SUPPLEMENT TO "ESTIMATING THE EFFECT OF TREATMENTS ALLOCATED BY RANDOMIZED WAITING LISTS"

CLÉMENT DE CHAISEMARTIN
Economics Department, University of California at Santa Barbara

LUC BEHAGHEL
Paris School of Economics, INRAE

S1. A SURVEY OF ARTICLES THAT HAVE USED RANDOMIZED WAITING LISTS

IN THIS SECTION, WE PRESENT a survey of articles using randomized waitlists to estimate treatment effects. This survey helps motivate the analytical framework we adopt in Section 3 of the paper. To gather a sample of such articles, we started from six articles using randomized waitlists. Four study the effects of U.S. charter schools. Those are Dobbie and Fryer (2011), Angrist, Pathak, and Walters (2013), Curto and Fryer (2014), and Dobbie and Fryer (2015). Two study the effects of youth training programs in Latin America and the Caribbean. Those are Attanasio, Kugler, and Meghir (2011) and Card, Ibarrarán, Regalia, Rosas-Shady, and Soares (2011). Then, we reviewed the 667 articles cited by, and citing on Google Scholar as of the end of June 2016, those six articles. Among those, we found 37 other articles that also use randomized waitlists, thus leaving us with 43 articles.[1] A list of those 43 articles can be found in Table SI. Twenty-seven are published and 16 are not. They estimate the effects of a variety of interventions, including U.S. charter schools, an agricultural training for ex-fighters in Liberia, or Turkey's vocational training program for the unemployed.

All the treatments considered by these articles have capacity constraints for various groups of applicants, typically defined by their gender, their school grade, or the course they apply for. In each group, a randomized waitlist takes place. Among the 24 articles that report the number of waitlists they use in their analysis, the median is 64. Among the 25 articles that report the average number of applicants per waitlist, the median is 55. Among the 13 articles that report the ratio of seats to applicants, the median is 0.56. Finally, among the 19 articles that report the share of applicants that decline a treatment offer, the median is 0.24.

Most articles estimate the effect of getting an offer on treatment, the first-stage effect (FS). Some articles estimate the intention to treat (ITT) effect of getting an offer only, other articles estimate the LATE only, and many articles estimate both. In the LATE estimations, not all articles use the same instrument. Twenty articles use an indicator for applicants getting an initial offer, the so-called IO instrument. Twenty-two articles use an indicator for applicants ever getting an offer, the so-called EO instrument. These include five articles that use the two instruments. Five articles use other instruments. For instance, one article uses an indicator for applicants receiving an IO and for the 10 applicants ranked below them in each waitlist. Another article uses the IO instrument, but discards

Clément de Chaisemartin: clementdechaisemartin@ucsb.edu
Luc Behaghel: luc.behaghel@ens.fr

[1]This methodology enabled us to find a large number of articles relatively quickly, though it precluded us from obtaining a sample representative of articles using randomized waitlists.

TABLE SI

ARTICLES USING RANDOMIZED WAITLISTS TO ESTIMATE CAUSAL EFFECTS

Article	# Waitlists	Applicants/Waitlist	Seats/Applicant	% Declining Offer	Estimator	Control for Waitlists
Abdulkadiroglu, Angrist, Dynarski, Kane, and Pathak (2011)	80	136	na	0.20	IO and EO	Risk set FE
Abdulkadiroglu, Angrist, Hull, and Pathak (2016)	3	321	na	na	IO and EO	Risk set FE
Acevedo, Cruces, Gertler, and Martinez (2017)	468	35	0.71	na	EO	Lottery FE
Angrist, Dynarski, Kane, Pathak, and Walters (2010)	na	na	na	na	EO	Lottery FE
Angrist, Dynarski, Kane, Pathak, and Walters (2012)	4	112	0.53	na	EO	Lottery FE
Angrist, Pathak, and Walters (2013)	na	na	na	na	EO	Risk set FE
Angrist, Cohodes, Dynarski, Pathak, and Walters (2016)	26	141	na	0.59	IO and EO	Risk set FE
Angrist, Hull, Pathak, and Walters (2017)	na	na	na	na	IO	na
Attanasio, Kugler, and Meghir (2011)	989	45	0.67	0.03	IO	Lottery FE
Behaghel, de Chaisemartin, and Gurgand (2017)	14	28	0.56	0.15	DREO	Reweighting
Blattman and Annan (2016)	70	18	0.48	0.29	EO	Lottery FE
Card, Ibarrarán, Regalia, Rosas-Shady, and Soares (2011)	245	30	0.67	0.17	EO	Lottery FE
Clark, Gleason, Tuttle, and Silverberg (2011)	na	na	na	0.22	EO	Reweighting
Cohodes (2016)	na	na	na	0.30	EO	Risk set FE
Cullen, Jacob, and Levitt (2006)	194	74	0.15	0.64	IO	Lottery FE
Curto and Fryer (2014)	4	55	na	na	EO	Lottery FE
Davis and Heller (2015)	na	na	na	na	Other	Lottery FE
Deming (2011)	na	na	na	0.15	IO	Lottery FE
Deming (2014)	118	22	na	na	IO	Lottery FE
Deming, Hastings, Kane, and Staiger (2014)	na	na	na	na	IO	Lottery FE
Dobbie and Fryer (2011)	7	120	na	0.34	EO and other	Lottery FE
Dobbie and Fryer (2013)	58	170	na	na	Other	Lottery FE
Dobbie and Fryer (2015)	2	300	na	0.37	Other	Lottery FE

(Continues)

TABLE SI—*Continued*

Article	# Waitlists	Applicants/Waitlist	Seats/Applicant	% Declining Offer	Estimator	Control for Waitlists
Furgeson et al. (2012)	na	na	na	na	IO	Reweighting
Gill et al. (2013)	na	na	na	na	IO	Reweighting
Gnagey and Lavertu (2016)	8	26	0.58	0.00	EO	Lottery FE
Hastings, Kane, and Staiger (2006)	na	na	na	na	IO	Lottery FE
Hastings, Kane, and Staiger (2009)	na	na	na	na	IO	Lottery FE
Hastings, Neilson, and Zimmerman (2012)	na	na	na	na	IO	Lottery FE
Hirshleifer, McKenzie, Almeida, and Ridao-Cano (2016)	457	13	na	na	Other	Lottery FE
Hoxby and Rockoff (2004)	na	na	na	na	Other	Lottery FE
Hoxby and Murarka (2009)	725	45	na	0.24	EO	Lottery FE
Ibarraran, Ripani, Taboada, Villa, and Garcia (2014)	295	35	0.57	0.17	IO	Lottery FE
Ibarrarán, Kluve, Ripani, and Rosas (2015)	295	35	0.57	0.17	IO	Lottery FE
Kraft (2014)	na	na	na	na	EO	Lottery FE
McClure, Strick, Jacob-Almeida, and Reicher (2005)	2	102	0.49	na	IO	na
Setren (2015)	161	226	0.31	0.45	IO and EO	Risk set FE
Strick (2012)	na	na	na	na	IO	na
Tuttle, Gleason, and Clark (2012)	na	139	0.36	0.50	EO	Reweighting
Tuttle et al. (2013)	19	62	na	0.28	IO and EO	Reweighting
Tuttle et al. (2015)	na	na	na	na	EO	Reweighting
Walters (2014)	na	na	na	na	EO	Risk set FE
West et al. (2016)	5	40	na	na	EO	Risk set FE

all the applicants that got an offer in a subsequent round.[2] Finally, one article uses the EO instrument in some specifications, and another instrument in other specifications.

Because they combine data from several waitlists, most articles use statistical methods ensuring they compare applicants within and not across waitlists. To do so, seven articles follow Hirano, Imbens, and Ridder (2003) and use propensity score reweighting. Twenty-five articles include waitlist fixed effects in their regressions. Finally, eight articles evaluating the effects of charter schools use a variation of waitlist fixed effects. Students can apply to several schools, and schools conduct their own separate randomized waitlists. As a result, only students applying to the same set of schools have the same probability of entering a charter. Therefore, the authors include fixed effects for each set of applications in their regressions. They refer to these fixed effects as "risk sets".

S2. FURTHER RESULTS

S2.1. *Additional Results on the Bias of $\widehat{\beta}_{\mathrm{FE}}^{E}$*

We extend the comparative statics on B performed in Section 3.3 to the case of heterogeneous waitlists. If, for all k, $N_k = N_0$ and

$$E\left(\left[\frac{1}{T_k}\sum_{i:D_{ik}(1)=1}Y_{ik}(0) - \frac{1}{N_k - T_k}\sum_{i:D_{ik}(1)=0}Y_{ik}(0)\right]\right) = \Delta_{Y(0)},$$

then

$$|B| = \frac{1/K\sum_{k=1}^{K}S_k\left(1 - \frac{t_k}{t_k + 1/N_0}(1 + 1/N_0)\right)}{1/K\sum_{k=1}^{K}S_k\left(1 - \frac{s_k}{t_k + 1/N_0}(1 + 1/N_0)\right)}|\Delta_{Y(0)}|,$$

where t_k and s_k respectively denote the proportion of takers and the ratio of applicants to seats in waitlist k. Then, the derivative of $|B|$ with respect to $1/N_0$ is equal to a ratio whose denominator is strictly positive and whose numerator is

$$\left[\left[1/K\sum_{k=1}^{K}S_k\left(-\frac{t_k}{t_k + 1/N_0} + \frac{t_k}{(t_k + 1/N_0)^2}(1 + 1/N_0)\right)\right]\right.$$
$$\times \left[1/K\sum_{k=1}^{K}S_k\left(1 - \frac{s_k}{t_k + 1/N_0}(1 + 1/N_0)\right)\right]$$
$$- \left[1/K\sum_{k=1}^{K}S_k\left(1 - \frac{t_k}{t_k + 1/N_0}(1 + 1/N_0)\right)\right]$$
$$\left.\times \left[1/K\sum_{k=1}^{K}S_k\left(-\frac{s_k}{t_k + 1/N_0} + \frac{s_k}{(t_k + 1/N_0)^2}(1 + 1/N_0)\right)\right]\right]|\Delta_{Y(0)}|.$$

---

[2]Relatedly, in another article, researchers randomly assign applicants to three groups: the treatment group, the control group, and the replacement group. Program implementers only pick non-takers' replacements from the replacement group. In the end, researchers compare the treatment and control groups and discard the replacement group.

One has

$$-\frac{t_k}{t_k + 1/N_0} + \frac{t_k}{(t_k + 1/N_0)^2}(1 + 1/N_0)$$

$$= \frac{t_k(1 - t_k)}{(t_k + 1/N_0)^2} > \frac{s_k(1 - t_k)}{(t_k + 1/N_0)^2} = -\frac{s_k}{t_k + 1/N_0} + \frac{s_k}{(t_k + 1/N_0)^2}(1 + 1/N_0) > 0.$$

Moreover,

$$1 - \frac{s_k}{t_k + 1/N_0}(1 + 1/N_0) > 1 - \frac{t_k}{t_k + 1/N_0}(1 + 1/N_0) \geq 0,$$

so $|B|$ is increasing in $1/N_0$, and therefore decreasing in N_0.

Similarly, if, for all k, $t_k = t_0$ and $E([\frac{1}{T_k}\sum_{i:D_{ik}(1)=1} Y_{ik}(0) - \frac{1}{N_k - T_k}\sum_{i:D_{ik}(1)=0} Y_{ik}(0)]) = \Delta_{Y(0)}$, then

$$|B| = \frac{1/K\sum_{k=1}^{K} S_k\left(1 - \frac{t_0}{t_0 + 1/N_k}(1 + 1/N_k)\right)}{1/K\sum_{k=1}^{K} S_k\left(1 - \frac{s_k}{t_0 + 1/N_k}(1 + 1/N_k)\right)}|\Delta_{Y(0)}|,$$

Then, the derivative of $|B|$ with respect to t_0 is equal to a ratio whose denominator is strictly positive and whose numerator is

$$\left[\left[1/K\sum_{k=1}^{K} -S_k\frac{1}{(t_0 + 1/N_k)^2}1/N_k(1 + 1/N_k)\right]\right.$$

$$\times \left[1/K\sum_{k=1}^{K} S_k\left(1 - \frac{s_k}{t_0 + 1/N_k}(1 + 1/N_k)\right)\right]$$

$$- \left[1/K\sum_{k=1}^{K} S_k\left(1 - \frac{t_0}{t_0 + 1/N_k}(1 + 1/N_k)\right)\right]$$

$$\left.\times \left[1/K\sum_{k=1}^{K} S_k\frac{s_k}{(t_0 + 1/N_k)^2}(1 + 1/N_k)\right]\right]|\Delta_{Y(0)}|,$$

which is strictly negative, so $|B|$ is decreasing in t_0.

Finally, if, for all k, $s_k = s_0$ and

$$E\left(\left[\frac{1}{T_k}\sum_{i:D_{ik}(1)=1} Y_{ik}(0) - \frac{1}{N_k - T_k}\sum_{i:D_{ik}(1)=0} Y_{ik}(0)\right]\right) = \Delta_{Y(0)},$$

then

$$|B| = \frac{1/K \sum_{k=1}^{K} \left(N_k - \frac{t_k}{t_k + 1/N_k}(N_k + 1) \right)}{1/K \sum_{k=1}^{K} \left(N_k - \frac{s_0}{t_k + 1/N_k}(N_k + 1) \right)} |\Delta_{Y(0)}|,$$

which is clearly increasing in s_0.

<h2 style="text-align:center">S2.2. Asymptotic Distributions of $\widehat{\beta}_{\mathrm{FE}}^{E}$ and $\widehat{\Delta} - \widehat{\beta}_{\mathrm{FE}}^{E}$</h2>

In the section, we derive the asymptotic distributions of $\widehat{\beta}_{\mathrm{FE}}^{E}$ and $\widehat{\Delta} - \widehat{\beta}_{\mathrm{FE}}^{E}$. Let

$$\beta_{\mathrm{FE},K}^{E} = \frac{\frac{1}{K} \sum_{k=1}^{K} E(RF_k^E)}{\frac{1}{K} \sum_{k=1}^{K} E(FS_k^E)}. \tag{31}$$

Let also $FS^E = \lim_{K \to +\infty} \frac{1}{K} \sum_{k=1}^{K} E(FS_k^E)$ and $\beta_{\mathrm{FE}}^{E} = \lim_{K \to +\infty} \beta_{\mathrm{FE},K}^{E}$, where Assumption 7 ensures that those limits exist. Finally, for all k, let $\Lambda_k^E = \frac{RF_k^E - \beta_{\mathrm{FE}}^{E} FS_k^E}{FS^E}$.

ASSUMPTION 6—Assumptions to Derive the Asymptotic Distributions of $\widehat{\beta}_{\mathrm{FE}}^{E}$ and $\widehat{\Delta} - \widehat{\beta}_{\mathrm{FE}}^{E}$: *The following sequences have finite limits when* $K \to +\infty$: (i) $\frac{1}{K} \sum_{k=1}^{K} V(RF_k^E)$, (ii) $\frac{1}{K} \sum_{k=1}^{K} V(FS_k^E)$, (iii) $\frac{1}{K} \sum_{k=1}^{K} E(RF_k^E FS_k^E)$, (iv) $\frac{1}{K} \sum_{k=1}^{K} V(\Lambda_k - \Lambda_k^E)$, (v) $\frac{1}{K} \sum_{k=1}^{K} E(RF_k \times RF_k^E)$, (vi) $\frac{1}{K} \sum_{k=1}^{K} E(RF_k FS_k^E)$, (vii) $\frac{1}{K} \sum_{k=1}^{K} E(FS_k RF_k^E)$, (viii) $\frac{1}{K} \sum_{k=1}^{K} E(FS_k FS_k^E)$, (ix) $\frac{1}{K} \sum_{k=1}^{K} E(|RF_k^E - E(RF_k^E)|^4)$, (x) $\frac{1}{K} \sum_{k=1}^{K} E(|FS_k^E - E(FS_k^E)|^4)$, (xi) $\frac{1}{K} \sum_{k=1}^{K} E(|\Lambda_k^E - E(\Lambda_k^E)|^4)$, *and* (xii) $\frac{1}{K} \sum_{k=1}^{K} E(|\Lambda_k - \Lambda_k^E - E(\Lambda_k - \Lambda_k^E)|^4)$.

Let

$$\sigma_E^2 = \lim_{K \to +\infty} \frac{1}{K} \sum_{k=1}^{K} V(\Lambda_k^E), \qquad \sigma_{E,+}^2 = \lim_{K \to +\infty} \left[\frac{1}{K} \sum_{k=1}^{K} E((\Lambda_k^E)^2) - \left(\frac{1}{K} \sum_{k=1}^{K} E(\Lambda_k^E) \right)^2 \right],$$

$$\widehat{\Lambda}_k^E = \frac{RF_k^E - \widehat{\beta}_{\mathrm{FE}}^{E} FS_k^E}{\frac{1}{K} \sum_{k=1}^{K} FS_k^E}, \qquad \widehat{\sigma}_{E,+}^2 = \frac{1}{K} \sum_{k=1}^{K} \left(\widehat{\Lambda}_k^E - \frac{1}{K} \sum_{j=1}^{K} \widehat{\Lambda}_j^E \right)^2,$$

$$\sigma_{\mathrm{diff}}^2 = \lim_{K \to +\infty} \frac{1}{K} \sum_{k=1}^{K} V(\Lambda_k - \Lambda_k^E),$$

$$\sigma_{\mathrm{diff},+}^2 = \lim_{K \to +\infty} \left[\frac{1}{K} \sum_{k=1}^{K} E((\Lambda_k - \Lambda_k^E)^2) - \left(\frac{1}{K} \sum_{k=1}^{K} E(\Lambda_k - \Lambda_k^E) \right)^2 \right],$$

$$\widehat{\sigma}_{\text{diff},+}^2 = \frac{1}{K}\sum_{k=1}^{K}\left(\widehat{\Lambda}_k - \widehat{\Lambda}_k^E - \frac{1}{K}\sum_{j=1}^{K}(\widehat{\Lambda}_j - \widehat{\Lambda}_j^E)\right)^2.$$

THEOREM S2.1: *If Assumptions* 1–4, 6, *and* 7 *hold*:
(a) $\sqrt{K}(\widehat{\beta}_{\text{FE}}^E - \beta_{\text{FE},K}^E) \xrightarrow{d} \mathcal{N}(0, \sigma_E^2)$ *and* $\widehat{\sigma}_{E,+}^2 \xrightarrow{p} \sigma_{E,+}^2 \geq \sigma_E^2$,
(b) $\sqrt{K}(\widehat{\Delta} - \widehat{\beta}_{\text{FE}}^E - (\Delta_K - \beta_{\text{FE},K}^E)) \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{diff}}^2)$ *and* $\widehat{\sigma}_{\text{diff},+}^2 \xrightarrow{p} \sigma_{\text{diff},+}^2 \geq \sigma_{\text{diff}}^2$.

Theorem S2.1 implies that $\widehat{\beta}_{\text{FE}}^E$ and $\widehat{\Delta} - \widehat{\beta}_{\text{FE}}^E$ are asymptotically normal. As in Theorem 3.1 in the paper, their asymptotic variances can only be conservatively estimated.

S2.3. *Which Fraction of the Bias of the EO Estimator Respectively Come From the Endogenous Stopping Rule and the Waitlist Fixed Effects?*

As explained in Section 2 of the paper, the inconsistency of $\widehat{\beta}_{\text{FE}}^E$ stems from the endogenous stopping rule in the offer process, as well as from the inclusion of waitlist fixed effects in the regression. One may then wonder which fraction of the total asymptotic bias these two issues respectively account for. To answer this question, we study the probability limit of

$$\widehat{\beta}_{\text{FE}}^D = \frac{\dfrac{1}{K}\displaystyle\sum_{k=1}^{K} N_k \dfrac{L_k}{N_k}\left(1 - \dfrac{L_k}{N_k}\right)\left(\dfrac{1}{L_k - 1}\displaystyle\sum_{i:Z_{ik}=1} w_{ik}Y_{ik} - \dfrac{1}{N_k - L_k}\displaystyle\sum_{i:Z_{ik}=0} Y_{ik}\right)}{\dfrac{1}{K}\displaystyle\sum_{k=1}^{K} N_k \dfrac{L_k}{N_k}\left(1 - \dfrac{L_k}{N_k}\right)\dfrac{1}{L_k - 1}\displaystyle\sum_{i:Z_{ik}=1} w_{ik}D_{ik}},$$

an estimator downweighting takers receiving an offer as the DREO estimator, but reweighting lotteries as the EO estimator. One can show that if there exist real numbers $\Delta_{Y(0)}$, N_0, T_0, and S_0 such that for every k,

$$E\left(\left[\frac{1}{T_k}\sum_{i:D_{ik}(1)=1} Y_{ik}(0) - \frac{1}{N_k - T_k}\sum_{i:D_{ik}(1)=0} Y_{ik}(0)\right]\right) = \Delta_{Y(0)},$$

$$N_k = N_0, \qquad T_k = T_0, \quad \text{and} \quad S_k = S_0,$$

then under Assumptions 1–4 and a technical condition similar to Assumption 7, $\widehat{\beta}_{\text{FE}}^D$ converges in probability to the sum of two terms when $K \to +\infty$. The first term is a weighted average of the LATEs of takers in each lottery, while the second term is a bias term equal to

$$\frac{\left(1 - \dfrac{1}{N_0}\right)\left(S_0 - 1 + \dfrac{T_0}{N_0}\right) - (T_0 - 1)\dfrac{S_0}{T_0 + 1}\dfrac{N_0 + 1}{N_0}}{\left(1 - \dfrac{1}{N_0}\right)\left(S_0 - 1 + \dfrac{T_0}{N_0}\right) - (S_0 - 1)\dfrac{S_0}{T_0 + 1}\dfrac{N_0 + 1}{N_0}}\Delta_{Y(0)}. \tag{32}$$

Using (32) above and (3) in the paper, one can decompose the asymptotic bias of $\widehat{\beta}_{\text{FE}}^E$ in any DGP with homogeneous lotteries. For instance, if $N_0 = 20$, $T_0 = 12$, $S_0 = 10$, $\Delta_{Y(0)} = -0.4$, and if the treatment effect is constant and equal to 0.2, the limits of $\widehat{\beta}_{\text{FE}}^E$ and $\widehat{\beta}_{\text{FE}}^D$ are respectively equal to 0.136 and 0.149, thus implying that in this DGP, the lottery fixed effects account for 79% of the total asymptotic bias of the EO estimator.

S2.4. *Assessing Whether Heterogeneous LATEs Across Waitlists Can Account for the Difference Between the EO and DREO Estimators*

Under Assumptions 1–7, Theorem 3.2 in the paper shows that $\widehat{\beta}_{\text{FE}}^{E}$ converges towards the sum of two terms. The first is a weighted average of the LATEs of takers in each waitlist. The second term, B, is a bias term. Even if $B = 0$, the probability limits of $\widehat{\beta}_{\text{FE}}^{E}$ and $\widehat{\Delta}$ may differ if takers' LATEs are heterogeneous across waitlists.

To assess whether heterogeneous LATEs across waitlists are likely to explain the difference between $\widehat{\beta}_{\text{FE}}^{E}$ and $\widehat{\Delta}$, one can compare $\widehat{\beta}_{\text{PS}}^{I}$ to $\widehat{\beta}_{\text{FE}}^{I}$, the coefficient of D_{ik} in a 2SLS regression of Y_{ik} on D_{ik} and waitlist fixed effects using Z'_{ik} as the instrument. $\widehat{\beta}_{\text{PS}}^{I}$ converges towards the LATE of takers, while $\widehat{\beta}_{\text{FE}}^{I}$ converges towards a weighted average of takers' LATEs in each waitlist, so any statistically significant difference between these estimators must come from heterogeneous LATEs across waitlists. If $\widehat{\beta}_{\text{PS}}^{I}$ and $\widehat{\beta}_{\text{FE}}^{I}$ are close, heterogeneous LATEs are unlikely to explain the difference between $\widehat{\beta}_{\text{FE}}^{E}$ and $\widehat{\Delta}$. This test remains suggestive, because the reweightings of waitlists attached to $\widehat{\beta}_{\text{FE}}^{I}$ and $\widehat{\beta}_{\text{FE}}^{E}$ are not the same.[3]

S2.5. *The DREO Estimator With Covariates*

Assume one wants to include covariates in the estimation, to increase statistical precision. The DREO estimator with covariates we propose is $\widehat{\Delta}^{X}$, the coefficient of Z_{ik} in a 2SLS regression of Y_{ik} on D_{ik} and X_{ik}, using Z_{ik} as the instrument and weighted by the weights w_{ik}^{DR} defined in the paper. One can show that $\sqrt{K}(\widehat{\Delta}^{X} - \Delta_{K})$ converges towards a normal distribution under Assumptions 1 and 3, and modified versions of Assumptions 2 and 4 accounting for the covariates.

S2.6. *The DREO Estimator With a Non-Binary Treatment*

Throughout the paper, we have assumed that treatment is binary. When the treatment takes a finite number of values $\{0, 1, \ldots, \overline{d}\}$, one can still use the DREO estimator, provided one replaces w_{ik} by $1 - Z_{ik}1\{D_{ik} > 0\}/S_{k}$ in its definition. Then, one can show that $\widehat{\Delta}$ consistently estimates the average causal response parameter defined in Angrist and Imbens (1995).

S3. MONTE CARLO SIMULATIONS

S3.1. *Comparing the EO, IO, and DREO Estimators*

In this section, we run simulations to compare the IO, EO, and DREO estimators. We consider designs with 120 and 240 waitlists. Each waitlist has 40 applicants (30 takers and 10 non-takers), and 20 seats for treatment. To choose these numbers, we used our survey in Section S1. For instance, 38% of articles therein use at least 120 waitlists in their analysis, and 29% use at least 240 waitlists. Then, we respectively draw values of $Y_{ik}(0)|D_{ik}(1) = 1$ and $Y_{ik}(0)|D_{ik}(1) = 0$ from $\mathcal{N}(0, 1)$ and $\mathcal{N}(0.4, 1)$ distributions, so the mean of $Y_{ik}(0)$ is 0.4 standard deviation (σ) larger for non-takers than takers. The treatment effect is constant across applicants and waitlists: $Y_{ik}(1) - Y_{ik}(0) = 0.2$. Once

---

[3]On the other hand, comparing the EO estimator with propensity score reweighting to $\widehat{\beta}_{\text{FE}}^{E}$ is uninformative as to LATEs' heterogeneity across waitlists.

potential treatments and outcomes have been drawn, applicants are randomly ranked, offers are made according to that ranking until all seats are filled, D_{ik} and Y_{ik} are determined accordingly, and we estimate the IO, EO, and DREO estimators. We repeat this procedure 500 times, and report the mean and standard error (SE) of each estimator, as well as the coverage rate of their 95% confidence interval (CI). To construct those CIs, we use the robust standard errors of the IO and EO estimators, and $\widehat{\sigma}_+/\sqrt{K-1}$ for the DREO estimator.[4] We also consider two other designs. In Design 2, there are 24 takers per waitlist, and all the other parameters are the same as in Design 1. In Design 3, there are 20 applicants and 10 seats per waitlist, 12 applicants are takers, and all the other parameters are the same as in Design 1. Note that in these designs, all waitlists have the same number of applicants, the same number of takers, and the same expectations of takers' and non-takers' potential outcomes, so the confidence interval attached to the DREO estimator should not be conservative.

Results are shown in Table SII below. In all designs, $\widehat{\beta}_{\mathrm{FE}}^E$ is a biased estimator of Δ_K. This is despite the fact that all these simulation designs have homogeneous waitlists: as discussed in the paper, the EO estimator is less biased with homogeneous than with heterogeneous waitlists. On the other hand, $\widehat{\beta}_{\mathrm{PS}}^I$ and $\widehat{\Delta}$ are not visibly biased. In Design 1, the bias of $\widehat{\beta}_{\mathrm{FE}}^E$ is too small to distort the coverage of its confidence interval. In Designs 2 and 3, this bias strongly distorts coverage, especially with 240 waitlists. On the contrary, the confidence intervals of $\widehat{\beta}_{\mathrm{PS}}^I$ and $\widehat{\Delta}$ have the desired coverage in all the panels of the table. Interestingly, in all the panels, the mean of $\widehat{\beta}_{\mathrm{FE}}^E$ is close to its probability limit. Indeed, using (3) in the paper, one can compute that this limit is respectively equal to 0.190, 0.164, and 0.136 in Designs 1 to 3. Across all panels, the standard error of $\widehat{\Delta}$ is 27.6 to 57.3% smaller than that of $\widehat{\beta}_{\mathrm{PS}}^I$. On the other hand, the standard error of $\widehat{\Delta}$ is 2.4 to 12.5% larger than that of $\widehat{\beta}_{\mathrm{FE}}^E$. The mean squared error of $\widehat{\Delta}$ is still lower than that of $\widehat{\beta}_{\mathrm{FE}}^E$ in all panels, but one can find DGPs where the converse holds.

S3.2. *Assessing the Robustness of the Asymptotic Formulas Derived in Theorem 3.3 to Violations of Assumption 5*

In this subsection, we provide simulations to assess the robustness of the asymptotic formulas in Theorem 3.3 to violations of points (c), (d) and (f) of Assumption 5. The simulations are based on the same design as in Panel E of Table SII, with homogeneous lotteries such that $N_k = 40$, $S_k = 20$, and $T_k = 24$ for all k. This implies that point (f) of Assumption 5 is not satisfied. In a first DGP (Panel A of Table SIII), we draw values of $Y_{ik}(0)|D_{ik}(1) = 1$ and $Y_{ik}(0)|D_{ik}(1) = 0$ from the same $\mathcal{N}(0,1)$ distribution, so that all points of Assumption 5 except (f) are satisfied. In a second DGP (Panel B), we draw values of $Y_{ik}(0)|D_{ik}(1) = 1$ and $Y_{ik}(0)|D_{ik}(1) = 0$ from different distributions ($\mathcal{N}(0,1)$ and $\mathcal{N}(0.4,1)$ distributions respectively, as in Panel E of Table SII), so that points (c) and (f) are violated. The third DGP (Panel C) has heterogeneous treatment effects with values of $Y_{ik}(1) - Y_{ik}(0)|D_{ik}(1) = 1$ drawn from a $\mathcal{N}(0.2,1)$ distribution, so that points (c), (d), and (f) are violated. For these three DGPs, Table SIII displays the mean and standard deviation of the simulated IO and DREO estimators over 500 replications (Columns 2 and 3 respectively), and the standard errors computed following the asymptotic formulas derived in Theorem 3.3 (Column 4). The results show that the approximations provided

---

[4]The coverage of the DREO estimator's CI in Table SII remains similar if we instead use robust standard errors.

TABLE SII

SIMULATION RESULTS[a]

	Average	SE	Coverage 95% CI
Simulations with 120 waitlists			
A. Design 1: $N_k = 40$, $S_k = 20$, $T_k = 30$, $\Delta_{Y(0)} = -0.4$, $\Delta_K = 0.2$			
$\widehat{\beta}^I_{PS}$	0.201	0.058	0.956
$\widehat{\beta}^E_{FE}$	0.190	0.041	0.932
$\widehat{\Delta}$	0.199	0.042	0.952
B. Design 2: 24 takers per lottery, otherwise same as Design 1			
$\widehat{\beta}^I_{PS}$	0.210	0.151	0.958
$\widehat{\beta}^E_{FE}$	0.164	0.062	0.900
$\widehat{\Delta}$	0.199	0.067	0.950
C. Design 3: Lotteries twice smaller, otherwise same as Design 2			
$\widehat{\beta}^I_{PS}$	0.210	0.199	0.970
$\widehat{\beta}^E_{FE}$	0.135	0.080	0.894
$\widehat{\Delta}$	0.198	0.090	0.956
Simulations with 240 waitlists			
D. Design 1: $N_k = 40$, $S_k = 20$, $T_k = 30$, $\Delta_{Y(0)} = -0.4$, $\Delta_K = 0.2$			
$\widehat{\beta}^I_{PS}$	0.198	0.039	0.968
$\widehat{\beta}^E_{FE}$	0.189	0.026	0.962
$\widehat{\Delta}$	0.199	0.027	0.964
E. Design 2: 24 takers per lottery, otherwise same as Design 1			
$\widehat{\beta}^I_{PS}$	0.207	0.103	0.944
$\widehat{\beta}^E_{FE}$	0.166	0.042	0.866
$\widehat{\Delta}$	0.201	0.044	0.950
F. Design 3: Lotteries twice smaller, otherwise same as Design 2			
$\widehat{\beta}^I_{PS}$	0.209	0.149	0.948
$\widehat{\beta}^E_{FE}$	0.137	0.060	0.778
$\widehat{\Delta}$	0.202	0.066	0.954

[a]The table simulates the IO, EO, and DREO estimators of Δ_K. Columns 2 to 4 display the mean, standard error, and coverage of the 95% confidence interval of the estimators in Column 1, computed from 500 replications. Designs in the top three and bottom three panels respectively have 120 and 240 waitlists. Design 1 has 30 takers, 10 non-takers, and 20 seats per waitlist. The treatment effect is constant and equal to 0.2, and the mean of $Y(0)$ is 0.4 standard deviation larger for non-takers than takers. Design 2 is the same, except that it has 24 takers per waitlist. Design 3 is the same, except that it has 20 applicants, 12 takers, and 10 seats per waitlist.

by Theorem 3.3 remain quite accurate when points (c) and (f) of Assumption 5 are violated. The approximation of $V(\widehat{\beta}^I_{PS})$ is slightly less accurate when point (d) is violated too, but that of $V(\widehat{\Delta})$ remains accurate.

S3.3. *Comparing the DREO Estimator to the Optimal 2SLS Estimator Using a Function of Applicants' Rank as an Instrument for Treatment*

In the paper, we discuss the fact that the IO estimator is a consistent estimator of the LATE of takers. Actually, any 2SLS estimator using a function of applicants' random numbers R_{ik} to instrument for D_{ik} is also consistent, provided that function has some

TABLE SIII

ROBUSTNESS OF STANDARD ERRORS IN THEOREM 3.3 TO
VIOLATIONS OF ASSUMPTION 5[a]

	Average	SE	Asympt. formula
	A. Violation of point (f)		
$\widehat{\beta}_{PS}^{I}$	0.195	0.096	0.102
$\widehat{\Delta}$	0.199	0.044	0.046
	B. Violations of points (c) and (f)		
$\widehat{\beta}_{PS}^{I}$	0.207	0.103	0.102
$\widehat{\Delta}$	0.201	0.044	0.046
	C. Violations of points (c), (d), and (f)		
$\widehat{\beta}_{PS}^{I}$	0.195	0.123	0.102
$\widehat{\Delta}$	0.200	0.047	0.046

[a]The table simulates the IO and DREO estimators of Δ_K. Columns 2 and 3 display the mean and standard error of the estimators in Column 1, computed from 500 replications. Column 4 displays the asymptotic approximations based on Theorem 3.3. All designs have 240 waitlists, 24 takers, 16 non-takers, and 20 seats per waitlist. The average treatment effect is equal to 0.2. The treatment effect is constant in Panels A and B, but drawn from a $\mathcal{N}(0.2, 1)$ distribution in Panel C. The mean of $Y(0)$ is the same for non-takers and takers in Panel A, but it is 0.4 standard deviation larger for non-takers than takers in Panels B and C.

predictive power for D_{ik}.[5] Newey (1990) showed that if the treatment effect is constant, the (infeasible) optimal estimator of the constant treatment effect within that class is that which uses $E(D_{ik}|R_{ik})$ to instrument for D_{ik}. Let $\widehat{\beta}_R^*$ denote that estimator. One can show that under Assumptions 1–3,

$$
E(D_{ik}|R_{ik}=r, \mathcal{P}_k) = 1\{1 \leq r \leq S_k\}\frac{T_k}{N_k}
$$

$$
+ 1\{S_k + 1 \leq r \leq S_k + N_k - T_k\} \sum_{j=\max(1, r-(N_k-T_k))}^{S_k} \frac{\binom{r-1}{j-1}\binom{N_k-r}{T_k-j}}{\binom{N_k}{T_k}}.[6]
$$

We use the formula above to compute $\widehat{\beta}_R^*$. Our simulation design is the same as that in Panel A of Table SII, where the treatment effect is constant. Results in Table SIV show that $\widehat{\Delta}$ has a lower variance than $\widehat{\beta}_R^*$, and the difference is highly significant (t-stat $= -5.54$). This shows that in DGPs with a constant treatment effect, $\widehat{\Delta}$ is not uniformly dominated by $\widehat{\beta}_R^*$, the optimal estimator among all 2SLS estimators using a function of R_{ik} to instrument for D_{ik}.

---

[5]$\widehat{\Delta}$ does not belong to that class of estimators, because Z_{ik} is a function of both R_{ik} and L_k.

[6]For $r \in \{S_k + 1, \ldots, S_k + N_k - T_k\}$, the formula corresponds to the sum of the probability of having a taker with $R_{ik} = r$ and $j - 1$ takers with $R_{ik} \leq r - 1$, for j going from $\max(1, r - (N_k - T_k))$ to S_k. Indeed, at least $\max(0, r - 1 - (N_k - T_k))$ takers must have $R_{ik} \leq r - 1$, and if more than $S_k - 1$ takers have $R_{ik} \leq r - 1$, the applicant with $R_{ik} = r$ does not get an offer.

TABLE SIV

COMPARING THE PERFORMANCES OF $\widehat{\Delta}$ AND $\widehat{\beta}_R^*$ [a]

	Average	SE	Coverage 95% CI
$\widehat{\Delta}$	0.203	0.040	0.964
$\widehat{\beta}_R^*$	0.204	0.043	0.964

[a]The table simulates the DREO estimator and $\widehat{\beta}_R^*$, the optimal estimator of the LATE among all 2SLS estimators using a function of applicants' ranks to instrument for the treatment. Columns 2 to 4 display the mean, standard error, and coverage of the 95% confidence interval of the estimators in Column 1. These statistics are computed from 500 replications. The simulation design is the same as in Panel A of Table SII.

S3.4. *Assessing the Number of Waitlists Needed for the Asymptotic Distribution in Theorem 3.1 in the Paper to Approximate the Distribution of $\widehat{\Delta}$ Well*

In this subsection, we assess the number of waitlists needed for the asymptotic distribution in Theorem 3.1 in the paper to approximate well the distribution of $\widehat{\Delta}$. We consider the same design as Design 1 in Section S3.1, with 40, 20, 10, and 5 waitlists. In each simulation, we draw 2,000 samples. Table SV below shows the coverage rate of the 95% CI attached to $\widehat{\Delta}$, and the 95% CI of this coverage rate. With 40 and 20 waitlists, the CI attached to $\widehat{\Delta}$ has the desired coverage rate. On the other hand, with 10 and 5 waitlists, its coverage is significantly lower than 0.95. This suggests that articles that use less than 20 waitlists in their analysis may not be able to use Theorem 3.1 in the paper for inference.

S3.5. *Assessing the Use of Robust Standard Errors in Applications With a Small Number of Waitlists*

In this subsection, we assess whether one can use robust standard errors for inference when the number of waitlists is too small to rely on Theorem 3.1 in the paper for inference. We first consider the same design as Design 1 in Section S3.1, except that the number of waitlists is equal to 10. Then, we consider a design with 20 applicants, 15 takers, and 10 seats per waitlist. Then, we consider another design with 10 applicants, 7 takers, and 5 seats per waitlist. Finally, we consider a design with 4 applicants, 3 takers, and 2 seats per waitlist. In each design, we draw 2,000 samples. To estimate the 95% CI attached to $\widehat{\Delta}$, we use the variance of the coefficient of D_{ik} in a 2SLS regression of Y_{ik} on D_{ik} using

TABLE SV

COVERAGE RATE OF THE 95% CI ATTACHED TO $\widehat{\Delta}$ [a]

	Coverage 95% CI
40 waitlists	0.944 ([0.933, 0.954])
20 waitlists	0.950 ([0.940, 0.959])
10 waitlists	0.929 ([0.918, 0.940])
5 waitlists	0.912 ([0.900, 0.925])

[a]The table shows the coverage rate of the confidence interval attached to $\widehat{\Delta}$, in the same design as Design 1 in Table SII, but with 40, 20, 10, and 5 waitlists. Each simulation has 2,000 replications. The 95% confidence interval of the coverage rate is reported in parentheses.

TABLE SVI

USING ROBUST STANDARD ERRORS IN APPLICATIONS
WITH 10 WAITLISTS[a]

	Coverage 95% CI
40 applicants, 30 takers, 20 seats	0.945 ([0.934, 0.955])
20 applicants, 15 takers, 10 seats	0.953 ([0.944, 0.962])
10 applicants, 7 takers, 5 seats	0.937 ([0.926, 0.948])
4 applicants, 3 takers, 2 seats	0.933 ([0.922, 0.944])

[a]The table shows the coverage rate of the confidence interval attached to $\widehat{\Delta}$ using robust standard errors, in the same design as Design 1 in Table SII, but with 40, 20, 10, and 5 applicants per waitlist. Each simulation has 2,000 replications. The 95% confidence interval of the coverage rate is reported in parentheses.

Z_{ik} as the instrument, where applicants are reweighted by w_{ik}^{DR}, using robust standard errors. Table SVI shows the coverage rate of the 95% CI attached to $\widehat{\Delta}$ using this robust variance estimator, and the 95% CI of this coverage rate. With 40 and 20 applicants per waitlist, the CI attached to $\widehat{\Delta}$ has the desired coverage rate. On the other hand, with 10 and 4 applicants, its coverage is significantly lower than 0.95. This suggests that articles that use less than 20 waitlists in their analysis may be able to use robust standard errors for inference, provided they have at least 20 applicants per waitlist.

S4. PROOFS OF SOME RESULTS IN THE PAPER NOT PROVEN THEREIN

Theorem 3.2 relies on Assumption 7 below. Let

$$RF_k^E = N_k \frac{L_k}{N_k}\left(1 - \frac{L_k}{N_k}\right)\left(\frac{1}{L_k}\sum_{i:Z_{ik}=1} Y_{ik} - \frac{1}{N_k - L_k}\sum_{i:Z_{ik}=0} Y_{ik}\right),$$

$$FS_k^E = N_k \frac{L_k}{N_k}\left(1 - \frac{L_k}{N_k}\right)\frac{1}{L_k}\sum_{i:Z_{ik}=1} D_{ik}.$$

ASSUMPTION 7—Technical Assumptions to Derive the Probability Limit of $\widehat{\beta}_{\text{FE}}^E$:

$$\frac{1}{K}\sum_{k=1}^K E\left(\frac{S_k\left(N_k - S_k\frac{N_k+1}{T_k+1}\right)}{N_k}\right),$$

$$\frac{1}{K}\sum_{k=1}^K E\left(\frac{S_k\left(N_k - T_k\frac{N_k+1}{T_k+1}\right)}{N_k}\left[\frac{1}{T_k}\sum_{i:D_{ik}(1)=1} Y_{ik}(0) - \frac{1}{N_k - T_k}\sum_{i:D_{ik}(1)=0} Y_{ik}(0)\right]\right),$$

and

$$\frac{1}{K} \sum_{k=1}^{K} E \left(\frac{S_k \left(N_k - S_k \dfrac{N_k + 1}{T_k + 1} \right)}{N_k} \frac{1}{T_k} \sum_{i:D_{ik}(1)=1} \left[Y_{ik}(1) - Y_{ik}(0) \right] \right)$$

have finite limits when $K \to +\infty$.

PROOF OF THEOREM 3.2: First,

$$\begin{aligned}
E(L_k | \mathcal{P}_k) &= \sum_{l=S_k}^{N_k - T_k + S_k} l \frac{\dbinom{l-1}{S_k - 1} \dbinom{N_k - l}{T_k - S_k}}{\dbinom{N_k}{T_k}} \\
&= S_k \frac{N_k + 1}{T_k + 1} \sum_{l=S_k}^{N_k - T_k + S_k} \frac{\dbinom{l}{S_k} \dbinom{N_k + 1 - (l+1)}{T_k + 1 - (S_k + 1)}}{\dbinom{N_k + 1}{T_k + 1}} \\
&= S_k \frac{N_k + 1}{T_k + 1}.
\end{aligned} \tag{33}$$

This derivation follows from arguments similar to those used when deriving (5).

Then, it follows from the fact that a 2SLS coefficient with one endogenous variable and one instrument is equal to the ratio of the reduced form and first-stage coefficients, from Equation (3.3.7) in Angrist and Pischke (2008), and from the definitions of RF_k^E and FS_k^E, that

$$\widehat{\beta}_{\text{FE}}^E = \frac{\dfrac{1}{K} \sum_{k=1}^{K} RF_k^E}{\dfrac{1}{K} \sum_{k=1}^{K} FS_k^E}. \tag{34}$$

For every k,

$$\begin{aligned}
& E\bigl(RF_k^E\bigr) \\
&= E\Biggl(\left(1 - \frac{L_k}{N_k} \right) \sum_{i=1}^{N_k} Y_{ik}\bigl(D_{ik}(1)\bigr) 1\{R_{ik} \leq L_k\} \\
&\qquad - \frac{L_k}{N_k} \sum_{i=1}^{N_k} Y_{ik}(0)\bigl(1 - 1\{R_{ik} \leq L_k\}\bigr) \Biggr) \\
&= E\Biggl(\left(1 - \frac{L_k}{N_k} \right) \sum_{i=1}^{N_k} Y_{ik}\bigl(D_{ik}(1)\bigr) E\bigl(1\{R_{ik} \leq L_k\} | L_k, \mathcal{P}_k\bigr)
\end{aligned}$$

$$- \frac{L_k}{N_k} \sum_{i=1}^{N_k} Y_{ik}(0)\big(1 - E\big(1\{R_{ik} \le L_k\}|L_k, \mathcal{P}_k\big)\big)\bigg)$$

$$= E\bigg(\bigg(1 - \frac{L_k}{N_k}\bigg) \sum_{i=1}^{N_k} Y_{ik}\big(D_{ik}(1)\big)\bigg(D_{ik}(1)\frac{S_k}{T_k} + \big(1 - D_{ik}(1)\big)\frac{L_k - S_k}{N_k - T_k}\bigg)$$

$$- \frac{L_k}{N_k} \sum_{i=1}^{N_k} Y_{ik}(0)\bigg(D_{ik}(1)\frac{T_k - S_k}{T_k} + \big(1 - D_{ik}(1)\big)\frac{N_k - T_k - L_k + S_k}{N_k - T_k}\bigg)\bigg)\bigg)$$

$$= E\bigg(\frac{(N_k - L_k)S_k}{N_k}\frac{1}{T_k} \sum_{i:D_{ik}(1)=1} Y_{ik}(1) - \frac{L_k(T_k - S_k)}{N_k}\frac{1}{T_k} \sum_{i:D_{ik}(1)=1} Y_{ik}(0)$$

$$+ \frac{(N_k - L_k)(L_k - S_k) - L_k(N_k - T_k - L_k + S_k)}{N_k}\frac{1}{N_k - T_k} \sum_{i:D_{ik}(1)=0} Y_{ik}(0)\bigg)$$

$$= E\bigg(\frac{(N_k - L_k)S_k}{N_k}\frac{1}{T_k} \sum_{i:D_{ik}(1)=1} \big[Y_{ik}(1) - Y_{ik}(0)\big]$$

$$+ \frac{N_k S_k - L_k T_k}{N_k}\bigg(\frac{1}{T_k} \sum_{i:D_{ik}(1)=1} Y_{ik}(0) - \frac{1}{N_k - T_k} \sum_{i:D_{ik}(1)=0} Y_{ik}(0)\bigg)\bigg)$$

$$= E\left(\frac{S_k\bigg(N_k - S_k\dfrac{N_k + 1}{T_k + 1}\bigg)}{N_k}\frac{1}{T_k} \sum_{i:D_{ik}(1)=1} \big[Y_{ik}(1) - Y_{ik}(0)\big]\right.$$

$$\left.+ \frac{S_k\bigg(N_k - T_k\dfrac{N_k + 1}{T_k + 1}\bigg)}{N_k}\bigg(\frac{1}{T_k} \sum_{i:D_{ik}(1)=1} Y_{ik}(0) - \frac{1}{N_k - T_k} \sum_{i:D_{ik}(1)=0} Y_{ik}(0)\bigg)\right). \quad (35)$$

The first equality follows from the definition of RF_k^E and some algebra. The second equality follows from the law of iterated expectations and the linearity of the conditional expectation. The third equality follows from (7) and (8). The fourth and fifth equalities follow from some algebra. The last equality follows from the law of iterated expectations, the linearity of the conditional expectation, and (33).

Similarly, one can show that for every k,

$$E\big(FS_k^E\big) = E\left(\frac{S_k\bigg(N_k - S_k\dfrac{N_k + 1}{T_k + 1}\bigg)}{N_k}\right). \quad (36)$$

Equations (35) and (36) combined with Assumption 7 imply that $\frac{1}{K}\sum_{k=1}^{K} E(RF_k^E)$ and $\frac{1}{K}\sum_{k=1}^{K} E(FS_k^E)$ converge towards finite limits when $K \to +\infty$. Then, one can use a rea-

soning similar to that used to prove (26) to show that

$$\widehat{\beta}_{\text{FE}}^E \xrightarrow{p} \frac{\lim\limits_{K \to +\infty} \frac{1}{K} \sum\limits_{k=1}^K E(RF_k^E)}{\lim\limits_{K \to +\infty} \frac{1}{K} \sum\limits_{k=1}^K E(FS_k^E)}. \tag{37}$$

The result follows from plugging (35) and (36) into (37).   *Q.E.D.*

PROOF OF THEOREM 3.3: *Proof of (a).*

The organization of the proof is as follows. We use the variance decomposition formula

$$KV(\widehat{\Delta}) = E\big(KV\big(\widehat{\Delta} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K}\big)\big) + KV\big(E\big(\widehat{\Delta} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K}\big)\big), \tag{38}$$

and then compute the limsup of the first term on the right-hand side, before showing that the second term is equal to 0.

Let $\widehat{FS} = \frac{1}{K} \sum_{k=1}^K \frac{N_k}{\overline{N}} \frac{S_k - 1}{L_k - 1}$:

$$KV\big(\widehat{\Delta} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K}\big)$$

$$= \frac{K}{\widehat{FS}^2} V\left(\frac{1}{K} \sum_{k=1}^K \frac{N_k}{\overline{N}} \left(\frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik}\big(Y_{ik}(0) + \tau D_{ik}(1)\big) \right. \right.$$

$$\left. \left. - \frac{1}{N_k - L_k} \sum_{i:Z_{ik}=0} Y_{ik}(0) \right) \,\middle|\, (\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K} \right)$$

$$= \frac{K}{\widehat{FS}^2} V\left(\frac{1}{K} \sum_{k=1}^K \frac{N_k}{\overline{N}} \left(\frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik} Y_{ik}(0) \right. \right.$$

$$\left. \left. - \frac{1}{N_k - L_k} \sum_{i:Z_{ik}=0} Y_{ik}(0) \right) \,\middle|\, (\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K} \right)$$

$$= \frac{1}{\widehat{FS}^2} \frac{1}{K} \sum_{k=1}^K \frac{N_k^2}{\overline{N}^2} \left(\frac{1}{(L_k - 1)^2} \sum_{i:Z_{ik}=1} w_{ik}^2 V\big(Y_{ik}(0) \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K}\big) \right.$$

$$\left. + \frac{1}{(N_k - L_k)^2} \sum_{i:Z_{ik}=0} V\big(Y_{ik}(0) \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K}\big) \right)$$

$$= \frac{\sigma_{Y(0)}^2}{\widehat{FS}^2} \frac{1}{K} \sum_{k=1}^K \frac{N_k^2}{\overline{N}^2} \left(\frac{1}{(L_k - 1)^2} \sum_{i:Z_{ik}=1} w_{ik}^2 + \frac{1}{N_k - L_k} \right)$$

$$\le \frac{\sigma_{Y(0)}^2}{\widehat{FS}^2} \frac{1}{K} \sum_{k=1}^K \frac{N_k^2}{\overline{N}^2} \left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k} \right). \tag{39}$$

The first equality follows from the fact that $\widehat{FS}$ is a function of $(\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K}$, and from point (d) of Assumption 5. The second equality follows from the fact that conditional on

$(\mathcal{D}_k, \mathcal{R}_k)_{1 \le k \le K}$, $\frac{1}{K} \sum_{k=1}^{K} \frac{N_k}{N} \frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik} D_{ik}(1)\tau$ is a constant. The third equality follows from point (a) of Assumption 5. The fourth equality follows from point (b) of Assumption 5. The inequality follows from the fact that $0 \le w_{ik} \le 1$ implies $0 \le w_{ik}^2 \le w_{ik}$, and $\sum_{i:Z_{ik}=1} w_{ik} = L_k - 1$.

Assumption 3 and point (b) of Assumption 4 guarantee that for all k,

$$\frac{N_k^2}{N^2}\left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k}\right) \le 2\left(\frac{N^+}{3}\right)^2. \tag{40}$$

Moreover, by point (a) of Assumption 4, these random variables are independent, so it follows from Gut (1992) that

$$\frac{1}{K}\sum_{k=1}^{K}\frac{N_k^2}{N^2}\left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k} - E\left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k}\right)\right) \xrightarrow{p} 0.$$

Using a reasoning similar to that used in the proof of Lemma A.1, one can show that

$$E\left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k}\right) = \frac{T_k}{N_k}\left(\frac{1}{S_k - 1} + \frac{1}{T_k - S_k}\right) = \frac{T_0}{N_0}\left(\frac{1}{S_0 - 1} + \frac{1}{T_0 - S_0}\right),$$

where the second equality follows from point (e) of Assumption 5. Combining the two preceding displays,

$$\frac{1}{K}\sum_{k=1}^{K}\frac{N_k^2}{N^2}\left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k}\right) \xrightarrow{p} \frac{T_0}{N_0}\left(\frac{1}{S_0 - 1} + \frac{1}{T_0 - S_0}\right). \tag{41}$$

Similarly, one can show that

$$\frac{1}{K}\sum_{k=1}^{K}\frac{N_k}{N}\frac{S_k - 1}{L_k - 1} \xrightarrow{p} \frac{T_0}{N_0}. \tag{42}$$

Then, it follows from (41), (42), and the continuous mapping theorem that

$$\frac{\sigma_{Y(0)}^2}{\widehat{FS}^2}\frac{1}{K}\sum_{k=1}^{K}\frac{N_k^2}{N^2}\left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k}\right) \xrightarrow{p} \sigma_{Y(0)}^2 \frac{\dfrac{1}{S_0 - 1} + \dfrac{1}{T_0 - S_0}}{\dfrac{T_0}{N_0}}. \tag{43}$$

Finally, Assumption 3 and point (b) of Assumption 4 imply that $\widehat{FS} \ge \frac{1}{N^+ - 2}$. Combined with (40), this implies that $\frac{\sigma_{Y(0)}^2}{\widehat{FS}^2}\frac{1}{K}\sum_{k=1}^{K}\frac{N_k^2}{N^2}(\frac{1}{L_k-1} + \frac{1}{N_k-L_k})$ is bounded. Then, it follows from Equation (43) and Theorem 2.20 in Van der Vaart (2000) that

$$\lim_{K \to +\infty} E\left(\frac{\sigma_{Y(0)}^2}{\widehat{FS}^2}\frac{1}{K}\sum_{k=1}^{K}\frac{N_k^2}{N^2}\left(\frac{1}{L_k - 1} + \frac{1}{N_k - L_k}\right)\right) = \sigma_{Y(0)}^2 \frac{\dfrac{1}{S_0 - 1} + \dfrac{1}{T_0 - S_0}}{\dfrac{T_0}{N_0}}. \tag{44}$$

Combined with Equation (39), Equation (44) implies that

$$\limsup_{K \to +\infty} E\big(KV\big(\widehat{\Delta} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \leq k \leq K}\big)\big) \leq \sigma_{Y(0)}^2 \frac{\frac{1}{S_0 - 1} + \frac{1}{T_0 - S_0}}{\frac{T_0}{N_0}}. \tag{45}$$

Then,

$$V\big(E\big(\widehat{\Delta} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \leq k \leq K}\big)\big)$$

$$= V\Bigg(\frac{1}{\widehat{FS}} E\Bigg(\frac{1}{K} \sum_{k=1}^{K} \frac{N_k}{\overline{N}} \Bigg(\frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik}\big(Y_{ik}(0) + \tau D_{ik}(1)\big)$$

$$- \frac{1}{N_k - L_k} \sum_{i:Z_{ik}=0} Y_{ik}(0) \Bigg) \,\Big|\, (\mathcal{D}_k, \mathcal{R}_k)_{1 \leq k \leq K} \Bigg) \Bigg)$$

$$= V\Bigg(\tau + \frac{1}{\widehat{FS}} \frac{1}{K} \sum_{k=1}^{K} \frac{N_k}{\overline{N}} \Bigg(\frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik} E\big(Y_{ik}(0) \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \leq k \leq K}\big)$$

$$- \frac{1}{N_k - L_k} \sum_{i:Z_{ik}=0} E\big(Y_{ik}(0) \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \leq k \leq K}\big) \Bigg) \Bigg)$$

$$= V\Bigg(\frac{1}{\widehat{FS}} \frac{1}{K} \sum_{k=1}^{K} \frac{N_k}{\overline{N}} \mu_k \Bigg(\frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik} - 1 \Bigg) \Bigg)$$

$$= 0. \tag{46}$$

The first equality follows from the fact that $\widehat{FS}$ is a function of $(\mathcal{D}_k, \mathcal{R}_k)_{1 \leq k \leq K}$, and from point (d) of Assumption 5. The second equality follows from $\frac{1}{K} \sum_{k=1}^{K} \frac{N_k}{\overline{N}} \frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik} \times D_{ik}(1) = \widehat{FS}$. The third equality follows from point (c) of Assumption 5. The last equality follows from $\sum_{i:Z_{ik}=1} w_{ik} = L_k - 1$.

Finally, the result follows from Equations (38), (45), and (46).

Proof of (b).

The proof is fairly similar to that of Point (a), so we just sketch it.

Let $\widehat{FS}_I = \frac{1}{K} \sum_{k=1}^{K} \big(\frac{\sum_{i:Z'_{ik}=1} D_{ik}(1)}{S_0} - \frac{S_0 - \sum_{i:Z'_{ik}=1} D_{ik}(1)}{N_0 - S_0} \big)$. One can show that under points (a), (b), (d), and (e) of Assumption 5,

$$KV\big(\widehat{\beta}_{\text{PS}}^I \mid (\mathcal{D}_k, \mathcal{R}_k)_{1 \leq k \leq K}\big) = \frac{\sigma_{Y(0)}^2 \Big(\frac{1}{S_0} + \frac{1}{N_0 - S_0} \Big)}{\widehat{FS}_I^2}. \tag{47}$$

Then, for every k,

$$E\Bigg(\sum_{i:Z'_{ik}=1} D_{ik}(1) \Bigg) = E\Bigg(\sum_{i=1}^{N_0} 1\{R_{ik} \leq S_0\} D_{ik}(1) \Bigg) = E\Bigg(\sum_{i=1}^{N_0} D_{ik}(1) \frac{S_0}{N_0} \Bigg) = \frac{S_0 T_0}{N_0}.$$

The second equality follows from the law of iterated expectations, Assumption 2, and from point (e) of Assumption 5. The third follows from point (e) of Assumption 5.

It follows from the previous display that for every k,

$$E\left(\frac{\sum_{i:Z'_{ik}=1} D_{ik}(1)}{S_0} - \frac{S_0 - \sum_{i:Z'_{ik}=1} D_{ik}(1)}{N_0 - S_0}\right) = \frac{T_0}{N_0} - \frac{S_0 - S_0\frac{T_0}{N_0}}{N_0 - S_0} = \frac{T_0 - S_0}{N_0 - S_0}. \qquad (48)$$

Combining Equations (47) and (48), one can then show that

$$KV\left(\widehat{\beta}^I_{\text{PS}} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1\le k\le K}\right) \xrightarrow{P} \frac{\sigma^2_{Y(0)}\left(\frac{1}{S_0} + \frac{1}{N_0 - S_0}\right)}{\left(\frac{T_0 - S_0}{N_0 - S_0}\right)^2}. \qquad (49)$$

Finally, $\widehat{FS}_I > \frac{S_0 - (N_0 - T_0)}{S_0} - \frac{N_0 - T_0}{N_0 - S_0}$, which is stricly positive under point (f) of Assumption 5. Together with Equation (47) and Assumption 3, this implies that $KV(\widehat{\beta}^I_{\text{PS}} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1\le k\le K})$ is bounded, so Equation (49) implies that

$$\lim_{K\to+\infty} E\left(KV\left(\widehat{\beta}^I_{\text{PS}} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1\le k\le K}\right)\right) = \frac{\sigma^2_{Y(0)}\left(\frac{1}{S_0} + \frac{1}{N_0 - S_0}\right)}{\left(\frac{T_0 - S_0}{N_0 - S_0}\right)^2}. \qquad (50)$$

Then, one can follow the steps used to prove Equation (46) to show that

$$V\left(E\left(\widehat{\beta}^I_{\text{PS}} \mid (\mathcal{D}_k, \mathcal{R}_k)_{1\le k\le K}\right)\right) = 0. \qquad (51)$$

Finally, the result follows from the variance decomposition formula and Equations (50) and (51).

Proof of the Sufficient Condition for $\limsup_{K\to+\infty} V(\sqrt{K}(\widehat{\Delta} - \Delta_K)) \le \lim_{K\to+\infty} V(\sqrt{K} \times (\widehat{\beta}^I_{\text{PS}} - \Delta_K))$.

One has $0 < \frac{T_0 - S_0}{N_0 - S_0} \le \frac{T_0}{N_0}$, so

$$\frac{\frac{1}{S_0 - 1} + \frac{1}{T_0 - S_0}}{\frac{T_0}{N_0}} \le \frac{\frac{1}{S_0 - 1} + \frac{1}{T_0 - S_0}}{\frac{T_0 - S_0}{N_0 - S_0}}.$$

Then,

$$\frac{\frac{1}{S_0 - 1} + \frac{1}{T_0 - S_0}}{\frac{T_0 - S_0}{N_0 - S_0}} \le \frac{\frac{1}{S_0} + \frac{1}{N_0 - S_0}}{\left(\frac{T_0 - S_0}{N_0 - S_0}\right)^2}$$

$$\Leftrightarrow \quad \frac{T_0 - S_0}{N_0 - S_0}\frac{1}{S_0 - 1} + \frac{1}{N_0 - S_0} \le \frac{1}{S_0} + \frac{1}{N_0 - S_0} \quad \Leftrightarrow \quad \frac{T_0 - S_0}{N_0 - S_0} \le 1 - \frac{1}{S_0}.$$

The result follows from the two preceding displays, and using again the fact that $\frac{T_0 - S_0}{N_0 - S_0} \leq \frac{T_0}{N_0}$.

Q.E.D.

S5. PROOFS OF THE RESULTS IN THE SUPPLEMENTAL MATERIAL

PROOF OF THEOREM S2.1: *Proof of Point 1.*
One has

$$\sqrt{K}\left(\widehat{\beta}_{\mathrm{FE}}^{E} - \beta_{\mathrm{FE},K}^{E}\right) = \sqrt{K}\left(\frac{\frac{1}{K}\sum_{k=1}^{K}RF_{k}^{E}}{\frac{1}{K}\sum_{k=1}^{K}FS_{k}^{E}} - \frac{\frac{1}{K}\sum_{k=1}^{K}E\left(RF_{k}^{E}\right)}{\frac{1}{K}\sum_{k=1}^{K}E\left(FS_{k}^{E}\right)}\right),$$

so the proof follows from the same arguments as that of Theorem 3.1.

Proof of Point 2.
Using the same reasoning as in the proof of Theorem 3.1, one can show that

$$\sqrt{K}\left(\widehat{\Delta} - \widehat{\beta}_{\mathrm{FE}}^{E} - \left(\Delta_{K} - \beta_{\mathrm{FE},K}^{E}\right)\right)$$

$$= \frac{FS}{E\left(\frac{1}{K}\sum_{k=1}^{K}FS_{k}\right)}\sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K}\left(\Lambda_{k} - E(\Lambda_{k})\right)\right)$$

$$- \frac{FS^{E}}{E\left(\frac{1}{K}\sum_{k=1}^{K}FS_{k}^{E}\right)}\sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K}\left(\Lambda_{k}^{E} - E\left(\Lambda_{k}^{E}\right)\right)\right) + o_{P}(1)$$

$$= \sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K}\left(\Lambda_{k} - \Lambda_{k}^{E} - E\left(\Lambda_{k} - \Lambda_{k}^{E}\right)\right)\right) + o_{P}(1) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{\mathrm{diff}}^{2}\right).$$

Then, to prove that $\widehat{\sigma}_{\mathrm{diff},+}^{2} \xrightarrow{P} \sigma_{\mathrm{diff},+}^{2} \geq \sigma_{\mathrm{diff}}^{2}$, one can use the same arguments as those used to show that $\widehat{\sigma}_{+}^{2} \xrightarrow{P} \sigma_{+}^{2} \geq \sigma^{2}$ in the proof of Theorem 3.1.   *Q.E.D.*

REFERENCES

ABDULKADIROGLU, A., J. ANGRIST, S. M. DYNARSKI, T. J. KANE, AND P. A. PATHAK (2011): "Accountability and Flexibility in Public Schools: Evidence From Boston's Charters and Pilots," *Quarterly Journal of Economics*, 126 (2), 699–748. [2]

ABDULKADIROGLU, A., J. D. ANGRIST, P. D. HULL, AND P. A. PATHAK (2016): "Charters Without Waitlists: Testing Takeovers in New Orleans and Boston," *American Economic Review*, 106 (7), 1878–1920. [2]

ACEVEDO, P., G. CRUCES, P. GERTLER, AND S. MARTINEZ (2017): "Living up to Expectations: How Job Training Made Women Better off and Men Worse off," Technical Report, National Bureau of Economic Research. [2]

ANGRIST, J. D., AND G. W. IMBENS (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90 (430), 431–442. [8]

ANGRIST, J. D., AND J.-S. PISCHKE (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ Princeton University Press. [14]

ANGRIST, J. D., S. R. COHODES, S. M. DYNARSKI, P. A. PATHAK, AND C. R. WALTERS (2016): "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice," *Journal of Labor Economics*, 34 (2), 275–318. [2]

ANGRIST, J. D., S. M. DYNARSKI, T. J. KANE, P. A. PATHAK, AND C. R. WALTERS (2010): "Inputs and Impacts in Charter Schools: Kipp Lynn," *American Economic Review*, 100 (2), 239–243. [2]

——— (2012): "Who Benefits From Kipp?" *Journal of Policy Analysis and Management*, 31 (4), 837–860. [2]

ANGRIST, J. D., P. D. HULL, P. A. PATHAK, AND C. R. WALTERS (2017): "Leveraging Waitlists for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 132 (2), 871–919. [2]

ANGRIST, J. D., P. A. PATHAK, AND C. R. WALTERS (2013): "Explaining Charter School Effectiveness," *American Economic Journal: Applied Economics*, 5 (4), 1–27. [1,2]

ATTANASIO, O., A. KUGLER, AND C. MEGHIR (2011): "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence From a Randomized Trial," *American Economic Journal: Applied Economics*, 3 (3), 188–220. [1,2]

BEHAGHEL, L., C. DE CHAISEMARTIN, AND M. GURGAND (2017): "Ready for Boarding? The Effects of a Boarding School for Disadvantaged Students," *American Economic Journal: Applied Economics*, 9 (1), 140–164. [2]

BLATTMAN, C., AND J. ANNAN (2016): "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment With High-Risk Men in a Fragile State," *American Political Science Review*, 110 (1), 1–17. [2]

CARD, D., P. IBARRARÁN, F. REGALIA, D. ROSAS-SHADY, AND Y. SOARES (2011): "The Labor Market Impacts of Youth Training in the Dominican Republic," *Journal of Labor Economics*, 29 (2), 267–300. [1,2]

CLARK, M. A., P. GLEASON, C. C. TUTTLE, AND M. K. SILVERBERG (2011): "Do Charter Schools Improve Student Achievement? Evidence From a National Randomized Study," Working Paper, Mathematica Policy Research, Inc. [2]

COHODES, S. R. (2016): "Teaching to the Student: Charter School Effectiveness in Spite of Perverse Incentives," *Education Finance and Policy*, 11, 1–42. [2]

CULLEN, J. B., B. A. JACOB, AND S. LEVITT (2006): "The Effect of School Choice on Participants: Evidence From Randomized Waitlists," *Econometrica*, 74 (5), 1191–1230. [2]

CURTO, V. E., AND R. G. FRYER (2014): "The Potential of Urban Boarding Schools for the Poor: Evidence From Seed," *Journal of Labor Economics*, 32 (1), 65–93. [1,2]

DAVIS, M., AND B. HELLER (2015): "'No Excuses' Charter Schools and College Enrollment: New Evidence From a High School Network in Chicago," Available at SSRN 2606059. [2]

DEMING, D. J. (2011): "Better Schools, Less Crime?" *Quarterly Journal of Economics*, 126 (4), 2063–2115. [2]

——— (2014): "Using School Choice Waitlists to Test Measures of School Effectiveness," *American Economic Review*, 104 (5), 406–411. [2]

DEMING, D. J., J. S. HASTINGS, T. J. KANE, AND D. O. STAIGER (2014): "School Choice, School Quality, and Postsecondary Attainment," *American Economic Review*, 104 (3), 991–1013. [2]

DOBBIE, W., AND R. G. FRYER JR. (2013): "Getting Beneath the Veil of Effective Schools: Evidence From New York City," *American Economic Journal: Applied Economics*, 5 (4), 28–60. [2]

——— (2015): "The Medium-Term Impacts of High-Achieving Charter Schools," *Journal of Political Economy*, 123 (5), 985–1037. [1,2]

DOBBIE, W., AND R. G. FRYER (2011): "Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence From the Harlem Children's Zone," *American Economic Journal: Applied Economics*, 3 (3), 158–187. [1,2]

FURGESON, J., B. GILL, J. HAIMSON, A. KILLEWALD, M. MCCULLOUGH, I. NICHOLS-BARRER, N. VERBITSKY-SAVITZ, B. TEH, M. BOWEN, A. DEMERITT et al. (2012): "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts," Mathematica Policy Research, Inc. and The Center on Reinventing Public Education. [3]

GILL, B., J. FURGESON, H. S. CHIANG, B.-R. TEH, J. HAIMSON, AND N. VERBITSKY-SAVITZ (2013): "Replicating Experimental Impact Estimates With Nonexperimental Methods in the Context of Control Crossover," Technical Report, Mathematica Policy Research, Inc. [3]

GNAGEY, J., AND S. LAVERTU (2016): "The Impact of Inclusive STEM High Schools on Student Achievement," *AERA Open*, 2 (2). [3]

GUT, A. (1992): "The Weak Law of Large Numbers for Arrays," *Statistics & Probability Letters*, 14 (1), 49–52. [17]

HASTINGS, J., T. J. KANE, AND D. STAIGER (2009): "Heterogeneous Preferences and the Efficacy of Public School Choice," NBER Working Paper 2145. [3]

HASTINGS, J. S., T. J. KANE, AND D. O. STAIGER (2006): "Gender and Performance: Evidence From School Assignment by Randomized Waitlist," *American Economic Review*, 96 (2), 232–236. [3]

HASTINGS, J. S., C. A. NEILSON, AND S. D. ZIMMERMAN (2012): "The Effect of School Choice on Intrinsic Motivation and Academic Outcomes," Technical Report, National Bureau of Economic Research. [3]

HIRANO, K., G. W. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71 (4), 1161–1189. [4]

HIRSHLEIFER, S., D. MCKENZIE, R. ALMEIDA, AND C. RIDAO-CANO (2016): "The Impact of Vocational Training for the Unemployed: Experimental Evidence From Turkey," *The Economic Journal*, 126, 2115–2146. [3]

HOXBY, C. M., AND S. MURARKA (2009): "Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement," Technical Report, National Bureau of Economic Research. [3]

HOXBY, C. M., AND J. E. ROCKOFF (2004): "The Impact of Charter Schools on Student Achievement." [3]

IBARRARÁN, P., J. KLUVE, L. RIPANI, AND D. ROSAS (2015): "Experimental Evidence on the Long-Term Impacts of a Youth Training Program." [3]

IBARRARAN, P., L. RIPANI, B. TABOADA, J. M. VILLA, AND B. GARCIA (2014): "Life Skills, Employability and Training for Disadvantaged Youth: Evidence From a Randomized Evaluation Design," *IZA Journal of Labor & Development*, 3 (1), 10. [3]

KRAFT, M. A. (2014): "How to Make Additional Time Matter: Integrating Individualized Tutorials Into an Extended Day," *Education Finance and Policy*, 10, 81–116. [3]

MCCLURE, L., B. STRICK, R. JACOB-ALMEIDA, AND C. REICHER (2005): "The Preuss School at UCSD: School Characteristics and Students' Achievement," The Center for Research on Educational Equity, Assessment and Teaching Effectiveness, UCSD. [3]

NEWEY, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837. [11]

SETREN, E. (2015): "Special Education and English Language Learner Students in Boston Charter Schools: Impact and Classification." [3]

STRICK, B. R. (2012): "Evidence for the Influence of School Context on College Access," *Education Policy Analysis Archives*, 20, 35. [3]

TUTTLE, C. C., B. GILL, P. GLEASON, V. KNECHTEL, I. NICHOLS-BARRER, AND A. RESCH (2013): "Kipp Middle Schools: Impacts on Achievement and Other Outcomes," Final Report, Mathematica Policy Research, Inc. [3]

TUTTLE, C. C., P. GLEASON, AND M. CLARK (2012): "Using Waitlists to Evaluate Schools of Choice: Evidence From a National Study of Charter Schools," *Economics of Education Review*, 31 (2), 237–253. [3]

TUTTLE, C. C., P. GLEASON, V. KNECHTEL, I. NICHOLS-BARRER, K. BOOKER, G. CHOJNACKI, T. COEN, AND L. GOBLE (2015): "Understanding the Effect of Kipp as It Scales: Volume I, Impacts on Achievement and Other Outcomes," Final Report of Kipp's Investing in Innovation Grant Evaluation, Mathematica Policy Research, Inc. [3]

VAN DER VAART, A. W. (2000): *Asymptotic Statistics*, Vol. 3. Cambridge: Cambridge University Press. [17]

WALTERS, C. (2014): "The Demand for Effective Charter Schools," NBER Working Paper w20640. [3]

WEST, M. R., M. A. KRAFT, A. S. FINN, R. E. MARTIN, A. L. DUCKWORTH, C. F. GABRIELI, AND J. D. GABRIELI (2016): "Promise and Paradox Measuring Students' Non-Cognitive Skills and the Impact of Schooling," *Educational Evaluation and Policy Analysis*, 38 (1), 148–170. [3]