ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

https://www.econometricsociety.org/

Econometrica, Vol. 88, No. 4 (July, 2020), 1453–1477

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

Econometrica, Vol. 88, No. 4 (July, 2020), 1453–1477

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

Oversubscribed treatments are often allocated using randomized waiting lists. Applicants are ranked randomly, and treatment offers are made following that ranking until all seats are filled. To estimate causal effects, researchers often compare applicants getting and not getting an offer. We show that those two groups are not statistically comparable. Therefore, the estimator arising from that comparison is inconsistent when the number of waitlists goes to infinity. We propose a new estimator, and show that it is consistent, provided the waitlists have at least two seats. Finally, we revisit an application, and we show that using our estimator can lead to a statistically significant difference with respect to the results obtained using the commonly used estimator.

KEYWORDS: Waiting lists, non-takers, non compliance, instrumental variable, local average treatment effect, randomized controlled trials.

1. INTRODUCTION

OFTENTIMES, SOME INDIVIDUALS WHO APPLY for a treatment are non-takers. They decline to get treated when they receive an offer, for instance because they then realize that their benefit from treatment is lower than they thought. When a treatment is oversubscribed but some applicants are non-takers, an appealing way of allocating the available seats is to use randomized waitlists. First, applicants are ranked randomly. Then, if S seats are available, an initial round of offers takes place, whereby the first S applicants get an offer. If r of them decline it, a subsequent round of offers takes place whereby the next r applicants get an offer. Offers stop when all the seats have been filled. This allocation method is fair: each taker has the same probability of being treated; it is also efficient: no seat for treatment remains unused, despite the presence of non-takers. Therefore, oversubscribed treatments with non-takers are often allocated by randomized waitlists. We conducted a survey, and found 43 articles studying treatments allocated by randomized waitlists, ranging from charter schools in the United States to agricultural trainings in Liberia. These treatments often have capacity constraints for various groups of applicants. For instance, a charter school may have 20 seats available in 7th grade and 25 seats in 8th grade. Then, a lottery takes place in each group.

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

This paper originated from a comment an anonymous referee made to us when he/she reviewed Behaghel, de Chaisemartin, and Gurgand (2017). We would like to thank him/her. We would also like to thank Hugo Botton for outstanding research assistance. We are very grateful to Chris Blattman and Jeannie Annan for making their data publicly available, and for answering all our questions. We are also very grateful to Josh Angrist, Bart Cockx, Xavier D'Haultfœuille, Thomas Le Barbanchon, Chang Lee, Heather Royer, Doug Steigerwald, Chris Walters, three anonymous referees, members of the UCSB econometrics research group, and seminar participants at: the 2017 California Econometrics conference, the 2017 Labor and Education workshop of the NBER Summer Institute, Louvain-la-Neuve, the Paris School of Economics, UC Santa Barbara, UC San Diego, and Warwick for their helpful comments. Chaisemartin thanks Jean-Yves Chevalier and Christian Lixi for teaching him combinatorics.

 https://doi.org/10.3982/ECTA16032

As applicants are ranked randomly, it may be possible to form two comparable groups with different likelihoods of getting an offer. One could then compare those two groups to estimate the effect of the treatment. In practice, researchers have used two types of comparisons. Some researchers have compared applicants getting and not getting an initial offer, thus giving rise to the so-called initial-offer (IO) estimators. Other researchers have compared applicants ever and never getting an offer, thus giving rise to the so-called ever-offer (EO) estimators. When several lotteries were conducted, as in the charter school example above, researchers have often included waitlist fixed effects in their specifications, to ensure they compare applicants within and not across waitlists. In our survey, 22 articles used the EO estimator, 20 used the IO estimator, and a handful used other estimators. Overall, practices are not standardized.

We start by showing that the expected proportion of takers is strictly greater among applicants ever getting an offer than among applicants never getting one. Intuitively, this is because offers continue until all seats are filled, so the last applicant getting an offer must by construction be a taker. Moreover, when waitlist fixed effects are included in the estimation, they induce an endogenous reweighting of waitlists that usually further increases this imbalance between the two groups, as we explain in more detail in Section 2.

Then, we show that due to this imbalance, the EO estimator is inconsistent when the number of waitlists goes to infinity. In our survey, the median of the number of waitlists divided by the number of applicants per waitlist is 1.9, and 25% of the articles have more than 100 waitlists and less than 40 applicants per waitlist. This motivates the asymptotic sequence we consider. By contrast, if the number of applicants and takers per waitlist goes to infinity, the asymptotic bias of the EO estimator goes to 0.

We show that dropping the last applicant getting an offer in each waitlist is sufficient to restore the comparability between applicants getting and not getting an offer. Based on this result, we propose a new estimator of the treatment effect. It is built out of comparisons of applicants that get and do not get an offer in each waitlist, downweighting applicants that accept their offer by an amount equivalent to dropping one of them. Then, our estimator takes a weighted average of those within-waitlist comparisons, with a weighting scheme that avoids the endogenous reweighting induced by the waitlist fixed effects. We refer to that estimator as the doubly-reweighted ever-offer (DREO) estimator. We show that our estimator is consistent and asymptotically normal when the number of waitlists goes to infinity.

Contrary to subsequent-round offers, initial offers are only a function of applicants' random ranks in the waitlist. Therefore, applicants getting and not getting an initial offer are statistically comparable, and the IO estimator is also consistent. However, we show that the asymptotic variance of that estimator is often much larger than that of the DREO estimator, so using it will often result in large efficiency losses.

We use our results to revisit Blattman and Annan (2016), who studied the effects of an agricultural training. The difference between the DREO estimator and the EO estimator computed by the authors is statistically significant for some of the outcomes they considered.[1]

The remainder of the paper is organized as follows. Section 2 uses a simple example to give the intuition of our results. Section 3 presents our main results. Section 4 presents our empirical application. Appendix 5 presents the proofs. In the Supplemental Material (de Chaisemartin and Behaghel (2020)), we present our survey of articles that have used randomized waitlists, and some additional results and proofs as well as some simulations.

[1]A Stata adofile computing the DREO estimator is available from the authors' website.

2. INTRODUCING THE RESULTS THROUGH A SIMPLE EXAMPLE

We start with a simple example. We consider a waitlist where five applicants compete for three seats. Four applicants are takers (T) and one is a non-taker (NT), meaning that she will refuse to get treated if she gets an offer. Applicants are randomly ranked, and treatment offers are made following that ranking until all seats are filled. There are 5! possible orderings of the applicants, that can be divided into five groups of 4! orderings, according to the rank of the non-taker. Table I displays those five groups of orderings, hereafter referred to, slightly abusively, as "orderings" rather than groups of orderings. For each ordering, applicants getting an offer are depicted in italics, while those not getting an offer are depicted in bold. In orderings 1 and 2, the first three applicants are takers, so offers stop after the third offer. In orderings 3, 4, and 5, one of the first three applicants is a non-taker, so a fourth offer is made; then the next applicant is a taker so offers stop, as the available seats have been filled.

The first issue with the EO estimator is that, on average, applicants getting an offer bear a higher proportion of takers than applicants not getting an offer. Each ordering has a 0.20 probability of being selected. Across the five orderings, the expected share of takers among applicants getting an offer is $0.2 \times (1 + 1 + 3/4 + 3/4 + 3/4) = 17/20$. On the other hand, the expected share of takers among applicants not getting an offer is $0.2 \times (1/2 + 1/2 + 1 + 1 + 1) = 4/5$. Intuitively, this imbalance arises because offers stop when sufficiently many takers have accepted an offer. This endogenous stopping rule creates a positive correlation between getting an offer and being a taker. When the average potential outcomes of takers and non-takers differ,[2] this imbalance implies that applicants getting and not getting an offer are not statistically comparable: those two groups have different average potential outcomes. The second issue with the EO estimator arises from the inclusion of fixed effects when pooling waitlists. Assume that one pools waitlists that all have four takers, one non-taker, and three seats. In some waitlists, the realized ordering of takers and non-takers is Ordering 1 in Table I; in other waitlists, the realized ordering is Ordering 2, etc. With several waitlists, it follows from, for example, Equation (3.3.7) in Angrist and Pischke (2008), that the EO estimator with waitlist fixed effects is a weighted average of the EO estimators in each waitlist, that gives more weight to waitlists where the share of applicants getting an offer is closer to 1/2. In our example, 3/5 of applicants get an offer in waitlists with Ordering 1 or 2, while 4/5 of applicants get an offer in waitlists with Ordering 3, 4, or 5. Accordingly, waitlists with Ordering 1 or 2 receive more weight. But those are precisely the waitlists where the proportion of takers among

TABLE I

APPLICANTS *GETTING* AND **NOT GETTING** AN OFFER IN AN EXAMPLE

Ordering 1	Ordering 2	Ordering 3	Ordering 4	Ordering 5
T	*T*	*T*	*T*	*NT*
T	*T*	*T*	*NT*	*T*
T	*T*	*NT*	*T*	*T*
T	**NT**	*T*	*T*	*T*
NT	**T**	**T**	**T**	**T**

[2]This is often the case. Abadie, Angrist, and Imbens (2002) and Crépon, Devoto, Duflo, and Parienté (2015) are just a few examples of the many papers that have found large differences between the average potential outcomes of takers and non-takers.

TABLE II

APPLICANTS *GETTING* AND **NOT GETTING** AN OFFER, DROPPING THE LAST TAKER GETTING AN OFFER

Ordering 1	Ordering 2	Ordering 3	Ordering 4	Ordering 5
T	*T*	*T*	*T*	*NT*
T	*T*	*T*	*NT*	*T*
		NT	*T*	*T*
T	**NT**			
NT	**T**	**T**	**T**	**T**

applicants getting an offer is the highest. Therefore, the reweighting of waitlists induced by the fixed effects aggravates the over-representation of takers among applicants getting an offer.[3]

The DREO estimator we propose addresses those two issues. First, in our example, dropping the last taker getting an offer is sufficient to solve the endogenous stopping rule issue. Table II shows that then, the expected share of takers among applicants getting an offer is equal to $0.2 \times (1 + 1 + 2/3 + 2/3 + 2/3) = 4/5$, the same as among applicants not getting an offer. Still, dropping the last taker getting an offer is arbitrary: dropping the first or the second would have the same effect. Besides, doing so reduces the sample size and statistical precision. Instead, one can give to the three of them a weight equal to 2/3: this reduces the expected share of takers among applicants getting an offer by the same amount as dropping one. Second, instead of using fixed effects to pool waitlists, we simply take an average of the estimators in each waitlist, weighting waitlists proportionally to their number of applicants. These weights are independent of how many offers one has to make to fill the available seats, which solves the second issue of the EO estimator. Table II shows that this second reweighting is necessary. Even after downweighting takers getting an offer, including waitlist fixed effects would still lead to over-represent takers among applicants getting an offer. Indeed, doing so gives more weight to waitlists with Ordering 1 or 2, where 1/2 of applicants get an offer, while those are the waitlists where the proportion of takers among applicants getting an offer is the highest.

3. MAIN RESULTS

3.1. *Assumptions and Parameter of Interest*

Throughout the paper, we consider the following setup.

ASSUMPTION 1—Setup: (a) *Applicants for a binary treatment are divided into K mutually exclusive waitlists. For every $k \in \{1, \dots, K\}$, N_k denotes the number of applicants in waitlist k. N_k is non-stochastic.*

(b) *In each waitlist, S_k seats are available, and are allocated as follows: applicants are ranked, and treatment offers are made following that order until S_k applicants have accepted to get treated or all applicants have received an offer. S_k is non-stochastic.*

(c) *Applicants that do not get an offer cannot get treated.*

[3]Note that the fixed effects do not always aggravate the over-representation of takers. In applications with few seats and many takers per waitlist, the waitlists with a share of offers closer to 1/2 are those where the most offers had to be made, which are also those with the lowest proportion of takers among applicants getting an offer. Then, the fixed effects would counteract the over-representation of takers due to the endogenous stopping rule, but it is very unlikely that they would exactly cancel it.

In a previous version of this paper (see de Chaisemartin and Behaghel (2018)), we considered various extensions of this setup. For instance, we showed that our results remain unchanged if we allow for the possibility that some applicants manage to get treated even if they do not receive an offer. Similarly, we allowed for the possibility that some applicants may participate in several waiting lists, or that the treatment may not be binary. Those extensions are not of essence, so we focus on the basic setup outlined in Assumption 1.

Then, we assume that ranks are randomly assigned to applicants. Let R_{ik} denote the rank assigned to applicant i in waitlist k, let $\mathcal{R}_k = (R_{1k}, \dots, R_{N_k k})$ denote the ranks assigned to applicants 1 to N_k in waitlist k, let L_k denote the number of applicants getting an offer in waitlist k, and let $Z_{ik} = 1\{R_{ik} \leq L_k\}$ denote whether applicant i gets an offer, the so-called ever-offer instrument. Let $D_{ik}(1)$ denote her potential treatment if she gets an offer, and let D_{ik} denote her observed treatment. Under point (c) of Assumption 1, $D_{ik} = Z_{ik} D_{ik}(1)$. For every $d \in \{0, 1\}$, let $Y_{ik}(d)$ denote her potential outcome if $D_{ik} = d$,[4] and let $Y_{ik} = Y_{ik}(D_{ik})$ denote her observed outcome. Let

$$\mathcal{P}_k = \big(\big(D_{1k}(1), Y_{1k}(0), Y_{1k}(1)\big), \dots, \big(D_{N_k k}(1), Y_{N_k k}(0), Y_{N_k k}(1)\big)\big)$$

be a vector stacking the potential treatments and outcomes of the applicants in waitlist k. We consider those vectors as random: for instance, students' test scores may be influenced by stochastic shocks affecting their health on the day of the exam. Thereafter, expectations are taken with respect to the distribution of $(\mathcal{P}_k, \mathcal{R}_k)_{1 \leq k \leq K}$. For any integer j, let Π_j denote the set of permutations of $\{1, \dots, j\}$.

ASSUMPTION 2—Randomly Assigned Ranks: *For all* $k \in \{1, \dots, K\}$ *and* $(r_1, \dots, r_{N_k}) \in \Pi_{N_k}$, $P(\mathcal{R}_k = (r_1, \dots, r_{N_k}) | \mathcal{P}_k) = \frac{1}{N_k!}$.

Assumption 2 requires that the ranks assigned to applicants be independent of their potential treatments and outcomes, and uniformly distributed on Π_{N_k}. It implies that each applicant has the same probability of being in the first, second, ..., or last rank.

Finally, we consider a last assumption. Let applicants with $D_{ik}(1) = 1$ (resp. $D_{ik}(1) = 0$) be referred to as takers (resp. non-takers). For every $k \in \{1, \dots, K\}$, let $T_k = \sum_{i=1}^{N_k} D_{ik}(1)$ denote the number of takers in waitlist k.

ASSUMPTION 3—Strictly More Takers Than Seats: *For every* $k \in \{1, \dots, K\}$, $2 \leq S_k < T_k$.

Assumption 3 requires that each waitlist have at least two seats, so waitlists with less than two seats have to be dropped from the analysis. Assumption 3 also requires that each waitlist have strictly more takers than seats. If there are as many takers as seats, offers will stop at the last taker, so there will not be any taker in the group of applicants that do not receive an offer, and dropping one taker from the group of applicants that receive an offer will not be sufficient to restore the balance between the two groups. When all the seats available in a waitlist get filled, it must be that $S_k \leq T_k$, but it is still possible that $S_k = T_k$: offers may have stopped at the last taker, and all applicants not getting an offer might be

[4]We implicitly assume that getting an offer does not have a direct effect on the outcome, the so-called exclusion restriction; see Angrist, Imbens, and Rubin (1996).

non-takers. Still, in a previous version of this paper (see de Chaisemartin and Behaghel (2018)), we proposed a statistical test of whether $S_k < T_k$ for all k, or $S_k = T_k$ for some k.

Let $T = \sum_{k=1}^{K} T_k$ denote the total number of takers. Our parameter of interest is

$$\Delta_K = E\left(\frac{1}{T} \sum_{(i,k):D_{ik}(1)=1} \left[Y_{ik}(1) - Y_{ik}(0)\right]\right),$$

the local average treatment effect of the takers.

3.2. *The Doubly Reweighted Ever-Offer Estimator*

Let $N = \sum_{k=1}^{K} N_k$ and $\overline{N} = \frac{N}{K}$ respectively denote the total number of applicants and the average number of applicants per waitlist. Let $\mathcal{I} = \{(i,k) \in \mathbb{N}^2 : i \in \{1,\dots,N_k\}, k \in \{1,\dots,K\}\}$, and for every $(i,k) \in \mathcal{I}$, let $w_{ik} = 1 - \frac{Z_{ik}D_{ik}}{S_k}$. w_{ik} is equal to $1 - \frac{1}{S_k}$ for applicants that get and accept an offer, and to 1 for everyone else. As S_k takers receive an offer in each waitlist, weighting applicants getting an offer by w_{ik} decreases the share of takers among them by the same amount as dropping one taker, as illustrated in the numerical example in Section 2.

The DREO estimator of Δ_K is defined as

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

Importantly, note that under Assumption 1, $S_k = \sum_{i=1}^{N_k} Z_{ik}D_{ik}$, so observing $(Z_{ik}, D_{ik}, Y_{ik})_{(i,k)\in\mathcal{I}}$ is sufficient to compute $\widehat{\Delta}$. $\widehat{\Delta}$ can be computed through a 2SLS regression. Let $L = \sum_{k=1}^{K} L_k$, and let

$$w_{ik}^{\mathrm{DR}} = w_{ik}\left(Z_{ik} \times \frac{L-K}{N-K} \times \frac{N_k}{L_k - 1} + (1 - Z_{ik}) \times \frac{N-L}{N-K} \times \frac{N_k}{N_k - L_k}\right)$$

be a weighting scheme combining w_{ik} with propensity score reweighting. One can show that $\widehat{\Delta}$ is equal to the coefficient of D_{ik} in a 2SLS regression of Y_{ik} on D_{ik} using Z_{ik} as the instrument, and weighted by w_{ik}^{DR}.

Our main result relies on the following lemma:

LEMMA 3.1: *If Assumptions 1–3 hold, then for all* $k \in \{1,\dots,K\}$,

(a)

$$E\left(\frac{1}{K}\sum_{k=1}^{K} \frac{N_k}{\overline{N}} \left(\frac{1}{L_k - 1} \sum_{i:Z_{ik}=1} w_{ik} Y_{ik} - \frac{1}{N_k - L_k} \sum_{i:Z_{ik}=0} Y_{ik}\right)\right)$$
$$= E\left(\frac{1}{N} \sum_{(i,k)\in\mathcal{I}} \left[Y_{ik}\big(D_{ik}(1)\big) - Y_{ik}(0)\right]\right),$$

(b)

$$E\left(\frac{1}{K}\sum_{k=1}^{K}\frac{N_k}{N}\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}D_{ik}\right)=E\left(\frac{1}{N}\sum_{(i,k)\in\mathcal{I}}D_{ik}(1)\right).$$

The intuition of point (a) of the lemma goes as follows. As the numerical example in Section 2 illustrates, one can show that in each waitlist, w_{ik}-reweighted applicants getting an offer are statistically comparable to applicants not getting an offer. Therefore, the only difference between these two groups is that one receives an offer and not the other one. Accordingly, $\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}Y_{ik}-\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}Y_{ik}$, the difference between the average outcome of the two groups, is an unbiased estimator of $E(\frac{1}{N_k}\sum_{i=1}^{N_k}[Y_{ik}(D_{ik}(1))-Y_{ik}(0)])$, the intention to treat (ITT) effect of getting an offer on applicants' outcome in waitlist k. The numerator of $\widehat{\Delta}$ is an average of those unbiased within-waitlist comparisons, that gives to each waitlist a weight proportional to its number of applicants. Therefore, this numerator is an unbiased estimator of $E(\frac{1}{N}\sum_{(i,k)\in\mathcal{I}}[Y_{ik}(D_{ik}(1))-Y_{ik}(0)])$, the intention to treat effect among all applicants. The intuition of point (b) is similar.

Lemma 3.1 implies that the numerator of $\widehat{\Delta}$ is an unbiased estimator of the intention to treat effect, while its denominator is an unbiased estimator of the first-stage effect. As usual with instrumental variables, this does not imply that $\widehat{\Delta}$ is an unbiased estimator of Δ_K, but we now show that $\widehat{\Delta}-\Delta_K$ goes to 0. In our survey of articles that have used randomized waitlists, the median number of waitlists used in the analysis is equal to 64. Therefore, we consider a sequence where K, the number of waitlists, goes to infinity. An alternative would be to consider a sequence where the number of applicants per waitlist goes to infinity, but in our survey the median of waitlists divided by applicants per waitlist is equal to 1.9, so the former asymptotic may be more appropriate in a majority of applications. For all $k\in\{1,\dots,K\}$, let $RF_k=\frac{N_k}{N}[\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}Y_{ik}-\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}Y_{ik}]$ and $FS_k=\frac{N_k}{N}\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}D_{ik}$. Let also $FS=\lim_{K\to+\infty}\frac{1}{K}\sum_{k=1}^{K}E(FS_k)$ and $\Delta=\lim_{K\to+\infty}\frac{1}{K}\sum_{k=1}^{K}E(RF_k)/FS$, where Assumption 4 below ensures that those limits exist. Finally, for all k, let $\Lambda_k=\frac{RF_k-\Delta FS_k}{FS}$.

ASSUMPTION 4—Technical Assumptions to Derive the Asymptotic Distribution of $\widehat{\Delta}$: (a) *The vectors* $(\mathcal{P}_k,\mathcal{R}_k)_{1\le k\le K}$ *are mutually independent*.

(b) *For every* k, $N_k\le N^+$, *for some integer* N^+.

(c) *There is a constant* M *such that for all* $k\in\mathbb{N}^*$, *for all* $(i_1,i_2,i_3,i_4)\in\{1,\dots,N_k\}^4$, *and for all* $(d_1,d_2,d_3,d_4)\in\{0,1\}^4$, $E(|Y_{i_1k}(d_1)||Y_{i_2k}(d_2)||Y_{i_3k}(d_3)||Y_{i_4k}(d_4)|)<M$.

(d) *The following sequences have finite limits when* $K\to+\infty$: (i) $\frac{1}{K}\sum_{k=1}^{K}E(RF_k)$, (ii) $\frac{1}{K}\sum_{k=1}^{K}E(FS_k)$, (iii) $\frac{1}{K}\sum_{k=1}^{K}V(RF_k)$, (iv) $\frac{1}{K}\sum_{k=1}^{K}V(FS_k)$, (v) $\frac{1}{K}\sum_{k=1}^{K}E(RF_kFS_k)$, (vi) $\frac{1}{K}\sum_{k=1}^{K}E(|RF_k-E(RF_k)|^4)$, (vii) $\frac{1}{K}\sum_{k=1}^{K}E(|FS_k-E(FS_k)|^4)$, (viii) $\frac{1}{K}\sum_{k=1}^{K}E(|\Lambda_k-E(\Lambda_k)|^4)$, (ix) $\frac{1}{K}\sum_{k=1}^{K}E(T_k)$, *(x)* $\frac{1}{K}\sum_{k=1}^{K}E(\sum_{i:D_{ik}=1}Y_{ik}(1)-Y_{ik}(0))$, (xi) $\frac{1}{K}\sum_{k=1}^{K}V(T_k)$, *and* (xii) $\frac{1}{K}\sum_{k=1}^{K}V(\sum_{i:D_{ik}=1}Y_{ik}(1)-Y_{ik}(0))$.

Typically, the lotteries determining applicants' ranks are independent across waitlists, so by design the vectors $(\mathcal{R}_k)_{1\le k\le K}$ are independent, and $(\mathcal{R}_k)_{1\le k\le K}$ is independent of $(\mathcal{P}_k)_{1\le k\le K}$. Then, point (a) of Assumption 4 only requires that the vectors $(\mathcal{P}_k)_{1\le k\le K}$ be independent. This is often plausible, for instance when the waitlists correspond to different schools. Point (b) requires that the number of applicants per waitlist be uniformly bounded by some constant N^+. Point (c) requires that the vectors $(Y_{ik}(0), Y_{ik}(1))_{1\leq i\leq N_k}$ have bounded fourth moments. One can show that this implies that $(RF_k)_{k\in\mathbb{N}^*}$ also has bounded fourth moments. Together with point (d), this then ensures that we can apply Liapunov's central limit theorem to $(RF_k)_{k\in\mathbb{N}^*}$, $(FS_k)_{k\in\mathbb{N}^*}$, and $(\Lambda_k)_{k\in\mathbb{N}^*}$, and the weak law of large numbers in Gut (1992) to $(RF_k)_{k\in\mathbb{N}^*}$, $(FS_k)_{k\in\mathbb{N}^*}$, $(RF_k^2)_{k\in\mathbb{N}^*}$, and $(FS_k^2)_{k\in\mathbb{N}^*}$.[5]

Let $\sigma^2 = \lim_{K\to+\infty} \frac{1}{K}\sum_{k=1}^K V(\Lambda_k)$, $\sigma_+^2 = \lim_{K\to+\infty}[\frac{1}{K}\sum_{k=1}^K E(\Lambda_k^2) - (\frac{1}{K}\sum_{k=1}^K E(\Lambda_k))^2]$, $\widehat{\Lambda}_k = \frac{RF_k - \widehat{\Delta} FS_k}{\frac{1}{K}\sum_{k=1}^K FS_k}$, and $\widehat{\sigma}_+^2 = \frac{1}{K}\sum_{k=1}^K(\widehat{\Lambda}_k - \frac{1}{K}\sum_{j=1}^K \widehat{\Lambda}_j)^2$.

THEOREM 3.1: *If Assumptions 1–4 hold*, $\sqrt{K}(\widehat{\Delta} - \Delta_K) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ *and* $\widehat{\sigma}_+^2 \xrightarrow{p} \sigma_+^2 \geq \sigma^2$.

Theorem 3.1 implies that $\widehat{\Delta}$ is an asymptotically normal estimator of Δ_K when the number of waitlists goes to infinity. As is usually the case for estimators constructed using independent but not identically distributed random variables (see, e.g., Liu and Singh (1995)), the asymptotic variance σ^2 of $\widehat{\Delta}$ can only be conservatively estimated: we provide a consistent estimator of σ_+^2, an upper bound of σ^2. That estimator can then be used to build conservative confidence intervals for Δ_K.[6] When all the Λ_k have the same expectation, something that for instance happens when all waitlists have the same number of applicants, the same expectation of the proportion of takers, and the same expectations of takers' and non-takers' potential outcomes, $\sigma_+^2 = \sigma^2$ so those confidence intervals are exact. Finally, in simulations shown in Section S3.4 of the Supplemental Material, we find that the asymptotic distribution in Theorem 3.1 approximates the distribution of $\widehat{\Delta}$ well if 20 waitlists or more are used in the analysis. This suggests that articles using more than 20 waitlists may rely on Theorem 3.1 for inference, while articles using less than 20 waitlists may not.

3.3. *Comparison With the Ever-Offer and Initial-Offer Estimators*

3.3.1. *Comparison With the Ever-Offer Estimator*

Let $\widehat{\beta}_{\text{FE}}^E$ be 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 for D_{ik}. We refer to $\widehat{\beta}_{\text{FE}}^E$ as the EO estimator. The derivation of its limit relies on Assumption 7, another technical assumption, that is stated in the proof of the theorem, in Section S4 of the Supplemental Material. Assumption 7

[5]To apply those theorems, one only needs that $(RF_k)_{k\in\mathbb{N}^*}$ has bounded moments of order $2+\delta$ for some strictly positive δ. However, applying the central limit theorem to $(RF_k)_{k\in\mathbb{N}^*}$ and $(FS_k)_{k\in\mathbb{N}^*}$ would only yield that $\sqrt{K}(\widehat{\Delta} - \frac{\frac{1}{K}\sum_{k=1}^K E(RF_k)}{\frac{1}{K}\sum_{k=1}^K E(FS_k)})$ is asymptotically normal. This is not the result stated in Theorem 3.1, as $\frac{\frac{1}{K}\sum_{k=1}^K E(RF_k)}{\frac{1}{K}\sum_{k=1}^K E(FS_k)} \neq \Delta_K$. Assuming fourth moments, one can show that $\lim_{K\to\infty}\sqrt{K}(\frac{\frac{1}{K}\sum_{k=1}^K E(RF_k)}{\frac{1}{K}\sum_{k=1}^K E(FS_k)} - \Delta_K) = 0$, thus ensuring that $\sqrt{K}(\widehat{\Delta} - \Delta_K)$ is asymptotically normal.

[6]Conservative variance estimators arise in articles studying treatment effect estimation in randomized experiments, when one does not assume that the experimental units are a random sample from an infinite population (see, e.g., Neyman (1990)).

ensures that the limits in the definition of w_k and B below exist. Let

$$w_k = \frac{\dfrac{S_k\left(N_k - S_k\dfrac{N_k+1}{T_k+1}\right)}{N_k}}{\lim\limits_{K\to+\infty}\dfrac{1}{K}\sum\limits_{j=1}^{K} E\left(\dfrac{S_j\left(N_j - S_j\dfrac{N_j+1}{T_j+1}\right)}{N_j}\right)},$$

$$B = \frac{\lim\limits_{K\to+\infty}\dfrac{1}{K}\sum\limits_{k=1}^{K} E\left(\dfrac{S_k\left(N_k - T_k\dfrac{N_k+1}{T_k+1}\right)}{N_k}\left[\dfrac{1}{T_k}\sum\limits_{i:D_{ik}(1)=1} Y_{ik}(0) - \dfrac{1}{N_k-T_k}\sum\limits_{i:D_{ik}(1)=0} Y_{ik}(0)\right]\right)}{\lim\limits_{K\to+\infty}\dfrac{1}{K}\sum\limits_{k=1}^{K} E\left(\dfrac{S_k\left(N_k - S_k\dfrac{N_k+1}{T_k+1}\right)}{N_k}\right)}.$$

THEOREM 3.2: *If Assumptions 1–4 and 7 hold,*

$$\widehat{\beta}^{E}_{\mathrm{FE}} \xrightarrow{p} \lim_{K\to+\infty}\frac{1}{K}\sum_{k=1}^{K} E\left(w_k\frac{1}{T_k}\sum_{i:D_{ik}(1)=1}\left[Y_{ik}(1) - Y_{ik}(0)\right]\right) + B. \tag{2}$$

Under Assumptions 1–4 and 7, $\widehat{\beta}^{E}_{\mathrm{FE}}$ converges towards the sum of two terms. The first is a weighted average of the LATEs of takers in each waitlist. If those LATEs vary across waitlists, this weighted average is not equal to the LATE of all takers, because it over-represents waitlists with a ratio of seats to takers closer to 1/2.[7] The second term, B, is a bias term. As explained in Section 2, this bias arises from the endogenous stopping of offers in each waitlist, and from the waitlist fixed effects.

We start by performing comparative statics on $|B|$, assuming that waitlists are homogeneous: there exist real numbers N_0, T_0, S_0, and $\Delta_{Y(0)}$ such that for all k, $N_k = N_0$, $T_k = T_0$, $S_k = S_0$, 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,

$$|B| = \frac{1-t_0}{1-s_0+N_0(t_0-s_0)}|\Delta_{Y(0)}|, \tag{3}$$

where $t_0 = T_0/N_0$ and $s_0 = S_0/N_0$ respectively denote the proportion of takers and the ratio of seats to applicants in the waitlist. The right-hand side of (3) is decreasing in N_0, decreasing in t_0, increasing in s_0, and increasing in $|\Delta_{Y(0)}|$.[8]

Then, we study how waitlists' heterogeneity affects $|B|$. Let $(S_0^a, S_0^b) \in \{2, \dots, T_0 - 1\}^2$, let $(T_0^a, T_0^b) \in \{3, \dots, N_0\}^2$, and let

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

---

[7]This can be seen from the fact that $\frac{S_k(N_k - S_k\frac{N_k+1}{T_k+1})}{N_k} = T_k\frac{S_k}{T_k}(1 - \frac{S_k}{(T_k+1)N_k/(N_k+1)}) \approx T_k\frac{S_k}{T_k}(1-\frac{S_k}{T_k})$.

[8]These results remain true if waitlists are heterogeneous; see Section S2.1 of the Supplemental Material.

The three following results hold:

1. If $(N_k, T_k, \Delta_{Y(0),k}) = (N_0, T_0, \Delta_{Y(0)})$ for all k, $|B|$ is larger if $\alpha\%$ of the waitlists have S_0^a seats and $(1-\alpha)\%$ have S_0^b seats than if all of them have $\alpha S_0^a + (1-\alpha)S_0^b$ seats.
2. If $(N_k, S_k, \Delta_{Y(0),k}) = (N_0, S_0, \Delta_{Y(0)})$ for all k, $|B|$ is larger if $\alpha\%$ of the waitlists have T_0^a takers and $(1-\alpha)\%$ have T_0^b takers than if all of them have $\alpha T_0^a + (1-\alpha)T_0^b$ takers.
3. If $(\frac{T_k}{N_k}, \frac{S_k}{N_k}, \Delta_{Y(0),k}) = (t_0, s_0, \Delta_{Y(0)})$ for all k, $|B|$ is larger if $\alpha\%$ of the waitlists have N_0^a applicants and $(1-\alpha)\%$ have N_0^b applicants than if all have $\alpha N_0^a + (1-\alpha)N_0^b$ applicants.

Overall, $|B|$ seems to be higher when waitlists have heterogeneous numbers of applicants, takers, and seats. The impact of waitlists' heterogeneity on $|B|$ can be large. For instance, if $(N_k, S_k, \Delta_{Y(0),k}) = (40, 20, \Delta_{Y(0)})$, $|B|$ is 17.1% larger if 50% of waitlists have 25 takers and 50% have 35 takers than if all have 30 takers.

3.3.2. *Comparison With the Initial-Offer Estimator*

Let $Z'_{ik} = 1\{R_{ik} \leq S_k\}$ be an indicator for applicants in the initial round of offers, the so-called initial-offer instrument. Let $S = \sum_{k=1}^K S_k$. Let $w_{ik}^I = Z'_{ik} \times \frac{S}{N} \times \frac{N_k}{S_k} + (1 - Z'_{ik}) \times \frac{N-S}{N} \times \frac{N_k}{N_k - S_k}$ be the propensity score weights attached to initial offers. Let $\widehat{\beta}_{\mathrm{PS}}^I$ be the coefficient of D_{ik} in a 2SLS regression of Y_{ik} on D_{ik}, using Z'_{ik} as the instrument, and weighted by w_{ik}^I. We call $\widehat{\beta}_{\mathrm{PS}}^I$ the IO estimator.[9]

Under Assumptions 1–2 and a technical condition similar to Assumption 4, $\sqrt{K}(\widehat{\beta}_{\mathrm{PS}}^I - \Delta_K)$ converges towards a normal distribution. Contrary to Z_{ik}, Z'_{ik} is only a function of applicants' random numbers and of the number of seats in their waitlist. Thus, it satisfies the exogeneity assumption in Imbens and Angrist (1994). Under Assumption 1, it also satisfies the monotonicity condition therein. Then, one can show that $\widehat{\beta}_{\mathrm{PS}}^I$ is an asymptotically normal estimator of the LATE of applicants complying with an initial offer. As those are a random subset of the takers, this LATE is equal to Δ_K.

However, using $\widehat{\beta}_{\mathrm{PS}}^I$ instead of $\widehat{\Delta}$ may result in a large loss of precision, as shown in Theorem 3.3. For every k, let $\mathcal{D}_k = (D_{1k}(1), \dots, D_{N_k k}(1))$.

ASSUMPTION 5—Assumptions to Compare the Asymptotic Variances of $\widehat{\Delta}$ and $\widehat{\beta}_{\mathrm{PS}}^I$: (a) *For every* $(i,k) \neq (i',k') \in \mathcal{I}^2$, $\mathrm{cov}(Y_{ik}(0), Y_{i'k'}(0) | (\mathcal{D}_k, \mathcal{R}_k)_{1\leq k\leq K}) = 0$.

(b) *For every* $(i,k) \in \mathcal{I}$, $V(Y_{ik}(0) \mid (\mathcal{D}_k, \mathcal{R}_k)_{1\leq k\leq K}) = \sigma^2_{Y(0)}$.

(c) *For every* $k \in \{1, \dots, K\}$, *for every* $i \in \{1, \dots, N_k\}$, $E(Y_{ik}(0) \mid (\mathcal{D}_k, \mathcal{R}_k)_{1\leq k\leq K}) = \mu_k$.

(d) *For every* $(i,k) \in \mathcal{I}$, $Y_{ik}(1) - Y_{ik}(0) = \tau$.

(e) *For every* $k \in \{1, \dots, K\}$, $N_k = N_0$, $S_k = S_0$, $T_k = T_0$, *for some integers* N_0, S_0, *and* T_0.

(f) $S_0(N_0 - S_0) - N_0(N_0 - T_0) > 0$.

Point (a) of Assumption 5 requires that the potential outcomes $Y_{ik}(0)$ be uncorrelated across applicants, point (b) requires that they be homoscedastic, and point (c) requires that in each waitlist, $E(Y_{ik}(0) \mid (\mathcal{D}_k, \mathcal{R}_k)_{1\leq k\leq K})$ be constant. Point (d) requires that the treatment effect be constant. Point (e) requires that all waitlists have the same number of applicants, takers, and seats. Point (f) ensures that the denominator of the IO estimator differs from 0 with probability 1. The conditions in Assumption 5 help simplify the

[9]In our survey of articles that have used randomized waitlists, three articles used the initial-offer instrument with propensity score weights. Other papers used that instrument with lottery fixed effects. We focus on the initial-offer estimator with propensity score weights, as it estimates the same parameter as the DREO estimator, so a comparison of their precision is meaningful. The initial-offer estimator with lottery fixed effects is consistent for a different causal effect.

formulas of $V(\widehat{\Delta})$ and $V(\widehat{\beta}^I_{\text{PS}})$, thus ensuring that these variances can be compared analytically. Some of these conditions may not be plausible in practice. For instance, point (c) fails if the expected value of $Y_{ik}(0)$ differs across takers and non-takers. However, in simulations shown in Section S3.2 of the Supplemental Material, we find that the variance formulas derived in Theorem 3.3 still provide good approximations of $V(\widehat{\Delta})$ and $V(\widehat{\beta}^I_{\text{PS}})$ when points (c), (d), and (f) of Assumption 5 are violated.

THEOREM 3.3: (a) *If Assumptions* 1–4 *and* (a)–(e) *of Assumption* 5 *hold*,

$$\limsup_{K\to+\infty} V\big(\sqrt{K}(\widehat{\Delta}-\Delta_K)\big) \leq \sigma^2_{Y(0)} \frac{\dfrac{1}{S_0-1}+\dfrac{1}{T_0-S_0}}{\dfrac{T_0}{N_0}}.$$

(b) *If Assumptions* 1–5 *hold*,

$$\lim_{K\to+\infty} V\big(\sqrt{K}(\widehat{\beta}^I_{\text{PS}}-\Delta_K)\big) = \sigma^2_{Y(0)} \frac{\dfrac{1}{S_0}+\dfrac{1}{N_0-S_0}}{\left(\dfrac{T_0-S_0}{N_0-S_0}\right)^2}.$$

Point (a) of Theorem 3.3 gives an upper bound of the limsup of $V(\sqrt{K}(\widehat{\Delta}-\Delta_K))$,[10] while point (b) gives the limit of $V(\sqrt{K}(\widehat{\beta}^I_{\text{PS}}-\Delta_K))$. In order to compare these parameters, note that $0<\frac{T_0}{N_0}\leq 1-\frac{1}{S_0}<1$ is a sufficient condition to have $\frac{\frac{1}{S_0-1}+\frac{1}{T_0-S_0}}{\frac{T_0}{N_0}} \leq \frac{\frac{1}{S_0}+\frac{1}{N_0-S_0}}{(\frac{T_0-S_0}{N_0-S_0})^2}$. This condition usually holds in practice. For instance, if $\frac{T_0}{N_0}=0.75$, the condition will hold as soon as there are more than 4 seats per waitlist. Then, the upper bound in point (a) is lower than the limit of $V(\sqrt{K}(\widehat{\beta}^I_{\text{PS}}-\Delta_K))$. In practice, using the IO rather than the DREO estimator can lead to large efficiency losses. For instance, if $N_0=40$, $T_0=30$, and $S_0=20$, the asymptotic variance of the IO estimator is 1.97 times larger than that of the DREO one. Still, there are also instances where Assumption 5 holds and the variance in point (a) of Theorem 3.3 is higher than that in point (b). For instance, that is the case if all waitlists have 3 seats, 38 takers, and 40 applicants.

4. APPLICATION TO BLATTMAN AND ANNAN (2016)

After the second Liberian civil war, some ex-fighters started engaging in illegal activities, and working abroad as mercenaries. Blattman and Annan (2016)[11] studied the effect of an agricultural training on their employment and on their social networks. By improving their labor market opportunities, the program hoped to reduce their interest in illegal and mercenary activities, and to sever their relationships with other ex-combatants. To

[10] Under technical conditions, for instance if one assumes that the potential outcomes have a bounded support, it follows from Theorem 2.20 in Van der Vaart (2000) and Theorem 3.1 that $\sqrt{K}(\widehat{\Delta}-\Delta_K)$ converges in L^2, so $\limsup_{K\to+\infty} V(\sqrt{K}(\widehat{\Delta}-\Delta_K))$ is actually a simple limit, and it is equal to σ^2, the asymptotic variance of $\widehat{\Delta}$.

[11] Blattman and Annan (2016) is one of the few articles in our survey in Section S1 of the Supplemental Material whose data are not proprietary.

allocate the treatment, the authors divided applicants into 70 waitlists, according to the training site they applied for, their former military rank, and their community. In each waitlist, they randomly ranked applicants, and offers were made following that ranking until the seats available were filled.

Blattman and Annan (2016) estimated the training's effect on 62 outcomes, that are either applicants' answers to survey questions, or indexes averaging their answers to several related questions. To preserve space, we only consider the main outcomes. Here are the rules we used to make our selection: we chose indexes rather than questions averaged into an index; among questions not averaged into an index, we discarded those asking applicants to give a subjective opinion; finally, we discarded a few measures the authors did not comment on in the paper. We end up with four measures of employment, one measure of applicants' interest in working as mercenaries, and five measures of their social network.

For each outcome, Table III shows the EO estimator computed by the authors, and the DREO estimator computed with the same controls as those used by the authors.[12] An estimate of $\widehat{\sigma}_+/\sqrt{K-1}$ is shown next to each DREO estimator.[13] In Theorem S2.1 in the Supplemental Material, we derive the asymptotic distribution of the EO estimator. Accordingly, an estimate of $\widehat{\sigma}^2_{E,+}/\sqrt{K-1}$, defined in Section S2.2 of the Supplemental Material, is shown next to each EO estimator. The table then shows the p-value of a t-test that the estimands of the EO and DREO estimators are equal,[14] also computed following Theorem S2.1 in the Supplemental Material. Finally, the table shows the estimated difference between the mean of $Y_{ik}(0)$ among non-takers and takers, denoted $\Delta_{Y(0)}$. The EO and DREO estimators are close for all employment outcomes, but they significantly

TABLE III

ESTIMATORS OF THE LATE IN BLATTMAN AND ANNAN (2016)[a]

	EO (s.e.)	DREO (s.e.)	EO = DREO	$\Delta_{Y(0)}$
Works in agriculture	0.155 (0.037)	0.167 (0.037)	0.214	0.020
Hours illegal work	−3.697 (1.822)	−3.188 (1.614)	0.183	−2.807
Hours farming work	4.090 (1.511)	4.319 (1.472)	0.468	3.070
Income index	0.157 (0.075)	0.169 (0.069)	0.400	−0.087
Interest mercenary work	−0.239 (0.140)	−0.361 (0.155)	0.010	0.307
Relations ex-combatants	0.073 (0.091)	0.050 (0.097)	0.388	−0.079
Relations ex-commanders	−0.154 (0.113)	−0.227 (0.109)	0.011	0.251
Social network quality	0.027 (0.068)	0.082 (0.068)	0.013	−0.041
Social support	0.188 (0.091)	0.161 (0.089)	0.166	−0.165
Relationships families	0.133 (0.075)	0.161 (0.079)	0.205	−0.059
N	1,025	1,016		

[a]Columns 2 and 3 show the EO and DREO estimators in Blattman and Annan (2016), for the outcome variables in Column 1, and with the same controls as in Blattman and Annan (2016). The EO estimators are computed using all the waitlists, while the DREO estimators are computed excluding two waitlists that had less than two seats. An estimate of $\widehat{\sigma}_+/\sqrt{K-1}$ is shown next to each DREO estimator, in parentheses. An estimate of $\widehat{\sigma}^2_{E,+}/\sqrt{K-1}$ (see Section S2.2 of the Supplemental Material) is shown next to each EO estimator, in parentheses. Column 4 shows the p-value of a t-test that the EO and DREO estimators are equal, where we follow Theorem S2.1 in the Supplemental Material to compute the standard error of the difference between the two estimators. Column 5 shows the estimated difference between the mean of $Y_{ik}(0)$ among takers and non-takers.

[12]The DREO estimator with controls is defined in Section S2.5 of the Supplemental Material.

[13]To account for the controls included in the estimation, Y_{ik} and D_{ik} are regressed on the controls, and then the residuals from those two regressions are used instead of Y_{ik} and D_{ik} in the computation of $\widehat{\sigma}_+$.

[14]Under homogeneous effects, this test is equivalent to testing the absence of bias in the EO estimator ($B = 0$).

differ for three of the other outcomes. For applicants' interest in mercenary work, the DREO estimator is 51.0% larger in absolute value than the EO one, and it is statistically significant while the EO estimator is only marginally significant; for applicants' relations with their ex-commanders, the DREO estimator is 47.4% larger, and it is statistically significant while the EO estimator is not; for applicants' social network quality, the DREO estimator is three times larger, but neither of the two estimators is statistically significant. For the first two outcomes, the estimated difference between the mean of $Y_{ik}(0)$ of takers and non-takers is large (30.7% and 25.1% of the standard deviation of these variables, respectively), which may explain why the EO and DREO estimators differ.

5. CONCLUSION

When the seats available for a treatment are allocated using randomized waitlists, we show that applicants getting and not getting an offer are not statistically comparable. Accordingly, a commonly used estimator of the treatment effect, the ever-offer estimator, is inconsistent when the number of waitlists goes to infinity. We propose a new estimator, the doubly-reweighted ever-offer (DREO) estimator, and we show that it is consistent and asymptotically normal. Finally, we show that the DREO estimator is often more efficient than another consistent estimator, the initial-offer estimator. Overall, we recommend that practitioners use the DREO estimator when they analyze randomized waitlists.

APPENDIX: PROOFS

This appendix contains the proofs of the paper's main results. Theorems 3.2 and 3.3 are proven in the Supplemental Material. The next lemma shows that the expectation of the average of any function of potential treatments and outcomes is the same among w_{ik}-reweighted applicants getting an offer and those not getting an offer. $\forall(i,k)\in\mathcal{I}$, let $P_{ik}=(D_{ik}(1),Y_{ik}(0),Y_{ik}(1))$.

LEMMA A.1: *If Assumptions 1–3 hold, then* $\forall k\in\{1,\dots,K\}$ *and for any function* $\phi:\mathbb{R}^3\mapsto\mathbb{R}$,

$$E\left(\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}\phi(P_{ik})\,\Big|\,\mathcal{P}_k\right)=E\left(\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}\phi(P_{ik})\,\Big|\,\mathcal{P}_k\right)=\frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik}).$$

PROOF: We start by showing that

$$E\left(\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}\phi(P_{ik})\,\Big|\,\mathcal{P}_k\right)=\frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik}). \tag{4}$$

First, we show that (4) holds when $\mathcal{P}_k$ is such that $T_k<N_k$. Then, we have

$$\begin{aligned}&E\left(\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}\phi(P_{ik})\,\Big|\,\mathcal{P}_k\right)\\&=E\left(\sum_{i=1}^{N_k}\frac{1}{L_k-1}\left(1-\frac{D_{ik}(1)}{S_k}\right)\phi(P_{ik})1\{R_{ik}\le L_k\}\,\Big|\,\mathcal{P}_k\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{N_k}\left(1-\frac{D_{ik}(1)}{S_k}\right)\phi(P_{ik})E\left(\frac{1}{L_k-1}1\{R_{ik}\leq L_k\}\,\Big|\,\mathcal{P}_k\right) \\
&= \sum_{i=1}^{N_k}\left(1-\frac{D_{ik}(1)}{S_k}\right)\phi(P_{ik})\sum_{l=S_k}^{N_k-T_k+S_k} P(L_k=l\mid\mathcal{P}_k)\frac{1}{l-1}E\big(1\{R_{ik}\leq l\}\mid L_k=l,\mathcal{P}_k\big) \\
&= \sum_{i=1}^{N_k}\left(1-\frac{D_{ik}(1)}{S_k}\right)\phi(P_{ik}) \\
&\quad\times\sum_{l=S_k}^{N_k-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}}{\binom{N_k}{T_k}}\frac{1}{l-1}E\big(1\{R_{ik}\leq l\}\mid L_k=l,\mathcal{P}_k\big) \\
&= \sum_{i=1}^{N_k}\left(1-\frac{D_{ik}(1)}{S_k}\right)\phi(P_{ik}) \\
&\quad\times\sum_{l=S_k}^{N_k-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}}{\binom{N_k}{T_k}}\frac{1}{l-1}\left(D_{ik}(1)\frac{S_k}{T_k}+\big(1-D_{ik}(1)\big)\frac{l-S_k}{N_k-T_k}\right) \\
&= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik})\left(D_{ik}(1)\sum_{l=S_k}^{N_k-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}\frac{S_k-1}{l-1}}{\binom{N_k}{T_k}\frac{T_k}{N_k}}\right. \\
&\quad\left.+\big(1-D_{ik}(1)\big)\sum_{l=S_k+1}^{N_k-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}\frac{l-S_k}{l-1}}{\binom{N_k}{T_k}\frac{N_k-T_k}{N_k}}\right) \\
&= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik})\left(D_{ik}(1)\sum_{l=S_k-1}^{N_k-T_k+S_k-1}\frac{\binom{l-1}{S_k-2}\binom{N_k-1-l}{T_k-S_k}}{\binom{N_k-1}{T_k-1}}\right. \\
&\quad\left.+\big(1-D_{ik}(1)\big)\sum_{l=S_k}^{N_k-1-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-1-l}{T_k-S_k}}{\binom{N_k-1}{T_k}}\right) \\
&= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik}). \qquad (5)
\end{aligned}
$$

The first equality follows from the definitions of w_{ik}, Z_{ik}, and D_{ik}. The second equality holds because $D_{ik}(1)$ and $\phi(P_{ik})$ are functions of $\mathcal{P}_k$, N_k and S_k are non-stochastic, and the conditional expectation is linear. The third follows from the law of iterated expectations, and the fact that L_k is included between S_k and $N_k - T_k + S_k$ under Assumptions 1 and 3.

Then, under Assumption 1, having $L_k = l$ is equivalent to having $S_k - 1$ takers with $R_{ik} \leq l - 1$, one with $R_{ik} = l$, and $T_k - S_k$ with $R_{ik} \geq l + 1$. $\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}T_k!(N_k - T_k)!$ possible values of $\mathcal{R}_k$ satisfy these constraints. Under Assumption 2, conditional on $\mathcal{P}_k$ each of those values has a probability $\frac{1}{N_k!}$ of being realized. Hence the fourth equality.

Then,

$$\begin{aligned}
&E\big(1\{R_{ik} \leq l\} \mid L_k = l, \mathcal{P}_k\big) \\
&\quad = D_{ik}(1)E\big(1\{R_{ik} \leq l\} \mid L_k = l, D_{ik}(1) = 1, \mathcal{P}_k \setminus D_{ik}(1)\big) \\
&\qquad + \big(1 - D_{ik}(1)\big)E\big(1\{R_{ik} \leq l\} \mid L_k = l, D_{ik}(1) = 0, \mathcal{P}_k \setminus D_{ik}(1)\big).
\end{aligned} \tag{6}$$

Conditional on $L_k = l$, S_k takers out of T_k satisfy $R_{ik} \leq l$, and Assumption 2 ensures that each taker has the same probability of satisfying this condition, so

$$E\big(1\{R_{ik} \leq l\} \mid L_k = l, D_{ik}(1) = 1, \mathcal{P}_k \setminus D_{ik}(1)\big) = \frac{S_k}{T_k}. \tag{7}$$

Similarly, conditional on $L_k = l$ and $T_k < N_k$, $l - S_k$ non-takers out of $N_k - T_k$ satisfy $R_{ik} \leq l$, and Assumption 2 ensures that each has the same probability of satisfying this condition, so

$$E\big(1\{R_{ik} \leq l\} \mid L_k = l, D_{ik}(1) = 0, \mathcal{P}_k \setminus D_{ik}(1)\big) = \frac{l - S_k}{N_k - T_k}. \tag{8}$$

Plugging (7) and (8) into (6) yields the fifth equality. The sixth and seventh equalities follow after some algebra.

Then, we prove the eighth equality. Before that, note that $T_k < N_k$ and Assumption 3 ensure that $1 \leq S_k - 1 \leq T_k - 1 \leq N_k - 1$ and $1 \leq S_k \leq T_k \leq N_k - 1$, thus ensuring that all the quantities that follow are well-defined. There are $\binom{N_k-1}{T_k-1}$ ways of distributing $T_k - 1$ units over $N_k - 1$ ranks. The rank of the $S_k - 1$th unit must be included between $S_k - 1$ and $N_k - T_k + S_k - 1$, and for every $l \in \{S_k - 1, \dots, N_k - T_k + S_k - 1\}$, there are $\binom{l-1}{S_k-2}\binom{N_k-1-l}{T_k-S_k}$ ways of distributing those $T_k - 1$ units while having that the $S_k - 1$th unit is at the lth rank. Therefore,

$$\sum_{l=S_k-1}^{N_k-T_k+S_k-1} \binom{l-1}{S_k-2}\binom{N_k-1-l}{T_k-S_k} = \binom{N_k-1}{T_k-1}. \tag{9}$$

Similarly, when distributing T_k units over $N_k - 1$ ranks, the rank of the S_kth unit must lie between S_k and $N_k - 1 - T_k + S_k$. For every $l \in \{S_k, \dots, N_k - 1 - T_k + S_k\}$, there are $\binom{l-1}{S_k-1}\binom{N_k-1-l}{T_k-S_k}$ ways of distributing those T_k units while having the S_kth unit at the lth rank. Thus,

$$\sum_{l=S_k}^{N_k-1-T_k+S_k} \binom{l-1}{S_k-1}\binom{N_k-1-l}{T_k-S_k} = \binom{N_k-1}{T_k}. \tag{10}$$

The eighth equality follows from (9) and (10). This concludes the proof of (5).

Second, we show that (4) holds when $\mathcal{P}_k$ is such that $T_k = N_k$. Then, we have

$$\begin{aligned} E\left(\frac{1}{L_k-1}\sum_{i:Z_{ik}=1} w_{ik}\phi(P_{ik}) \,\middle|\, \mathcal{P}_k\right) &= E\left(\sum_{i=1}^{N_k}\phi(P_{ik})\frac{1}{S_k}1\{R_{ik}\le S_k\} \,\middle|\, \mathcal{P}_k\right) \\ &= \sum_{i=1}^{N_k}\phi(P_{ik})\frac{1}{S_k}E\big(1\{R_{ik}\le S_k\} \mid \mathcal{P}_k\big) \\ &= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik}). \end{aligned} \tag{11}$$

The first equality follows from the definition of w_{ik} and from the fact that if $T_k = N_k$, $L_k = S_k$. The second equality holds because $\phi(P_{ik})$ is a function of $\mathcal{P}_k$, N_k and S_k are non-stochastic, and the conditional expectation is linear. The third equality follows from the fact that under Assumption 2, if $T_k = N_k$, then, conditional on $\mathcal{P}_k$, each applicant has a probability $\frac{S_k}{N_k}$ of having $R_{ik} \le S_k$. This proves (11). Equations (5) and (11) prove (4).

We then show that

$$E\left(\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}\phi(P_{ik}) \,\middle|\, \mathcal{P}_k\right) = \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik}). \tag{12}$$

First, we show that (12) holds when $\mathcal{P}_k$ is such that $T_k < N_k$. Then, we have

$$\begin{aligned} &E\left(\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}\phi(P_{ik}) \mid \mathcal{P}_k\right) \\ &\quad= \sum_{i=1}^{N_k}\phi(P_{ik})E\left(\frac{1}{N_k-L_k}1\{R_{ik}>L_k\} \mid \mathcal{P}_k\right) \\ &\quad= \sum_{i=1}^{N_k}\phi(P_{ik})\sum_{l=S_k}^{N_k-T_k+S_k} \\ &\qquad\times \frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}}{\binom{N_k}{T_k}}\frac{1}{N_k-l}E\big(1\{R_{ik}>l\} \mid L_k=l, \mathcal{P}_k\big) \\ &\quad= \sum_{i=1}^{N_k}\phi(P_{ik})\sum_{l=S_k}^{N_k-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}}{\binom{N_k}{T_k}} \\ &\qquad\times \frac{1}{N_k-l}\left(D_{ik}(1)\frac{T_k-S_k}{T_k} + \big(1-D_{ik}(1)\big)\frac{N_k-T_k-l+S_k}{N_k-T_k}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik})\left(D_{ik}(1)\sum_{l=S_k}^{N_k-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}\frac{T_k-S_k}{N_k-l}}{\binom{N_k}{T_k}\frac{T_k}{N_k}}\right.\\
&\qquad\left.+\big(1-D_{ik}(1)\big)\sum_{l=S_k}^{N_k-1-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-l}{T_k-S_k}\frac{N_k-T_k-l+S_k}{N_k-l}}{\binom{N_k}{T_k}\frac{N_k-T_k}{N_k}}\right)\\
&= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik})\left(D_{ik}(1)\sum_{l=S_k}^{N_k-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-1-l}{T_k-1-S_k}}{\binom{N_k-1}{T_k-1}}\right.\\
&\qquad\left.+\big(1-D_{ik}(1)\big)\sum_{l=S_k}^{N_k-1-T_k+S_k}\frac{\binom{l-1}{S_k-1}\binom{N_k-1-l}{T_k-S_k}}{\binom{N_k-1}{T_k}}\right)\\
&= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik}).
\end{aligned}
\tag{13}
$$

This derivation follows from arguments similar to those used when deriving (5). We only prove the last equality. Note that Assumption 3 ensures that $1 \le S_k \le T_k - 1 \le N_k - 1$, thus ensuring that all the quantities that follow are well-defined. There are $\binom{N_k-1}{T_k-1}$ ways of distributing $T_k - 1$ units over $N_k - 1$ ranks. The rank of the S_kth unit must be included between S_k and $N_k - T_k + S_k$, and for every $l \in \{S_k, \dots, N_k - T_k + S_k\}$, there are $\binom{l-1}{S_k-1}\binom{N_k-1-l}{T_k-1-S_k}$ ways of distributing those $T_k - 1$ units while having that the S_kth unit is at the lth rank. Therefore,

$$
\sum_{l=S_k}^{N_k-T_k+S_k}\binom{l-1}{S_k-1}\binom{N_k-1-l}{T_k-1-S_k} = \binom{N_k-1}{T_k-1}. \tag{14}
$$

The last equality in the derivation of (13) follows from (10) and (14).

Second, we show that (12) holds when $\mathcal{P}_k$ is such that $T_k = N_k$. Then, we have

$$
\begin{aligned}
E\left(\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}\phi(P_{ik})\,\middle|\,\mathcal{P}_k\right) &= \sum_{i=1}^{N_k}\phi(P_{ik})\frac{1}{N_k-S_k}E\big(1\{R_{ik}>S_k\}\mid\mathcal{P}_k\big)\\
&= \frac{1}{N_k}\sum_{i=1}^{N_k}\phi(P_{ik}).
\end{aligned}
\tag{15}
$$

This derivation follows from arguments similar to those used when deriving (11). Equations (13) and (15) prove (12). *Q.E.D.*

PROOF OF LEMMA 3.1: We only prove point (a); point (b) follows from a similar argument:

$$\begin{aligned}
&E\left(\frac{1}{K}\sum_{k=1}^{K}\frac{N_k}{\overline{N}}\left(\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}Y_{ik}-\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}Y_{ik}\right)\right)\\
&=\frac{1}{K}\sum_{k=1}^{K}\frac{N_k}{\overline{N}}E\left(E\left(\frac{1}{L_k-1}\sum_{i:Z_{ik}=1}w_{ik}Y_{ik}\big(D_{ik}(1)\big)\,\Big|\,\mathcal{P}_k\right)\right.\\
&\quad\left.-E\left(\frac{1}{N_k-L_k}\sum_{i:Z_{ik}=0}Y_{ik}(0)\,\Big|\,\mathcal{P}_k\right)\right)\\
&=\frac{1}{K}\sum_{k=1}^{K}\frac{N_k}{\overline{N}}E\left(\frac{1}{N_k}\sum_{i=1}^{N_k}\big[Y_{ik}\big(D_{ik}(1)\big)-Y_{ik}(0)\big]\right)\\
&=E\left(\frac{1}{N}\sum_{(i,k)\in\mathcal{I}}\big[Y_{ik}\big(D_{ik}(1)\big)-Y_{ik}(0)\big]\right).
\end{aligned}$$

The first equality follows from the linearity of the expectation, from the fact N_k and $\overline{N}$ are not stochastic, from point (c) of Assumption 1 and the definitions of Y_{ik} and D_{ik}, from the law of iterated expectations, and from the linearity of the conditional expectation. The second equality follows from Lemma A.1, with $\phi(P_{ik})=Y_{ik}(D_{ik}(1))$ for the first conditional expectation, and $\phi(P_{ik})=Y_{ik}(0)$ for the second one. The third equality follows after some algebra. *Q.E.D.*

The proof of Theorem 3.1 makes use of the following lemma, where $O_p(1)$ (resp. $o_p(1)$) stands for a sequence of random variables bounded in probability (resp. converging towards 0 in probability); see, for example, Van der Vaart (2000).

LEMMA A.2: *Let $(A_K)_{K\in\mathbb{N}}$ and $(B_K)_{K\in\mathbb{N}}$ be two sequences of real numbers such that, for every K, $B_K\geq C$ for some real number $C>0$, and $\frac{A_K}{B_K}$ converges towards a finite limit. Let $(\widehat{A}_K)_{K\in\mathbb{N}}$ and $(\widehat{B}_K)_{K\in\mathbb{N}}$ be two sequences of random variables such that $\sqrt{K}(\widehat{A}_K-A_K)=O_p(1)$ and $\sqrt{K}(\widehat{B}_K-B_K)=O_p(1)$. Then,*

$$\sqrt{K}\left(\frac{\widehat{A}_K}{\widehat{B}_K}-\frac{A_K}{B_K}\right)=\sqrt{K}\frac{1}{B_K}\left((\widehat{A}_K-A_K)-\frac{A_K}{B_K}(\widehat{B}_K-B_K)\right)+o_P(1).$$

PROOF: $\sqrt{K}(\widehat{A}_K-A_K)=O_p(1)$ and $\sqrt{K}(\widehat{B}_K-B_K)=O_p(1)$ imply that $\widehat{A}_K-A_K=o_p(1)$ and $\widehat{B}_K-B_K=o_p(1)$. Therefore, with probability approaching 1, $\max(\widehat{A}_K-A_K,\widehat{B}_K-B_K)\leq\frac{C}{2}$. Then, Lemma S3 in de Chaisemartin and D'Haultfœuille (2018) implies that with probability approaching 1,

$$\begin{aligned}
&\left|\sqrt{K}\left(\frac{\widehat{A}_K}{\widehat{B}_K}-\frac{A_K}{B_K}\right)-\sqrt{K}\frac{1}{B_K}\left((\widehat{A}_K-A_K)-\frac{A_K}{B_K}(\widehat{B}_K-B_K)\right)\right|\\
&\quad\leq\frac{2\left(1+\frac{A_K}{B_K}\right)}{C^2}\max\big(\sqrt{K}(\widehat{A}_K-A_K),\sqrt{K}(\widehat{B}_K-B_K)\big)\max(\widehat{A}_K-A_K,\widehat{B}_K-B_K).
\end{aligned}$$

The right-hand side of the inequality in the previous display is an $o_p(1)$. With probability approaching 1, the left-hand side is bounded by an $o_p(1)$, so it is itself an $o_p(1)$. *Q.E.D.*

PROOF OF THEOREM 3.1: *1. Proof of the asymptotic normality of* $\widehat{\Delta}$.

Let $\overline{ITT}_K = \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{N_k}[Y_{ik}(D_{ik}(1)) - Y_{ik}(0)]$, let $\overline{T}_K = \frac{1}{K}\sum_{k=1}^{K} T_k$, and let $\tilde{\Delta}_K = \frac{E(\overline{ITT}_K)}{E(\overline{T}_K)}$. We prove the asymptotic normality of $\widehat{\Delta}$ in two steps. First, we prove that $\sqrt{K}(\widehat{\Delta} - \tilde{\Delta}_K)$ is asymptotically normal. Second, we show that $\sqrt{K}(\tilde{\Delta}_K - \Delta_K)$ converges towards 0.

Proof that $\sqrt{K}(\widehat{\Delta} - \tilde{\Delta}_K) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

First, notice that

$$\tilde{\Delta}_K = \frac{E\Big(\frac{1}{N}\sum_{(i,k)\in\mathcal{I}}\big[Y_{ik}\big(D_{ik}(1)\big) - Y_{ik}(0)\big]\Big)}{E\Big(\frac{1}{N}\sum_{(i,k)\in\mathcal{I}} D_{ik}(1)\Big)} = \frac{E\Big(\frac{1}{K}\sum_{k=1}^{K} RF_k\Big)}{E\Big(\frac{1}{K}\sum_{k=1}^{K} FS_k\Big)}. \tag{16}$$

The first equality follows from some algebra, and from point (a) of Assumption 1. The second equality follows from points (a) and (b) of Lemma 3.1 and the definitions of RF_k and FS_k.

Then,

$$\sqrt{K}\Bigg(\frac{1}{K}\sum_{k=1}^{K} RF_k - E\Bigg(\frac{1}{K}\sum_{k=1}^{K} RF_k\Bigg)\Bigg) = \frac{\sum_{k=1}^{K}\big(RF_k - E(RF_k)\big)}{\sqrt{\sum_{k=1}^{K} V(RF_k)}}\sqrt{\frac{1}{K}\sum_{k=1}^{K} V(RF_k)}. \tag{17}$$

Point (a) of Assumption 1 and point (a) of Assumption 4 ensure that the RF_k's are independent. It follows from a few lines of algebra and point (c) of Assumption 4 that the fourth moment of RF_k is bounded for all k, thus ensuring that its expectation and variance exist. Finally, points (d.iii) and (d.vi) of Assumption 4 ensure that $(RF_k)_{k\in\mathbb{N}^*}$ satisfies the Liapunov condition for $\delta = 2$. Then, the Liapunov central limit theorem (see, e.g., Billingsley (1986, Theorem 27.3)) implies that

$$\frac{\sum_{k=1}^{K}\left(RF_k - E(RF_k)\right)}{\sqrt{\sum_{k=1}^{K} V(RF_k)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{18}$$

Point (d.iii) of Assumption 4 ensures that $\sqrt{\frac{1}{K}\sum_{k=1}^{K} V(RF_k)}$ has a finite limit, denoted σ_{RF}. Therefore, combining (17), (18), and the Slutsky lemma,

$$\sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K}\left(RF_k - E\left(\frac{1}{K}\sum_{k=1}^{K} RF_k\right)\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{RF}^2\right). \tag{19}$$

Similarly, let σ_{FS} be the limit of $\sqrt{\frac{1}{K}\sum_{k=1}^{K} V(FS_k)}$. One can show that

$$\sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K}\left(FS_k - E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{FS}^2\right). \tag{20}$$

Finally,

$$\begin{aligned}
&\sqrt{K}(\widehat{\Delta} - \tilde{\Delta}_K) \\
&= \sqrt{K}\left(\frac{\frac{1}{K}\sum_{k=1}^{K} RF_k}{\frac{1}{K}\sum_{k=1}^{K} FS_k} - \frac{E\left(\frac{1}{K}\sum_{k=1}^{K} RF_k\right)}{E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)}\right) \\
&= \sqrt{K}\frac{1}{E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)}\left(\frac{1}{K}\sum_{k=1}^{K} RF_k - E\left(\frac{1}{K}\sum_{k=1}^{K} RF_k\right)\right. \\
&\quad \left. - \frac{E\left(\frac{1}{K}\sum_{k=1}^{K} RF_k\right)}{E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)}\left(\frac{1}{K}\sum_{k=1}^{K} FS_k - E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)\right)\right) + o_P(1) \\
&= \sqrt{K}\frac{1}{E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)}\left(\frac{1}{K}\sum_{k=1}^{K} RF_k - E\left(\frac{1}{K}\sum_{k=1}^{K} RF_k\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&- \Delta\left(\frac{1}{K}\sum_{k=1}^{K} FS_k - E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)\right)\Bigg) + o_P(1) \\
&= \frac{FS}{E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)} \sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K}\bigl(\Lambda_k - E(\Lambda_k)\bigr)\right) + o_P(1) \xrightarrow{d} \mathcal{N}\bigl(0, \sigma^2\bigr).
\end{aligned}
$$

The first equality follows from the definitions of FS_k and RF_k and from (16).

The second equality follows from the fact $E(\frac{1}{K}\sum_{k=1}^{K} RF_k)$, $E(\frac{1}{K}\sum_{k=1}^{K} FS_k)$, $\frac{1}{K}\sum_{k=1}^{K} RF_k$, and $\frac{1}{K}\sum_{k=1}^{K} FS_k$ satisfy the assumptions of Lemma A.2. Indeed, point (b) of Lemma 3.1, point (b) of Assumption 4, and Assumption 3 imply that $E(\frac{1}{K}\sum_{k=1}^{K} FS_k) \geq \frac{3}{N^+} > 0$. Moreover, points (d.i) and (d.ii) of Assumption 4 imply that $E(\frac{1}{K}\sum_{k=1}^{K} RF_k)/E(\frac{1}{K}\sum_{k=1}^{K} FS_k)$ converges towards a finite limit. Finally, it follows from (19), (20), and the fact that convergence in distribution implies boundedness in probability, that

$$
\sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K} RF_k - E\left(\frac{1}{K}\sum_{k=1}^{K} RF_k\right)\right) = O_p(1),
$$

$$
\sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K} FS_k - E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)\right) = O_p(1).
$$

Points (d.i) and (d.ii) of Assumption 4 and (20) ensure that

$$
\frac{1}{E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)}\left(\Delta - \frac{E\left(\frac{1}{K}\sum_{k=1}^{K} RF_k\right)}{E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)}\right)\sqrt{K}\left(\frac{1}{K}\sum_{k=1}^{K} FS_k - E\left(\frac{1}{K}\sum_{k=1}^{K} FS_k\right)\right) = o_P(1),
$$

hence the third equality. The fourth equality follows from the definition of Λ_k. The convergence in distribution arrow follows from a reasoning similar to that used to prove (19), and from the Slutsky lemma and the definition of FS.

Proof that $\lim_{K\to\infty} \sqrt{K}(\tilde{\Delta}_K - \underset{\sim}{\Delta_K}) = 0$.

Let $\boldsymbol{D(1)} = (D_{i,k}(1))_{(i,k)\in\mathcal{I}}$, $\widetilde{\overline{ITT}}_K = E(\overline{ITT}_K|\boldsymbol{D(1)})$, and $I = \mathbb{1}\{|\overline{T}_K - E(\overline{T}_K)| < \varepsilon_K\}$, where $\varepsilon_K > 0$ will be specified below. By point (d.ix) of Assumption 4 and Assumption 3, $\lim_{K\to\infty} E(\overline{T}_K) > 0$. Thus, it suffices to prove that $\sqrt{K}(E(\overline{T}_K)\Delta_K - E(\overline{ITT}_K)) \to 0$. Because $\Delta_K = E(\overline{ITT}_K/\overline{T}_K)$, we have

$$
\begin{aligned}
&\sqrt{K}E(\overline{T}_K)\Delta_K - E(\overline{ITT}_K) \\
&\quad = \sqrt{K}E\bigl[\widetilde{\overline{ITT}}_K\bigl(E(\overline{T}_K) - \overline{T}_K\bigr)/\overline{T}_K\bigr] \\
&\quad = \sqrt{K}E\bigl[\widetilde{\overline{ITT}}_K I\bigl(E(\overline{T}_K) - \overline{T}_K\bigr)/\overline{T}_K\bigr] + \sqrt{K}E\bigl[\widetilde{\overline{ITT}}_K(1-I)\bigl(E(\overline{T}_K) - \overline{T}_K\bigr)/\overline{T}_K\bigr].
\end{aligned} \tag{21}
$$

We first show that the second term on the right-hand side goes to 0. By applying twice the Cauchy–Schwarz inequality, we get

$$\begin{aligned} &\sqrt{K}\big|E\big[\widetilde{\overline{ITT}}_K(1-I)\big(E(\overline{T}_K)-\overline{T}_K\big)/\overline{T}_K\big]\big| \\ &\quad \leq \sqrt{K}V(\overline{T}_K)^{1/2}\big[E\big[(\widetilde{\overline{ITT}}_K/\overline{T}_K)^4\big]\Pr(I=0)\big]^{1/4}. \end{aligned} \tag{22}$$

By points (a) and (d.xi) of Assumption 4, $\sqrt{K}V(\overline{T}_K)^{1/2}$ converges towards a finite limit. Thus, it suffices to show that the term in brackets in the right-hand side tends to 0. Note that

$$E\big[\overline{ITT}_K|\boldsymbol{D}(\mathbf{1})\big]=\frac{1}{K}\sum_{(i,k)\in\mathcal{I}}D_{ik}(1)E\big(Y_{ik}(1)-Y_{ik}(0)|\boldsymbol{D}(\mathbf{1})\big).$$

Then,

$$\begin{aligned} E\bigg[\bigg(\frac{\widetilde{\overline{ITT}}_K}{\overline{T}_K}\bigg)^4\bigg] &\leq E\left[\left(\frac{\sum_{(i,k)\in\mathcal{I}}D_{ik}(1)\big|E\big(Y_{ik}(1)-Y_{ik}(0)|\boldsymbol{D}(\mathbf{1})\big)\big|}{\sum_{(i,k)\in\mathcal{I}}D_{ik}(1)}\right)^4\right] \\ &\leq E\Big[\Big(\max_{(i,k)\in\mathcal{I}}\big(E\big(Y_{ik}(1)-Y_{ik}(0)|\boldsymbol{D}(\mathbf{1})\big)\big)^4\Big)\Big] \\ &\leq E\bigg[\sum_{(i,k)\in\mathcal{I}}\big(E\big(Y_{ik}(1)-Y_{ik}(0)|\boldsymbol{D}(\mathbf{1})\big)\big)^4\bigg] \\ &\leq E\bigg[\sum_{(i,k)\in\mathcal{I}}E\big(\big(Y_{ik}(1)-Y_{ik}(0)\big)^4|\boldsymbol{D}(\mathbf{1})\big)\bigg] \\ &= \sum_{(i,k)\in\mathcal{I}}E\big(\big(Y_{ik}(1)-Y_{ik}(0)\big)^4\big) \\ &\leq C_1N^+K \end{aligned} \tag{23}$$

for some constant C_1. The first inequality follows from the fact $x\mapsto x^4$ is increasing in $|x|$ and from the triangle inequality. The second inequality follows from the fact that a weighted average of some numbers is smaller than the largest of those numbers. The fourth inequality follows from Jensen's inequality. The equality follows from the law of iterated expectations. The last inequality follows from points (b) and (c) of Assumption 4.

Moreover, by Hoeffding's inequality, point (a) of Assumption 4, Assumption 3, and point (b) of Assumption 4,

$$\Pr(I=0)\leq 2\exp\bigg(\frac{-2K\varepsilon_K^2}{(N^+)^2}\bigg).$$

Let $\varepsilon_K=(C\ln(K)/K)^{1/2}$, for some $C>\frac{(N^+)^2}{2}$. Then, $K\Pr(I=0)\to 0$. Combining this with (22) and (23), we obtain that the second term of the right-hand side of (21) tends to zero.

Now, let us move to the first term of the right-hand side of (21). We have

$$\begin{aligned}
&\sqrt{K}\big|E\big[\widetilde{\overline{ITT}}_K I\big(E(\overline{T}_K)-\overline{T}_K\big)/\overline{T}_K\big]\big|\\
&\quad=\sqrt{K}\big|E\big[I\big(\widetilde{\overline{ITT}}_K-E(\widetilde{\overline{ITT}}_K)\big)\big(E(\overline{T}_K)-\overline{T}_K\big)/\overline{T}_K\big]\\
&\qquad+E(\widetilde{\overline{ITT}}_K)\big(E(\overline{T}_K)E(I/\overline{T}_K)-E(I)\big)\big|\\
&\quad\le E(\overline{T}_K)\sqrt{K}V(\widetilde{\overline{ITT}}_K)^{1/2}\big(E\big[I\big(1/\overline{T}_K-1/E(\overline{T}_K)\big)^2\big]\big)^{1/2}\\
&\qquad+\sqrt{K}\big|E(\overline{ITT}_K)\big|E(\overline{T}_K)\big|E(I/\overline{T}_K)-E(I)/E(\overline{T}_K)\big|,
\end{aligned} \tag{24}$$

where the inequality follows from the triangle and Cauchy–Schwarz inequalities.

We start by proving that the first term on the right-hand side of (24) converges to zero. By definition of $\widetilde{\overline{ITT}}_K$, $KV(\widetilde{\overline{ITT}}_K)\le KV(\overline{ITT}_K)$. Moreover, $KV(\overline{ITT}_K)$ and $E(\overline{T}_K)$ converge towards finite limits by points (a), (d.xii) and (d.ix) of Assumption 4. Therefore, to prove the result, it will be sufficient to prove that $E[I(1/\overline{T}_K-1/E(\overline{T}_K))^2]$ goes to 0. By a Taylor expansion of $x\mapsto 1/x$ around $E(\overline{T}_K)$, there exists T_1 in the interval between $\overline{T}_K$ and $E(\overline{T}_K)$ such that

$$\frac{1}{\overline{T}_K}=\frac{1}{E(\overline{T}_K)}-\frac{\overline{T}_K-E(\overline{T}_K)}{T_1^2}. \tag{25}$$

Then, for K large enough,

$$\begin{aligned}
E\big[I\big(1/\overline{T}_K-1/E(\overline{T}_K)\big)^2\big]&=E\left[I\left(\frac{\overline{T}_K-E(\overline{T}_K)}{T_1^2}\right)^2\right]\\
&\le\frac{V(\overline{T}_K)}{\big(E(\overline{T}_K)-\varepsilon_K\big)^4}\to 0.
\end{aligned}$$

The equality follows from (25). The inequality follows from the fact that when $I=1$, $|T_1-E(\overline{T}_K)|<\varepsilon_K$, so $T_1>E(\overline{T}_K)-\varepsilon_K$, and $E(\overline{T}_K)-\varepsilon_K>0$ for K large enough, as $\lim_{K\to\infty}E(\overline{T}_K)>0$ as shown above. Finally, convergence to 0 follows from the fact $\lim_{K\to\infty}V(\overline{T}_K)=0$ by points (a) and (d.xi) of Assumption 4, and $\lim_{K\to\infty}E(\overline{T}_K)>0$.

We now prove that the second term on the right-hand side of (24) converges to zero. As $E(\overline{T}_K)$ and $E(\overline{ITT}_K)$ converge towards finite limits by points (d.ix) and (d.x) of Assumption 4, it is sufficient to show that $\sqrt{K}|E(I/\overline{T}_K)-E(I)/E(\overline{T}_K)|$ goes to 0. Multiplying (25) by I and taking expectations, we obtain, for K large enough,

$$\begin{aligned}
\sqrt{K}\big|E(I/\overline{T}_K)-E(I)/E(\overline{T}_K)\big|&=\sqrt{K}\left|E\left[\frac{I\big(\overline{T}_K-E(\overline{T}_K)\big)}{T_1^2}\right]\right|\\
&\le\left|\frac{\big[K\Pr(I=0)V(\overline{T}_K)\big]^{1/2}}{\big(E(\overline{T}_K)-\varepsilon_K\big)^2}\right|\to 0.
\end{aligned}$$

The inequality follows from the fact that when $I=1$, $T_1>E(\overline{T}_K)-\varepsilon_K>0$ for K large enough, from $E[I(\overline{T}_K-E(\overline{T}_K))]=-E[(1-I)(\overline{T}_K-E(\overline{T}_K))]$, and from the Cauchy–Schwarz inequality. Convergence to 0 follows from the fact that $K\Pr(I=0)\to 0$ as shown above, $V(\overline{T}_K) \to 0$ by points (a) and (d.xi) of Assumption 4, and $E(\overline{T}_K)$ converges toward a finite limit by point (d.ix) of Assumption 4.

2. Proof that $\widehat{\sigma}^2_+ \xrightarrow{p} \sigma^2_+ \geq \sigma^2$.

By point (a) of Assumption 1, point (a) of Assumption 4, the fact that, for all k, RF_k has a fourth moment by point (c) of Assumption 4, the weak law of large numbers in Gut (1992), and points (d.i) and (d.ii) of Assumption 4,

$$\frac{1}{K}\sum_{k=1}^{K} RF_k \xrightarrow{p} \lim_{K\to+\infty} \frac{1}{K}\sum_{k=1}^{K} E(RF_k),$$
$$\frac{1}{K}\sum_{k=1}^{K} FS_k \xrightarrow{p} \lim_{K\to+\infty} \frac{1}{K}\sum_{k=1}^{K} E(FS_k). \tag{26}$$

Then, (26) and the continuous mapping theorem imply that

$$\widehat{\Delta} \xrightarrow{p} \Delta. \tag{27}$$

Then, (26), (27), and the continuous mapping theorem imply that

$$\frac{1}{K}\sum_{k=1}^{K} \widehat{\Lambda}_k = \frac{1}{\frac{1}{K}\sum_{k=1}^{K} FS_k}\left(\frac{1}{K}\sum_{k=1}^{K} RF_k - \widehat{\Delta}\frac{1}{K}\sum_{k=1}^{K} FS_k\right) \xrightarrow{p} \lim_{K\to+\infty} \frac{1}{K}\sum_{k=1}^{K} E(\Lambda_k). \tag{28}$$

Similarly, one can show that

$$\frac{1}{K}\sum_{k=1}^{K} \widehat{\Lambda}^2_k \xrightarrow{p} \lim_{K\to+\infty} \frac{1}{K}\sum_{k=1}^{K} E(\Lambda^2_k). \tag{29}$$

Then, (28), (29), and the continuous mapping theorem imply that

$$\widehat{\sigma}^2_+ = \frac{1}{K}\sum_{k=1}^{K} \widehat{\Lambda}^2_k - \left(\frac{1}{K}\sum_{k=1}^{K} \widehat{\Lambda}_k\right)^2 \xrightarrow{p} \sigma^2_+. \tag{30}$$

Finally, the convexity of $x \mapsto x^2$ implies that $\frac{1}{K}\sum_{k=1}^{K} E(\Lambda_k)^2 \geq (\frac{1}{K}\sum_{k=1}^{K} E(\Lambda_k))^2$, so $\sigma^2_+ \geq \sigma^2$. *Q.E.D.*

REFERENCES

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70 (1), 91–117. [1455]

ANGRIST, J. D., AND J.-S. PISCHKE (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press. [1455]

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91 (434), 444–455. [1457]

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. [1453]

BILLINGSLEY, P. (1986): *Probability and Measure* (Second Ed.). New York: Wiley. [1472]

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. [1454,1463, 1464]

CRÉPON, B., F. DEVOTO, E. DUFLO, AND W. PARIENTÉ (2015): "Estimating the Impact of Microcredit on Those Who Take It Up: Evidence From a Randomized Experiment in Morocco," *American Economic Journal: Applied Economics*, 7 (1), 123–150. [1455]

DE CHAISEMARTIN, C., AND L. BEHAGHEL (2018): "Estimating the Effect of Treatments Allocated by Randomized Waiting Lists," Preprint. Available at arXiv:1511.01453v6. [1457,1458]

——— (2020): "Supplement to 'Estimating the Effect of Treatments Allocated by Randomized Waiting Lists'," *Econometrica Supplemental Material*, 88, https://doi.org/10.3982/ECTA16032. [1454]

DE CHAISEMARTIN, C., AND X. D'HAULTFŒUILLE (2018): "Fuzzy Differences-in-Differences," *Review of Economic Studies*, 85 (2), 999–1028. [1470]

GUT, A. (1992): "The Weak Law of Large Numbers for Arrays," *Statistics & Probability Letters*, 14 (1), 49–52. [1460,1476]

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62 (2), 467–475. [1462]

LIU, R. Y., AND K. SINGH (1995): "Using i.i.d. Bootstrap Inference for General Non-i.i.d. Models," *Journal of Statistical Planning and Inference*, 43 (1–2), 67–75. [1460]

NEYMAN, J. (1990): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science*, 5 (4), 465–472. [1460]

VAN DER VAART, A. W. (2000): *Asymptotic Statistics*, Vol. 3. Cambridge: Cambridge University Press. [1463, 1470]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 22 January, 2018; final version accepted 5 February, 2020; available online 7 February, 2020.