lambda_{20}(y_0), \dots$ are implicitly functions of y_1 and y_0 . In finite data, these bounds only require the researcher to compute eigenvalues for at most $N(N+1)$ values of y_1 and y_0 where N is the number of sampled agents. Optimizing over a smaller set gives valid but potentially wider bounds.

3.2.2 Definition of spectral treatment effects

We propose a matrix analog of the QTE. Let $\{\sigma_{rt}\}_{r=1}^R$ be the R largest (in absolute value) eigenvalues of Y_t ordered to be decreasing and $\{\phi_r\}_{r=1}^R$ be any orthogonal basis in $L^2([0, 1])$.

Definition 1: The Spectral Treatment Effects parameter (STE) is

$$STE(u, v; \phi) := \lim_{R \rightarrow \infty} \sum_{r=1}^R (\sigma_{r1} - \sigma_{r0}) \phi_r(u) \phi_r(v). \quad (11)$$

The STE is similar to the diagonalized difference in the eigenvalues of Y_1 and Y_0 , but its exact values depend on a choice of basis. Two natural choices are the eigenfunctions of Y_1 and Y_0 , denoted $\{\phi_{r1}\}_{r=1}^R$ and $\{\phi_{r0}\}_{r=1}^R$ respectively. We call $STE(\phi_1)$ and $STE(\phi_0)$ the Spectral Treatment Effects on the Treated (STT) and Untreated (STU). In words, the STT takes the observed matrix Y_1 and subtracts the counterfactual formed by keeping the eigenfunctions of Y_1 and inserting the eigenvalues of Y_0 . That is,

$$\begin{aligned} STT(u, v) &= Y_1(u, v) - \lim_{R \rightarrow \infty} \sum_{r=1}^R \sigma_{r0} \phi_{r1}(u) \phi_{r1}(v). \\ &= Y_1(u, v) - \int \int Y_0(s, t) W(u, v; s, t) ds dt \end{aligned}$$

where $W(u, v; s, t) = \sum_r \phi_{r1}(u) \phi_{r1}(v) \phi_{r0}(s) \phi_{r0}(t)$. The second line suggests an alternative interpretation of the STT where the counterfactual outcome for a pair of agents assigned to treatment 1 is formed by a W -weighted average of the outcomes of agent pairs assigned to treatment 0. The weights are generally extrapolative in the sense that they may be negative and do not necessarily integrate to 1. In some cases the researcher may wish to explicitly restrict the weights so that they satisfy these properties. We describe one way to do this in Online Appendix Section D.3, but leave a formal study to future work. The weights will also necessarily satisfy these properties under the rank invariance condition below.

The STT is analogous to the QTE which imputes a counterfactual for an agent assigned to treatment 1 by using the outcome of a similarly ranked agent assigned to treatment 0. In this analogy, the eigenfunctions serve the role of the rank of an agent and the eigenvalues serve the role of the quantiles. We first show that the STE satisfies conditions analogous to Standard Results 3 and 4 in Section 2.2.2. We then discuss some behavioral motivations for the parameter using examples from the economics literature.

3.2.3 Point identification of the DTE under rank invariance

Like the QTE, the STE is also a more conservative notion of the effect of treatment than $Y_1^* - Y_0^*$ as measured by mean squared error. Our third result is

Proposition 3: For any orthogonal basis $\{\phi_r\}_{r=1}^\infty$

$$\int \int STE(u, v; \phi)^2 dudv \leq \int \int (Y_1^*(u, v) - Y_0^*(u, v))^2 dudv. \quad (12)$$

In addition, under a rank invariance assumption, the STT, STU, and $Y_1^* - Y_0^*$ all have the same distribution. To define rank invariance in the double randomization setting, we extend the notion of a nondecreasing function to matrices following Horn and Johnson (1991), Chapter 6.1. For any $f : \mathbb{R} \rightarrow \mathbb{R}$ that admits the representation $f(x) = \sum_{r=1}^\infty c_r x^r$ and square matrix A (or function $A : [0, 1]^2 \rightarrow \mathbb{R}$), we define the matrix lift of f to be $f(A) = \sum_{r=1}^\infty c_r A^r$ where A^r is the r th matrix (or function) power of A , i.e. $A_{ij}^r = \sum_{t_1} \sum_{t_2} \dots \sum_{t_{r-1}} A_{it_1} A_{t_1 t_2} \dots A_{t_{r-1} j}$. Our matrix generalization of rank invariance is then

Definition 2: A treatment effect is rank invariant if $Y_1^* = g(Y_0^*)$ where g is the matrix lift of some nondecreasing $g : \mathbb{R} \rightarrow \mathbb{R}$.

We call Definition 2 a matrix generalization of rank invariance because it implies the definition from Section 2 when Y_1^* and Y_0^* are scalars. Our fourth result is

Proposition 4: Under rank invariance,

$$\Delta(y) = \int \int \mathbb{1}\{STT(u, v) \leq y\} dudv = \int \int \mathbb{1}\{STU(u, v) \leq y\} dudv. \quad (13)$$

Intuitively, if we think of the treatment working by taking in Y_0^* and producing $Y_1^* = g(Y_0^*)$, then rank invariance implies that the treatment affects the eigenvalues but not the eigenfunctions of the outcome matrix Y_0^* . This is analogous to rank invariance in the single randomization setting, where the treatment affects the quantiles but not the ranks of the outcome vector. We give examples where this assumption may be reasonable below.

3.2.4 Interpretation of the STE and rank invariance

Though rank invariance is a strong assumption, many economic models imply rank invariant treatment effects. We provide examples from the literature on information diffusion, social interaction, and community detection in Appendix Section B.

We also interpret the STE and rank invariance assumption using an orthogonal factor model. In this model, the outcome between two agents is determined by a weighted combination of latent orthogonal factors $Y_t(u, v) = \sum_{r=1}^R \rho_{rt} f_{rt}(u) f_{rt}(v)$. For example, the existence of a risk sharing connection between two households might depend on household size, socioeconomic status, physical location, etc. The weight that a household with type u places on the r th such factor is given by $f_{rt}(u)$ where $\int f_{rt}(u) f_{st}(u) du = 0$ if $r \neq s$. Intuitively, this orthogonality condition says that the factors are uncorrelated across the agent types. The model is linear in the product of the agent pairs. That is, the marginal impact of a unit change in the product $f_{rt}(u) f_{rt}(v)$ on the outcome Y_t is described by the factor weight ρ_{rt} .

The STE contrasts the factor weights ρ_{rt} across the two treatment groups, ignoring any difference in the corresponding factors $f_{rt}(u)$. In some cases, a policy maker may only be interested in the factor weights and so this focus is justified, even if rank invariance does not hold. For example, in the setting of Golub and Jackson (2010), under certain conditions the rate of convergence of the beliefs of a collection of naïve learners to a consensus is given by the second largest eigenvalue of a matrix of interactions. By comparing the factor weights of the adjacency matrices under various treatments, the policy maker can learn the impact of the treatment on this particular outcome.

Rank invariance justifies this focus on the difference in factor weights $\rho_{r1} - \rho_{r0}$, because it implies that the treatment works by shifting the factor weights but not the factors themselves. For example, a microfinance treatment might not alter household size, socioeconomic status, physical location, etc. Instead, it only changes how important these factors are in determining the risk sharing connections. For example, the treatment might crowd out some risk sharing between some wealthy households and so decrease the weight associated with the socioeconomic factor. If the ranks of the factor weights are not changed, then this treatment effect would satisfy rank invariance.

4 Sketch and discussion of the proof of Proposition 1

We show some of the main technical ideas of the paper by sketching the proof of Proposition 1. To simplify arguments we consider a finite population along the lines of Whitt (1976);

Heckman et al. (1997), Section 3. A full proof can be found in Appendix A.

4.1 Finite approximation

To simplify our sketch, we assume in this section that Y_t^* has finite dimension. For the single randomized experiment, we assume that Y_t^* is an N dimensional vector, the DPO is $\frac{1}{N} \sum_{i=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{i,t}^* \leq y_t\}$, and $Y_{i,t} = \sum_{j=1}^N Y_{j,t}^* P_{ij,t}$ is observed where P_t is an unknown permutation matrix. That is, P_t has $\{0, 1\}$ valued entries. Rows and columns sum to 1.

Intuitively, there are N types of agents. One agent of each type is assigned to treatment 1 and one agent of each type is assigned to treatment 0. Our task is to compare the outcomes of agents of the same type but different treatment assignment, but we do not know which agent is of which type. Bounds on the DPO are given by maximizing and minimizing $\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{j,t} \leq y_t\} P_{ij,t}$ over all permutation matrices P_0 and P_1 . That this problem is a good approximation to (4) is outlined in Section 2 of Whitt (1976). See also Heckman et al. (1997), Section 3.

For the double randomized experiment, we assume that Y_t^* is an $N \times N$ dimensional matrix, the DPO is $\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{ij,t}^* \leq y_t\}$, and $Y_{ij,t} = \sum_{k=1}^N \sum_{l=1}^N Y_{kl,t}^* P_{ik,t} P_{jl,t}$ is observed where P_t is an unknown permutation matrix. The intuition is the same as in the single randomized experiment. There are N types of agents, one agent of each type is assigned to each treatment, and though we would like to compare the outcomes of agents with the same type but different treatment assignments, we do not know which agent is of which type. The bounds on the DPO formed by maximizing and minimizing $\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{kl,t} \leq y_t\} P_{ik,t} P_{jl,t}$ over P_0 and P_1 are intractable “Quadratic Assignment Problems” or QAPs, see generally Cela (2013). Our bounds are instead based on conservative but tractable relaxations of these bounds.

4.2 Standard Result 1

The DPO for the finite approximation to the single randomized experiment is $F_N(y_1, y_0) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{j,t} \leq y_t\} P_{i,j,t}$. We show that

$$\max(F_{N1}(y_1) + F_{N0}(y_0) - 1, 0) \leq F_N(y_1, y_0) \leq \min(F_{N1}(y_1), F_{N0}(y_0)).$$

where $F_{Nt}(y_t) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{Y_{i,t} \leq y_t\}$ following Whitt (1976), Theorem 2.1. Our sketch relies on the following “rearrangement inequality” due to Hardy et al. (1952), Theorem 368.

Theorem 368 (Hardy-Littlewood-Polya): For any $m \in \mathbb{N}$ and $g, h \in \mathbb{R}^m$ we have $\sum_r g_{(r)} h_{(m-r+1)} \leq \sum_r g_r h_r \leq \sum_r g_{(r)} h_{(r)}$ where $g_{(r)}$ is the r th order statistic of g .

4.2.1 Sketch of proof

Theorem 368 implies that

$$\sum_{i=1}^N \mathbb{1}\{Y_{(N-i+1),0} \leq y_0\} \mathbb{1}\{Y_{(i),1} \leq y_1\} \leq N F_N(y_1, y_0) \leq \sum_{i=1}^N \mathbb{1}\{Y_{(i),0} \leq y_0\} \mathbb{1}\{Y_{(i),1} \leq y_1\}$$

where $Y_{(i),t}$ is the i th order statistic of Y_t . The upper bound follows

$$\sum_{i=1}^N \mathbb{1}\{Y_{(i),0} \leq y_0\} \mathbb{1}\{Y_{(i),1} \leq y_1\} \leq \min_{t \in \{0,1\}} \sum_{i=1}^N \mathbb{1}\{Y_{i,t} \leq y_t\}.$$

The lower bound follows

$$\begin{aligned} \sum_{i=1}^N \mathbb{1}\{Y_{(N-i+1),0} \leq y_0\} \mathbb{1}\{Y_{(i),1} \leq y_1\} &= \sum_{i=1}^N (1 - \mathbb{1}\{Y_{(N-i+1),0} > y_0\}) \mathbb{1}\{Y_{(i),1} \leq y_1\} \\ &\geq \sum_{i=1}^N \mathbb{1}\{Y_{i,1} \leq y_1\} - \min \left(\sum_{i=1}^N \mathbb{1}\{Y_{i,1} \leq y_1\}, \sum_{i=1}^N \mathbb{1}\{Y_{i,0} > y_0\} \right) \\ &= \max \left(\sum_{i=1}^N \mathbb{1}\{Y_{i,1} \leq y_1\} + \sum_{i=1}^N \mathbb{1}\{Y_{i,0} \leq y_0\} - N, 0 \right). \end{aligned}$$

4.3 Proposition 1

The DPO for the finite approximation to the double randomized experiment is

$$F_N(y_1, y_0) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{kl,t} \leq y_t\} P_{ik,t} P_{jl,t}.$$

We show that

$$\begin{aligned} \max \left(\sum_{r=1}^N (\lambda_{r1}^2 + \lambda_{r0}^2) - N^2, \sum_{r=1}^N \lambda_{r1} \lambda_{s_N(r)0}, 0 \right) &\leq N^2 F_N(y_1, y_2) \\ &\leq \min \left(\sum_{r=1}^N \lambda_{r1}^2, \sum_{r=1}^N \lambda_{r0}^2, \sum_{r=1}^N \lambda_{r1} \lambda_{r0} \right) \end{aligned}$$

where λ_{rt} is the r th eigenvalue of $\mathbb{1}\{Y_t \leq y_t\}$ and $s_N(r) = N - r + 1$. Our sketch has two parts. The second part relies on the following result due to Birkhoff (1946); von Neumann (1953). We say that a matrix is doubly stochastic if its entries are positive and every row and column sums to 1.

Theorem (Birkhoff): If M is doubly stochastic then there exist an $m \in \mathbb{N}$, $\alpha_1, \dots, \alpha_m > 0$, and permutation matrices P_1, \dots, P_m such that $\sum_{t=1}^m \alpha_t = 1$ and $M_{ij} = \sum_{t=1}^m \alpha_t P_{ij,t}$.

4.3.1 Sketch of proof, part 1

We first show $\max \left(\sum_{r=1}^N (\lambda_{r1}^2 + \lambda_{r0}^2) - N^2, 0 \right) \leq N^2 F_N(y_1, y_0) \leq \min \left(\sum_{r=1}^N \lambda_{r1}^2, \sum_{r=1}^N \lambda_{r0}^2 \right)$. Write $N^2 F_N(y_1, y_0) = \sum_{r=1}^{N^2} \sum_{s=1}^{N^2} \prod_{t \in \{0,1\}} \mathbb{1}\{\tilde{Y}_{r,t} \leq y_t\} \tilde{P}_{rs,t}$ where $i_r = \lfloor \frac{r-1}{N} \rfloor + 1$, $j_r = r - N \lfloor \frac{r-1}{N} \rfloor$, $\tilde{Y}_{r,t} = Y_{i_r j_r, t}$, and $\tilde{P}_{rs,t} = P_{i_r i_s, t} P_{j_r j_s, t}$. In words, \tilde{Y}_t and \tilde{P}_t are vectorized versions of Y_t and $P_t \times P_t$ formed by appending their rows. Theorem 368 implies that

$$\sum_{r=1}^{N^2} \mathbb{1}\{\tilde{Y}_{(N^2-r+1),0} \leq y_0\} \mathbb{1}\{\tilde{Y}_{(r),1} \leq y_1\} \leq N^2 F_N(y_1, y_0) \leq \sum_{r=1}^{N^2} \mathbb{1}\{\tilde{Y}_{(r),0} \leq y_0\} \mathbb{1}\{\tilde{Y}_{(r),1} \leq y_1\}$$

and so following the arguments of Section 4.2.1, we have

$$\max \left(\sum_{r=1}^{N^2} \mathbb{1}\{\tilde{Y}_{r,1} \leq y_1\} + \sum_{r=1}^{N^2} \mathbb{1}\{\tilde{Y}_{r,0} \leq y_0\} - N^2, 0 \right) \leq N^2 F_N(y_1, y_0) \leq \min_{t \in \{0,1\}} \sum_{r=1}^{N^2} \mathbb{1}\{\tilde{Y}_{r,t} \leq y_t\}.$$

The bounds follow since

$$\sum_{r=1}^{N^2} \mathbb{1}\{\tilde{Y}_{r,t} \leq y_t\} = \sum_{i=1}^N \sum_{j=1}^N \mathbb{1}\{Y_{ij,t} \leq y_t\} = \sum_{r=1}^N \lambda_{rt}^2.$$

4.3.2 Sketch of proof, part 2

We now show $\sum_{r=1}^N \lambda_{r1} \lambda_{s_N(r)0} \leq N^2 F_N(y_1, y_2) \leq \sum_{r=1}^N \lambda_{r1} \lambda_{r0}$. These bounds follow

$$\sum_{i,j=1}^N \sum_{k,l=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{kl,t} \leq y_t\} P_{ik,t} P_{jl,t} = \sum_{i,j=1}^N \left[\sum_{r=1}^N \lambda_{r1} \phi_{ir,1} \phi_{jr,1} \right] \left[\sum_{s=1}^N \lambda_{s0} \phi_{is,0} \phi_{js,0} \right] = \sum_{r,s=1}^N \lambda_{r1} \lambda_{s0} W_{rs}^\phi$$

where $(\lambda_{rt}, \phi_{rt})$ is the r th eigenvalue and eigenvector pair of $\sum_{k,l=1}^N \mathbb{1}\{Y_{kl,t} \leq y_t\} P_{ik,t} P_{jl,t}$ and W^ϕ , a matrix with rs th entry $W_{rs}^\phi = \left[\sum_{i=1}^N \phi_{ir,1} \phi_{is,0} \right]^2$, is the Hadamard square of the product of two orthogonal matrices and so is doubly stochastic. Birkhoff's Theorem implies

$$\sum_{i,j=1}^N \sum_{k,l=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{kl,t} \leq y_t\} P_{ik,t} P_{jl,t} = \sum_{r,s=1}^N \lambda_{r1} \lambda_{s0} W_{rs}^\phi = \sum_{k=1}^K \alpha_k^\phi \sum_{r,s=1}^N \lambda_{r1} \lambda_{s0} P_{rs,k}^\phi.$$

where $\alpha_1^\phi, \dots, \alpha_K^\phi > 0$, $P_1^\phi, \dots, P_K^\phi$ are permutation matrices, and $W_{rs}^\phi = \sum_{k=1}^K \alpha_k^\phi P_{rs,k}^\phi$. The claim follows from Theorem 368, which implies that for any permutation matrix P

$$\sum_{r=1}^N \lambda_{r1} \lambda_{s_N(r)0} \leq \sum_{r,s=1}^N \lambda_{r1} \lambda_{s0} P_{rs} \leq \sum_{r=1}^N \lambda_{r1} \lambda_{r0}.$$

4.3.3 Discussion

Proposition 1 follows by intersecting the bounds from parts 1 and 2. Each part describes a different relaxation of the intractable QAP. Take for instance the upper bound

$$\max_{P \in \mathcal{P}_N} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{kl,t} \leq y_t\} P_{ik,t} P_{jl,t}. \quad (14)$$

where \mathcal{P}_N is the set of all $N \times N$ permutation matrices. Part 1 bounds it from above with

$$\max_{P \in \mathcal{P}_{N^2}} \frac{1}{N^2} \sum_{i,j=1}^N \sum_{k,l=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{kl,t} \leq y_t\} P_{r(i,j)r(k,l),t} \quad (15)$$

where \mathcal{P}_{N^2} describes permutations of pairs of agents and $r(i, j) = N(i - 1) + j$. Intuitively, this relaxation treats the $N \times N$ outcome matrices as vectors of length N^2 and uses the fact that $\{P_{ij}\}_{i,j=1}^N \in \mathcal{P}_N$ implies that $\{P_{ik}P_{jl}\}_{i,j,k,l=1}^N \in \mathcal{P}_{N^2}$. Whereas the (14) quadratic, (15) is linear and can be bounded using Theorem 368.

Part 2 bounds the QAP from above with

$$\max_{O \in \mathcal{O}_N} \frac{1}{N^2} \sum_{i,j=1}^N \sum_{k,l=1}^N \prod_{t \in \{0,1\}} \mathbb{1}\{Y_{kl,t} \leq y_t\} O_{ik,t} O_{jl,t} \quad (16)$$

where \mathcal{O}_N is the set of orthogonal $N \times N$ matrices. This is an upper bound because $\mathcal{P}_N \subset \mathcal{O}_N$. The insight, see Finke et al. (1987), is to use Birkhoff's Theorem to rewrite the problem as $\max_{W \in \mathcal{D}_N^+} \sum_{r,s} \lambda_{r1} \lambda_{rs} W_{rs}$ where \mathcal{D}_N^+ is the set of doubly stochastic $N \times N$ matrices. This problem is also linear and can be bounded using Birkhoff's Theorem and Theorem 368.

Our proof of Proposition 1 as stated in Section 3 is complicated by the fact that the infinite dimensional analog of W^ϕ is not doubly stochastic and so Birkhoff's Theorem cannot be directly applied. We address this problem by first approximating the function $\mathbb{1}\{Y_t \leq y_t\}$ with a fixed rank matrix, applying the arguments of Part 2, and then showing convergence as the rank is taken to infinity along the lines of Lovász (2012). A similar method of proof is used to demonstrate Theorem 2.1 of Whitt (1976) (i.e. our Standard Result 1).

These bounds use two of many possible relaxations of the intractable QAP, see broadly Cela (2013), Section 2. We chose these relaxations because they are straightforward to compute, characterize, and appear to work well in practice. Intersecting our bounds with others may lead to smaller identified sets for the DPO and DTE, but potentially at the cost of greater computational complexity or statistical uncertainty. We leave this to future work.

5 Extensions

We sketch some extensions to the Section 3 framework. Details are in Appendix Section D.

5.1 Asymmetric outcome matrices

Asymmetric matrices or matrices indexed by two different populations can be handled in the following way. A population of workers and firms are indexed by latent types in $[0, 1]$ and $[2, 3]$ respectively. Workers and firms are independently randomized to a binary treatment $t \in \{0, 1\}$ and outcomes are measured for every pair of workers and firms assigned to the same treatment. The bounded measurable function $(Y_0^*, Y_1^*) : S \rightarrow \mathbb{R}^2$ records the potential outcome for every type pair and treatment where $S = [0, 1] \times [2, 3]$. For example, $Y_t^*(u, v)$ may describe the wage that a worker of type u would earn at a firm of type v in the case that the worker and firm are in a market that was affected ($t = 1$) or not affected ($t = 0$) by a trade shock. We assume that the researcher observes Y_1 and Y_0 where $Y_t(\phi_t(u), \psi_t(v)) = Y_t^*(u, v)$ for unknown measure preserving ϕ_t, ψ_t . The DPO, for example, is $\int \int \prod_{t \in \{0, 1\}} \mathbb{1}\{Y_t(\phi_t(u), \psi_t(v)) \leq y_t\} dudv$.

We symmetrize the potential outcomes as in Auerbach (2019). Let $S^2 = ([0, 1] \cup [2, 3]) \times ([0, 1] \cup [2, 3])$ and define $(Y_0^\dagger, Y_1^\dagger) : S^2 \rightarrow \mathbb{R}^2$ so that

$$Y_t^\dagger(u, v) = \begin{cases} Y_t(u, v) & \text{if } (u, v) \in [0, 1] \times [2, 3] \\ Y_t(v, u) & \text{if } (u, v) \in [2, 3] \times [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

and $\varphi_t(u) = \phi_t(u)\mathbb{1}\{u \in [0, 1]\} + \psi_t(u)\mathbb{1}\{u \in [2, 3]\}$. Then the DPO is also equal to $\int \int \prod_{t \in \{0, 1\}} \mathbb{1}\{Y_t^\dagger(\varphi_t(u), \varphi_t(v)) \leq y_t\} dudv/2$. Since Y^\dagger is symmetric and defined on one population, the logic of Section 3 can be applied the bound the DPO and DTE. One can similarly define the STE using the eigenvalues of Y_t^\dagger .

5.2 Row and column heterogeneity

Spectral methods may perform poorly when there is significant heterogeneity in the rows or columns of the outcome matrices. To address this issue we follow Finke et al. (1987) and decompose $\mathbb{1}\{Y_t^*(u, v) \leq y_t\} = \alpha_t(u) + \gamma_t(v) + \epsilon_t(u, v)$ where $\int \epsilon_t(u, v) du = \int \epsilon_t(u, v) dv = 0$. The DPO can then be written

$$\begin{aligned} F(y_1, y_0) &= \int \int \prod_{t \in \{0,1\}} (\alpha_t(\varphi_t(u)) + \gamma_t(\varphi_t(v)) + \epsilon_t(\varphi_t(u), \varphi_t(v))) dudv \\ &= \int \int \prod_{t \in \{0,1\}} (\alpha_t(\varphi_t(u)) + \gamma_t(\varphi_t(v))) dudv + \int \int \prod_{t \in \{0,1\}} \epsilon_t(\varphi_t(u), \varphi_t(v)) dudv. \end{aligned}$$

The first summand $\int \int \prod_{t \in \{0,1\}} (\alpha_t(\varphi_t(u)) + \gamma_t(\varphi_t(v))) dudv$ can be bounded using arguments from Section 2. The second summand $\int \int \prod_{t \in \{0,1\}} \epsilon_t(\varphi_t(u), \varphi_t(v)) dudv$ can be bounded using arguments from Section 3. One can similarly decompose $Y_t^*(u, v)$ and redefine the STE by matching on the quantiles of α_t , γ_t , and the eigenvalues of ϵ_t .

5.3 Estimation and inference

If the researcher observes only a noisy signal of Y_1 and Y_0 due to random sampling, missing data, or measurement error, they can estimate the STE or bounds on the DPO and DTE by replacing the eigenvalues of the Y_t with their empirical analogs. We give sufficient conditions for consistency and propose inference based on Weyl's inequality in Appendix D.

An alternative design based approach to inference exploits the random assignment of agents to treatment. We show how the researcher may, for example, test the null hypothesis of no treatment effects or bound the variance of any treatment effect in Appendix D.

5.4 Interference

One motivation for implementing a double randomized experimental design is to characterize social interactions, market externalities, or other spillovers. Our bounds and estimates of the treatment effect can be used to characterize various heterogeneous spillover effects. The kinds of spillover effects that are identified depends on the experimental design, see Bajari

et al. (2021). The following illustrative example is based on their Section 6.

Consider the setting of Example 4, and suppose the researcher is interested in characterizing how an information treatment assigned to the buyers (sellers) affects their transactions with the untreated sellers (buyers). To do this, they separately and independently assign the buyers and sellers to four groups. In the first group both the buyers and the sellers are treated, in the second group the buyers and not the sellers are treated, in the third group the sellers and not the buyers are treated, and in the fourth group neither the sellers nor the buyers are treated. Bajari et al. (2021) define the average buyer (seller) spillover effect to be the average difference in the outcomes in the second (third) and fourth groups. Following the symmetrization argument of Section 5.1, the arguments of Section 3 can be used to identify and estimate the distribution of such spillover effects.

6 Two empirical demonstrations

We revisit Examples 1 and 3 from Section 1.1. We find policy relevant heterogeneity in the effect of treatment that otherwise might be missed by focusing exclusively on average effects.

6.1 Example 1: Risk sharing

Our first demonstration uses data from Comola and Prina (2021). Households in nineteen villages are randomly provided with a savings account. The authors argue that this treatment alters incentives for some agents to form risk sharing links.¹ Let treatment 1 be the event that both households are provided with savings accounts, treatment 0 be the event that neither household is provided with a savings account, and $Y_{ij,t}$ indicate whether household pair ij reports a risk sharing link under treatment t .

Table 1 shows our bounds on the joint distribution of risk sharing links for villages 3 and 7. We find positive lower bounds on $P(Y_{ij,1} = 1, Y_{ij,0} = 0)$ for both villages, implying that the treatment created some links. This is consistent with the positive average treatment effect found by Comola and Prina (2021). However, we also find a positive lower bound on $P(Y_{ij,1} = 0, Y_{ij,0} = 1)$ for village 7. This implies that the treatment destroyed some links.

¹The data can be found at...

	Village 3		Village 7	
	Lower	Upper	Lower	Upper
$P(Y_{ij,1} = 1, Y_{ij,0} = 1)$	0.0000	0.0048	0.0000	0.0128
$P(Y_{ij,1} = 1, Y_{ij,0} = 0)$	0.0012	0.0059	0.0009	0.0139
$P(Y_{ij,1} = 0, Y_{ij,0} = 1)$	0.0000	0.0048	0.0023	0.0153
$P(Y_{ij,1} = 0, Y_{ij,0} = 0)$	0.9893	0.9941	0.9708	0.9838

Table 1: Bounds on the joint distribution of risk sharing links in villages 3 and 7 of Comola and Prina (2021).

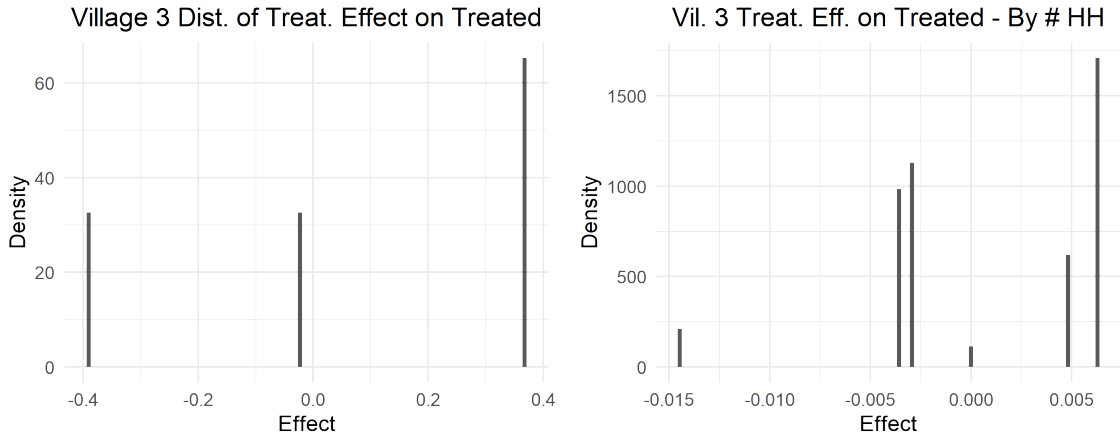
Figure 1 shows the distribution of spectral treatment effects on the treated for villages 3 and 7 in panels (a) and (c). For reference, we also show the distribution of average treatment effects conditional on household size. Specifically, we bin households by household size, compute the fraction of links between households for each treatment and pair of bins, and show the distribution of the differences in averages across treatment in panels (b) and (d). These CATEs are small and economically insignificant for every household pair. Our STTs, in contrast, are economically significant with changes on the order of thirty to forty percent for approximately one to five percent of the population. The strong response of some agent pairs to treatment may be missed by a researcher only looking at average effects.

6.2 Example 3: Auction format

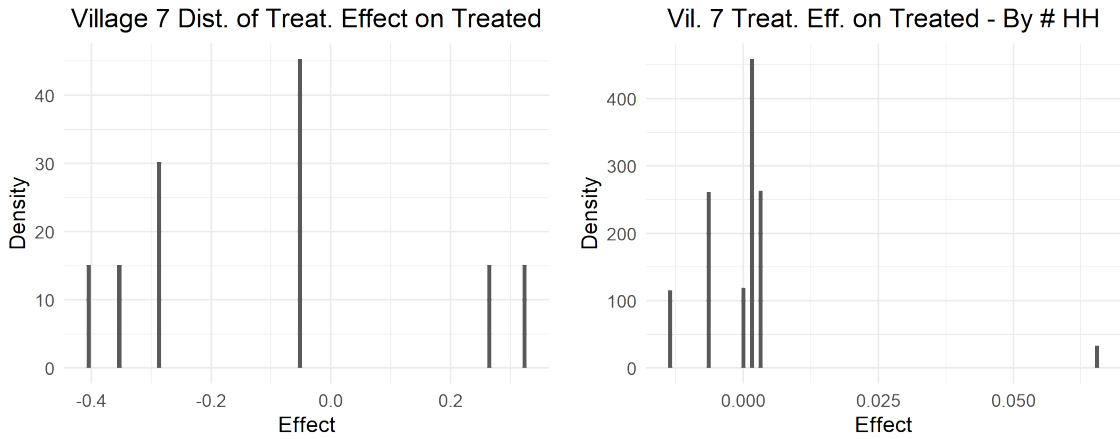
Our second demonstration uses data from Schuster and Niccolucci (1994) (see also Athey et al. 2011). These authors study US Forest Service (USFS) timber auctions where tracts of forest land were sold by either open or sealed bid auctions. They argue that participation is higher for some firms in the sealed bid format.² Let treatment 1 be the event that a bid was made under the sealed bid format and treatment 0 be the event that the bid was made under the open auction format.

We focus on two outcomes. The first is about entry and is indexed by firms and tracts. $E_{ij,t}$ describes whether firm i bid on tract j in auction format t . The second outcome matrix is about coentry and is indexed by pairs of tracts. $C_{ij,t} = \sum_k E_{ik,t} E_{jk,t}$ describes the number of firms who submit bids on both tracts i and j . Intuitively, the first outcome describes the

²The data can be found at...



(a) Spectral treatment effects on the treated for village 3. Not shown is a mass of 0.993 at 0. (b) Average treatment effects for village 3 conditional on household size.



(c) Spectral treatment effects on the treated for village 7. Not shown is a mass of 0.956 at 0. (d) Average treatment effects for village 7 conditional on household size.

Figure 1: This figure shows two characterizations of the distribution of treatment effects for villages 3 and 7 of Comola and Prina (2021). Panels (a) and (c) show the distribution of spectral treatment effects on the treated, ignoring a point mass at 0. Panels (b) and (d) show the distribution of average treatment effects conditional on household size.

amount of participation in the auction. The second outcome describes how substitutable the tracts are. A policy maker interested in encouraging firms to make more bids may be interested in the first outcome. A policy maker interested in encouraging firms to bid on different types of tracts may be interested in the second outcome.

Table 2 shows our bounds on the joint distribution of entry decisions. We find a strictly positive lower bound on $P(E_{ij,1} = 1, E_{ij,0} = 0)$ implying that the treatment caused some entry. This is consistent with a conclusion of Athey et al. (2011). However, our bounds do not rule out the possibility that the treatment only effects a small fifth of a percent of firm and tract pairs.

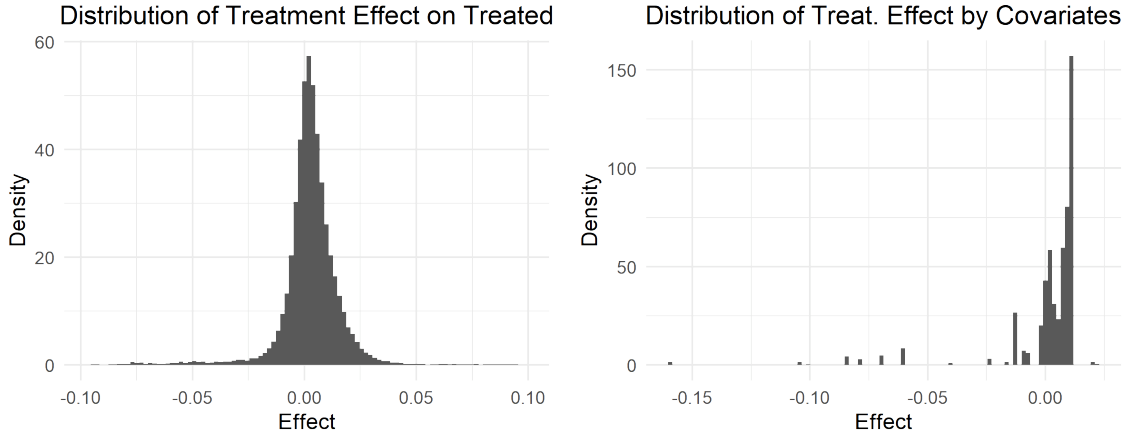
	Lower	Upper
$P(E_{ij,1} = 1, E_{ij,0} = 1)$	0.0000	0.0166
$P(E_{ij,1} = 1, E_{ij,0} = 0)$	0.0020	0.0186
$P(E_{ij,1} = 0, E_{ij,0} = 1)$	0.0000	0.0166
$P(E_{ij,1} = 0, E_{ij,0} = 0)$	0.9647	0.9814

Table 2: Bounds on the joint distribution of entry using data from Schuster and Niccolucci (1994).

Figure 2 shows the distribution of spectral treatment effects on the treated in panel (a). For reference, we also show the distribution of average treatment effects conditional on firm size and tract location in panel (b). The distribution of the STE matches the variance of the CATEs, but not the skewness. Both distributions predict a nontirival mass of firms and tracts actually have less participation under the sealed bid format.

Figure 3 shows our bounds on the joint distribution of treatment effects for coentry. We first note that $\Delta(0) \geq 0.25$. That is, the sealed bid format treatment decreases coentry for at least a quarter of tract pairs. We interpret this result as demonstrating higher levels of substitutability under the open auction format.

The distribution of spectral treatment effects and average treatment effects conditioning on tract location can be found in Figure 4. In this case, our method matches the skewness but not the variance of the conditional average treatment effects. Both methods suggest that the sealed bid format treatment decreases substitutability for most tract pairs.



(a) Spectral treatment effects on the treated for entry. Not shown is a mass of 0.074 at 0. (b) Average treatment effects for entry conditional on firm size and tract location.

Figure 2: Panel (a) shows the distribution of spectral treatment effects on the treated for entry, ignoring a point mass at 0. Panel (b) shows the distribution of average treatment effects for entry conditional on firm size and tract location.

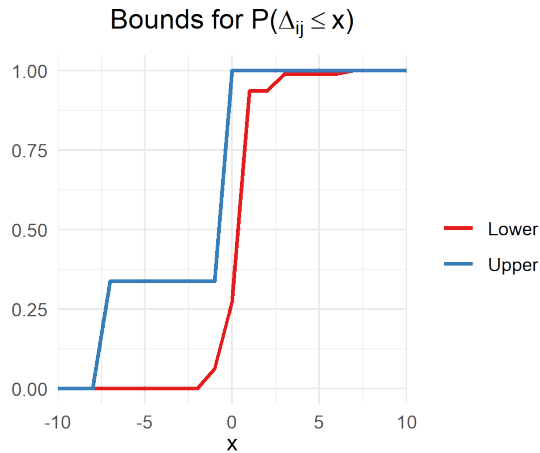
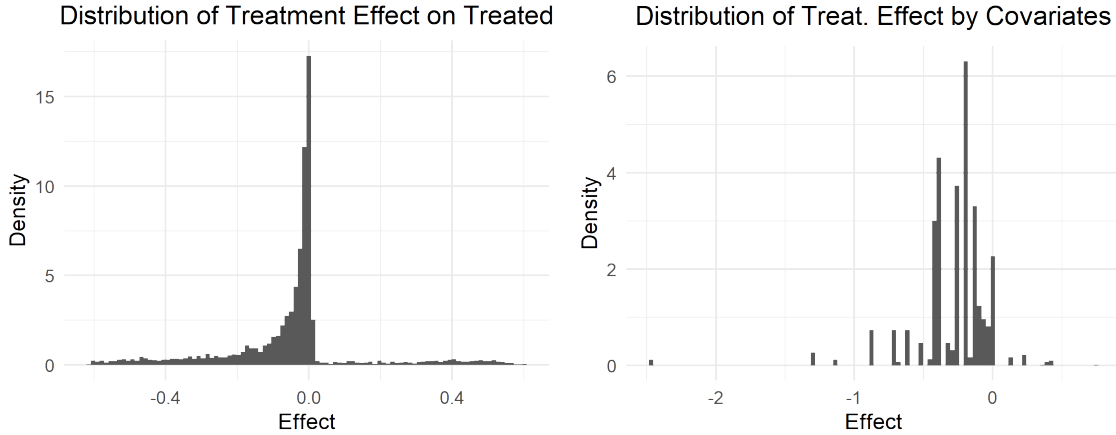


Figure 3: Bounds on the distribution of treatment effects for coentry using data from Schuster and Niccolucci (1994).

7 Conclusion

We characterize the distribution of treatment effects in a setting where each treatment is associated with a matrix of outcomes. We propose bounds on the distribution of treatment effects and a matrix analog of quantile treatment effects. Our proposal is based on a new matrix analog of the Fréchet-Hoeffding bounds that play a key role in the standard theory. We illustrate our methodology with two empirical demonstrations and find policy relevant



(a) Spectral treatment effects on the treated for coentry. Not shown is a mass of 0.140 at 0. (b) Average treatment effects for coentry conditional on tract location.

Figure 4: Panel (a) shows the distribution of spectral treatment effects on the treated for coentry, ignoring a point mass at 0. Panel (b) shows the distribution of average treatment effects for coentry conditional on tract location.

heterogeneity that might be missed by focusing exclusively on averages.

References

- Abadie, Alberto and Matias D Cattaneo, “Econometric methods for program evaluation,” *Annual Review of Economics*, 2018, 10, 465–503.
- , Joshua Angrist, and Guido Imbens, “Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings,” *Econometrica*, 2002, 70 (1), 91–117.
- Arellano, Manuel and Stéphane Bonhomme, “Recovering latent variables by matching,” *Journal of the American Statistical Association*, 2021, pp. 1–14.
- Athey, Susan, Jonathan Levin, and Enrique Seira, “Comparing open and sealed bid auctions: Evidence from timber auctions,” *The Quarterly Journal of Economics*, 2011, 126 (1), 207–257.
- Auerbach, Eric, “Testing for Differences in Stochastic Network Structure,” *arXiv preprint arXiv:1903.11117*, 2019.

- Azoulay, Pierre, Joshua S Graff Zivin, and Jialan Wang**, “Superstar Extinction,” *The Quarterly Journal of Economics*, 2010, *125* (2), 549–589.
- Bajari, Patrick, Brian Burdick, Guido W Imbens, Lorenzo Masoero, James McQueen, Thomas Richardson, and Ido M Rosen**, “Multiple Randomization Designs,” *arXiv preprint arXiv:2112.13495*, 2021.
- Banerjee, Abhijit, Emily Breza, Arun G Chandrasekhar, Esther Duflo, Matthew O Jackson, and Cynthia Kinnan**, “Changes in social network structure in response to exposure to formal credit markets,” Technical Report, National Bureau of Economic Research 2021.
- Birkhoff, Garrett**, “Three observations on linear algebra,” *Univ. Nac. Tacuman, Rev. Ser. A*, 1946, *5*, 147–151.
- Bitler, Marianne P, Jonah B Gelbach, and Hilary W Hoynes**, “What mean impacts miss: Distributional effects of welfare reform experiments,” *American Economic Review*, 2006, *96* (4), 988–1012.
- Carrier, Guillaume, Victor Chernozhukov, and Alfred Galichon**, “Vector quantile regression: an optimal transport approach,” *The Annals of Statistics*, 2016, *44* (3), 1165–1192.
- Cela, Eranda**, *The quadratic assignment problem: theory and algorithms*, Vol. 1, Springer Science & Business Media, 2013.
- Chernozhukov, Victor and Christian Hansen**, “An IV model of quantile treatment effects,” *Econometrica*, 2005, *73* (1), 245–261.
- Comola, Margherita and Silvia Prina**, “Treatment effect accounting for network changes,” *Review of Economics and Statistics*, 2021, *103* (3), 597–604.
- Doksum, Kjell**, “Empirical probability plots and statistical inference for nonlinear models in the two-sample case,” *The annals of statistics*, 1974, pp. 267–277.

- Dudley, Richard M**, *Real analysis and probability*, Vol. 74, Cambridge University Press, 2002.
- Fan, Yanqin and Marc Henry**, “Vector copulas,” *Journal of Econometrics*, 2022.
- **and Sang Soo Park**, “Sharp bounds on the distribution of treatment effects and their statistical inference,” *Econometric Theory*, 2010, *26* (3), 931–951.
- Finke, Gerd, Rainer E Burkard, and Franz Rendl**, “Quadratic assignment problems,” in “North-Holland Mathematics Studies,” Vol. 132, Elsevier, 1987, pp. 61–82.
- Firpo, Sergio**, “Efficient semiparametric estimation of quantile treatment effects,” *Econometrica*, 2007, *75* (1), 259–276.
- **and Geert Ridder**, “Partial identification of the treatment effect distribution and its functionals,” *Journal of Econometrics*, 2019, *213* (1), 210–234.
- Fréchet, Maurice**, “Sur les tableaux de corrélation dont les marges sont données,” *Ann. Univ. Lyon, 3^e e serie, Sciences, Sect. A*, 1951, *14*, 53–77.
- Galichon, Alfred**, *Optimal transport methods in economics*, Princeton University Press, 2016.
- Golub, Benjamin and Matthew O Jackson**, “Naive learning in social networks and the wisdom of crowds,” *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–49.
- Graham, Bryan S**, “Identifying social interactions through conditional variance restrictions,” *Econometrica*, 2008, *76* (3), 643–660.
- , “Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers,” *Handbook of social economics*, 2011, *1*, 965–1052.
- , “Network data,” in “Handbook of Econometrics,” Vol. 7, Elsevier, 2020, pp. 111–218.
- , **Guido W Imbens, and Geert Ridder**, “Complementarity and aggregate implications of assortative matching: A nonparametric analysis,” *Quantitative Economics*, 2014, *5* (1), 29–66.

- Hardy, Godfrey Harold, John Edensor Littlewood, George Pólya, György Pólya et al.**, *Inequalities*, Cambridge university press, 1952.
- Heckman, James J, Jeffrey Smith, and Nancy Clements**, “Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts,” *The Review of Economic Studies*, 1997, 64 (4), 487–535.
- Hoeffding, Wassilij**, “Masstabinvariante korrelationstheorie,” *Schriften des Mathematischen Instituts und Instituts für Angewandte Mathematik der Universität Berlin*, 1940, 5, 181–233.
- Hoffman, AJ and HW Wielandt**, “The variation of the spectrum of a normal matrix,” *Duke Math J*, 1953, 20, 37–39.
- Horn, Roger A and Charles R Johnson**, “Topics in matrix analysis, 1991,” *Cambridge University Presss, Cambridge*, 1991, 37, 39.
- Johari, Ramesh, Hannah Li, Inessa Liskovich, and Gabriel Y Weintraub**, “Experimental design in two-sided platforms: An analysis of bias,” *Management Science*, 2022.
- Kowalski, E**, “Spectral theory in Hilbert spaces (ETH Zürich, FS 09).”
- Lehmann, Erich L**, *Nonparametrics: statistical methods based on ranks.*, Holden-day, 1975.
- Lovász, László**, *Large networks and graph limits*, Vol. 60, American Mathematical Soc., 2012.
- Makarov, GD**, “Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed,” *Theory of Probability & its Applications*, 1982, 26 (4), 803–806.
- Manski, Charles F**, “The mixing problem in programme evaluation,” *The Review of Economic Studies*, 1997, 64 (4), 537–553.
- , *Partial identification of probability distributions*, Vol. 5, Springer, 2003.

Schuster, Ervin G and Michael J Niccolucci, “Sealed-bid versus oral-auction timber offerings: implications of imperfect data,” *Canadian Journal of Forest Research*, 1994, 24 (1), 87–91.

von Neumann, John, “A certain zero-sum two-person game equivalent to the optimal assignment problem,” *Contributions to the Theory of Games*, 1953, 2 (0), 5–12.

Whitt, Ward, “Bivariate distributions with given marginals,” *The Annals of statistics*, 1976, 4 (6), 1280–1289.

A Appendix: proof of Propositions 1-4

A.1 Definitions and lemmas

A.1.1 Graph embedding

Any finite dimensional matrix admits a function representation along the lines of Section 3.1, often called a graph embedding. See generally Lovász (2012). Let Y_t^* refer to an $N \times N$ matrix of potential outcomes with ij th entry $Y_{ij,t}^*$. The graph embedding of $Y_t^* : [0, 1]^2 \rightarrow \mathbb{R}$ is $Y_t^*(u, v) = Y_{\lceil Nu \rceil \lceil Nv \rceil, t}^*$ for $u, v \in [0, 1]$. Intuitively, the graph embedding assigns the mass of types in the region $S_i := \left(\frac{i-1}{N}, \frac{i}{N}\right]$ to observation i . Similarly, any permutation matrix Π_t can be written as a measure preserving transformation $\varphi_t(u) = \lceil Nu \rceil - Nu + \Pi_t(\lceil Nu \rceil)$ where $\Pi_t(k) = \{l \in \mathbb{N} : \Pi_{kl} = 1\}$. Intuitively, if $\Pi_{kl} = 1$, φ_t maps $\left(\frac{k-1}{N}, \frac{k}{N}\right]$ to $\left(\frac{l-1}{N}, \frac{l}{N}\right]$.

Eigenvalues are defined a bit differently for matrices and their graph embeddings. In particular, if $\lambda_1, \dots, \lambda_N$ are the nonzero eigenvalues of a matrix, then the nonzero eigenvalues of its graph embedding will be $\lambda_1/N, \dots, \lambda_N/N$.

A.1.2 Kernel operators

Any bounded symmetric measurable function $f : [0, 1]^2 \rightarrow \mathbb{R}$ defines a compact Hilbert-Schmidt integral operator. As a result, it has a countable set of nonzero real eigenvalues with 0 as the only limit point. It also admits the spectral decomposition $\sum_r \lambda_r \phi_r(u) \phi_r(v)$ where ϕ_r is the eigenvalue associated with eigenvalue λ_r , see Section 7.5 of Lovász (2012).

The functions $\{\phi_r\}_{r=1}^\infty$ form an orthonormal basis, so that $\int \phi_r(u)^2 du = 1$ and $\int (\phi_r(u) - \phi_s(u))^2 du = 2$ if $r \neq s$.

A.1.3 Definitions

Let $f_t(u, v)$ refer to either $Y_t(\varphi_t(u), \varphi_t(v))$ or $\mathbb{1}\{Y_t(\varphi_t(u), \varphi_t(v)) \leq y_t\}$ for some $y_t \in \mathbb{R}$ and $\varphi_t \in \mathcal{M} := \{\phi : [0, 1] \rightarrow [0, 1] \text{ with } |\phi^{-1}(A)| = |A| \text{ for any measurable } A \subseteq [0, 1]\}$. For any $n \in \mathbb{N}$ let $S_i := (\frac{i-1}{n}, \frac{i}{n}]$, F_t^n be an $n \times n$ matrix with $F_{ij,t}^n \in \mathbb{R}$ as its ij th entry, and $f_t^n(u, v) = \sum_{ij} F_{ij,t}^n \mathbb{1}\{u \in S_i, v \in S_j\}$ such that $\int \int (f_t(u, v) - f_t^n(u, v))^2 dudv \rightarrow 0$ as $n \rightarrow \infty$. In words, F_t^n is an $n \times n$ matrix approximation of f_t and f_t^n is its graph embedding. The existence of such a sequence of matrices F_t^n follows Lemma 1 below.

Let $\{\lambda_{rt}^n\}_{r \in [R]}$ be the eigenvalues of f_t^n where $[R] := \{1, \dots, R\}$ for some $R \in \mathbb{N} \cup \{\infty\}$. Let \mathcal{P}_n be the set of $n \times n$ permutation matrices ($\{0, 1\}$ valued entries with row and column sums equal to 1), \mathcal{D}_n^+ be the set of $n \times n$ doubly stochastic matrices (positive entries with row and column sums equal to 1), and \mathcal{O}_n be the set of $n \times n$ orthogonal matrices (any two rows or any two columns have inner product 1 if they are the same or 0 otherwise).

A.1.4 Lemmas

Lemma 1: For every bounded measurable $g : [0, 1]^2 \rightarrow \mathbb{R}$ there exists sequences $\{G^n\}_{n \in \mathbb{N}}$ and $\{g^n\}_{n \in \mathbb{N}}$ where G^n is an $n \times n$ matrix with ij th entry G_{ij}^n and $g^n : [0, 1]^2 \rightarrow \mathbb{R}$ with $g^n(u, v) = \sum_{ij} G_{ij}^n \mathbb{1}\{u \in S_i, v \in S_j\}$ and $S_i := (\frac{i-1}{n}, \frac{i}{n}]$ such that for every $\varepsilon > 0$ there exists an $m \in \mathbb{N}$ such that $\int \int (g(u, v) - g^n(u, v))^2 dudv \leq \varepsilon$ for every $n > m$.

Proof of Lemma 1: Fix an arbitrary $\varepsilon > 0$. Lusin's Theorem (see Appendix Section B.1.1) implies that for any measurable $g : [0, 1]^2 \rightarrow \mathbb{R}$ and $\epsilon > 0$, there exists a compact $E_g^\epsilon \subseteq [0, 1]^2$ of measure at least $1 - \epsilon$ such that g is continuous when restricted to E_g^ϵ .

Let $G_{ij}^{n\epsilon} = \frac{\int \int_{(u,v) \in E_g^\epsilon} g_t(u,v) \mathbb{1}\{u \in S_i, v \in S_j\} dudv}{\int \int_{(u,v) \in E_g^\epsilon} \mathbb{1}\{u \in S_i, v \in S_j\} dudv}$ if $\int \int_{(u,v) \in E_g^\epsilon} \mathbb{1}\{u \in S_i, v \in S_j\} dudv > 0$ and $G_{ij}^{n\epsilon} = 0$ otherwise. Also let $g^{n\epsilon}(u, v) = \sum_{ij} G_{ij}^{n\epsilon} \mathbb{1}\{u \in S_i, v \in S_j\}$ and $\epsilon = \theta(\varepsilon) := \varepsilon / (8\bar{g})$ where $\bar{g} := \sup_{(u,v) \in [0,1]^2} |g(u, v)|^2$.

Since g is continuous on $E_g^{\theta(\varepsilon)}$ there exists an $m(\varepsilon) \in \mathbb{N}$ such that

$$\int \int_{(u,v) \in E_g^{\theta(\varepsilon)}} (g(u,v) - g^{n\theta(\varepsilon)}(u,v))^2 dudv \leq \varepsilon/2 \text{ for every } n > m(\varepsilon). \text{ In addition,}$$

$$\int \int_{(u,v) \notin E_g^{\theta(\varepsilon)}} (g(u,v) - g^{n\theta(\varepsilon)}(u,v))^2 dudv \leq 4\bar{g}\theta(\varepsilon) = \varepsilon/2 \text{ for every } n. \text{ So}$$

$$\int \int_{(u,v) \in [0,1]^2} (g(u,v) - g^{n\theta(\varepsilon)}(u,v))^2 dudv \leq \varepsilon \text{ for every } n > m(\varepsilon).$$

It follows that $\mu^\dagger(n) := \inf_{e>0} \{m(e) \leq n\} \rightarrow 0$ as $n \rightarrow \infty$. Let $G^n = G^{n\theta(\mu^\dagger(n))}$ and $g^n = g^{n\theta(\mu^\dagger(n))}$. Then $\int \int_{(u,v) \in [0,1]^2} (g(u,v) - g^n(u,v))^2 dudv \leq \mu^\dagger(n)$ for $n > \mu^\dagger(n)$ by construction and the claim follows by taking m sufficiently large so that $\mu^\dagger(m) < \varepsilon$. \square

Lemma 2: $\sum_{r \in [n]} \lambda_{s_n(r)0}^n \lambda_{r1}^n \leq \int \int f_1^n(u,v) f_0^n(u,v) dudv \leq \sum_{r \in [n]} \lambda_{r0}^n \lambda_{r1}^n$ where $s_n(r) = n - r + 1$.

Proof of Lemma 2: By construction $\int \int f_1^n(u,v) f_0^n(u,v) dudv = \frac{1}{n^2} \sum_{ij} F_{ij,1}^n F_{ij,0}^n$ so it is sufficient to show that $n^2 \sum_{r \in [n]} \lambda_{s_n(r)0}^n \lambda_{r1}^n \leq \sum_{ij} F_{ij,1}^n F_{ij,0}^n \leq n^2 \sum_{r \in [n]} \lambda_{r0}^n \lambda_{r1}^n$. Recall that $\{n\lambda_{rt}^n\}_{r \in [n]}$ are the eigenvalues of $F_{ij,t}^n$.

Since F_t^n is square and symmetric, the spectral theorem (see Appendix Section B.1.1) implies that $F_{ij,t}^n = n \sum_{r \in [n]} \lambda_{rt}^n \phi_{ir,t}^n \phi_{jr,t}^n$ where $\phi_{ir,t}^n$ is the eigenvector of $F_{ij,t}^n$ associated with eigenvalue $n\lambda_{rt}^n$. As a result $\sum_{ij} F_{ij,1}^n F_{ij,0}^n = n^2 \sum_{r,s \in [n]} \lambda_{r1}^n \lambda_{s0}^n \left[\sum_i \phi_{ir,1}^n \phi_{is,0}^n \right]^2$.

The matrix $\left[\sum_i \phi_{ir,1}^n \phi_{is,0}^n \right]^2$ is an element of \mathcal{D}_n^+ and so Birkhoff's Theorem (see Appendix Section B.1.1) implies that

$$\sum_{r,s \in [n]} \lambda_{r1}^n \lambda_{s0}^n \left[\sum_i \phi_{ir,1}^n \phi_{is,0}^n \right]^2 = \sum_{r,s \in [n]} \lambda_{r1}^n \lambda_{s0}^n \sum_{t \in [m]} \alpha_t P_{ij,t} = \sum_{t \in [m]} \alpha_t \sum_{r,s \in [n]} \lambda_{r1}^n \lambda_{s0}^n P_{ij,t}$$

where $m \in \mathbb{N}$, $\alpha_1, \dots, \alpha_m > 0$, with $\sum_{t \in [m]} \alpha_t = 1$, and $P_1, \dots, P_m \in \mathcal{P}_n$.

The Hardy-Littlewood Theorem (see Appendix Section B.1.1) implies that

$$\sum_{r \in [n]} \lambda_{r1}^n \lambda_{s_n(r)0}^n \leq \sum_{r,s \in [n]} \lambda_{r1}^n \lambda_{s0}^n P_{ij} \leq \sum_{r \in [n]} \lambda_{r1}^n \lambda_{r0}^n$$

for any $P \in \mathcal{P}_n$ and so

$$\sum_{r \in [n]} \lambda_{r1}^n \lambda_{s_n(r)0}^n \leq \sum_{t \in [m]} \alpha_t \sum_{r, s \in [n]} \lambda_{r1}^n \lambda_{s0}^n P_{ij,t} \leq \sum_{r \in [n]} \lambda_{r1}^n \lambda_{r0}^n$$

because $\sum_{t \in [m]} \alpha_t = 1$. \square

Lemma 3: For every $\varepsilon > 0$ there exists an $m \in \mathbb{N}$ such that

- i. $|\int \int f_1^n(u, v) f_0^n(u, v) dudv - \int \int f_1(u, v) f_0(u, v) dudv| \leq \varepsilon$, and
- ii. $|\sum_{r \in [n]} \lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \sum_r \lambda_{\sigma(r)0} \lambda_{r1}| \leq \varepsilon$,

for every $n > m$ where $\sum_r \lambda_{\sigma(r)t} \lambda_{r1}$ refers to $\sum_{r \in [\bar{R}]} \lambda_{\sigma_{\bar{R}}(r)t} \lambda_{rt'}$ for $t, t' \in \{0, 1\}$, $\sigma_x(r)$ refers to r or $s_x(r) := x - r + 1$, and $\bar{R} = \max\{R \in \mathbb{N} \cup \infty : \max_{t \in \{0, 1\}} \min_{r \in [R]} |\lambda_{rt}| > 0\}$.

Proof of Lemma 3: Fix an arbitrary $\varepsilon > 0$. Part i. follows from

$$\begin{aligned} & \int \int f_1^n(u, v) f_0^n(u, v) dudv - \int \int f_1(u, v) f_0(u, v) dudv \\ &= \int \int (f_1^n(u, v) - f_1(u, v)) f_0^n(u, v) dudv + \int \int (f_0^n(u, v) - f_0(u, v)) f_1(u, v) dudv \\ &\leq \left(\int \int (f_1^n(u, v) - f_1(u, v))^2 dudv \right)^{1/2} \bar{f}_0^n + \left(\int \int (f_0^n(u, v) - f_0(u, v))^2 dudv \right)^{1/2} \bar{f}_1 \\ &\leq \varepsilon (\bar{f}_0^n + \bar{f}_1) \text{ for } n > m(\varepsilon) \text{ where } m(\varepsilon) \text{ is from the hypothesis of Lemma 1} \\ &\leq \varepsilon \text{ for } n > m(\varepsilon) \text{ for } \varepsilon < \varepsilon / (\bar{f}_0^n + \bar{f}_1) \end{aligned}$$

where $\bar{f}_0^n = (\int \int f_0^n(u, v)^2 dudv)^{1/2}$ and $\bar{f}_1 = (\int \int f_1(u, v)^2 dudv)^{1/2}$, the first inequality is due to Cauchy-Schwarz, and the second is due to Lemma 1.

To demonstrate Part ii, we first bound $\sum_{r \in [n]} \lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1}$ where the sum $\sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1}$ is over the product of the n largest eigenvalues of f_0 and f_1 in absolute

value (see Section 3.2.1).

$$\begin{aligned}
& \sum_{r \in [n]} \lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1} = \sum_{r \in [n]} (\lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \lambda_{\sigma_n(r)0} \lambda_{r1}) \\
& = \sum_{r \in [n]} (\lambda_{\sigma_n(r)0}^n - \lambda_{\sigma_n(r)0}) \lambda_{r1}^n + \sum_{r \in [n]} (\lambda_{r1}^n - \lambda_{r1}) \lambda_{\sigma_n(r)0} \\
& \leq \left(\sum_{r \in [n]} (\lambda_{r0}^n - \lambda_{r0})^2 \right)^{1/2} \left(\sum_{r \in [n]} (\lambda_{r1}^n)^2 \right)^{1/2} + \left(\sum_{r \in [n]} (\lambda_{r1}^n - \lambda_{r1})^2 \right)^{1/2} \left(\sum_{r \in [n]} (\lambda_{r0})^2 \right)^{1/2} \\
& \leq \left(\sum_{r \in [n]} (\lambda_{r0}^n - \lambda_{r0})^2 \right)^{1/2} \bar{f}_1^n + \left(\sum_{r \in [n]} (\lambda_{r1}^n - \lambda_{r1})^2 \right)^{1/2} \bar{f}_0
\end{aligned}$$

where the first inequality is due to Cauchy-Schwarz. Since f_t^n and f_t are uniformly bounded then for every $\epsilon > 0$ there exists a $R, m \in \mathbb{N}$ such that $\sum_{r \in [n]-[R]} (\lambda_{rt}^n)^2 < \epsilon$ and $\sum_{r \in [n]-[R]} (\lambda_{rt})^2 < \epsilon$ for every $n > m$ and $t \in \{0, 1\}$. As a result,

$$\begin{aligned}
& \sum_{r \in [n]} \lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1} \\
& \leq \left(\sum_{r \in [n]} (\lambda_{r0}^n - \lambda_{r0})^2 \right)^{1/2} \bar{f}_1^n + \left(\sum_{r \in [n]} (\lambda_{r1}^n - \lambda_{r1})^2 \right)^{1/2} \bar{f}_0 \\
& \leq \left(\sum_{r \in [R]} (\lambda_{r0}^n - \lambda_{r0})^2 \right)^{1/2} \bar{f}_1^n + \left(\sum_{r \in [R]} (\lambda_{r1}^n - \lambda_{r1})^2 \right)^{1/2} \bar{f}_0 + 2\sqrt{\epsilon}(\bar{f}_1^n + \bar{f}_0) \\
& \leq \sqrt{R} \left(\int (f_0^n(u, v) - f_0(u, v))^2 \right)^{1/2} \bar{f}_1^n + \sqrt{R} \left(\int (f_1^n(u, v) - f_1(u, v))^2 \right)^{1/2} \bar{f}_0 + 2\sqrt{\epsilon}(\bar{f}_1^n + \bar{f}_0) \\
& \leq (\sqrt{R}\tilde{\epsilon} + 2\sqrt{\epsilon})(\bar{f}_1^n + \bar{f}_0) \text{ for } n > m(\tilde{\epsilon}) \text{ where } m(\tilde{\epsilon}) \text{ is from the hypothesis of Lemma 1} \\
& \leq \epsilon/2 \text{ for } n > m(\epsilon/2)
\end{aligned}$$

where the third inequality follows because the eigenvalues of compact Hermitian operators are Lipschitz continuous (see Appendix Section B.1.1) and the last inequality follows if ϵ and R are chosen so that $\epsilon < \epsilon^2/(8\bar{f}_1^n + 8\bar{f}_0)^2$ and then $\tilde{\epsilon}$ and m are chosen so that $\tilde{\epsilon} < \epsilon/(4\sqrt{R}\bar{f}_1^n + 4\sqrt{R}\bar{f}_0)$.

Finally, to bound $\left| \sum_{r \in [n]} \lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \sum_r \lambda_{\sigma(r)0} \lambda_{r1} \right|$, we consider two cases. In the case that $\bar{R} = \infty$ we have $\sum_r \lambda_{\sigma(r)0} \lambda_{r1} = \lim_{n \rightarrow \bar{R}} \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1}$ and so can choose m so that $|\sum_r \lambda_{\sigma(r)0} \lambda_{r1} - \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1}| < \varepsilon/2$. In the case that $\bar{R} < \infty$ we can choose $m > \bar{R}$ so that $\sum_r \lambda_{\sigma(r)0} \lambda_{r1} = \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1}$. It follows that

$$\begin{aligned} \left| \sum_{r \in [n]} \lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \sum_r \lambda_{\sigma(r)0} \lambda_{r1} \right| &\leq \left| \sum_{r \in [n]} \lambda_{\sigma_n(r)0}^n \lambda_{r1}^n - \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1} \right| + \left| \sum_{r \in [n]} \lambda_{\sigma_n(r)0} \lambda_{r1} - \sum_r \lambda_{\sigma(r)0} \lambda_{r1} \right| \\ &\leq \varepsilon/2 + \varepsilon/2 = \varepsilon \end{aligned}$$

where the first inequality is due to the triangle inequality. \square

Lemma 4: If f_0^n and f_1^n take values in $[0, 1]$ then $\max \left(\sum_{r \in [n]} ((\lambda_{r0}^n)^2 + (\lambda_{r1}^n)^2) - 1, 0 \right) \leq \int \int f_1^n(u, v) f_0^n(u, v) dudv \leq \min \left(\sum_{r \in [n]} (\lambda_{r0}^n)^2, \sum_{r \in [n]} (\lambda_{r1}^n)^2 \right)$.

Proof of Lemma 4: The upper bound follows

$$\int \int f_1^n(u, v) f_0^n(u, v) dudv \leq \min_{t \in \{0,1\}} \int \int f_t^n(u, v) dudv = \min_{t \in \{0,1\}} \sum_{r \in [n]} (\lambda_{rt}^n)^2.$$

The lower bound follows

$$\begin{aligned} \int \int f_1^n(u, v) f_0^n(u, v) dudv &= \int \int f_1^n(u, v) (1 - (1 - f_0^n(u, v))) dudv \\ &\geq \int \int f_1^n(u, v) dudv - \min \left(\int \int f_1^n(u, v) dudv, \int \int (1 - f_0^n(u, v)) dudv \right) \\ &= \max \left(0, \int \int f_1^n(u, v) dudv + \int \int f_0^n(u, v) dudv - 1 \right) \\ &= \max \left(\sum_{r \in [n]} ((\lambda_{r0}^n)^2 + (\lambda_{r1}^n)^2) - 1, 0 \right). \end{aligned}$$

Both inequalities are because f_0^n and f_1^n take values in $[0, 1]$. \square

A.2 Proposition 1

Let $f_t(u, v) = \mathbb{1}\{Y_t^*(u, v) \leq y_t\}$. For any $\epsilon > 0$ there exists an $m \in \mathbb{N}$ such that for every $n > m$

$$\begin{aligned} \int \int f_1(u, v) f_0(u, v) dudv &< \int f_1^n(u, v) f_0^n(u, v) dudv + \epsilon \\ &\leq \min \left(\sum_r \lambda_{r1}^n \lambda_{r0}^n, \sum_r \lambda_{r1}^{2n}, \sum_r \lambda_{r0}^{2n} \right) + \epsilon \\ &< \min \left(\sum_r \lambda_{r1} \lambda_{r0}, \sum_r \lambda_{r1}^2, \sum_r \lambda_{r0}^2 \right) + 2\epsilon \end{aligned}$$

where the first inequality is due to Part i of Lemma 3, the second inequality is the intersections of the upper bounds in Lemmas 2 and 4, and the third inequality is due to Part ii of Lemma 3. Similarly,

$$\begin{aligned} \int \int f_1(u, v) f_0(u, v) dudv &> \int f_1^n(u, v) f_0^n(u, v) dudv - \epsilon \\ &\geq \max \left(\sum_r \lambda_{r1}^n \lambda_{s(r)0}^n, \sum_r (\lambda_{r0}^{n2} + \lambda_{r1}^{n2}) - 1, 0 \right) - \epsilon \\ &> \max \left(\sum_r \lambda_{r1} \lambda_{s(r)0}, \sum_r (\lambda_{r0}^2 + \lambda_{r1}^2) - 1, 0 \right) - 2\epsilon. \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the claim follows. \square

A.3 Proposition 2

For any $y_1, y_0 \in \mathbb{R}$ such that $y_1 - y_0 = y$ we have

$$\begin{aligned} P(Y_1^* - Y_0^* \leq y) &\geq P(Y_1^* \leq y_1 \cap -Y_0^* < -y_0) \\ &= P(Y_1^* \leq y_1) - P(Y_1^* \leq y_1 \cap Y_0^* \leq y_0) \\ &\geq \sum_r \lambda_{r1}^2 - \min \left(\sum_r \lambda_{r1}^2, \sum_r \lambda_{r0}^2, \sum_r \lambda_{r1} \lambda_{r0} \right) \\ &= \max \left(\sum_r (\lambda_{r1}^2 - \lambda_{r0}^2), \sum_r (\lambda_{r1}^2 - \lambda_{r1} \lambda_{r0}), 0 \right) \end{aligned}$$

and

$$\begin{aligned}
P(Y_1^* - Y_0^* \leq y) &\leq P(Y_1^* \leq y_1 \cup -Y_0^* < -y_0) \\
&= 1 + P(Y_1^* \leq y_1 \cap Y_0^* \leq y_0) - P(Y_0^* \leq y_0) \\
&\leq 1 + \min \left(\sum_r \lambda_{r1}^2, \sum_r \lambda_{r0}^2, \sum_r \lambda_{r1} \lambda_{r0} \right) - \sum_r \lambda_{r0}^2 \\
&= 1 + \min \left(\sum_r (\lambda_{r1}^2 - \lambda_{r0}^2), \sum_r (\lambda_{r1} \lambda_{r0} - \lambda_{r0}^2), 0 \right)
\end{aligned}$$

where the the first inequality in both systems is due to the fact that for any $u, v \in [0, 1]$,

$$\begin{aligned}
\mathbb{1}\{Y_1^*(u, v) \leq y_1\} \mathbb{1}\{-Y_0^*(u, v) < -y_0\} &\leq \mathbb{1}\{Y_1^*(u, v) - Y_0^*(u, v) \leq y\} \\
&\leq \max(\mathbb{1}\{Y_1^*(u, v) \leq y_1\}, \mathbb{1}\{-Y_0^*(u, v) < -y_0\})
\end{aligned}$$

and the second inequality in both systems is due to the upper bound in Proposition 1. Since these inequalities holds for any such y_1, y_0 , the claim follows. \square

A.4 Proposition 3

This result is an infinite dimensional version of the Hoffman-Wielandt inequality (see Appendix Section B.1.1). Let $f_t(u, v) = Y_t^*(u, v)$ and σ_{rt} denote the r th eigenvalue of f_t (ordered to be decreasing). For any $\epsilon > 0$ there exists an $m \in \mathbb{N}$ such that for every $n > m$

$$\begin{aligned}
&\int \int (f_1(u, v) - f_0(u, v))^2 dudv \\
&= \int \int f_1(u, v)^2 dudv + \int \int f_0(u, v)^2 dudv - 2 \int \int f_1(u, v) f_0^*(u, v) dudv \\
&\geq \int \int f_1(u, v)^2 dudv + \int \int f_0(u, v)^2 dudv - 2 \int \int f_1^n(u, v) f_0^n(u, v) dudv - \epsilon \\
&\geq \sum_r \sigma_{r1}^2 + \sum_r \sigma_{r0}^2 - 2 \sum_r \sigma_{r1}^n \sigma_{r0}^n - \epsilon \\
&\geq \sum_r \sigma_{r1}^2 + \sum_r \sigma_{r0}^2 - 2 \sum_r \sigma_{r1} \sigma_{r0} - 2\epsilon \\
&= \sum_r (\sigma_{r1} - \sigma_{r0})^2 - 2\epsilon
\end{aligned}$$

where the first inequality is due to Part i of Lemma 3, the second inequality is due to the upper bound of Lemma 2, and the third inequality is due to Part ii of Lemma 3.

The claim then follows from the fact that $\int \int STE(u, v; \phi)^2 dudv = \sum_r (\sigma_{r1} - \sigma_{r0})^2$ for any choice of orthogonal basis $\{\phi_r\}_{r=1}^\infty$. Specifically,

$$\begin{aligned} \int \int STE(u, v; \phi)^2 dudv &= \int \int \sum_{r,s} (\sigma_{r1} - \sigma_{r0})(\sigma_{s1} - \sigma_{s0}) \phi_r(u) \phi_r(v) \phi_s(u) \phi_s(v) dudv \\ &= \sum_{r,s} (\sigma_{r1} - \sigma_{r0})(\sigma_{s1} - \sigma_{s0}) \left[\int \phi_r(u) \phi_s(u) du \right]^2 \\ &= \sum_r (\sigma_{r1} - \sigma_{r0})^2. \end{aligned}$$

The last equality is because $\{\phi_r\}_{r=1}^\infty$ is orthogonal and so $[\int \phi_r(u) \phi_s(u) du]^2 = \mathbb{1}\{r = s\}$. \square

A.5 Proposition 4

Let $\sum_s c_s x^s$ be the series representation of $g(x)$ and $(\sigma_{rt}, \phi_{rt}^*)$ be the r th eigenvalue-eigenfunction pair of Y_t^* where the eigenvalues are ordered to be decreasing. Then

$$Y_1^*(u, v) = g(Y_0^*(u, v)) = \sum_s c_s Y_0^*(u, v)^s = \sum_{r,s} c_s \sigma_{r0}^s \phi_{r0}^*(u) \phi_{r0}^*(v) = \sum_r g(\sigma_{r0}) \phi_{r0}^*(u) \phi_{r0}^*(v)$$

where the third equality follows from the identity $h^s = \sum_r \rho_r^s \psi_r \psi_r$ where (ρ_r, ψ_r) is the r th eigenvalue-eigenfunction pair of h . Since $Y_1^* = \sum_r \sigma_{1r} \phi_{r1}^* \phi_{r1}^*$, it follows the assumption that g is increasing that $\sigma_{1r} = g(\sigma_{0r})$, $\phi_{r1}^* = \phi_{r0}^*$, and so

$$Y_1^* - Y_0^* = \sum_r (\sigma_{r1} - \sigma_{r0}) \phi_{r1}^* \phi_{r1}^* = \sum_r (\sigma_{r1} - \sigma_{r0}) \phi_{r0}^* \phi_{r0}^*.$$

Since $STT(u, v) = \sum_r (\sigma_{r1} - \sigma_{r0}) \phi_{r1}(u) \phi_{r1}(v)$ and $STU(u, v) = \sum_r (\sigma_{r1} - \sigma_{r0}) \phi_{r0}(u) \phi_{r0}(v)$ where $\phi_{r1}^*(u) = \phi_{r1}(\varphi_1(u))$ and $\phi_{r0}^*(u) = \phi_{r0}(\varphi_0(u))$, we have

$$Y_1^*(u, v) - Y_0^*(u, v) = STT(\varphi_1(u), \varphi_1(v)) = STU(\varphi_0(u), \varphi_0(v)).$$

and so

$$\begin{aligned} \int \int \mathbb{1} \{Y_1^*(u, v) - Y_0^*(u, v) \leq y\} dudv &= \int \int \mathbb{1} \{STT(u, v) \leq y\} dudv \\ &= \int \int \mathbb{1} \{STU(u, v) \leq y\} dudv \end{aligned}$$

as claimed. \square

B Additional claims and details

B.1 Auxiliary lemmas

B.1.1 Lemmas used in Appendix Section A

Lemma A1 (Lusin): For every measurable $f : [0, 1]^2 \rightarrow \mathbb{R}$ and $\epsilon > 0$ there exists a compact $E_\epsilon \subseteq [0, 1]^2$ with measure at least $1 - \epsilon$ such that f is continuous when restricted to E_ϵ .

See Dudley (2002) Theorem 7.5.2.

Lemma A2 (Spectral): Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_H$ and f be a compact normal operator on H . Then there exist a sequence of eigenvalues λ_r with 0 as its only limit point and orthonormal basis ϕ_r such that $f(\phi_r) = \lambda_r \phi_r$ and $f(\psi) = \sum_r \lambda_r \langle \psi, \phi_r \rangle \phi_r$ for any $\psi \in H$.

See Kowalski (n.d.) Theorem 2.5.

Lemma A3 (Continuity): Let H be a Hilbert space and f, g be positive compact operators such that $\|f - g\|_{L(H)} \leq \epsilon$. Then $|\lambda_k(f) - \lambda_k(g)| \leq \epsilon$ for all $k \geq 1$.

See Kowalski (n.d.) Corollary 2.16(2).

Lemma A4 (Birkhoff-von Neumann): For every $M \in \mathcal{D}_n^+$ there exists an $m \in \mathbb{N}$, $\alpha_1, \dots, \alpha_m > 0$, and $P_1, \dots, P_m \in \mathcal{P}_n$ such that $\sum_{t=1}^m \alpha_t = 1$ and $M_{ij} = \sum_{t=1}^m \alpha_t P_{ij,t}$.

See Birkhoff (1946).

Lemma A5 (Hardy-Littlewood-Polya): For any $m \in \mathbb{N}$ and $g, h \in \mathbb{R}^m$ we have

$$\sum_r g_{(r)} h_{(m-r+1)} \leq \sum_r g_{(r)} h_r \leq \sum_r g_{(r)} h_{(r)}$$
 where $g_{(r)}$ is the r th order statistic of g .

See Hardy et al. (1952), Section 10.2, Theorem 368.

Lemma A6 (Hoffman-Wielandt): Let $\{\lambda_r^g\}_{r \in \mathbb{N}}$ and $\{\lambda_r^h\}_{r \in \mathbb{N}}$ be the ordered eigenvalues of two $N \times N$ real symmetric matrices G and H . Then $\sum_r (\lambda_r^g - \lambda_r^h)^2 \leq \sum_{i,j} (G_{ij} - H_{ij})^2$.

See Hoffman and Wielandt (1953).