Identification and Estimation of Spillover Effects in Randomized Experiments: Supplemental Appendix

Gonzalo Vazquez-Bare*

January 19, 2022

Abstract

This supplemental appendix provides additional discussions and results not included in the paper to conserve space.

Contents

*Department of Economics, University of California, Santa Barbara. gvazquez@econ.ucsb.edu

A1 Endogenous Effects and Structural Models

Consider the structural model:

$$Y_{ig} = \phi(D_{ig}, \mathbf{D}_{(i)g}) + \gamma \bar{Y}_g^{(i)} + u_{ig}.$$

which assumes additive separability between the functions that depend on treatment assignments and on outcomes. In this model, $\phi(1, \mathbf{d}_g) - \phi(0, \mathbf{d}_g)$, measures the direct effect of the treatment, $\phi(d, \mathbf{d}_g) - \phi(d, \tilde{\mathbf{d}}_g)$ measures the spillover effects of peers' treatments, commonly known as exogenous or contextual effects, and γ measures the endogenous effect.

Suppose that the assumptions from Corollary 1 hold, so that the treatment vector is randomly assigned, peers are exchangeable and spillover effects are linear. By exchangeability,

$$\begin{aligned}
\phi(D_{ig}, \mathbf{D}_{(i)g}) = \phi(D_{ig}, S_{ig}) &= \sum_{s=0}^{n_g} \tilde{\beta}_s \mathbb{1}(S_{ig} = s)(1 - D_{ig}) + \sum_{s=0}^{n_g} \tilde{\delta}_s \mathbb{1}(S_{ig} = s) D_{ig} \\
&= \tilde{\beta}_0 + (\tilde{\delta}_0 - \tilde{\beta}_0) D_{ig} + \sum_{s=1}^{n_g} (\tilde{\beta}_s - \tilde{\beta}_0) \mathbb{1}(S_{ig} = s)(1 - D_{ig}) \\
&+ \sum_{s=1}^{n_g} (\tilde{\delta}_s - \tilde{\delta}_0) \mathbb{1}(S_{ig} = s) D_{ig}
\end{aligned}$$

where the second equality is without loss of generality because all the variables are discrete, and where $\tilde{\beta}_s = \phi(0, s), \quad \tilde{\delta}_s = \phi(1, s)$. Let $\alpha = \tilde{\beta}_0$, $\beta = \tilde{\delta}_0 - \tilde{\beta}_0$, $\gamma_s^0 = \tilde{\beta}_s - \tilde{\beta}_0$, $\gamma_s^1 = \tilde{\delta}_s - \tilde{\delta}_0$ and rewrite the above model as:

$$\phi(D_{ig}, S_{ig}) = \alpha + \beta D_{ig} + \sum_{s=1}^{n_g} \gamma_s^0 \mathbb{1}(S_{ig} = s)(1 - D_{ig}) + \sum_{s=1}^{n_g} \gamma_s^1 \mathbb{1}(S_{ig} = s) D_{ig}.$$

Next, by linearity of spillover effects, $\gamma_0^d = \kappa_d s$ and $\sum_{s=1}^{n_g} \gamma_s^d \mathbb{1}(S_{ig} = s) = \kappa_d \sum_{s=1}^{n_g} s \mathbb{1}(S_{ig} = s) = \kappa_d S_{ig}$. Therefore,

$$Y_{ig} = \alpha + \beta D_{ig} + \kappa_0 S_{ig}(1 - D_{ig}) + \kappa_1 S_{ig} D_{ig} + \gamma \bar{Y}_g^{(i)} + u_{ig}.$$

In addition, suppose that contextual effects are equal between treated and controls so that $\kappa_0 = \kappa_1 = \kappa$. The model then reduces to:

$$Y_{ig} = \alpha + \beta D_{ig} + \kappa S_{ig} + \gamma \bar{Y}_g^{(i)} + u_{ig} = \alpha + \beta D_{ig} + \kappa n_g \bar{D}_g^{(i)} + \gamma \bar{Y}_g^{(i)} + u_{ig}.$$

Noting that κ can be a function of n_g, $\kappa = \kappa(n_g)$, let $\theta = \kappa(n_g) n_g$ where the dependence on n_g is left implicit, so that:

$$Y_{ig} = \alpha + \beta D_{ig} + \theta \bar{D}_g^{(i)} + \gamma \bar{Y}_g^{(i)} + u_{ig}$$

which is a standard LIM model where β is the direct effect of the treatment, θ is the exogenous or contextual effect and γ is the endogenous effect.

Next, note that $\bar{Y}_g^{(i)} = \frac{n_g+1}{n_g}\bar{Y}_g - \frac{Y_{ig}}{n_g}$ which implies that:

$$Y_{ig}\left(1+\frac{\gamma}{n_g}\right) = \alpha + \beta D_{ig} + \theta \bar{D}_g^{(i)} + \gamma\left(\frac{n_g+1}{n_g}\right)\bar{Y}_g + u_{ig}$$

and

$$\bar{Y}_g = \alpha + \beta \bar{D}_g + \theta \bar{D}_g + \gamma \bar{Y}_g + \bar{u}_g.$$

The last equation implies that, as long as $\gamma \neq 1$,

$$\begin{aligned}\bar{Y}_g &= \frac{\alpha}{1-\gamma} + \frac{\beta+\theta}{1-\gamma}\bar{D}_g + \frac{\bar{u}_g}{1-\gamma} \\ &= \frac{\alpha}{1-\gamma} + \frac{\beta+\theta}{1-\gamma}\left(\frac{1}{n_g+1}\right)D_{ig} + \frac{\beta+\theta}{1-\gamma}\left(\frac{n_g}{n_g+1}\right)\bar{D}_g^{(i)} + \frac{\bar{u}_g}{1-\gamma}\end{aligned}$$

so plugging back:

$$\begin{aligned}Y_{ig}\left(1+\frac{\gamma}{n_g}\right) = \alpha &+ \gamma\left(\frac{n_g+1}{n_g}\right)\frac{\alpha}{1-\gamma} \\ &+ \beta D_{ig} + \gamma\left(\frac{n_g+1}{n_g}\right)\frac{\beta+\theta}{1-\gamma}\left(\frac{1}{n_g+1}\right)D_{ig} \\ &+ \theta \bar{D}_g^{(i)} + \gamma\left(\frac{n_g+1}{n_g}\right)\frac{\beta+\theta}{1-\gamma}\left(\frac{n_g}{n_g+1}\right)\bar{D}_g^{(i)} \\ &+ u_{ig} + \gamma\left(\frac{n_g+1}{n_g}\right)\frac{\bar{u}_g}{1-\gamma}\end{aligned}$$

After some simplifications,

$$\begin{aligned}Y_{ig}\left(1+\frac{\gamma}{n_g}\right) &= \left[1+\left(\frac{n_g+1}{n_g}\right)\frac{\gamma}{1-\gamma}\right]\alpha + \left[\beta + \frac{\gamma}{1-\gamma}\cdot\frac{\beta+\theta}{n_g}\right]D_{ig} \\ &+ \left[\theta + \gamma\cdot\frac{\beta+\theta}{1-\gamma}\right]\bar{D}_g^{(i)} + u_{ig} + \gamma\left(\frac{n_g+1}{n_g}\right)\frac{\bar{u}_g}{1-\gamma}\end{aligned}$$

and thus

$$Y_{ig} = \alpha^* + \beta^* D_{ig} + \theta^* \bar{D}_g^{(i)} + u_{ig}^*$$

where

$$\begin{aligned}\alpha^* &= \left[1+\left(\frac{n_g+1}{n_g}\right)\frac{\gamma}{1-\gamma}\right]\left(1+\frac{\gamma}{n_g}\right)^{-1}\alpha \\ \beta^* &= \left[\beta + \frac{\gamma}{1-\gamma}\cdot\frac{\beta+\theta}{n_g}\right]\left(1+\frac{\gamma}{n_g}\right)^{-1} \\ \theta^* &= \left[\theta + \gamma\cdot\frac{\beta+\theta}{1-\gamma}\right]\left(1+\frac{\gamma}{n_g}\right)^{-1} \\ u_{ig}^* &= u_{ig}\left(1+\frac{\gamma}{n_g}\right)^{-1} + \gamma\left(\frac{n_g+1}{n_g}\right)\left(1+\frac{\gamma}{n_g}\right)^{-1}\frac{\bar{u}_g}{1-\gamma}.\end{aligned}$$

In this context, random assignment of the treatment implies that $\mathbb{E}[u_{ig}|D_{ig}, \mathbf{D}_{(i)g}] = 0$ and hence the reduced-form parameters $(\alpha^*, \beta^*, \theta^*)$ are identified. As in any structural LIM model, however, the structural parameters $(\alpha, \beta, \theta, \gamma)$ are not identified without further assumptions.

A2 Assignment Mechanism for 2SR-FM

In a 2SR-FM assignment mechanism, given a group size $n+1$ groups are assigned to receive $0, 1, 2, \ldots, n+1$ treated units with probabilities $q_0, q_1, \ldots, q_{n+1}$. Treatment assignments in this case are given by $\mathbf{A}_{ig} = (D_{ig}, T_g)$ where $D_{ig} \in \{0, 1\}$ and $T_g \in \{0, 1, \ldots, n+1\}$, and $\pi(\mathbf{a}) = \mathbb{P}[D_{ig} = d|T_g = t]q_t = q_t \left(\frac{t}{n+1}\right)^d \left(1 - \frac{t}{n+1}\right)^{1-d}$. When $n+1$ is odd, the choice of q_t is determined by the following system of equations:

$$q_j = q_{n+1-j}, \quad j \leq \frac{n}{2}$$
$$q_j = \frac{(n+1)q_0}{j}, \quad 0 < j \leq \frac{n}{2}$$
$$\sum_j q_j = 1.$$

The first set of equations imposes symmetry, that is, $\mathbb{P}[T_g = 0] = \mathbb{P}[T_g = n+1]$ and so on. The second set of equations makes the expected sample size in the smallest assignment in each group (untreated units in high-intensity treatment groups and vice versa) equal to the expected sample size of pure controls. The solution to this system is given by:

$$q_0 \left(1 + (n+1) \sum_{j=1}^{\frac{n}{2}} \frac{1}{j}\right) = \frac{1}{2}$$

and the remaining probabilities are obtained from the previous relationships. If $n+1$ is even, the system of equations is given by:

$$q_j = q_{n+1-j}, \quad j \leq \frac{n-1}{2}$$
$$q_j = \frac{(n+1)q_0}{j}, \quad 0 < j \leq \frac{n-1}{2}$$
$$\sum_j q_j = 1.$$

and the solution is:

$$q_0 \left(2 + (n+1) \sum_{j=1}^{\frac{n+1}{2}-1} \frac{1}{j}\right) = \frac{1}{2}.$$

A3 Implications for Experimental Design

Theorem 4 shows that the accuracy of the standard normal to approximate the distribution of the standardized statistic depends on the treatment assignment mechanism through $\underline{\pi}_n$. The intuition behind this result is that the amount of information to estimate each $\mu(\mathbf{a})$ depends on the number of observations facing assignment $\mathbf{a}$, and this number depends on $\pi(\mathbf{a})$. When the goal is to estimate all the $\mu(\mathbf{a})$ simultaneously, the binding factor will be the number of observations in the smallest cell, controlled by $\underline{\pi}_n$. When an assignment sets a value of $\underline{\pi}_n$ that is very close to zero, the normal distribution may provide a poor approximation to the distribution of the estimators.

When designing an experiment to estimate spillover effects, the researcher can choose distribution of treatment assignments $\pi(\cdot)$. Theorem 4 provides a way to rank different assignment mechanisms based on their rate of the approximation, which gives a principled way to choose between different assignment mechanisms.

To illustrate these issues, consider the case of an exchangeable exposure mapping $\mathcal{A}_n = \{(d,s) : d = 0,1, s = 0,1,\ldots,n\}$. The results below compare two treatment assignment mechanisms: simple random assignment (SR) and two-stage randomization with fixed margins (2SR-FM). See Section B for further details on this design.

Corollary A1 (SR) *Under simple random assignment, if:*

$$\frac{n+1}{\log G} \to 0, \tag{1}$$

then $\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n} \to 0$ *and* $\frac{|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n} = O(1)$.

Corollary A2 (2SR-FM) *Under the 2SR-FM mechanism described in Section C, if:*

$$\frac{\log(n+1)}{\log G} \to 0, \tag{2}$$

then $\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n} \to 0$ *and* $\frac{|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n} = O(1)$.

In qualitative terms, both results imply that estimation and inference for spillover effects requires group size to be small relative to the total number of groups. Thus, these results formalize the requirement of "many small groups" that is commonly invoked, for example, when estimating LIM models.

Corollary A1 shows that when the treatment is assigned using simple random assignment, group size has to be small relative to $\log G$. Given the concavity of the log function, this is a strong requirement. Hence, groups have to be very small relative to the sample size for inference to be asymptotically valid. The intuition behind this result is that under a SR, the probability of the tail assignments $(0,0)$ and $(1,n)$ decreases exponentially fast with group size.

On the other hand, Corollary A2 shows that a 2SR-FM mechanism reduces the requirement to $\log(n+1)/\log G \approx 0$, so now the log of group size has to be small compared to the log of the number of groups. This condition is much more easily satisfied, which in practical terms implies that a 2SR-FM mechanism can handle larger groups compared to SR. The intuition behind this result is that, by fixing the number of treated units in each group, a 2SR-FM design has better control on how small the probabilities of each assignment can be, hence facilitating the estimation of the tail assignments. Also note that Condition (2) can be replaced by $n \log n/G \to 0$, $n^2/G = O(1)$.

A4 Unequally-Sized Groups

To explicitly account for different group sizes, let n (the total number of peers in each group) take values in $\mathcal{N} = \{n_1, n_2, \ldots, n_K\}$ where $n_k \geq 1$ for all k and $n_1 < n_2 < \ldots < n_K$. Let the potential outcome be $Y_{ig}(n, d, s(n))$ where $n \in \mathcal{N}$ and $s(n) \in \{0, 1, 2, \ldots, n\}$. Let N_g be the observed value of n, $S_{ig}(n) = \sum_{j \neq i}^{n} D_{jg}$ and $S_{ig} = \sum_{k=1}^{K} S_{ig}(n_k)\mathbb{1}(N_g = n_k)$. The independence assumption can be modified to hold conditional on group size:

$$\{Y_{ig}(n, d, s(n)) : d = 0, 1, s(n) = 0, 1, \ldots, n\}_{i=1}^{n} \perp\!\!\!\perp \mathbf{D}_g(n))|N_g = n$$

where $\mathbf{D}_g(n)$ is the vector of all treatment assignments when the group size is $n+1$.

Under this assumption, we have that for $n \in \mathcal{N}$ and $s \leq n$,

$$\mathbb{E}[Y_{ig}|D_{ig} = d, S_{ig} = s, N_g = n] = \mathbb{E}[Y_{ig}(n, d, s)].$$

The average observed outcome conditional on $N_g = n$ can be written as:

$$\begin{aligned}\mathbb{E}[Y_{ig}|D_{ig}, S_{ig}, N_g = n] = \mathbb{E}[Y_{ig}(n, 0, 0)] &+ \tau_0(n) D_{ig} \\ &+ \sum_{s=1}^{n} \theta_0(s, n)\mathbb{1}(S_{ig} = s)(1 - D_{ig}) \\ &+ \sum_{s=1}^{n} \theta_1(s, n)\mathbb{1}(S_{ig} = s) D_{ig}\end{aligned}$$

The easiest approach is to simply run separate analyses for each group size and estimate all the effects separately. In this case, it is possible to test whether spillover effects are different in groups with different sizes. The total number of parameters in this case is given by $\sum_{k=1}^{K}(n_k + 1)$.

In practice, however, there may be cases in which group size has a rich support with only a few groups at each value n, so separate analyses may not be feasible. In such a setting, a possible solution is to impose an additivity assumption on group size. According to this assumption, the average direct and spillover effects do not change with group size. For example, the spillover effect of having one treated neighbor is the same in a group with two or three units. Under this assumption,

$$\begin{aligned}\mathbb{E}[Y_{ig}|D_{ig}, S_{ig}, N_g] = &\sum_{n\in\mathcal{N}_g} \alpha(n)\mathbb{1}(N_g = n_g) + \tau_0 D_{ig} \\ &+ \sum_{s=1}^{N_g} \theta_0(s)\mathbb{1}(S_{ig} = s)(1 - D_{ig}) \\ &+ \sum_{s=1}^{N_g} \theta_1(s)\mathbb{1}(S_{ig} = s)D_{ig}\end{aligned}$$

where the first sum can be seen in practice as adding group-size fixed effects. Then, the identification results and estimation strategies in the paper are valid after controlling for group-size fixed effects. Note that in this case the total number of parameters to estimate is $n_K + K - 1$ where n_K is the size of the largest group and K is the total number of different group sizes.

Another possibility is to assume that for any constant $c \in \mathbb{N}$, $Y_{ig}(c\cdot n, d, c\cdot s) = Y_{ig}(n, d, s)$. This assumption allows us to rewrite the potential outcomes as a function of the ratio of treated peers, $Y_{ig}(d, s/n)$. Letting $P_{ig} = S_{ig}/N_g$, all the parameters can be estimated by running a regression including D_{ig}, $\mathbb{1}(P_{ig} = p)$ for all possible values of $p > 0$ (excluding $p = 0$ to avoid perfect collinearity) and interactions. In this case, the total number of parameters can be bounded by $n_1 + \sum_{k=2}^{K}(n_k - 1)$. Note that assuming that the potential outcomes depend only on the proportion of treated siblings does not justify including the variable P_{ig} linearly, as commonly done in linear-in-means models.

A5 Including Covariates

There are several reasons why one may want to include covariates when estimating direct and spillover effects. First, pre-treatment characteristics may help reduce the variability of the estimators and decrease small-sample bias, which is standard practice when analyzing randomly assigned programs. Covariates can also help get valid inference when the assignment mechanisms stratifies on baseline covariates. This can be done by simply augmenting Equation (8) with a vector of covariates $\boldsymbol{\gamma}'\mathbf{x}_{ig}$ which can vary at the unit or at the group level. The covariates can also be interacted with the treatment assignment indicators to explore effect heterogeneity across observable characteristics (for example, by separately estimating effects for males and females.

Second, exogenous covariates can be used to relax the mean-independence assumption in observational studies. More precisely, if $\mathbf{X}_g$ is a matrix of covariates, a conditional mean-independence assumption would be $\mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g, \mathbf{D}_g] = \mathbb{E}[Y_{ig}(d, \mathbf{d}_g)|\mathbf{X}_g]$ which is a version of the standard unconfoundeness condition. The vector of covariates can include both individual-level and group-level characteristics.

Third, covariates can be included to make an exposure mapping more likely to be correctly specified. For instance, the exchangeability assumption can be relaxed by assuming it holds after conditioning on covariates, so that for any pair of treatment assignments $\mathbf{d}_g$ and $\tilde{\mathbf{d}}_g$ with the same number of ones, $\mathbb{E}[Y_{ig}(d,\mathbf{d}_g)|\mathbf{X}_g] = \mathbb{E}[Y_{ig}(d,\tilde{\mathbf{d}}_g)|\mathbf{X}_g]$. As an example, exchangeability can be assumed to hold for all siblings with the same age, gender or going to the same school.

All the identification results in the paper can be adapted to hold after conditioning on covariates. In terms of implementation, when the covariates are discrete the parameters of interest can be estimated at each possible value of the matrix $\mathbf{X}_g$, although this strategy can worsen the dimensionality problem. Alternatively, covariates can be included in a regression framework after imposing parametric assumptions, for example, assuming the covariates enter linearly.

A6 Proofs of Technical Lemmas

Proof of Lemma A1 Take $\varepsilon > 0$, then

$$|\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})} - 1\right| > \varepsilon\right] = |\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[|N(\mathbf{a}) - \mathbb{E}[N(\mathbf{a})]| > \varepsilon\mathbb{E}[N(\mathbf{a})]\right]$$

Now, $N(\mathbf{a}) - \mathbb{E}[N(\mathbf{a})] = \sum_g \sum_i \mathbb{1}_{ig}(\mathbf{a}) - G(n+1)\pi(\mathbf{a}) = \sum_g W_g$ where $W_g = \sum_i \mathbb{1}_{ig}(\mathbf{a}) - (n+1)\pi(\mathbf{a}) = N_g(\mathbf{a}) - \mathbb{E}[N_g(\mathbf{a})]$. Note that the W_g are independent, $\mathbb{E}[W_g] = 0$ and:

$$\begin{aligned}
|W_g| &\leq (n+1)\max\{\pi(\mathbf{a}), 1-\pi(\mathbf{a})\} \\
\mathbb{V}[W_g] &= \mathbb{V}\left[\sum_i \mathbb{1}_{ig}(\mathbf{a})\right] = \sum_i \mathbb{V}[\mathbb{1}_{ig}(\mathbf{a})] + 2\sum_i\sum_{j>i}\mathrm{Cov}(\mathbb{1}_{ig}(\mathbf{a}), \mathbb{1}_{jg}(\mathbf{a})) \\
&= (n+1)\pi(\mathbf{a})(1-\pi(\mathbf{a})) + (n+1)(n+2)\{\mathbb{E}[\mathbb{1}_{ig}(\mathbf{a})\mathbb{1}_{jg(\mathbf{a})}] - \pi(\mathbf{a})^2\} \\
&\leq (n+1)\pi(\mathbf{a})(1-\pi(\mathbf{a})) + (n+1)(n+2)\pi(\mathbf{a})(1-\pi(\mathbf{a})) \\
&= (n+1)(n+3)\pi(\mathbf{a})(1-\pi(\mathbf{a})).
\end{aligned}$$

Then, by Bernstein's inequality,

$$\begin{aligned}
\mathbb{P}\left[|W_g|| > \varepsilon\mathbb{E}[N(\mathbf{a})]\right] &\leq 2\exp\left\{-\frac{\mathbb{E}[N(\mathbf{a})]^2\varepsilon^2}{\sum_g \mathbb{V}[W_g] + \frac{1}{3}(n+1)\max\{\pi(\mathbf{a}),1-\pi(\mathbf{a})\}\mathbb{E}[N(\mathbf{a})]\varepsilon}\right\} \\
&= 2\exp\left\{-\frac{\frac{1}{2}G^2(n+1)^2\pi(\mathbf{a})^2\varepsilon^2}{G(n+1)(n+3)\pi(\mathbf{a})(1-\pi(\mathbf{a})) + \frac{1}{3}G(n+1)^2\pi(\mathbf{a})\max\{\pi(\mathbf{a}),1-\pi(\mathbf{a})\}\varepsilon}\right\} \\
&= 2\exp\left\{-\frac{\frac{1}{2}G\pi(\mathbf{a})\varepsilon^2}{\frac{n+3}{n+1}(1-\pi(\mathbf{a})) + \frac{1}{3}\max\{\pi(\mathbf{a}),1-\pi(\mathbf{a})\varepsilon\}}\right\} \\
&\leq 2\exp\left\{-\frac{\frac{1}{2}G\pi(\mathbf{a})\varepsilon^2}{\frac{n+3}{n+1} + \frac{\varepsilon}{3}}\right\}
\end{aligned}$$

Therefore,

$$|\mathcal{A}_n| \max_{\mathbf{a}\in\mathcal{A}_n} \mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\varepsilon\right] \leq 2\exp\left\{-G\underline{\pi}_n\left(\frac{\frac{1}{2}\varepsilon^2}{\frac{n+3}{n+1}+\frac{\varepsilon}{3}}-\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n}\right)\right\} \to 0.$$

as required. □

Proof of Lemma A2 Take $\varepsilon > 0$, then:

$$\mathbb{P}\left[\max_{\mathbf{a}\in\mathcal{A}_n}\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\varepsilon\right] \leq \sum_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\varepsilon\right] \leq |\mathcal{A}_n|\max_{\mathbf{a}\in\mathcal{A}_n}\mathbb{P}\left[\left|\frac{\hat{\pi}(\mathbf{a})}{\pi(\mathbf{a})}-1\right|>\varepsilon\right] \to 0$$

by Lemma A1. □

A7 Proofs of Additional Results

Proof of Corollary A1 Under exchangeability $\pi(\mathbf{a}) = \pi(d,s) = p^d(1-p)^{1-d}\binom{n}{s}p^s(1-p)^{n-s} = \binom{n}{s}p^{s+d}(1-p)^{n+1-s-d}$. This function is minimized at $\underline{\pi}_n = \underline{p}^{n+1}$ where $\underline{p} = \min\{p, 1-p\}$. Thus,

$$\frac{\log|\mathcal{A}_n|}{G\underline{p}^{n+1}} = \exp\left\{-\log G\left(1+\frac{n+1}{\log G}\log\underline{p}-\frac{\log\log|\mathcal{A}_n|}{\log G}\right)\right\}$$

and since $|\mathcal{A}_n| = 2(n+1)$, this term converges to zero when $(n+1)/\log G \to 0$. On the other hand,

$$\frac{|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n} = \frac{2}{G\underline{p}^{n+1}} \leq \frac{2\log|\mathcal{A}_n|}{G\underline{p}^{n+1}} \to 0$$

under the same condition. □

Proof of Corollary A2 Under exchangeability, $\pi(\mathbf{a}) = \pi(d,s) = q_{d+s}\left(\frac{s+1}{n+1}\right)^d \times \left(1-\frac{s}{n+1}\right)^{1-d}$. Under the assignment mechanism in Section B, $\underline{\pi}_n = q_0$ and $q_0 \geq \frac{1}{2(n+3)}$ and thus:

$$\frac{\log|\mathcal{A}_n|}{G\underline{\pi}_n} \leq \frac{2(n+3)\log(2(n+2))}{G} = \exp\left\{-\log G\left(1-\frac{\log(2(n+2))}{\log G}-\frac{\log\log 2(n+1)}{\log G}\right)\right\} \to 0$$

if $\log(n+1)/\log G \to 0$. Finally,

$$\frac{|\mathcal{A}_n|}{G(n+1)\underline{\pi}_n} \leq \frac{4(n+1)(n+3)}{G} = \exp\left\{-\log G\left(1-\frac{\log(n+1)}{\log G}-\frac{\log(4(n+3))}{\log G}\right)\right\} \to 0$$

under the previous condition. □