Online Appendix: Heterogeneous Treatment Effects for Networks, Panels, and other Outcome Matrices

Eric Auerbach* Yong Cai †

October 2, 2022

This is the online appendix for Auerbach and Cai (2022). Section B states auxiliary lemmas used to prove Propositions 1-4 in the main text. Section C contains examples of experiments with rank invariant treatment effects and heterogeneous spillover effects. Section D discusses extensions to row and column heterogeneity, estimation, and inference.

B Auxiliary lemmas

The following auxiliary lemmas are used to demonstrate Propositions 1-4 in the main text.

B.1 Lemmas used in Appendix Section A

Lemma B1 (Lusin): For every measurable $f : [0,1]^2 \to \mathbb{R}$ and $\epsilon > 0$ there exists a compact $E_\epsilon \subseteq [0,1]^2$ with Lebesgue measure at least $1-\epsilon$ such that f is continuous when restricted to E_ϵ. See Dudley (2002) Theorem 7.5.2.

Lemma B2 (Spectral): Let $f : [0,1]^2 \to \mathbb{R}$ be a bounded symmetric measurable function and $T_f : L_2[0,1] \to L_2[0,1]$ the associated integral operator $(T_f g)(u) = \int f(u,\tau)g(\tau)d\tau$. T_f admits the spectral decomposition $f(u,v) = \sum_{r=1}^{\infty} \lambda_r \phi_r(u), \phi_r(v)$ in the sense that $(T_f g)(u) = \int f(u,\tau)g(\tau)d\tau = \sum_{r=1}^{\infty} \lambda_r \phi_r(u) \int \phi_r(\tau)g(\tau)d\tau$ for any $g \in L_2[0,1]$. Each (λ_r, ϕ_r) pair satisfies $\int f(u,\tau)\phi_r(\tau)d\tau = \lambda_r \phi_r(u)$ where $\{\lambda_r\}_{r=1}^{\infty}$ is a multiset of bounded

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

real numbers with 0 as its only limit point and $\{\phi_r\}_{r=1}^{\infty}$ is an orthogonal basis of $L_2[0,1]$. See Birman and Solomjak (2012) equation (5) preceding Theorem 4 in Chapter 9.2.

The spectral decomposition described in Lemma B2 is related to but distinct from a decomposition by the same name for matrices. Specifically, if Y is an $N \times N$ dimensional symmetric real-valued matrix then it admits the spectral decomposition $Y_{ij} = \sum_{r=1}^{N} \lambda_r \phi_{ir} \phi_{jr}$. Each (λ_r, ϕ_r) pair satisfies $\sum_{j=1}^{N} Y_{ij} \phi_{jr} = \lambda_r \phi_{ir}$ where $\{\lambda_r\}_{r=1}^{N}$ is a multiset of real numbers and $\{\phi_{ir}\}_{i,r=1}^{N}$ is an orthogonal matrix with rth column denoted by the $N \times 1$ dimensional vector ϕ_r.

Lemma B3 (Continuity): Let $f, g : [0,1]^2 \to \mathbb{R}$ be bounded symmetric measurable functions with positive eigenvalues $\{\lambda_r^+(f), \lambda_r^+(g)\}_r$ and negative eigenvalues $\{\lambda_r^-(f), \lambda_r^-(g)\}_r$ both ordered to be decreasing in absolute value. Suppose $\left(\int \int (f(u,v) - g(u,v))^2 \, dudv\right)^{1/2} \leq \epsilon$. Then $|\lambda_i^+(U) - \lambda_i^+(W)| \leq \epsilon$ and $|\lambda_i^-(U) - \lambda_i^-(W)| \leq \epsilon$. See Birman and Solomjak (2012), equation (19) following Theorem 8 in Chapter 9.2.

In Lemma 3 of Appendix Section A.4 in the main text, we use the following corollary of this result and Theorem 368 of Hardy et al. (1952) (Lemma B5 below). That is, $\left(\sum_{r \in [R]} (\lambda_r(f) - \lambda_r(g))^2\right)^{1/2} \leq \sqrt{R} \left(\int \int (f(u,v) - g(u,v))^2 \, dudv\right)^{1/2}$ where $\{\lambda_r(f), \lambda_r(g)\}_{r \in [R]}$ are the R largest (in absolute value) eigenvalues of f and g ordered to be decreasing.

Lemma B4 (Birkhoff): For every $M \in \mathcal{D}_n^+$ there exists an $m \in \mathbb{N}$, $\alpha_1, ..., \alpha_m > 0$, and $P_1, ..., P_m \in \mathcal{P}_n$ 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 B5 (Hardy-Littlewood-Polya Theorem 368): For any $m \in \mathbb{N}$ and $g, h \in \mathbb{R}^m$ we have $\sum_{r=1}^{m} g_{(r)} h_{(m-r+1)} \leq \sum_{r=1}^{m} g_r h_r \leq \sum_{r=1}^{m} g_{(r)} h_{(r)}$ where $g_{(r)}$ is the rth order statistic of g. See Hardy et al. (1952), Section 10.2, Theorem 368.

Lemma B6 (Hoffman-Wielandt): Let $\{\lambda_r(F)\}_{r \in [n]}$ and $\{\lambda_r(G)\}_{r \in [n]}$ be the eigenvalues of two $n \times n$ real symmetric matrices F and G, ordered to be decreasing. Then $\sum_{r=1}^{n} (\lambda_r(F) - \lambda_r(G))^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{n} (F_{ij} - G_{ij})^2$. See Hoffman and Wielandt (1953).

C Examples

C.1 Examples of rank invariant treatment effects

We provide four concrete examples of heterogeneous treatment effects with outcome matrices from the economics literature. We show that under certain conditions implied by economic theory the treatment effect is rank invariant in the sense of Definition 2 of Section 3.2.3 of the main text. As a result of our Proposition 4, the distributions of the STT and STU are both equal to the DTE. In these examples, we take the population of agents to be finite, although our results also apply to the infinite population case.

C.1.1 Information diffusion

This example follows Banerjee et al. (2013); Cruz et al. (2017); Bramoullé and Genicot (2018). N agents are linked in a social network as described by an $N \times N$ symmetric binary adjacency matrix G. $G_{ij} = 1$ if agents i and j are linked and $G_{ij} = 0$ otherwise. Information about a new product or social program diffuses over the social network in T discrete time periods. In an initial period 0, one agent receives the relevant piece of information. For periods $t = 1, ..., T$, agents who received the information in period $t-1$, transmit it to their neighbors in period t. Specifically, if an agent receives the information M times from their neighbors in period $t-1$, the number of times they transmit the information to each of their neighbors in time period t is the sum of M independent Bernoulli(α) trials. The parameter α describes the probability that an agent will transmit information to their neighbors once they receive it.

The outcome of interest is the expected number of times agent j receives the information in the T time periods when agent i is initially informed in period 0. Proposition 1 of Bramoullé and Genicot (2018) implies that it is given by

$$E[Y_{ij}] = \sum_{t=1}^{T} \alpha^t [G^t]_{ij}.$$

where $[G^t]_{ij} = \sum_{s_1} \sum_{s_2} ... \sum_{s_{t-1}} G_{is_1} G_{s_1 s_2} ... G_{s_{t-1} j}$ is the ijth entry of the tth operator power of G.

Now consider an intervention that increases α, the probability of information transmission between agents. For example, the intervention may be a new advertisement campaign. Let $\alpha(s)$ denote the transmission probability and $E[Y_{ij}](s)$ the resulting outcome matrix with and without the campaign as indexed by $s \in \{0,1\}$. Then

$$E[Y_{ij}](s) = \sum_{r=1}^{n} \left(\sum_{t=1}^{T} \alpha(s)^t \lambda_r^t \right) \phi_{ir} \phi_{jr}$$

where (λ_r, ϕ_r) are the eigenvalue and eigenvector pairs of G. Under certain conditions on T and $\alpha(1), \alpha(0)$, the treatment effect is rank invariant in the sense of Definition 2 of Section 3.2.3 that $E[Y_{ij}](1) = g(E[Y_{ij}](0))$ where g is the matrix lift of a nondecreasing function. For example, if $\alpha(1) > \alpha(0)$ and T is taken to infinity then $g(x) = \frac{\alpha(1)x}{\alpha(0) - (\alpha(1) - \alpha(0))x}$ which is increasing in x. More generally, if $\alpha(0) < \alpha(1) < \left(\frac{1}{T}\right)^{1/(T-2)}$ (for example, $\alpha(1), \alpha(0) < .5$ and $T > 3$) then one can show that such a g exists and is nondecreasing, although to our knowledge it does not have a tractable analytical representation.

C.1.2 Factor model

See generally Bai et al. (2008); Stock and Watson (2016). We focus on the setting of a "classic factor model" to simplify the exposition. One can extend the example to more general settings, but we leave this to future work.

The returns on N assets in T time periods are described by the factor model

$$X_{it} = \sum_{r=1}^{R} \lambda_{ir} F_{tr} + e_{it}$$

where $R \leq \min(N,T)$, F_{tr} describes the return of factor r in time period t, λ_{ir} describes the exposure of asset i to factor r, and e_{it} describes the idiosyncratic return. The classic normalizations are $\sum_{t=1}^{T} F_{tr} F_{ts} = \mathbb{1}\{r = s\}$, $\sum_{i=1}^{N} \lambda_{ir} \lambda_{is} = \nu_r \mathbb{1}\{r = s\}$, and $E\left[e_{it} e_{jt}\right] = \sigma_{it}^2 \mathbb{1}\{i = j\}$. F_{tr} and λ_{ir} are fixed; e_{it} is stochastic with mean zero.

The outcome of interest is the covariance matrix of asset returns

$$E\left[Y_{ij}\right] = E\left[\sum_{t=1}^{T} X_{it} X_{jt}\right] = \sum_{r=1}^{R} \lambda_{ir} \lambda_{jr} + \sum_{t=1}^{T} \sigma_{it}^2 \mathbb{1}\{i = j\}.$$

Now consider an intervention that scales σ_{it}^2, the variance of the idiosyncratic error. For example, the intervention may be the announcement of an economic policy to be implemented in the future. The effects of the policy are uncertain. Let $\sigma_{it}^2(s)$ denote the variance of e_{it}, $\sigma_i^2(s) = \sum_{t=1}^{T} \sigma_{it}^2(s)$, and $E\left[Y_{ij}\right](s)$ the resulting outcome matrix before and after the announcement as indexed by $s \in \{0, 1\}$. Then

$$E\left[Y_{ij}\right](s) = \sum_{r=1}^{R} \lambda_{ir}\lambda_{jr} + \sum_{t=1}^{T} \sigma_{it}^2(s)\mathbb{1}\{i = j\}.$$

and by the normalization on λ_{ir}

$$E\left[\tilde{Y}_{ab}\right](s) := \sum_{i=1}^{N}\sum_{j=1}^{N} E\left[Y_{ij}\right](s)\lambda_{ia}\lambda_{jb} - \nu_a^2\mathbb{1}\{a = b\} = \sum_{i=1}^{N} \sigma_i^2(s)\lambda_{ia}\lambda_{ib}$$

for $a, b = 1, ..., R$. Then the treatment effect is rank invariant for the rotated and shifted outcome matrix $E\left[\tilde{Y}_{ab}\right]$ if $\sigma_i^2(1) = g(\sigma_i^2(0))$ for some nondecreasing function. For example, if $\sigma_i^2(s) = \rho(s)\sigma_i^2$ for some scalars $\rho(1) > \rho(0)$. This may be the case if the announcement scales the idiosyncratic variation in returns for all assets, regardless of their factor structure.

Practically, this example suggests an identification strategy where the researcher first identifies λ_{ir} and ν_r from Y, constructs $\tilde{Y}$, uses rank invariance to identify the distribution of treatment effects for the outcome matrix $\tilde{Y}$, and then uses λ_{ir} and ν_r to identify the distribution of treatment effects for Y.

C.1.3 Social interaction

This example follows Ballester et al. (2006); Calvó-Armengol et al. (2009). N agents are linked in a social network as described by an $N \times N$ symmetric binary adjacency matrix G. $G_{ij} = 1$ if agents i and j are linked and $G_{ij} = 0$ otherwise. Agents take K real-valued actions. The kth action of agent i is described by A_{ik}. It may describe, for example, how much agent i smokes or invests in a risky venture. The utility agent i receives from choosing action A_{ik} depends on the total amount of the action taken by their peers. Specifically,

$$U_i(A_k) = \eta_{ik} A_{ik} - \frac{1}{2} A_{ik}^2 + \beta \sum_{j=1}^{N} G_{ij} A_{jk} A_{ik}$$

where $\eta_{ik} \sim_{iid} (0, \sigma_k^2)$ is an idiosyncratic shock. The parameter β describes the size of the peer effect. That is, how much agents are influenced by their peers. Under the assumption that $I - \beta G$ is invertible, there exists a unique Nash equilibrium

$$A_k = (I - \beta G)^{-1} \eta_k$$

where $A_k = (A_{1k}, ..., A_{Nk})$.

The outcome of interest is the correlation of actions between agent pairs. It is given by

$$E[Y_{ij}] = E\left[\sum_{k=1}^{K} A_{ik} A_{jk}\right] = \left[(I - \beta G)^{-1} \Sigma \left(I - \beta G\right)^{-1}\right]_{ij}$$

where $\Sigma_{ij} = E[\sum_{k=1}^{K} \eta_{ik}\eta_{jk}]$ is a diagonal matrix with $\Sigma_{ii} = \sum_{k=1}^{K} \sigma_k^2 := \sigma^2$.

Now consider an intervention that increases β, the peer effect size parameter. For example, the intervention may be a school program that better informs students about their peers' actions. Let $\beta(s)$ denote the peer effects parameter and $E[Y_{ij}](s)$ the resulting outcome matrix with and without the program as indexed by $s \in \{0, 1\}$. Then

$$E[Y_{ij}](s) = \sum_{r=1}^{N} \left(\frac{\sigma^2}{(1 - \beta(s)\lambda_r)^2}\right) \phi_{ir}\phi_{jr}$$

where (λ_r, ϕ_r) are the eigenvalue and eigenvector pairs of G. If $\beta(1) > \beta(0) > 0$ then the treatment effect is rank invariant in the sense of Definition 2 of Section 3.2.3 that $E[Y_{ij}](1) = g(E[Y_{ij}](0))$ where g is the matrix lift of a nondecreasing function. Specifically, one can verify that $g(x) = \sigma^2 \left(1 - \frac{\beta(1)}{\beta(0)}\left(1 - \left(\frac{\sigma^2}{x}\right)\right)^{1/2}\right)^{-2}$ which is well defined and nondecreasing for x in the range of outcome matrices Y_0 that satisfy the condition that $I - \beta(0)G$ is invertible.

C.1.4 Link formation

This example follows Graham (2017). N agents are linked in a social network as described by an $N \times N$ symmetric binary adjacency matrix G. $G_{ij} = 1$ if i and j are linked and $G_{ij} = 0$ otherwise. The marginal transferable utility agents i and j receive from forming a link depends on homophily. That is, their proximity in a K-dimensional social characteristic space. Specifically,

$$U_{ij}(G_{ij} = 1) - U_{ij}(G_{ij} = 0) = \alpha_i + \alpha_j - \beta \sum_{k=1}^{K} (x_{ik} - x_{jk})^2 + \eta_{ij}$$

where the fixed effect α_i describes agent i's degree heterogeneity or popularity, x_{ik} describes the kth social characteristic of agent i, and the idiosyncratic error η_{ij} is iid logistic. The parameter β describes the size of the homophily effect. That is, how much link formation is influenced by agent proximity in the social characteristic space.

The conditional probability that utility-maximizing agents i and j form a link is given by

$$E[G_{ij}] = \Lambda(\alpha_i + \alpha_j - \beta \sum_{k=1}^{K} (x_{ik} - x_{jk})^2).$$

where Λ is the standard logistic distribution function. We define the conditional logit function

$$Y_{ij} = \Lambda^{-1}(E[G_{ij}]) = (\alpha_i + \beta \sum_k x_{ik}^2) + (\alpha_j + \beta \sum_k x_{jk}^2) + 2\beta \sum_k x_{ik} x_{jk}.$$

To simplify our exposition, we take as our outcome of interest the demeaned conditional logit function

$$E\left[\tilde{Y}_{ij}\right] = 2\beta \tilde{X}_{ij}$$

where $\tilde{Y}_{ij} = Y_{ij} - \frac{1}{N} \sum_{i=1}^{N} Y_{ij} - \frac{1}{N} \sum_{j=1}^{N} Y_{ij}$ and $\tilde{X}_{ij} = \sum_k x_{ik} x_{jk} - \frac{1}{N} \sum_{i,k=1}^{N} x_{ik} x_{jk} - \frac{1}{N} \sum_{j,k=1}^{N} x_{ik} x_{jk}$. It is straightforward to also demonstrate rank invariance for the undemeaned conditional logit function, following the logic of Section D.1 below.

Now consider an intervention that increases β, the homophily size parameter. For example, the intervention may be a technology that decreases the costs of communication between locations. Let $\beta(s)$ denote the homophily parameter and $E\left[\tilde{Y}_{ij}\right](s)$ the resulting outcome matrix with and without the communication technology as indexed by $s \in \{0,1\}$. Then

$$\tilde{Y}_{ij}(s) = \sum_{r=1}^{N} \left(2\beta(s)\lambda_r\right) \phi_{ir}\phi_{jr}$$

where (λ_r, ϕ_r) are the eigenvalue and eigenvector pairs of $\tilde{X}$. If $\beta(1) > \beta(0) > 0$ then the treatment effect is rank invariant in the sense of Definition 2 of Section 3.2.3 that $E[Y_{ij}](1) = g(E[Y_{ij}](0))$ where g is the matrix lift of a nondecreasing function. Specifically, one can verify that $g(x) = \frac{\beta(1)}{\beta(0)}x$ which is increasing.

C.2 Examples of heterogeneous spillover effects

We provide two additional concrete examples of settings with heterogeneous spillover effects.

C.2.1 Treatment spillovers

This example follows Bajari et al. (2021). Consider a buyer-seller experiment where pairs of buyers and sellers are assigned to an information treatment. For example, the setting may be an online marketplace where a buyer-seller pair is treated if the online platform explicitly recommends the seller's product to the buyer. The outcome of interest Y_{ij} is the size of the transaction between buyer i and seller j. Let $X_{ij} = 1$ if buyer i and seller j are assigned to treatment and $X_{ij} = 0$ otherwise.

Following Bajari et al. (2021), we assume local interference (their Assumption 5.4). That is, the outcome Y_{ij} between buyer i and seller j depends on whether i and j are treated, the number of sellers l for which i and l are treated, and the number of buyers k for which k and j are treated. For example, buyer i may be more likely to buy from seller j if the platform recommends one of seller j's products, all else equal. Buyer i may be less likely to buy from seller j if the platform recommends products from seller j's competitors, all else equal.

Formally, we model

$$Y_{ij} = f_{ij}\left(X_{ij}, \sum_l X_{il}, \sum_k X_{kj}\right).$$

Let $f_{ij}\left(x_b, x_s\right)$ be the expected outcome for agent pair ij when $\sum_l X_{il} = x_b$ and $\sum_k X_{kj} = x_s$ for some $x_b, x_s \in \mathbb{Z}_+$, i.e. $\sum_{t\in\{0,1\}} f_{ij}\left(t, x_b, x_s\right) P(X_{ij} = t)$. In this example, our parameter of interest is the distribution of treatment spillover effects

$$\frac{1}{NM}\sum_{i\in[N],j\in[M]} \mathbb{1}\{f_{ij}\left(x_b^1, x_s^1\right) - f_{ij}\left(x_b^0, x_s^0\right) \leq y\}.$$

In words, it is the fraction of buyer-seller pairs whose change in outcome, after altering the number of relevant treated agent pairs for i and j from (x_b^0, x_s^0) to (x_b^1, x_s^1) is less than y.

To identify the distribution of treatment spillover effects, we propose the following experiment. First randomly assign the treatment to pairs of buyers and sellers. Then form two groups. The first group collects all of the buyers that belong to exactly x_b^1 treated buyer-seller pairs and all of the sellers that belong to exactly x_s^1 treated buyer-seller pairs. The second group similarly collects all of the buyers and sellers that belong to x_b^0 and x_s^0 treated pairs. The next step is to use the matrix of outcomes associated with each group $Y_t := \{f_{ij}\left(X_{ij}, x_b^t, x_s^t\right)\}_{i,j\in \text{ group } t}$ to compute $\bar{Y}_t := \{f_{ij}\left(x_b^t, x_s^t\right)\}_{i,j\in \text{ group } t}$ in the case that $P(X_{ij} = t)$ is fixed by the researcher and estimate $\bar{Y}_t$ in the case that $P(X_{ij} = t)$ is to be recovered from data (see Section D.2 below). Finally, after symmetrization as in Section 5.1, the distribution of spillover effects can be characterized exactly as in Section 3.

C.2.2 Market externalities

Consider the setting of a market economy with N agents and L goods. For a fixed price $p \in \mathbb{R}^{L-1}$, agent i's demand for the lth good is given by the function $Q_{il}(p)$ with $Q_i(p) = \{Q_{i1}(p), ..., Q_{iL}(p)\} \in \mathbb{R}^L$. Q_{il} may be negative in which case i is a supplier of good l. An equilibrium market price $p^*(0)$ is assumed to satisfy the market clearing condition $\sum_{i=1}^N Q_i^*(0) = 0$ where $Q_i^*(0) = Q_i(p^*(0))$ is agent i's equilibrium demand. Absent any market intervention, the equilibrium price and quantity $(p^*(0), Q_1^*(0), ..., Q_N^*(0))$ is realized.

We are interested in understanding the impact of a market intervention such as a price floor on the equilibrium demand matrix between agents and goods. An equilibrium market price $p^*(1)$ is assumed to satisfy the market clearing condition $\sum_{i=1}^N Q_i^*(1) = 0$ and restriction $p^*(1) \geq c$ where $Q_i^*(1) = Q_i(p^*(1))$ is agent i's equilibrium demand under the price floor and $c \in \mathbb{R}^{L-1}$ is chosen by the policy maker. Under the price floor intervention, the equilibrium price and quantity $(p^*(1), Q_1^*(1), ..., Q_N^*(1))$ is realized.

Our interest is in the distribution of treatment effects

$$\frac{1}{NL} \sum_{i \in [N], l \in [L]} \mathbb{1}\{Q_{il}^*(1) - Q_{il}^*(0) \leq y\}.$$

In words, it is the fraction of agent and good pairs whose difference in equilibrium demand with and without the price ceiling is less than y. There are market externalities in this example because while price floor may only nominally restrict one agent or item, the equilibrium condition implies that implementing the policy may result in changes in the equilibrium demand matrix for any agent and item.

To identify the distribution treatment effects with market externalities, we suppose that the researcher is given data from the following natural experiment. They observe a matrix of equilibrium demand for a population of agents and goods from a region without the price floor. They observe another matrix of equilibrium demand for a different population of agents and goods from a region with the price floor. The two regions may not have any agents or goods in common, but they are assumed to be comparable in the sense that the distribution of potential outcomes (here equilibrium demand) is the same for both regions. For example, the regions may be located nearby each other but under different political jurisdictions. After symmetrization as in Section 5.1, the two outcome matrices matrices can be used to characterize the distribution of equilibrium treatment effects exactly as in Section 3.

D Extensions

We provide additional details about some of the extensions described in Section 5 of the main text.

D.1 Row and column heterogeneity

While the bounds from Section 3 are valid for any symmetric outcome matrices, they may be uninformative when there is substantial variation in the row and column variances. In such cases, we recommend allowing for row and column heterogeneity following ideas of Finke et al. (1987).

D.1.1 Bounds on the DPO and DTE

In Section 5.2 of the main text we write the DPO as

$$F(y_1, y_0) = \int\int \prod_{t\in\{0,1\}} \left(\alpha_t(\varphi_t(u)) + \alpha_t(\varphi_t(v))\right) dudv + \int\int \prod_{t\in\{0,1\}} \epsilon_t(\varphi_t(u), \varphi_t(v)) dudv$$

where $\mathbb{1}\{Y_t^*(u,v) \leq y_t\} = \alpha_t(u) + \alpha_t(v) + \epsilon_t(u,v)$ and $\int \epsilon_t(\tau, v)d\tau = \int \epsilon_t(u,\tau)d\tau = 0$ for every $u, v \in [0,1]$. We bound the two summands separately. Specifically, the upper bound is constructed by

$$F(y_1, y_0) \leq \max_{\varphi_1,\varphi_0\in\mathcal{M}} \left[\int\int \prod_{t\in\{0,1\}} \left(\alpha_t(\varphi_t(u)) + \alpha_t(\varphi_t(v))\right) dudv + \int\int \prod_{t\in\{0,1\}} \epsilon_t(\varphi_t(u), \varphi_t(v)) dudv\right]$$
$$\leq \max_{\varphi_1,\varphi_0\in\mathcal{M}} \left[\int\int \prod_{t\in\{0,1\}} \left(\alpha_t(\varphi_t(u)) + \alpha_t(\varphi_t(v))\right) dudv\right] + \max_{\varphi_1,\varphi_0\in\mathcal{M}} \left[\int\int \prod_{t\in\{0,1\}} \epsilon_t(\varphi_t(u), \varphi_t(v)) dudv\right].$$

The first summand is bounded from above by

$$2 \max_{\varphi_1,\varphi_0\in\mathcal{M}} \left[\int \alpha_1(\varphi_1(u))\alpha_0(\varphi_0(u))du\right] + 2\alpha_1\alpha_0 \leq 2\int \alpha_1^+(u)\alpha_0^+(u)du + 2\alpha_1\alpha_0$$

where $\alpha_t = \int \alpha_t(u)du$ and α_t^+ is the "increasing rearrangement" or quantile function of α_t (i.e. α_t^+ is nondecreasing and equal to α_t up to a measure preserving transformation). See Theorem 378 of Hardy et al. (1952) or the second proof of Theorems 2.1 and 2.5 of Whitt (1976) for details. Following Proposition 1 of Section 3, the second summand is bounded from above by

$$\min\left(\sum_r \lambda_{r1}^2, \sum_r \lambda_{r0}^2, \sum_r \lambda_{r1}\lambda_{r0}\right)$$

where λ_{rt} refers to the eigenvalues of ϵ_t and the sums are defined as in Section 3.2.1 of the main text. Together, the two bounds imply that

$$F(y_1, y_0) \leq 2\int \alpha_1^+(u)\alpha_0^+(u)du + 2\alpha_1\alpha_0 + \min\left(\sum_r \lambda_{r1}^2, \sum_r \lambda_{r0}^2, \sum_r \lambda_{r1}\lambda_{r0}\right).$$

By the same logic, the lower bound on the DPO is

$$F(y_1, y_0) \geq 2\int \alpha_1^+(u)\alpha_0^+(1-u)du + 2\alpha_1\alpha_0 + \max\left(\sum_r \left(\lambda_{r1}^2 + \lambda_{r0}^2\right) - 1, \sum_r \lambda_{r1}\lambda_{s(r)0}, 0\right).$$

Bounds on the DTE can be constructed from those on the DPO following exactly the logic of Proposition 2 in Section 3 of the main text.

D.1.2 Spectral treatment effects

Suppose the rank invariance assumption that $\alpha_1(\varphi_1(u)) = g_\alpha(\alpha_0(\varphi_0(u))$ and $\epsilon_1(\varphi_1(u), \varphi_1(v)) = g_\epsilon(\epsilon_0(\varphi_0(u), \varphi_0(v))$ for every $u, v \in [0,1]$ where g_α is a nondecreasing function and g_ϵ is the matrix lift of a nondecreasing function as in Definition 2 of Section 3.2.3. Define the spectral treatment effect with row and column heterogeneity to be

$$STE(u, v; \phi) = \left(\alpha_1^+(u) - \alpha_0^+(u)\right) + \left(\alpha_1^+(v) - \alpha_0^+(v)\right) + \sum_r (\sigma_{r1} - \sigma_{r0})\phi_r(u)\phi_r(v)$$

where $\{\phi_r\}$ is any orthogonal basis in $L^2([0,1])$ and $\{\sigma_{rt}\}$ are the eigenvalues of ϵ_t. Similarly define $STT(u,v) = STE(u,v;\phi_1)$ and $STU(u,v) = STE(u,v;\phi_0)$ where ϕ_1 and ϕ_0 refer to the eigenfunctions of ϵ_1 and ϵ_0 respectively.

Then by the logic of Standard Result 4 in Section 2 and Proposition 4 in Section 3

$$
\begin{aligned}
Y_1^*(u,v) - Y_0^*(u,v) &= (\alpha_1(\varphi_1(u)) - \alpha_0(\varphi_0(u))) + (\alpha_1(\varphi_1(v)) - \alpha_0(\varphi_0(v))) \\
&\quad + (\epsilon_1(\varphi_1(u), \varphi_1(v)) - \epsilon_0(\varphi_0(u), \varphi_0(v))) \\
&= (g_\alpha(\alpha_0(\varphi_0(u))) - \alpha_0(\varphi_0(u))) + (g_\alpha(\alpha_0(\varphi_0(v))) - \alpha_0(\varphi_0(v)) \\
&\quad + (g_\epsilon(\epsilon_0(\varphi_0(u), \varphi_0(v))) - \epsilon_0(\varphi_0(u), \varphi_0(v))) \\
&= (g_\alpha(\alpha_0(\varphi_0(u))) - \alpha_0(\varphi_0(u))) + (g_\alpha(\alpha_0(\varphi_0(v))) - \alpha_0(\varphi_0(v)) \\
&\quad + \sum_r (g_\epsilon(\sigma_{r0}) - \sigma_{r0}) \phi_{r0}(\varphi_0(u)) \phi_{r0}(\varphi_0(v)) \\
&= STE(\varphi_0(u), \varphi_0(v); \phi_0))
\end{aligned}
$$

Since ε_1 and ε_0 are rank invariant, they have the same eigenfunctions (see the proof of Proposition 4 in Section A.5), and so $STE(\varphi_0(u), \varphi_0(v); \phi_0)) = STE(\varphi_0(u), \varphi_0(v); \phi_1))$. It follows that under the above rank invariance assumption $Y_1^* - Y_0^*$, STT, and STU all have the same distribution.

D.2 Randomization inference

Our paper is mainly about identification, but for completeness we also sketch how one can conduct randomization-based inference about the impact of the treatment in three of the motivating examples from Section 1.1 of the main text. We focus on the global point null of no treatment effect, take the populations to be finite and the potential outcomes to be fixed, and do not explicitly consider network interference. One can, however, similarly consider infinite populations, test other such hypotheses, or invert the tests to construct point estimates and confidence intervals in the sense of Hodges and Lehmann (2012); Rosenbaum (2002), see for instance Athey et al. (2018); Basse et al. (2019a;b). Whereas Auerbach (2022b) compares outcome matrices defined on the same set of agents, this section compares outcome matrices defined on two (or more) different sets of agents.

D.2.1 Matched pair design

This example is based on Azoulay et al. (2010), see Example 2 in Section 1.1 of the main text. The data consists of a collection of V matched pairs of research groups. One group from each pair is affected by the death of a superstar ($t = 1$), the other group is not ($t = 0$). We assume that the treatment is (as good as) randomly assigned so that its ex-ante probability is the same for each group in a given pair. The treatment assignment is also independent across pairs of groups.

Let $Y_{ij,sv,t}$ be the potential research productivity of researchers i and j in group $s \in \{1, 2\}$ of pair $v \in [V]$ under treatment $t \in \{0, 1\}$. $\tilde{Y}_{ij,v,t}$ is the observed research productivity for i and j in the group that is actually assigned treatment t in pair v. That is, $\tilde{Y}_{ij,v,1} = Y_{ij,1v,1}$ if the first group in pair v is treated, $\tilde{Y}_{ij,v,1} = Y_{ij,2v,1}$ if the second group in pair v is treated, etc. N_{sv} is the number of researchers in group s of pair v and $\tilde{N}_{tv}$ is the number of researchers in the group that is actually assigned treatment t in pair v.

The null hypothesis is that the research productivity between pairs of researchers in each group are completely unaffected by the death of the superstar. That is,

$$H_0 : Y_{ij,sv,t} = Y_{ij,sv,t'} \text{ for every } i, j \in [N_{sv}], s \in \{1, 2\}, v \in [V] \text{ and } t, t' \in \{0, 1\}.$$

For a test statistic, we propose the difference in the eigenvalues of the thresholded outcome matrices associated with the treated and untreated groups in each pair

$$T = \max_{v \in [V]} \sup_{y \in \mathbb{R}} \sum_r \left(\lambda_{r,v,1}(y) - \lambda_{r,v,0}(y)\right)^2$$

where $\lambda_{r,v,t}(y)$ is the rth eigenvalue of the thresholded outcome matrix $\mathbb{1}\{\tilde{Y}_{v,t} \leq y\}/\tilde{N}_{tv}$ and $\tilde{Y}_{v,t}$ is a $\tilde{N}_{tv} \times \tilde{N}_{tv}$ matrix with ijth entry equal to $\tilde{Y}_{ij,v,t}$.

For a reference distribution, we propose re-randomizing the treatment assignments within each pair. For $A \in \mathbb{N}$ and $a \in [A]$, let $\rho_{v,a}$ be a collection of independent Bernoulli (1/2) random variables,

$$\tilde{Y}^a_{v,1} = \tilde{Y}_{v,1}\rho_{v,a} + \tilde{Y}_{v,0}(1-\rho_{v,a}),$$
$$\tilde{Y}^a_{v,0} = \tilde{Y}_{v,1}(1-\rho_{v,a}) + \tilde{Y}_{v,0}\rho_{v,a},$$

and

$$T^a = \max_{v\in[V]} \sup_{y\in\mathbb{R}} \sum_r \left(\lambda^a_{r,v,1}(y) - \lambda^a_{r,v,0}(y)\right)^2$$

where $\lambda^a_{r,v,t}(y)$ is the rth eigenvalue of the thresholded outcome matrix $\mathbb{1}\{\tilde{Y}^a_{v,t} \leq y\}/\tilde{N}^a_{tv}$, $\tilde{N}^a_{1v} = \tilde{N}_{1v}\rho_{v,a} + \tilde{N}_{0v}(1-\rho_{v,a})$, and $\tilde{N}^a_{0v} = \tilde{N}_{1v}(1-\rho_{v,a}) + \tilde{N}_{0v}\rho_{v,a}$.

By Lehmann and Romano (2006) Theorem 15.2.1, the test that rejects H_0 whenever

$$(A+1)^{-1}\left(1 + \sum_{a\in[A]} \mathbb{1}\{T^a \geq T\}\right) \leq \alpha$$

is level α. It is powered to detect deviations in the eigenvalues of the thresholded outcome matrices associated with each treatment. That such deviations detect a large class of heterogeneous treatment effects follows Proposition 2 in the main text.

D.2.2 Double randomization with uncensored outcomes

This example is based on the conjunctive simple multiple randomization design of Bajari et al. (2021), see Example 4 in Section 1.1 of the main text. A group of B buyers and S sellers are independently randomized to one of two groups. The probability that any buyer or seller is assigned to group 1 is $\pi \in (0,1)$. Every buyer-seller pair where both the buyer and the seller of that pair are assigned to group 1 is given an information treatment.

$Y_{ij,st}$ records the potential transaction between buyer i and seller j in the event that i is assigned to group $s \in \{1,2\}$ and j is assigned to group $t \in \{1,2\}$. $\tilde{Y}_{ij}$ is the observed transaction for buyer i and seller j under their realized group assignments. We call this example uncensored because the researcher observes transactions for every pair of agents. To simplify arguments, we assume that the potential transactions for a buyer-seller pair do not depend on exactly which other buyers or sellers are assigned to groups 1 and 2. This may be the case when the number of buyers and sellers assigned to each group is large.

The null hypothesis is that the group assignments have no effect on the potential transactions between buyers and sellers in the marketplace. That is,

$$H_0 : Y_{ij,st} = Y_{ij,s't'} \text{ for every } i \in [B], j \in [S], s, s', t, t' \in \{1, 2\}.$$

For a test statistic, we propose the difference in the thresholded eigenvalues of the outcome matrices associated with each treatment

$$T = \max_{s,s',t,t' \in \{1,2\}} \sup_{y \in \mathbb{R}} \sum_r \left(\lambda_{r,st}(y) - \lambda_{r,s't'}(y)\right)^2 .$$

where $\lambda_{r,st}(y)$ is the rth eigenvalue of the symmetrized (see Section 5.4 of the main text) thresholded outcome matrix $\{\mathbb{1}\{\tilde{Y}_{ij}^{\dagger} \leq y\}/N_{st}\}_{i,j \in B(s) \cup S(t)}$, $B(s) = \{i \in B : i \text{ is assigned to group } s\}$ and $S(t) = \{j \in S : i \text{ is assigned to group } t\}$, and $N_{st} = |B(s)| + |S(t)|$.

For a reference distribution, we propose re-randomizing the individual treatment assignments. For $A \in \mathbb{N}$ and $a \in [A]$, let $\rho_{i,a}^B$ and $\rho_{j,a}^S$ be a collection of independent Bernoulli(π) random variables, $B^a(s) = \{i \in B : \rho_{i,a}^B + 1 = s\}$ and $S^a(t) = \{j \in S : \rho_{j,a}^S + 1 = t\}$ be the set of buyers and sellers re-randomized to group s and t respectively,

$$\tilde{Y}_{st}^a = \{\tilde{Y}_{ij}\}_{i \in B^a(s), j \in S^a(t)}$$

and

$$T^a = \max_{s,s',t,t' \in \{1,2\}} \sup_{y \in \mathbb{R}} \sum_r \left(\lambda_{r,st}^a(y) - \lambda_{r,s't'}^a(y)\right)^2$$

where $\lambda_{r,st}^a(y)$ is the rth eigenvalue of the symmetrized thresholded outcome matrix $\mathbb{1}\{\tilde{Y}_{st}^{a\dagger} \leq y\}/\tilde{N}_{st}^a$ and $\tilde{N}_{st}^a = |B^a(s)| + |S^a(t)|$.

By Lehmann and Romano (2006) Theorem 15.2.1, the test that rejects H_0 whenever

$$(A+1)^{-1}\left(1+\sum_{a\in[A]}\mathbb{1}\{T^a\geq T\}\right)\leq\alpha$$

is level α. It is powered to detect deviations in the eigenvalues of the thresholded outcome matrices associated with each treatment. That such deviations detect a large class of heterogeneous treatment effects follows Proposition 2 in the main text.

D.2.3 Double randomization with censored outcomes

This example is based on Comola and Prina (2021), see Example 1 in Section 1.1 of the main text. A collection of households are randomly chosen to participate in a savings program. Each household is assigned to participate independently with probability $\pi\in(0,1)$.

$Y_{ij,t}$ records the potential risk sharing link for households i and j when both ($t=1$) or neither ($t=0$) participate in the program. $\tilde{Y}_{ij,t}$ records the observed risk sharing link for households i and j that were actually assigned to treatment t. $\tilde{N}_t$ is the number of households actually assigned to treatment t. We call this example censored because the researcher only observes the potential risk sharing links for pairs of agents assigned to the same treatment. To simplify arguments, we assume that the number of participants is small relative to the number of non-participants (i.e. $\pi\approx 0$).

The null hypothesis is that participation in the savings program has no effect on the potential risk sharing links between pairs of households. That is,

$$H_0: Y_{ij,t}=Y_{ij,t'}\text{ for every }t,t'\in\{0,1\}.$$

For a test statistic, we propose the difference in the eigenvalues of the thresholded outcome matrices associated with the treated and untreated household pairs

$$T=\sup_{y\in\mathbb{R}}\sum_r\left(\lambda_{r,1}(y)-\lambda_{r,0}(y)\right)^2$$

where $\lambda_{r,t}(y)$ is the rth eigenvalue of the thresholded outcome matrix $\mathbb{1}\{\tilde{Y}_t\leq y\}/\tilde{N}_t$.

For a reference distribution, we propose re-randomizing the individual treatment assignments. For $A \in \mathbb{N}$ and $a \in [A]$, let $\rho_{i,a}$ be a collection of independent Bernoulli(π) random variables,

$$\tilde{Y}_1^a = \{Y_{ij,0}\}_{\rho_{i,a}=1, \rho_{j,a}=1},$$

and

$$T^a = \sup_{y \in \mathbb{R}} \sum_r \left(\lambda_{r,1}^a(y) - \lambda_{r,0}(y)\right)^2$$

where $\lambda_{r,1}^a$ is the rth eigenvalue of the thresholded outcome matrix $\mathbb{1}\{\tilde{Y}_1^a \leq y\}/\tilde{N}_1^a$ where $\tilde{N}_1$ is the number of households re-randomized to treatment 1.

By Lehmann and Romano (2006) Theorem 15.2.1, the test that rejects H_0 whenever

$$(A+1)^{-1} \left(1 + \sum_{a \in [A]} \mathbb{1}\{T^a \geq T\}\right) \leq \alpha$$

is level α. It is powered to detect deviations in the eigenvalues of the thresholded outcome matrices associated with each treatment. That such deviations detect a large class of heterogeneous treatment effects follows Proposition 2 in the main text.

D.3 Estimation and large sample inference

Our paper is mainly about identification, but for completeness we also sketch how one can estimate and conduct inference about the bounds on the DTE, DPO, and the distribution of STE using sampled, mismeasured, or missing data. The data is assumed to be drawn from an infinite population. Our sketch relies on kernel density smoothing along the lines of Horowitz (1992). Alternative strategies may potentially lead to more accurate inferences in practice, but we leave their study to future work.

D.3.1 Assumptions about the data generating process

The researcher does not observe the outcome functions $Y_t : [0,1]^2 \to \mathbb{R}$ for $t \in \{0,1\}$. They have instead a stochastic approximation $\hat{Y}_t : [0,1]^2 \to \mathbb{R}$ whose accuracy depends on a sample size N. We give more information about the relationship between $\hat{Y}_t$ and Y_t below.

For example, the researcher may observe the $N \times N$ matrix M_t with ijth entry $M_{ij,t} = (Y_t(w_{i,t}, w_{j,t}) + \epsilon_{ij,t})\, \eta_{ij,t}$. The agent types are sampled from a population as described by $w_{i,t} \sim_{iid} F_w$. The outcomes are observed with measurement error as described by $\epsilon_{ij,t} \sim_{iid} F_\epsilon$. Some outcomes are missing as described by $\eta_{ij,t} \sim_{iid} Bernoulli(p_t)$ with $p_t \in (0,1)$. To construct $\hat{Y}_t$, the researcher first estimates the entries of $Y_t(w_{i,t}, w_{j,t})$ conditional on $\{w_{i,t}\}_{i,\in[N]}$. This may be done by local averaging, k-means clustering, linear regression, principal components analysis, spectral thresholding, etc. See for instance Bai et al. (2008); Bonhomme and Manresa (2015); Chatterjee (2015); Stock and Watson (2016); Jochmans and Weidner (2019); Graham (2020). $\hat{Y}_t$ is then the function embedding (see Appendix Section A.1.1) of this matrix of estimates, reweighted by the inverse density of $w_{i,t}$, see Hsieh et al. (2018).

To demonstrate consistency of our estimators (specified below), our main assumption is that $\hat{Y}_t$ is consistent in MSE. That is,

Assumption D1:

$$MSE(\hat{Y}) := \max_{t\in\{0,1\}} \int\int \left(\hat{Y}_t(u,v) - Y_t(u,v)\right)^2 dudv \to_p 0 \text{ as } N \to \infty.$$

Nearly all of the methods proposed in the literature, including those referenced above, satisfy this property under certain regularity conditions. We show that under Assumption 1 and additional regularity conditions below, the rate of convergence of our estimators decreases with the MSE of $\hat{Y}_t$. As a result, Assumption D1 implies that our estimators are consistent.

To construct a distribution for large sample inference, our main assumption is that Y_t is determined by a linear model. Specifically,

Assumption D2: For $t \in \{0,1\}$, $Y_t(u,v) = X_t(u,v)\beta_t$ and $\hat{Y}_t(u,v) = X_t(u,v)\hat{\beta}_t$ with $\beta_t, \hat{\beta}_t \in \mathbb{R}^K$ for some $K \in \mathbb{N}$, $\left(\hat{\beta}_t - \beta_t\right) \to_d \mathcal{N}(0, V_t)$ as $N \to \infty$ for some covariates $X_t(u,v)$ observed up to a measure preserving transformation (see Section 3.1.4.), V_t can be consistently estimated, and $\hat{\beta}_1$ and $\hat{\beta}_0$ have independent entries.

Linear models are common in economics, see for instance Bonhomme and Manresa (2015); Jochmans and Weidner (2019); Auerbach (2022a). We believe the arguments of this section can also be applied to other classes of models, but at the cost of a more complicated characterization. We leave this to future work.

D.3.2 Additional regularity conditions

We rely on the following regularity conditions, see relatedly Assumptions K1, K2, 6, and 9 of Horowitz (1992). They are modified to fit our setting.

Assumption D3:

i $K : \mathbb{R} \to \mathbb{R}$ is everywhere twice differentiable with $|K|$, $|K'|$, and $|K''|$ uniformly bounded, $\lim_{u\to\infty} K(u) = 0$ and $\lim_{u\to-\infty} K(u) = 1$. $\int [K'(u)]^4 \, du$, $\int [K''(u)]^2 \, du$, and $\int [u^2 K''(u)]^4 \, du$ are finite. For some $P \in \mathbb{N}$, $P \geq 2$, and $p \in [P]$ let $\int |u^p K'(u)| du$ be finite with $\int u^p K'(u) du = 0$.

ii $h(N)$ is a bandwidth sequence such that $h \to 0$, $h^{p-P-1} \int_{|hu|>\eta} |u^p K'(u)| du \to 0$, $h^{-1} \int_{|hu|>\eta} |K''(u)| du \to 0$, and $h^{-P} MSE(\hat{Y}) \to \infty$ as $N \to \infty$ for any $\eta > 0$ and $p \in [P]$.

iii $F_{STE}(y; \phi) = \int \int \mathbb{1}\{STE(u, v; \phi) \leq y\} du dv$ is everywhere smooth with uniformly bounded derivatives.

Assumption D4:

i $K : \mathbb{R} \to \mathbb{R}$ is everywhere twice differentiable with $|K|$, $|K'|$, and $|K''|$ uniformly bounded, $\lim_{u\to\infty} K(u) = 0$, $K(0) = 1$, and $\lim_{u\to-\infty} K(u) = 1$. $\int [K'(u)]^4 \, du$, $\int [K''(u)]^2 \, du$, and $\int [u^2 K''(u)]^4 \, du$ are finite. For some $P \in \mathbb{N}$, $P \geq 2$ and $p \in [P]$ let $\int |u^p K'(u)| du$ be finite with $\int_{u\leq 0} u^p K'(u) du = 0 = \int u^p K(u) K'(u) du = 0$ for $p \in [P]$.

ii $h(N)$ is a bandwidth sequence such that $h \to 0$, $h^{p-P-1} \int_{|hu|>\eta} |u^p K'(u)| du \to 0$, $h^{-1} \int_{|hu|>\eta} |K''(u)| du \to 0$, and $h^{-P} MSE(\hat{Y})^{1/2} \to \infty$ as $N \to \infty$ for any $\eta > 0$ and $p \in [P]$.

iii $F_t(y_t) = \int \int \mathbb{1}\{Y_t(u,v) \leq y_t\} dudv$ is everywhere smooth with uniformly bounded derivatives.

A consequence of Assumption D4(i) is that $2\int K(u)K'(u)du = -1$ and $\int_{u\leq 0} K'(u) = 0$. We use these restrictions in our proofs below.

D.3.3 Consistent estimation of the bounds on the DPO and DTE

Let $\{\lambda_{rt}\}_{r\in[R]}$ and $\{\hat{\lambda}_{rt}\}_{r\in[R]}$ denote the R largest eigenvalues of the functions $\mathbb{1}\{Y_t \leq y_t\}$ and $K\left(\left(\hat{Y}_t - y_t\right)/h\right)$ respectively in absolute value, ordered to be decreasing. To estimate the bounds on the DPO and DTE, we propose using $\sum_{r\in\mathbb{N}} \hat{\lambda}_{rt}\hat{\lambda}_{rt'}$ to estimate $\sum_{r\in\mathbb{N}} \lambda_{rt}\lambda_{rt'}$ for $t, t' \in \{0, 1\}$. We show that

Proposition D1: Under Assumptions D1 and D4

$$\sup_{y_t, y_{t'}\in\mathbb{R}} \left| \sum_{r\in\mathbb{N}} \hat{\lambda}_{rt}\hat{\lambda}_{rt'} - \sum_{r\in\mathbb{N}} \lambda_{rt}\lambda_{rt'} \right| = O_p\left(MSE\left(\hat{Y}\right)\right).$$

Sketch of proof of Proposition D1: Write

$$\left|\sum_{r\in\mathbb{N}} \hat{\lambda}_{rt}\hat{\lambda}_{rt'} - \sum_{r\in\mathbb{N}} \lambda_{rt}\lambda_{rt'}\right| \leq \left|\sum_{r\in\mathbb{N}} \left(\hat{\lambda}_{rt} - \lambda_{rt}\right)\lambda_{rt'}\right| + \left|\sum_{r\in\mathbb{N}} \left(\hat{\lambda}_{rt'} - \lambda_{rt'}\right)\lambda_{rt}\right| + r_N$$

where $r_N = \left|\sum_{r\in\mathbb{N}} \left(\hat{\lambda}_{rt} - \lambda_{rt}\right)\left(\hat{\lambda}_{rt'} - \lambda_{rt'}\right)\right|$ is asymptotically negligible relative to the first two terms. The first two summands are bounded

$$\left|\sum_{r} \left(\hat{\lambda}_{rt} - \lambda_{rt}\right)\lambda_{rt'}\right| \leq \left[\sum_{r} \left(\hat{\lambda}_{rt} - \lambda_{rt}\right)^2\right]^{1/2} F_{t'}(y_{t'})^{1/2}$$

by Cauchy-Schwarz since $\left(\sum_r \lambda_{rt}^2\right)^{1/2} = F_t(y_t)^{1/2}$. The term $\left[\sum_r \left(\hat{\lambda}_{rt} - \lambda_{rt}\right)^2\right]^{1/2}$ is further bounded

$$\left[\sum_r \left(\hat{\lambda}_{rt} - \lambda_{rt}\right)^2\right]^{1/2} \leq \left[\int\int \left(K\left(\frac{\hat{Y}_t(u,v) - y_t}{h}\right) - \mathbb{1}\{Y_t(u,v) \leq y_t\}\right)^2 dudv\right]^{1/2}$$

$$\leq \left[\int\int \left(K\left(\frac{\hat{Y}_t(u,v) - y_t}{h}\right) - K\left(\frac{Y_t(u,v)}{h}\right)\right)^2 dudv\right]^{1/2}$$

$$+ \left[\int\int \left(K\left(\frac{Y_t(u,v) - y_t}{h}\right) - \mathbb{1}\{Y_t(u,v) \leq y_t\}\right)^2 dudv\right]^{1/2}.$$

We show that Assumption D4 implies that the second summand is $o_p\left(h^{\frac{P}{2}}\right)$ in Section D.3.6, Lemma D1 below, see also Horowitz (1992), Lemma 5. The first term in the block of equations is $O_p\left(MSE\left(\hat{Y}\right)\right)$ because

$$\left[\int\int \left(K\left(\frac{\hat{Y}_t(u,v) - y_t}{h}\right) - K\left(\frac{Y_t(u,v) - y_t}{h}\right)\right)^2 dudv\right]^{1/2} =$$

$$\left[\int\int \left(K'\left(\frac{Y_t(u,v) - y_t}{h}\right)\left[\frac{\hat{Y}_t(u,v) - Y_t(u,v)}{h}\right]\right)^2 dudv\right]^{1/2} + s_N$$

where s_N is asymptotically negligible since K is smooth. The claim follows since p is chosen in Assumption D4 so that $h^{P/2}$ is $o_p\left(MSE\left(\hat{Y}\right)\right)$. $\square$

D.3.4 Inference on the bounds on the DPO and DTE

Proposition D2: Under Assumptions D1,D2, and D4

$$P\left(\left|\sum_{r\in\mathbb{N}} \hat{\lambda}_{rt}\hat{\lambda}_{rt'} - \sum_{r\in\mathbb{N}} \lambda_{rt}\lambda_{rt'}\right| > \epsilon\right) \leq P\left(\left(\xi_t' A_t \xi_t\right)^{1/2} F_{t'}(y_{t'})^{1/2} + \left(\xi_{t'}' A_{t'} \xi_{t'}\right)^{1/2} F_t(y_t)^{1/2} > \epsilon\right) + o_p(1)$$

where $A_t = V_t^{1/2}\Omega_t V_t^{1/2}$, $\Omega_t = \int\int \frac{1}{h^2} K'\left(\frac{Y_t(u,v)-y_t}{h}\right)^2 X_t(u,v)' X_t(u,v) dudv$, and $(\xi_1, \xi_0) \sim \mathcal{N}\left(0, I_{dim(\beta_1)+dim(\beta_0)}\right)$.

Sketch of proof of Proposition D2: From the proof sketch of Proposition D1 in Section

D.3.3, Assumptions D1 and D3 imply that

$$\left|\sum_{r\in\mathbb{N}}\hat{\lambda}_{rt}\hat{\lambda}_{rt'}-\sum_{r\in\mathbb{N}}\lambda_{rt}\lambda_{rt'}\right|\leq\left[\int\int\left(K'\left(\frac{Y_t(u,v)-y_t}{h}\right)\left[\frac{\hat{Y}_t(u,v)-Y_t(u,v)}{h}\right]\right)^2dudv\right]^{1/2}F_{t'}(y_{t'})^{1/2}$$
$$+\left[\int\int\left(K'\left(\frac{Y_{t'}(u,v)-y_{t'}}{h}\right)\left[\frac{\hat{Y}_{t'}(u,v)-Y_{t'}(u,v)}{h}\right]\right)^2dudv\right]^{1/2}F_t(y_t)^{1/2}+o_p(1).$$

The claim then follows from Assumption D2, since

$$\int\int\left(K'\left(\frac{Y_t(u,v)-y_t}{h}\right)\left[\frac{\hat{Y}_t(u,v)-Y_t(u,v)}{h}\right]\right)^2dudv$$
$$=\left(\hat{\beta}_t-\beta_t\right)'\left[\int\int\frac{1}{h^2}K'\left(\frac{Y_t(u,v)-y_t}{h}\right)^2X_t(u,v)'X_t(u,v)dudv\right]\left(\hat{\beta}_t-\beta_t\right)\to_d\xi_t'A_t\xi_t.\ \square$$

Once can make inferences about $\sum_{r\in\mathbb{N}}\lambda_{rt}\lambda_{rt'}$ and the bounds on the DPO and DTE in practice by replacing A_t with the estimator $\hat{A}=\hat{V}_t^{1/2'}\hat{\Omega}_t\hat{V}_t^{1/2}$ where $\hat{V}_t$ is a consistent estimator of V_t, $\hat{\Omega}_t=\int\int\frac{1}{h^2}K'\left(\frac{\hat{Y}_t(u,v)-y_t}{h}\right)^2X_t(u,v)'X_t(u,v)dudv$, and $\hat{F}_t(y_t)=\int\int\mathbb{1}\{\hat{Y}_t(u,v)\leq y_t\}dudv$. One can use the right-hand side to construct critical values or confidence intervals in the usual way. Replacing $\frac{1}{h^2}K'\left(\frac{\hat{Y}_t(u,v)-y_t}{h}\right)^2$ with $\sup_u|\frac{1}{h^2}K'\left(\frac{u}{h}\right)^2|$ and $\hat{F}_{t'}(y_{t'})$ with 1 allows for uniformly valid inferences over $y_t,y_{t'}\in\mathbb{R}$.

D.3.5 Consistent estimation of the distribution of spectral treatment effects

Let $\{\sigma_{rt}\}_{r\in[R]}$ and $\{\hat{\sigma}_{rt}\}_{r\in[R]}$ denote the R largest eigenvalues of the functions Y_t and $\hat{Y}_t$ respectively in absolute value, ordered to be decreasing. We propose the estimator $S\hat{T}E(u,v;\phi)=\sum_{r\in[R]}(\hat{\sigma}_{r1}-\hat{\sigma}_{r0})\,\phi_r(u)\phi_r(v)$ for the parameter $STE(u,v;\phi)=\sum_{r\in[R]}(\sigma_{r1}-\sigma_{r0})\,\phi_r(u)\phi_r(v)$ and the estimator $\int\int K\left(\frac{S\hat{T}E(u,v;\phi)-y}{h}\right)dudv$ for the distribution of STE, $\int\int\mathbb{1}\{STE(u,v;\phi)\leq y\}$.

Proposition D3: Under Assumptions D1 and D3

$$\sup_{y\in\mathbb{R}}\left|\int\int K\left(\frac{S\hat{T}E(u,v;\phi)-y}{h}\right)dudv-\int\int\mathbb{1}\{STE(u,v;\phi)\leq y\}dudv\right|=o_p\left(1\right).$$

Sketch of proof of Proposition D3: Write

$$
\begin{aligned}
&\left| \int\int K\left(\frac{S\hat{T}E(u,v;\phi)-y}{h}\right)dudv - \int\int \mathbb{1}\{STE(u,v;\phi)\leq y\}dudv \right| \\
&\leq \left| \int\int K\left(\frac{S\hat{T}E(u,v;\phi)-y}{h}\right)dudv - \int\int K\left(\frac{STE(u,v;\phi)-y}{h}\right)dudv \right| \\
&\qquad + \left| \int\int K\left(\frac{STE(u,v;\phi)-y}{h}\right)dudv - \int\int \mathbb{1}\{STE(u,v;\phi)\leq y\}dudv \right|.
\end{aligned}
$$

We show that Assumption D3 implies that the second summand is $o_p\left(h^P\right)$ in Section D.3.6 below, see also Horowitz (1992)'s Lemma 5. The first summand is $o_p(1)$ because

$$
\begin{aligned}
&\left| \int\int K'\left(\frac{STE(u,v;\phi)-y}{h}\right)\left[\frac{S\hat{T}E(u,v;\phi)-STE(u,v;\phi)}{h}\right]dudv \right| \\
&= \left| \sum_{r\in[R]} \left((\hat{\sigma}_{r1}-\hat{\sigma}_{r0})-(\sigma_{r1}-\sigma_{r0})\right)W_r \right| \leq \left[\sum_{s\in\{0,1\}} ||\hat{Y}_s - Y_s||_2\right]\left(\sum_{r\in[R]} W_r^2\right)^{1/2}
\end{aligned}
$$

plus an asymptotically negligible term, where $W_r = \int\int \frac{1}{h}K'\left(\frac{STE(u,v;\phi)-y}{h}\right)\phi_r(u)\phi_r(v)dudv$ and $\sum_r W_r^2$ is finite because K' is square integrable by assumption. The claim follows since P is chosen in Assumption D3 so that h^P is $o_p\left(MSE\left(\hat{Y}\right)\right)$. □

D.3.6 Inference on the distribution of spectral treatment effects

Proposition D4: Under Assumptions D1-D3

$$
\begin{aligned}
&P\left(\left| \int\int K\left(\frac{S\hat{T}E(u,v;\phi)-y}{h}\right)dudv - \int\int \mathbb{1}\{STE(u,v;\phi)\leq y\}dudv \right| > \epsilon\right) \\
&\qquad\qquad \leq P\left(\left[(\xi_1' B_1 \xi_1)^{1/2} + (\xi_0' B_0 \xi_0)^{1/2}\right]\left(\sum_r W_r^2\right)^{1/2} > \epsilon\right)
\end{aligned}
$$

where $W_r = \int\int \frac{1}{h}K'\left(\frac{STE(u,v;\phi)-y}{h}\right)\phi_r(u)\phi_r(v)dudv$,
$B_t = V_t^{1/2}\left[\int\int X_t(u,v)'X_t(u,v)dudv\right]V_t^{1/2}$, and $(\xi_1\ \xi_0) \to \mathcal{N}\left(0, I_{dim(\beta_1)+dim(\beta_0)}\right)$.

Sketch of proof of Proposition D4: From the proof sketch of Proposition D3 in Section

D.3.5, Assumptions D1 and D3 imply that

$$\left| \int \int K \left(\frac{S\hat{T}E(u,v;\phi) - y}{h} \right) dudv - \int \int \mathbb{1}\{STE(u,v;\phi) \leq y\} dudv \right|$$
$$\leq \left[\sum_{s \in \{0,1\}} ||\hat{Y}_s - Y_s||_2 \right] \left(\sum_{r \in [R]} W_r^2 \right)^{1/2} + o_p \left(MSE \left(\hat{Y} \right) \right)$$

The claim then follows from Assumption D2, since

$$\int \int \left(\hat{Y}_t(u,v) - Y_t(u,v) \right)^2 dudv = \left(\hat{\beta}_t - \beta_t \right)' \left[\int \int X_t(u,v)' X_t(u,v) \right] \left(\hat{\beta}_t - \beta_t \right) \to_d \xi_t' B_t \xi_t. \square$$

One can make inferences about $\int \int \mathbb{1}\{STE(u,v;\phi) \leq y\} dudv$ in practice by replacing W_r with the estimator $\hat{W}_r = \frac{1}{h} \int \int K' \left(\frac{S\hat{T}E(u,v;\phi) - y}{h} \right) \phi_r(u) \phi_r(v) dudv$ and B_t with $\hat{B}_t = \hat{V}_t^{1/2} \left[\int \int X_t(u,v)' X_t(u,v) dudv \right] \hat{V}_t^{1/2}$. One can use the right-hand side to construct critical values or confidence intervals in the usual way.

D.3.7 Additional details about some of the calculations

Lemma D1: Under Assumption D4,

$$\left[\int \int \left(K \left(\frac{Y_t(u,v) - y_t}{h} \right) - \mathbb{1}\{Y_t(u,v) \leq y_t\} \right)^2 dudv \right] = o_p \left(h^P \right).$$

Sketch of proof of Lemma D1: Write

$$\int \int \left(K \left(\frac{Y_t(u,v) - y_t}{h} \right) - \mathbb{1}\{Y_t(u,v) \leq y_t\} \right)^2 dudv$$
$$= \int \int K \left(\frac{Y_t(u,v) - y_t}{h} \right)^2 dudv + F_t(y_t) - 2 \int \int K \left(\frac{Y_t(u,v) - y_t}{h} \right) \mathbb{1}\{Y_t(u,v) \leq y_t\} dudv.$$

The first summand

$$\int\int K\left(\frac{Y_t(u,v)-y_t}{h}\right)^2 dudv = -2\int K(\tau)K'(\tau)F_t(y_t+h\tau)d\tau$$
$$= -2\int K(\tau)K'(\tau)\left[F_t(y_t)+f_t(y_t)h\tau+...+\frac{1}{P!}f_t^P(y_t)h^P\tau^P+o_p\left(h^P\right)\right]d\tau$$
$$= F_t(y_t)+o_p\left(h^P\right)$$

where the first equality is due to a change in variables and integration by parts, the second equality is Taylor's Theorem, and the last equality is by the choice of K in Assumption D4: $2\int K(\tau)K'(\tau)d\tau=-1$ and $\int K(\tau)K'(\tau)\tau^p d\tau=0$ for $p\in[P]$.

Similarly, the third summand

$$\int\int K\left(\frac{Y_t(u,v)-y_t}{h}\right)\mathbb{1}\{Y_t(u,v)\leq y_t\}dudv = \int K'(\tau)\mathbb{1}\{\tau\leq 0\}F_t(y_t+h\tau)d\tau$$
$$= \int K'(\tau)\mathbb{1}\{\tau\leq 0\}\left[F_t(y_t)+f_t(y_t)h\tau+...+\frac{1}{P!}f_t^P(y_t)h^P\tau^P+o_p\left(h^P\right)\right]d\tau$$
$$= o_p\left(h^P\right)$$

where the last equality is also by the choice of K in Assumption D4: $\int_{\tau\leq 0}K'(\tau)\tau^P d\tau=0$ for $p=0$ and $p\in[P]$. The claim follows. $\square$

Lemma D2: Under Assumption D3,

$$\left|\int\int K\left(\frac{STE(u,v;\phi)-y}{h}\right)dudv-\int\int\mathbb{1}\{STE(u,v;\phi)\leq y\}dudv\right| = o_p\left(h^P\right).$$

Sketch of proof of Lemma D2: Write

$$\int\int K\left(\frac{STE(u,v;\phi)-y}{h}\right)dudv = \int K'(\tau)F_t(y_t+h\tau)d\tau$$
$$= \int K'(\tau)\left[F_t(y_t)+f_t(y_t)h\tau+...+\frac{1}{P!}f_t^P(y_t)h^P\tau^P+o_p\left(h^P\right)\right]d\tau$$
$$= F_t(y_t)+o_p\left(h^P\right)$$

where the first equality is due to a change in variables and integration by parts, the second equality is Taylor's Theorem, and the last equality is by the choice of K in Assumption D3: $\int K'(\tau)d\tau = -1$ and $\int K'(\tau)\tau^p d\tau = 0$ for $p \in [P]$. $\square$

D.4 Restricting the weights of the STT to be non-extrapolative

In Section 3.2.2 we interpret the the STT as the difference between Y_1 and a counterfactual formed by a weighted average of Y_0. The weights are potentially extrapolative in the sense that they may be negative and not necessarily integrate to 1 (although they necessarily do satisfy these properties under our matrix generalization of rank invariance in Definition 2 of the main text). In this section, we describe a modified version of the STT that is not extrapolative in the sense that the weights are necessarily nonnegative and integrate to 1, even when the rank invariance condition does not hold.

To do this, we first show that the STT is arbitrarily well approximated by the solution to a quadratic programming problem where the weights are optimized over a set of orthogonal matrices. We then propose a modified version of the problem where the weights are instead optimized over the set of doubly stochastic matrices. By construction, these weights are nonnegative and sum to one.

D.4.1 The STT is well-approximated by the solution to a quadratic programming problem

By definition of the STT in Section 3.2.2, we have that for $W_R(u,s) = \sum_{r=1}^{R} \phi_{r1}(u)\phi_{r0}(s)$,

$$STT(u,v) = Y_1(u,v) - \lim_{R\to\infty} \int \int Y_0(s,t) W_R(u,s) W_R(v,t) ds dt.$$

To motivate our modification of the STT, we approximate Y_1 and Y_0 with the function embeddings of their matrix approximations Y_1^N and Y_0^N (see Lemma 1 of Appendix Section A.1.3 for a definition). Specifically, we define

$$STT^N(u,v) = Y^N_{\lceil Nu\rceil\lceil Nv\rceil,1} - \int\int Y^N_{\lceil Ns\rceil\lceil Nt\rceil,0} W^N_{\lceil Nu\rceil\lceil Ns\rceil} W^N_{\lceil Nv\rceil\lceil Nt\rceil} ds dt$$

where for $i, j \in [N]$, $Y^N_{ij,t}$ is the ijth entry of the $N \times N$ matrix approximation to Y_t, $\{\sigma^N_{rt}, \phi^N_{rt}\}_{r\in[N]}$ are the eigenvalue-eigenvector pairs of Y^N_t, and $W^N_{ik} = \sum_{r\in[N]} \phi^N_{ir1}\phi^N_{kr0}$. Following Lemma 1 in Appendix A.1.3 of the main text,

$$\int\int \left(STT(u,v) - STT^N(u,v)\right)^2 dudv \to 0 \text{ as } N \to \infty.$$

The weights W^N can be derived as the solution to a quadratic programming problem. Specifically,

$$W^N = \text{argmin}_{O\in\mathcal{O}_N} \sum_{i,j\in[N]} \left(Y^N_{ij,1} - \sum_{k,l\in[N]} Y^N_{kl,0} O_{ik} O_{jl} \right)^2 \tag{1}$$

since

$$\begin{aligned} \sum_{i,j\in[N]} \left(Y^N_{ij,1} - \sum_{k,l\in[N]} Y^N_{kl,0} O_{ik} O_{jl} \right)^2 &= \sum_{i,j\in[N]} \left(Y^N_{ij,1}\right)^2 + \sum_{k,l\in[N]} \left(Y^N_{kl,0}\right)^2 - 2 \sum_{i,j,k,l\in[N]} Y^N_{ij,1} Y^N_{kl,0} O_{ik} O_{jl} \\ &= \sum_{r\in[N]} \left(\sigma^N_{r1}\right)^2 + \sum_{s\in[N]} \left(\sigma^N_{s0}\right)^2 - 2 \sum_{r,s\in[N]} \sigma^N_{r1}\sigma^N_{s0} \left[O_{ik}\phi^N_{ir1}\phi^N_{ks0}\right]^2 \\ &\geq \sum \left(\sigma^N_{r1} - \sigma^N_{r0}\right)^2 \end{aligned}$$

where the inequality follows from the fact that $\sum_{r,s\in[N]} \sigma^N_{r1}\sigma^N_{s0} \left[O_{ik}\phi^N_{ir1}\phi^N_{ks0}\right]^2 \leq \sum_{r\in[N]} \sigma^N_{r1}\sigma^N_{r0}$ (see Section 4.3.2). The inequality holds with equality when $O = W^N$, which demonstrates (1).

D.4.2 An alternative non-extrapolative treatment effects parameter

In the quadratic programming problem derivation of W^N, the weight matrix is only restricted to be orthogonal, so it generally may have entries that are negative and do not sum to one. One can impose these restrictions by altering the space of weight matrices to be $\mathcal{D}^+_N$ in (1) instead of $\mathcal{O}_N$. That is,

$$\tilde{X}^N = \text{argmin}_{D \in \mathcal{D}_N^+} \sum_{i,j \in [N]} \left(Y_{ij,1}^N - \sum_{k,l \in [N]} Y_{kl,0}^N D_{ik} D_{jl} \right)^2 \tag{2}$$

and consider the alternative treatment effects parameter

$$Y_{\lceil Nu \rceil \lceil Nv \rceil,1}^N - \int \int Y_{\lceil Ns \rceil \lceil Nt \rceil,0}^N X_{\lceil Nu \rceil \lceil Ns \rceil}^N X_{\lceil Nv \rceil \lceil Nt \rceil}^N ds dt$$

While this problem does not, to our knowledge, have a closed form analytical solution, it is a convex programming problem and so straightforward to solve using standard tools.

References

Athey, Susan, Dean Eckles, and Guido W Imbens, "Exact p-values for network interference," *Journal of the American Statistical Association*, 2018, *113* (521), 230–240.

Auerbach, Eric, "Identification and estimation of a partially linear regression model using network data," *Econometrica*, 2022, *90* (1), 347–365.

_ , "Testing for Differences in Stochastic Network Structure," *Econometrica*, 2022, *90* (3), 1205–1223.

_ **and Yong Cai**, "Heterogeneous Treatment Effects for Networks, Panels, and other Outcome Matrices," *arXiv preprint arXiv:2205.01246*, 2022.

Azoulay, Pierre, Joshua S Graff Zivin, and Jialan Wang, "Superstar Extinction," *The Quarterly Journal of Economics*, 2010, *125* (2), 549–589.

Bai, Jushan, Serena Ng et al., "Large dimensional factor analysis," *Foundations and Trends® in Econometrics*, 2008, *3* (2), 89–163.

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.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou, "Who's who in networks. wanted: the key player," *Econometrica*, 2006, *74* (5), 1403–1417.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, "The diffusion of microfinance," *Science*, 2013, *341* (6144), 1236498.

Basse, Guillaume, Peng Ding, Avi Feller, and Panos Toulis, "Randomization tests for peer effects in group formation experiments," *arXiv preprint arXiv:1904.02308*, 2019.

Basse, Guillaume W, Avi Feller, and Panos Toulis, "Randomization tests of causal effects under interference," *Biometrika*, 2019, *106* (2), 487–494.

Birkhoff, Garrett, "Three observations on linear algebra," *Univ. Nac. Tacuman, Rev. Ser. A*, 1946, *5*, 147–151.

Birman, Michael Sh and Michael Z Solomjak, *Spectral theory of self-adjoint operators in Hilbert space*, Vol. 5, Springer Science & Business Media, 2012.

Bonhomme, Stéphane and Elena Manresa, "Grouped patterns of heterogeneity in panel data," *Econometrica*, 2015, *83* (3), 1147–1184.

Bramoullé, Yann and Garance Genicot, "Diffusion centrality: Foundations and extensions," 2018.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer effects and social networks in education," *The Review of Economic Studies*, 2009, *76* (4), 1239–1267.

Chatterjee, Sourav, "Matrix estimation by universal singular value thresholding," *The Annals of Statistics*, 2015, *43* (1), 177–214.

Comola, Margherita and Silvia Prina, "Treatment effect accounting for network changes," *Review of Economics and Statistics*, 2021, *103* (3), 597–604.

Cruz, Cesi, Julien Labonne, and Pablo Querubin, "Politician family networks and electoral outcomes: Evidence from the Philippines," *American Economic Review*, 2017, *107* (10), 3006–37.

Dudley, Richard M, *Real analysis and probability*, Vol. 74, Cambridge University Press, 2002.

Finke, Gerd, Rainer E Burkard, and Franz Rendl, "Quadratic assignment problems," in "North-Holland Mathematics Studies," Vol. 132, Elsevier, 1987, pp. 61–82.

Graham, Bryan S, "An econometric model of network formation with degree heterogeneity," *Econometrica*, 2017, *85* (4), 1033–1063.

_ , "Network data," in "Handbook of Econometrics," Vol. 7, Elsevier, 2020, pp. 111–218.

Hardy, Godfrey, John Littlewood, and George Pólya, *Inequalities*, Cambridge university press, 1952.

Hodges, Joseph L and Erich L Lehmann, "Estimates of location based on rank tests," in "Selected Works of EL Lehmann," Springer, 2012, pp. 287–300.

Hoffman, AJ and HW Wielandt, "The variation of the spectrum of a normal matrix," *Duke Math J*, 1953, *20*, 37–39.

Horowitz, Joel L, "A smoothed maximum score estimator for the binary response model," *Econometrica*, 1992, pp. 505–531.

Hsieh, Chih-Sheng, Stanley IM Ko, Jaromír Kovářík, and Trevon Logan, "Non-randomly sampled networks: Biases and corrections," Technical Report, National Bureau of Economic Research 2018.

Jochmans, Koen and Martin Weidner, "Fixed-effect regressions on network data," *Econometrica*, 2019, *87* (5), 1543–1560.

Lehmann, Erich L and Joseph P Romano, *Testing statistical hypotheses*, Springer Science & Business Media, 2006.

Rosenbaum, P.R., *Observational Studies*, Springer, 2002.

Stock, James H and Mark W Watson, "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," in "Handbook of macroeconomics," Vol. 2, Elsevier, 2016, pp. 415–525.

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