Working Paper Series 686
(ISSN 1211-3298)

LATE Estimators under Costly Non-compliance in Student-College Matching Markets

Marin Drlje
Štěpán Jurajda

CERGE-EI
Prague, February 2021

ISBN 978-80-7343-493-9 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-582-9 (Národohospodářský ústav AV ČR, v. v. i.)

LATE Estimators under Costly Non-compliance in Student-College Matching Markets

Marin Drlje and Štěpán Jurajda *

1 · 31 · 2021

Abstract

A growing literature exploits a feature of centralized college admission systems where students with similar admission scores in a neighborhood of a school's admission threshold are or are not offered admission based on small quasi-random differences in admission scores. Assuming that the students at the margin of admission differ only in the treatment assignment, this literature relies on admission scores to instrument for admission or graduation. We point out that non-compliance with the centralized matching assignment typically corresponds to enrolling in one's preferred program a year after the initial assignment, introducing significant non-compliance costs. We show that with costly non-compliance, the *exclusion restriction*, the key assumption of the LATE theorem, is violated, leading to biased estimates when instrumenting for graduation, i.e., for a treatment taking place after non-compliance costs are incurred. We use data from a student-college matching market in Croatia to illustrate the empirical importance of this potential source of bias and propose a method inspired by Lee (2009), which recovers the treatment effect bounds under the assumption that the costs of non-compliance are not related to the treatment assignment.

*CERGE-EI, a joint workplace of Center for Economic Research and Graduate Education, Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, P.O. Box 882, 111 21 Prague 1, Czech Republic.

E-mail: marin.drlje@cerge-ei.cz stepan.jurajda@cerge-ei.cz. Jurajda is a Research Fellow at IZA, Bonn. We thank Thomas Le Barbanchon, Michal Bauer, Alena Bicakova, Henry Farber, Randall K. Filer, Vasily Korovkin, Dejan Kovac, David Lee, Andreas Menzel, Nikolas Mittag, Daniel Munich, Christopher Neilson, Christian Ochsner, Seth Zimmerman, Jan Zápal, Zhuan Pei and Kresimir Zigic for their many useful comments. We are also grateful to the Ministry of Education of Croatia and ASHE (AZVO) for access to their administrative data.

1. Introduction

The instrumental variable (IV) estimator is widely used to account for unmeasured confounding factors and to identify causal effects (Angrist and Krueger, 1991). It is predominantly implemented in the form of the 2SLS estimator, which, under certain assumptions, identifies the local average treatment effect (LATE) for individuals whose treatment is manipulated by (quasi-) random instrumental variation—the so-called *compliers*. In this paper, we consider the properties of the 2SLS estimator in a setup where non-compliance with quasi-random treatment assignment is costly, which violates the *exclusion restriction*, one of the crucial assumptions necessary for the causal interpretation of the 2SLS estimator. We build on the LATE theorem (Imbens and Angrist, 1994) to show that in the case of costly non-compliance, the IV estimator can be interpreted as LATE only after assuming that both the costs and the probabilities of non-compliance do not depend on the instrument's value. Intuitively, if the costs depend on the instrument's value, the instrument affects non-compliers through the costs of non-compliance, and becomes correlated with the outcome not only for compliers, but also for non-compliers, which biases the 2SLS estimator.

We apply this insight to the growing literature exploiting a feature of centralized college admission systems where students with similar admission scores in a neighbourhood of a school's admission threshold are or are not offered admission based on small differences in admission scores. Assuming that the students at the margin of admission differ only in their treatment assignment, this literature relies on an indicator of whether a student is above the school-specific admission threshold (admission score cutoff) to instrument for graduation or admission. The LATE theorem is then invoked to interpret these IV estimates (e.g., Kirkeboen et al. (2016)).

A basic feature of centralized college admission systems (as operated, e.g., in Chile, Croatia, Norway, and Sweden) is that a student who intends to not comply with his school assignment can choose to drop out of the system or can accept the initial admission offer, but apply to and enrol in his preferred school in the following year(s). The former happens rarely as it typically means not enrolling in any college in a given year; the latter happens frequently and it delays graduation and labor market entry by at least a year. Hence, in centralized matching markets of this type, non-compliance costs arise naturally, at least for always takers, i.e., those ultimately enrolling in a given school regardless of the initial application outcome.

Our analysis implies that when the admission offer is used to instrument for graduation (as in, e.g., Kirkeboen et al., 2016), these non-compliance costs originate before treatment status (graduation) is resolved, and therefore bias the LATE estimator. A plausible strategy to solve the problem of non-compliance is to change the treatment of interest. In the context of school-program evaluation, instead of estimating the effect of graduation, this would correspond to estimating the effect of admission into the first year of a given program. This strategy may trade off gains in terms of identification credibility with economic relevance of the treatment effect. When the admission offer is used to instrument for admission (as in, e.g., Altmejd et al., 2019), non-compliance costs originate only after initial-application admission treatment status is determined. In this case, there is no bias since the instrument-treatment mapping occurs before costs are realized, and is thus unaffected by the non-compliance costs. Nevertheless, when studies in the literature interpret admission as extended attendance, the interpretation of the treatment effect is similarly impaired as in the case of graduation effects.

As a prime example of this literature, consider Kirkeboen et al. (2016), who estimate the returns to graduating in different fields of education in Norway by instrumenting for graduation with the initial quasi-random admission offer, and by measuring labor market returns eight years after the initial application. Enrolling in a program other than the one initially assigned a year or more after the initial application (we refer to such situation as 're-enrolling') results in deferred graduation and thus reduces labor market experience as labor market returns are measured eight years after the initial application regardless of the actual graduation date. This in turn implies costly non-compliance.[1] Therefore, according to the results provided here, the estimates in Kirkeboen et al. (2016) can be interpreted as returns to fields of study only if the costs of foregoing labor market experience are not field-specific and if the probability of non-compliance with the initial assignment does not depend on the initial assignment. Using data from the centralized college-student matching market in Croatia spanning the period from 2012 to 2018, we show this is not the case by documenting that the probabilities of non-compliance do depend on the initial assignment.

Using the Croatian data, we consider the same instrument as Kirkeboen et al. (2016) and document a sizeable re-application rate.[2] Importantly, the rate of applying to programs other than the one initially assigned within two years of the initial application (referred to as re-applying) for those just below the treatment program's admission score cutoff is 18.3% compared with 12% for those just above

[1] Non-compliance implies net costs if the negative effect of lower labor market experience outweighs the potential benefits of temporary enrolment in a non-preferred program. Providing evidence on this issue is beyond the scope of our analysis; for the purpose of our analysis, we assume the benefits are small.

[2] In order to re-enroll in the year(s) after the initial college application, a student needs to re-apply. Therefore, we analyze the re-application rate (intent to non-comply) and the re-enrollment rate (non-compliance) separately.

the treatment program's cutoff.[3] This discontinuity in the re-application rate at the cutoff translates into discontinuity in the non-compliance (re-enrollment) rate at the cutoff: there are 14.6% of non-compliers just below the cutoff, compared with 10.1% just above the cutoff. The higher share of non-compliers just below the cutoff compared with non-compliers just above the cutoff breaks the exclusion restriction, as the instrument now affects the outcome through channels other than the treatment assignment since it also affects the non-compliers due to non-compliance costs.

To deal with this issue, we propose a method inspired by Lee (2009), which recovers the treatment effect bounds under the *homogenous non-compliance costs* assumption, i.e., when all non-compliers pay the same cost. The method consists of two steps. The first involves trimming the data (excluding observations) until the non-compliance rates for those assigned and those not assigned to the treatment program are the same. For example, suppose that the fraction of always takers (those ultimately receiving the treatment regardless of the instrument's value from the initial application year) is larger than the fraction of never takers (those never getting the treatment) so that the fraction of non-compliers who were assigned not to get the treatment is disproportionally large. The first step of the proposed method balances the fraction of always takers assigned away from the treatment and the fraction of never takers assigned to the treatment by dropping a fraction of always takers assigned away from the treatment. Due to the homogenous non-compliance costs assumption, the effect of the non-compliance costs of the remaining always takers is then offset by (the same amount of) non-compliance costs of never takers.

However, excluding the non-compliers based on their treatment assignment and treatment indicator induces selection bias (as selection into non-complying can generally be non-random). By selectively excluding only the always takers who are not assigned to the treatment, the instrument now gains predictive power over outcomes of non-compliers—the probability of an individual being an always taker becomes higher for those assigned to the treatment in the trimmed sample, compared to the original sample, and always takers may have different outcomes than never takers. Therefore, the second step of the proposed procedure accounts for the sample selection by adapting the Lee (2009) treatment effect bounds, additionally trimming individuals in order to ensure that the instrument does not predict the outcome for non-compliers; in our case, this involves trimming individuals who were assigned to the treatment. The final ingredient of the method

[3]Following Dustan (2018), Fernandez (2019) and Kirkeboen et al. (2016), we define the *treatment program* for a particular applicant as the program for which he was close to the admission score cutoff, i.e., either just above the cutoff or just below the cutoff.

is to select individuals for trimming in each stage from the upper/lower tails of the outcome distribution in order to ensure the most conservative treatment effect bounds.

The homogenous costs assumption is plausible in the school choice setting since non-compliance costs here originate in large part in reduced labor market exerience due to re-enrolling in another program a year or more after the initial quasi-random assignment, and therefore postponing graduation. [4] Whether these costs are homogenous can be tested empirically by asking whether the slopes of experience wage profiles of always takers and never takers who did not comply with the treatment assignment are similar.

This paper contributes to several strands of the literature. First, it is relevant to the literature employing 2SLS-type estimators in centralized school-student matching markets, in which non-compliance costs arise naturally. Using 2SLS near admission cutoffs or the closely related regression discontinuity design (RDD) estimators, Kirkeboen et al. (2016) analyze school-specific labor-market returns, Lucas and Mbiti (2014) and Abdulkadiroglu et al. (2014) study school-specific attendance achievement effects (measured through standardized test scores), Kaufmann et al. (2013) study marriage market returns, while Dustan (2018), Fernandez (2019) and Altmejd et al. (2019) analyze the role of family ties in school choice. These applications are potentially affected by the non-compliance cost issue.

More generally, this approach can be applied in other empirical settings. For example, when programs are offered through a randomized list and applicants can apply to several lotteries (de Chaisemartin and Behaghel (2020)), or in college applications without matching markets (see e.g. Zimmerman, 2014, Goodman et al. (2017), Goodman et al. (2020) and Kozakowski (2020)).

Second, it adds to the literature on exclusion restriction violation. Heckman (1997) establishes that any selection into treatment based on individual-specific unobserved characteristics breaks the exclusion restriction and results in economically un-interesting parameters. Similarly, Jones (2015) identifies economically plausible potential violations of exclusion restriction for infra-marginal individuals (always takers and never takers) in cases where treatment may change their outcomes, which loosely fits our framework. However, Jones (2015) only constructs isolated theoretical examples, in which the exclusion restriction is likely violated, without presenting empirical content or developing a solution, while we develop a general non-compliance setup and tie it directly to a large literature. We also provide

[4] Postponing graduation could also produce certain gains (i.e maturation effect), or different types of costs (i.e. the (cognitive) costs of preparing and re-taking the state exam). In this paper, we interpret the *net costs* after "aggregating" all gains and costs, thus abstracting away from potential cost breakdowns.

an alternative estimator, which addresses the underlying issue. Moreover, in our setup the cost is generated endogenously to the IV model - by the decision of agents to not comply - and not by external spillovers of treatment assignment as in Jones (2015).

The remainder of this paper is structured as follows. In the next section, we develop the procedure for bounding the treatment effect in the case of costly non-compliance. In the third section, we demonstrate that the Croatian college-student matching market is subject to differing probabilities of complying depending on the college assignment. The fourth section concludes.

2. Treatment Effect Bounds

In this section, we develop a general framework that supports the assumption of costly non-compliance and analyzes the behavior of the LATE estimator. We use the typical LATE notation and introduce an additional parameter γ, which denotes non-compliance costs. As a result, we produce a practical framework that can be straightforwardly used in typical LATE applications.

Our illustrative empirical school-choice analysis presented in the next section is based on a dynamic setup where the costs of non-compliance are embodied in the time needed to alter the treatment assignment by re-enrolling at another school. A more complicated, structural model could attempt to elicit the gains (i.e., maturation effects) and losses (i.e., foregone labor-market experiences) from this non-compliance process. Our model collapses the net non-compliance costs into the parameter γ, and applies the newly developed LATE framework. Such an approach allows one to divide the analysis into two steps. First, to analyze the components of the non-compliance costs embodied in the parameter γ, and second, to analyze the LATE conditional on a specific value of the non-compliance costs γ.

We show that in presence of non-compliance costs, the exclusion restriction is likely violated, thus biasing the LATE estimator. We address this issue by developing a treatment-bounds method inspired by Lee (2009), and discuss the assumptions needed to recover treatment effect bounds.

Suppose we are interested in the causal effect of treatment D_i on the outcome y_i. Denote with Y_{1i} (Y_{0i}) potential outcomes of individual i when $D_i = 1$ ($D_i = 0$). An instrument $Z_i = \{0, 1\}$ (treatment assignment) is assumed to shift the treatment indicator D_i. In particular, denote with D_{1i} (D_{0i}) the treatment indicator of individual i when $Z_i = 1$ ($Z_i = 0$). The outcome of interest y_i is now indexed against two variables, the value of the treatment indicator D_i and the value of the instrument Z_i as $y_i = Y_i(D_i, Z_i)$.

Define an indicator t_i describing an individual i's type as:

$$t_i = \begin{cases} N \text{ if } D_{1i} = 0 \text{ and } D_{0i}{=}0 \ (\textit{Never taker}), \\ A \text{ if } D_{1i} = 1 \text{ and } D_{0i}{=}1 \ (\textit{Always taker}), \\ C \text{ if } D_{1i} = 1 \text{ and } D_{0i}{=}0 \ (\textit{Complier}), \\ D \text{ if } D_{1i} = 0 \text{ and } D_{0i}{=}1 \ (\textit{Defier}), \end{cases}$$

and denote with $P(t_i = x)$ the probability that individual i's type is x. The LATE theorem of Imbens and Angrist (1994) is widely used to identify local average treatment effects in (quasi-) experimental studies:

Theorem 1. *Assume the following LATE assumptions:*

- *Independence - The instrument is independent:*

$$\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_i(1), D_i(0)\} \perp Z_i$$

- *Exclusion restriction - The instrument affects the outcome only through the treatment indicator:*

$$Y_i(d, 0) = Y_i(d, 1) \equiv Y_{di} \textit{ for } d = 0, 1$$

- *First stage - The instrument has predictive power over assignment:*

$$E[D_{1i} - D_{0i}] \neq 0$$

- *Monotonicity - There are no defiers:*

$$D_{1i} - D_0 \geq 0 \textit{ or vice versa}, \forall i$$

Then, the Wald estimator equals the average treatment effect on the treated:

$$\frac{E[y_i|Z = 1] - E[y_i|Z = 0]}{E[D_i|Z = 1] - E[D_i|Z = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0]$$

Proof. See Imbens and Angrist (1994). ∎

Under the LATE assumptions, the Wald estimator equals the average treatment effect for compliers (individuals with $t_i = C$). Intuitively, non-compliers, i.e., always takers (those with $t_i = A$) and never takers (those with $t_i = N$), do not contribute to the IV estimator for two reasons. First, this is due to the exclusion restriction as the instrument does not change their treatment assignment. Second, this is due to the *independence* assumption, as the instrument is independent from their treatment decisions D_i. Therefore, the instrument has no predictive power over the outcomes of non-compliers.

In contrast, if non-compliance with the quasi-random treatment assignment is costly, non-compliers generally do contribute to the IV estimator of LATE. For example, if always takers with $Z = 0$ have to pay a cost to get treatment, they are no longer the same as the always takers with $Z = 1$ (the exclusion restriction does not hold). Generally, this implies predictive power of the instrument over the outcome for the non-compliers, which violates the assumptions of the LATE theorem.

Proposition 1. *Assume that Independence, First stage and Monotonicity assumptions from Theorem 1 hold and assume heterogenous non-compliance costs accross* t*:*

$$E[Y_i(1,1) - Y_i(1,0)] = \gamma_A \text{ and } E[Y_i(0,1) - Y_i(0,0)] = \gamma_N, \gamma_A \neq \gamma_N.$$

Let $\bar{\gamma} = \frac{\gamma_A + \gamma_N}{2}$*. The Wald estimator now equals:*

$$\begin{aligned}
\frac{E[y_i|Z=1] - E[y_i|Z=0]}{E[D_i|Z=1] - E[D_i|Z=0]} =& E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0] \\
&+ \frac{\bar{\gamma} \cdot (P(t_i = A) - P(t_i = N))}{P(t_i = C)} \\
&+ \frac{\frac{P(t_i=A)+P(t_i=N)}{2} \cdot (\gamma_A - \gamma_N)}{P(t_i = C)}.
\end{aligned}$$

Proof. Applying the Independence and Monotonicity assumption to the first term of the Wald estimator we obtain:

$$\begin{aligned}
E[y_i|Z=1] = & E[Y_{1i}|D_{1i} = 1, D_{0i} = 1] \cdot \overbrace{P[D_{1i} = 1, D_{0i} = 1]}^{\text{Always takers, } t_i = \text{A}} \\
&+ E[Y_{1i}|D_{1i} = 1, D_{0i} = 0] \cdot \overbrace{P[D_{1i} = 1, D_{0i} = 0]}^{\text{Compliers, } t_i = \text{C}} \\
&+ E[Y_{0i}|D_{1i} = 0, D_{0i} = 0] \cdot \overbrace{P[D_{1i} = 0, D_{0i} = 0]}^{\text{Never takers, } t_i = \text{N}}.
\end{aligned}$$

After performing an analogous decomposition of $E[y_i|Z=0]$, and using the *Heterogenous non-compliance costs* assumption, the numerator of the Wald estimator, after some algebra, becomes:

$$\begin{aligned}
E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0] + & \bar{\gamma} \cdot (P(t_i = A) - P(t_i = N)) \\
&+ \frac{P(t_i = A) - P(t_i = N)}{2} \cdot (\gamma_A - \gamma_N).
\end{aligned}$$

A similar argument shows that

$$E[D_i|Z=1] - E[D_i|Z=0] = E[D_{1i} - D_{0i}] = P[D_{1i} = 1, D_{0i} = 1]) = P[t_i = C].$$

■

Proposition 1 says that under costly non-compliance with the (quasi-) random treatment assignment, the Wald estimator equals the average treatment effect for compliers if the costs as well as the probabilities of non-compliance are the same for always takers and never takers (i.e., if $\gamma_A = \gamma_N$ and $P(t_i = A) = P(t_i = N)$).

In the remainder of this section we propose sharp bounds of the LATE[5] for the simple homogenous non-compliance costs case (i.e, $\gamma_A = \gamma_N$).[6] In the next section, we apply the bounding procedure to the Croatian centralized student-school matching system, arguing that in these types of settings assuming homogenous non-compliance costs may be reasonable.

At an intuitive level, the proposed bounding method mechanically equates the probabilities $P(t_i = A)$ and $P(t_i = N)$ by excluding individuals leading to the highest upper (lowest lower) bound. Suppose, WLOG, that $P(t_i = A) > P(t_i = N)$ - there are more always takers than never takers. Therefore, to calculate the upper LATE bound, we trim a proportion of always takers (individuals with $D = 1$ and $Z = 0$) until $P(t_i = A) = P(t_i = C)$, starting with those with the highest Y values (to obtain the highest possible value of the Wald estimator). This solves the problem of differing probabilities of non-compliance for always takers and never takers, but it also introduces a selection problem by selectively excluding always takers with $Z = 0$ values. Intuitively, in the new sample, individuals with $Z = 0$ are less likely to be always takers than individuals with $Z = 1$. Therefore, in addition to predicting treatment, the instrument now predicts the non-compliance types, and potentially also the outcome (if the selection into non-compliance is non-random), which breaks the exclusion restriction.

To account for this, we aim to drop the same number of always takers who were assigned to treatment (i.e., $Z = 1$). The problem is that among the individuals with $Z = 1$, we cannot distinguish compliers from always takers — both of them accept the treatment assignment. However, by trimming individuals with the lowest Y values (of those with $Z = 1$), we generate the upper LATE bound. This result is formalized in the following adaptation of Proposition 1 from Lee (2009).

Proposition 2. *Let Y be a continuous random variable. Assume that Independence, First stage and Monotonicity assumptions from Theorem 1 hold and assume Homogenous non-compliance costs:*

$$E[Y_i(1,1) - Y_i(1,0)] = \gamma = E[Y_i(0,1) - Y_i(0,0)]$$

Assume, WLOG, that $P(t_i = A) > P(t_i = N)$ and introduce $R = \frac{P(t_i=A)-P(t_i=N)}{P(t_i=A)}$. Next, set $y_{q|E} = G^{-1}(q)$, where G is the cdf of Y conditional on an event E, which defines the value of treatment D_i and instrument Z_i. Under these assumptions,

[5] The bounds are sharp in the Lee (2009) sense that they are the largest (smallest) lower (upper) LATE bounds consistent with the data.

[6] The homogenous costs assumption in the school choice setting is testable with data on labor market outcomes since the costs originate in large part in the reduced labor market experience due to re-enrolling in another program. One can test the equality of slopes of the experience profiles of always takers and never takers who did not comply with the initial treatment assignment by comparing their realized experience curves.

Δ_{LB} and Δ_{UB} are sharp lower and upper bounds for the average LATE effect $E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0]$:

$$\Delta_{LB} = \frac{E\left[Y|Z=1, Y \leq y_{1-R\cdot p(t_i=A)|Z=1}\right] - E\left[Y|(Z=0, D=0) \cup (Z=0, D=1, Y \geq y_{R|(Z=1,D=0)})\right]}{P_L(T_i = C)},$$

$$\Delta_{UB} = \frac{E\left[Y|Z=1, Y \geq y_{R\cdot p(t_i=A)|Z=1}\right] - E\left[Y|(Z=0, D=0) \cup (Z=1, D=0, Y \leq y_{1-R|(Z=1,D=0)})\right]}{P_U(T_i = C)}$$

and P_L (P_U) is a probability measure evaluated on the trimmed sample used when calculating Δ_{LB} (Δ_{UB}). The bounds are sharp in the sense that Δ_{LB} (Δ_{UB}) is the largest (smallest) lower (upper) bound that is consistent with the observed data.

Proof. First, draw a random proportion R of individuals with $Z = 0$ and $D = 1$ and assign them values $S_{0ir} = 0$, where r indexes the random seed generating this variable. Assign the remaining individuals with values $S_{0ir} = 1$. To simplify notation, assume that the variable $S_{1ir} = 1$ for each individual i and introduce:

$$\begin{aligned} S_{ir} =& S_{1ir}Z + S_{01r}(1 - Z) \\ Y_i^* =& S_{ir} \cdot \{Y_{1i}Z + Y_{0i}(1 - Z)\} \end{aligned} \tag{1}$$

Next, assume that the variable Y_i^* is only observed when $S_{ir} = 1$ and is, in that case, equal to Y_i. In other words, model 1 treats S_{ir} as a sample selection indicator. Denote with Y_{1i}^* (Y_{0i}^*) the outcome of the individual i when $Z_i = 1$ ($Z_i = 0$). According to the Lee (2009) theorem, the sharp lower ($\Delta_{LB,r}$) and upper ($\Delta_{UB,r}$) bounds for the intention to treat estimator ($E[Y_i|Z = 1, S_{1i} = 1, S_{0i} = 1] - E[Y_i|Z = 0, S_{1i} = 1, S_{0i} = 1]$) are:

$$\begin{aligned} \Delta_{LB,r} &= E\left[Y|Z=1, S=1, Y^* \leq y^*_{1-R\cdot p(t_i=A)}\right] - E\left[Y|Z=0, S=1\right], \\ \Delta_{UB,r} &= E\left[Y|Z=1, S=1, Y^* \geq y^*_{R\cdot p(t_i=A)}\right] - E\left[Y|Z=0, S=1\right]. \end{aligned}$$

We index the bounds with r to emphasize the dependence on the seed corresponding to the random draw of R individuals.

Note that on the sample of individuals with $S_{ir} = 1$, $P(t_i = A) = P(t_i = N)$. Therefore, according to Proposition 1:

$$\frac{\Delta_{LB,r}}{P_L(T_i = C)} \leq E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0, S_{1ir} = 1, S_{0ir} = 1] \leq \frac{\Delta_{UB,r}}{P_U(T_i = C)}$$

$$\Longrightarrow$$

$$\min_r \left(\frac{\Delta_{LB,r}}{P_L(T_i = C)}\right) \leq E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0] \leq \max_r \left(\frac{\Delta_{UB,r}}{P_U(T_i = C)}\right),$$

where P_L (P_U) is the probability evaluated on the trimmed sample used when calculating Δ_{LB} (Δ_{UB}). Finally, note that the treatment bounds depend on the random draw R only through the outcome values of randomly sampled individuals with $Z = 1$ and $D = 0$ (i.e., they do not the depend on the randomly sampled subset of those with $D = 1$). The proposition now follows from observing that, for example, the lowest $\Delta_{LB,r}$ is achieved when trimming those individuals with $Z = 1$ and $D = 0$ who have the highest y values. ■

To demonstrate the value of Proposition 2, we compare the 2SLS estimator to the proposed LATE bounds on a simulated dataset. We generate N individuals according to the following steps:[7]

- The type of individual i is drawn from the following distribution:

$$t_i = \begin{cases} A \text{ with probability } p_a, \\ N \text{ with probability } p_n, \\ C \text{ with probability } 1 - p_a - p_n. \end{cases} \tag{2}$$

- The treatment assignment Z_i is a Bernoulli random variable with parameter 0.5.

- The outcome of interest y_i is defined as:

$$y_i = \epsilon - \gamma \cdot 1_{Z_i \neq D_i},$$

  where ϵ is $N(0, 1)$.

The procedure generates a population with no treatment effect (i.e., the treatment effect is zero) where assignment to treatment is equiprobable for each individual. Individuals differ only with respect to their type, which defines their attitude towards treatment assignment (i.e., $t_i = A$ individuals always get treated regardless of the assignment status, $t_i = N$ never get treated and $t_i = C$ comply with the assignment).

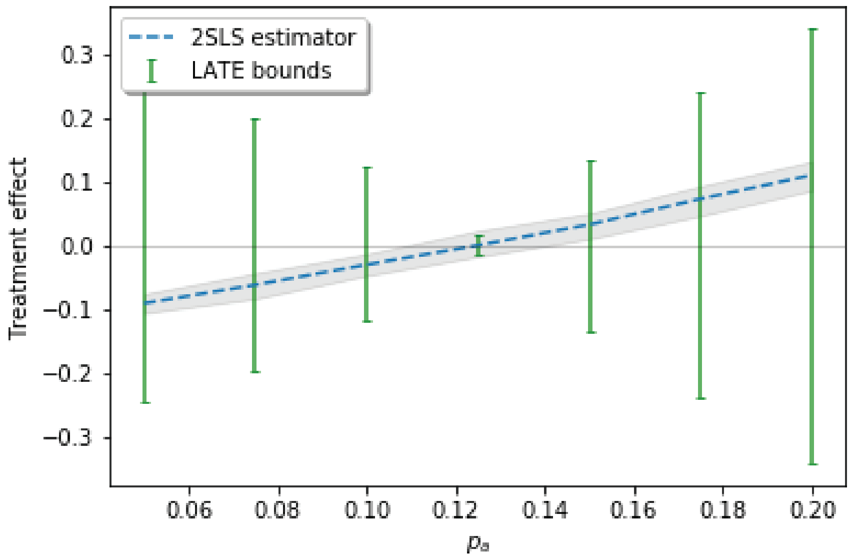
We conduct two exercises: First asking about the performance of the 2SLS estimator and of the treatment-effect bounds under fixed costs of non-compliance and varying gaps between p_a and p_n, second allowing the cost of non-compliance to vary but keeping non-compliance probabilities fixed. Specifically, in the first exercise we set the non-compliance cost to equal the outcome standard deviation ($\gamma = 1$). Next, we set p_n at 0.12[8] and for each value $p_a \in \{0.05, 0.075, \ldots, 0.175, 0.2\}$, we generate 1,000 independent populations and plot the average LATE bounds and the average of the 2SLS estimates in Figure 1. Even though the treatment effect is 0 by construction, the 2SLS estimator reflects the asymptotic bias $\frac{\gamma \cdot (p_a - p_n)}{1 - p_a - p_n}$ and would lead one to reject the zero treatment effect even for small differences between

[7] N is set at 50,000 to resemble Kirkeboen et al. (2016), where $N = 50,083$.

[8] This probability correspond to an empirical estimate obtained in section 3.

p_a and p_n, while the LATE bounds correctly include 0 and remain smaller than one half of the treatment standard deviation even for large differences between p_a and p_n.

Figure 1: LATE bounds vs. 2SLS estimates - varying non-compliance probabilities



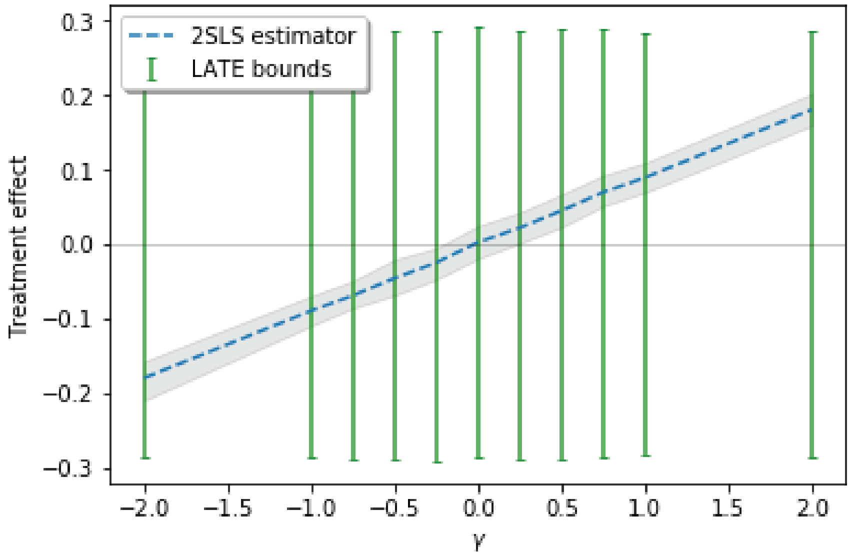
Note: The figure plots 2SLS estimates and LATE bounds (y-axis) against the probability (of being an always taker) p_a, holding the probability (of being a never taker) p_n fixed at 0.125 on a series of simulated datasets. For each parameter value $p_a \in \{0.05, 0.075, \dots, 0.175, 0.2\}$ we generate 1,000 independent populations using parameters $\gamma = 1, N = 50000$, $p_n = 0.125$ under no treatment effect (LATE= 0), and plot the average LATE bounds from Proposition 2 and the average 2SLS estimates as well as the corresponding average 95% confidence intervals.

In the second exercise presented in Figure 2, we vary non-compliance costs γ while holding p_a fixed at 18.3% and p_n at 12%.[9] Again, the 2SLS estimator coincides with its asymptotic bias and reports significant estimates even for reasonably small values of γ, while the LATE bounds correctly include 0.[10]

[9] Again, these probabilities correspond to empirical estimates obtained in section 3.

[10] The LATE bounds do not depend on the γ value, since they neutralize the effect of the non-compliance cost by trimming enough individuals so that the costs of always takers and never takers cancel.

Figure 2: LATE bounds vs. 2SLS estimates - varying γ



Note: The figure plots 2SLS estimates and LATE bounds (y-axis) against the costs of non-compliance γ, while holding fixed the probability (of being a never taker) $p_n = 0.12$ as well as the probability (of being an always taker) $p_a = 0.183$ on a series of simulated datasets. For each parameter value $\gamma \in \{-2, -1, 5, \ldots, 2\}$ we generate 1,000 independent populations using parameters $p_a = 0.183, p_n = 0.12, N = 50000$ under no treatment effect (LATE= 0), and plot the average LATE bounds from Proposition 2 and the average 2SLS estimates as well as the corresponding average 95% confidence intervals.

3. Empirical Application to Croatian College Matching Market

In a recent study, Kirkeboen et al. (2016) use RDD to instrument for graduation and estimate returns to graduating in different fields of education in Norway by instrumenting for the graduation with the initial admission offer and measuring labor market returns eight years after the initial college application. In such a setup, re-enrolling in another field, potentially years after the initial application, results in deferred graduation and reduces labor market experience. In the likely case that the length of labor market experience affects labor market returns, Proposition 1 implies that Kirkeboen et al. (2016) identify unbiased returns to fields only in the homogenous non-compliance costs case and if the probabilities of non-compliance do not depend on the initial treatment-program assignment. In this section, we show that the latter is not the case in Croatia, where the probabilities of non-compliance do depend on the initial assignment. Therefore, the LATE bounds from Proposition 2 should be applied when estimating LATE effects in the Croatian matching market. A similar issue arises naturally in studies that rely on quasi-random admission offers to instrument for graduation or other outcomes occuring years after the initial offer of admission (e.g. Hastings et al., 2014; Lucas and Mbiti, 2014; Abdulkadiroglu et al., 2014; Kaufmann et al., 2013; Dustan, 2018; Fernandez, 2019; Altmejd et al., 2019).

We begin the section with a brief summary of the estimation strategy employed in Kirkeboen et al. (2016) and similar student-school assignment studies. We proceed with a note on the institutional setup in Croatia, followed by a subsection rejecting equal probabilities of non-compliance for students who were or were not (quasi-randomly) offered admission to their treatment program (i.e., the program where they were just below or just above the program-specific admission cutoff). We conclude the section with a discussion of the homogeneous non-compliance costs assumption.

3.1. Empirical Strategy

The literature studying school-student centralized matching markets frequently exploits a feature of these systems in which students with similar admission scores in a neighborhood of a school's admission threshold are or are not offered admission to the schools based on small differences in admission scores. Taking advantage of these discontinuities, the literature typically uses regression discontinuity design (RDD) to instrument for admission/graduation , assuming that students around the cutoff are 'the same' in every aspect except the assigned school (program). The assigned school is assumed to be deterministically linked to the school-specific admission score (i.e., a student is offered admission if and only if his admission score is above the school-specific admission score cutoff). For schools ranked below the assigned school, this deterministic link between admission score and the assignment is broken — the student is never considered for admission even if he is above the cutoff for these schools. For this reason, applications to these schools are not included in the *RDD estimation sample*, which consists of applications at the margin of admission, i.e., within a bandwidth neighbourhood of school-specific admission score cutoffs.

More formaly, let c_{jt} be the admission score cutoff of program j in year t. If program j is ranked above program j' in student i's preference list, we write $(j) \succ_i (j')$. Denote the school-j-specific application score of individual i as a_{ijt}. Student i's application to program j belongs to the RDD estimation sample if student i:

(i) listed program j as his choice, such that all programs preferred to j had a higher cutoff score than c_{jt} (otherwise assignment to j is impossible):
$c_{jt} < c_{j't} \ \forall \ (j') \succ_i (j)$,

(ii) had a score a_{ijt} sufficiently close to j's cutoff score to be within a given bandwidth bw around the cutoff:
$|a_{ijt} - c_{jt}| \leq bw$.

The following regression, applied to the RDD estimation sample, is a typical specification used in the school-choice literature to estimate various graduation effects:

$$y_{i\tau} = \beta \cdot graduated_{ijt} + f(a_{ijt};\gamma) + \mu_\tau + \mu_{jt} + \varepsilon_{ijt} \tag{3}$$

where $y_{i\tau}$ is the outcome of interest measured at time $\tau > t$ of student i who was near the program j admission cutoff in year t, $graduated_{ijt}$ is an indicator variable that takes value 1 if student i graduated from program j where he initally applied in year t, $f(a_{ijt};\gamma)$ is a function of the application score of student i for program j in year t, μ_{jt} and $\mu_\tau$are fixed effects corresponding to application year-program combinations and outcome years, respectively, and where ε_{ijt} is an error term. Since $graduated_{ijt}$ is likely influenced by various unobserved, potentially endogenous factors, researchers typically use admission offer $(1_{a_{ijt} \geq c_{jt}})$ to instrument for graduation. In the language of the previous section, being just above the cutoff corresponds to the instrument value $Z = 1$, and being just below the cutoff corresponds to the instrument value $Z = 0$.

3.2. Institutional Setup

In Croatia, admissions to all college programs are implemented through a national online platform. Since its introduction in 2010, this platform operates a deferred acceptance (DA) algorithm that ranks students based on their high-school grades and subject-specific elective national-level exams that take place in June, a month after high-school graduation. Students register on the platform in early spring of their high-school graduation year when universities also list on the platform their program-specific admission quotas along with program-specific weights of subject-specific grades and exams. Students are free to submit their ranked priority lists of up to 10 programs as of registration and update these preference rankings until the system closes for clearing at a predetermined date in mid-July (i.e., in 2019, the final deadline was 2 pm on July 15$^{\text{th}}$).

Students first receive information on their position in various admission queues one week before the final deadline, immediately after receiving their admission scores. Admission scores, which are a function of student's high school grades and national exam scores, are the only factor determining admission rankings. The DA algorithm is then regularly updated to show students their current admission rankings. Students update their preference rankings continuously until the system closes for clearing in mid-July.

During the application period applicants often drop their previously highly ranked alternatives they are unlikely to get admitted to.[11] Therefore, in order to study a case similar to the typical centralized college admission system, where students are not able to get signals on the current school-specific demand, we analyze admission outcomes implied by the first preference submissions after receiving the national exam scores (5 days before the system closes), when students are fully aware of their admission scores but are not yet able to learn about the market demand structure. We thus consider that a student applied to a particular program if this program was on the student's preference list five days before the admission deadline.

In centralized college admission systems, it is not feasible for always takers to not comply with their initial assignment out of their treatment program within the year of initial application. They can, however, apply to their preferred program in the following years. Further, in Croatia, there is only limited scope for never takers to not comply with their initial-application assignment to their treatment program.[12] Therefore, since we do not observe enrollment, we assume that the

[11] Due to the dynamic nature of the admission system, students can get hourly updates about their admission rankings, and therefore resolve a significant part of the admission uncertainty. They can do this only after they receive their admission scores, approximately 1 week before the admission deadline.

[12] According to the Ministry of Science and Education, 95% of Croatian college applicants comply with their DA admission assignment, enrolling at their assigned program. If they decide not to comply, they lose their tuition waiver, otherwise covered by the Ministry. This introduces an additional (constant) cost of non-compliance.

final admission offer translates to enrollment one-to-one. Hence, we abstract from non-compliance within the year of initial application and focus on non-compliance through re-applications in years following the initial college application.

In sum, we analyze applications (based on the ranking lists submitted 5 days before the system closes) which resemble the applications at typical centralized college admission systems, and enrollments (based on the final ranking lists) separately. We consider that a student re-applied (attempted non-compliance) if he applied to a program different from the one initially assigned in the two years following the initial application year. While we observe re-applications, we do not observe re-enrollment, so again, we assume that a re-applying student always re-enrolled in a particular program if this program was his final DA admission assignment.

3.3. Data and Results

We use complete administrative data corresponding to the Croatian centralized college admission system from years 2012-2018. In these data, we consider a student to be a non-complier if, following a re-application, he was assigned by the DA algorithm to a program different from the one initially assigned at most two years after his initial college enrollment. As the re-application window is two years, we exclude the boundary years of the data[13] and generate the RDD estimation sample using applications from 2014-2016 that are at most 0.4 standard deviations (60 admission score points) away from program-specific admission cutoffs.[14]

Table 1 shows basic summary statistics for the Croatian DA matching market and the RDD estimation sample defined by a 60 admission score points bandwidth, throughout 2014-2016. Annually, approximately 35,000 students enter the system, choosing between about 600 programs belonging to 49 distinct universities. The RDD estimation sample appears to have similar average characteristics to the unrestricted sample.

Using the RDD estimation sample, we estimate the following regression:

$$y_i = \alpha_0 \cdot a_{ij} + \alpha_1 \cdot a_{ij} 1_{a_{ij} \geq c_j} + \delta \cdot 1_{a_{ij} \geq c_j} + f(a_{ij}) + \mu_j + \varepsilon_i, \quad (4)$$

where y_i is a non-compliance indicator for applicant i (i.e., a dummy variable taking

[13] We exclude the first two years to ensure that we work with only initial college applicants who have not applied in previous years. We exclude the last two years to observe re-applications following initial applications.

[14] Each cutoff is defined as the admission score of the applicant with the lowest admission score who was offered admission. The optimal bandwidth according to Imbens and Kalyanaraman (2012) is 60 admission points. We replicated the analysis for numerous bandwidth values (10, 20, 30, 40, 50, 60, 70, 80, 90 and 100) and obtained similar results.

the value 1 if the applicant i re-enrolled into a program different from the initially quasi-randomly assigned program within two years of his initial enrollment), a_{ij} is the initial-application admission score of applicant i at program j, c_j is the cutoff of program j, $f(a_{ij})$ is a polynomial in admission scores, and μ_j are program fixed effects. The time index, which should denote the year of the applicant's first (initial) college application, is surpressed. We study not only re-enrollment, but also re-application (non-compliance intent) by estimating a version of regression (4) with the dependent variable y_i indicating if applicant i re-applied after his initial enrollment. These regressions are also estimated on subsamples where program j is (or is not) the applicant's first priority, and where the applicant re-applies to program j (or not). A significant estimate of δ is interpreted as evidence that the probabilities of non-complying (re-applying) depend on the initial assignment.

The first column of Table 2 shows statistically as well as economically significant estimates of δ both when considering re-application (-6.2 p.p. compared to the baseline of 18.3%) and re-enrollment (-4.5 p.p. compared to the baseline of 14.6%).[15] Hence, there are 14.6% of non-compliers just below the admission cutoff, compared with 10.1% just above the cutoff. Approximately half of these non-compliance gaps is attributable to students who re-apply to the same program after they were marginally declined at their initial application (i.e., always takers). The effects are most pronounced when students are around the cutoff at their initial-application-ranking top-priority program (-7.9 p.p. when considering re-applications and -6.4 p.p. when considering re-enrollment). These results can also be seen in Figure 3 (Figure 4), which plots the re-application (re-enrollment) probability against the application score distance from the initial-application cutoff.[16]

In sum, being just below the admission score cutoff of a program during one's initial college application disproportionately incentivizes students to re-apply, and subsequently re-enroll, relative to students just above an initial-application program cutoff. If Croatian students are subject to non-complying (re-application and re-enrollment) costs, Proposition 1 implies that RDD induced estimates cannot be interpreted simply as graduation treatment effects.

3.4. Discussion

In the Croatian case, the probabilities of non-compliance for applicants just above the cutoff ($Z = 1$) are significantly lower (4.5 p.p.) than for those just below the cutoff ($Z = 0$). This, according to Proposition 1 violates the LATE theorem,

[15] On average, around 70% of the re-applying students succeed in changing their initial school assignment, such that the re-application effects largely translate into re-enrollment effects.

[16] The distance from cutoff is defined as admission score centered around the cutoff.

invoked in, e.g., Kirkeboen et al. (2016). In order to apply the LATE bounds from Proposition 2, one needs to assume the homogenous non-compliance costs assumption. In our case, the costs of non-compliance originate in the reduced labor market experience due to re-enrolling in another program.[17] For example, an always taker with $Z = 1$ is expected to graduate five years after admission, while an always taker with $Z = 0$ is going to use at least one additional year due to re-enrollment. Therefore, the homogenous costs assumption translates into assuming equal slopes of the experience wage profiles of always takers and never takers who did not comply with the treatment assignment—this can be tested empirically by directly comparing experience profiles of always takers and never takers who did not comply. If the gradients of these experience profiles are not the same, one can assume that the experience profile is multiplicative, and perform the same test using the logarithm of returns (or some other transformation of the outcome variable)

4. Conclusion

In this paper, we consider a quasi-experimental intention-to-treat setup where non-compliance with treatment assignment is costly (affects the outcome), which violates the exclusion restriction — one of the crucial LATE assumptions. We generalize the LATE theorem to include the case of costly non-compliance and show that the IV estimator can be interpreted as LATE only under the strong assumption that both the costs and the probability of non-compliance do not depend on treatment assignment. We recover treatment effect bounds with an alternative method, inspired by Lee (2009), under the homogenous non-compliance costs assumption, i.e., if the costs do not depend on the initial assignment.

To illustrate the relevance of this design, we consider the recent study by Kirkeboen et al. (2016), who estimate returns to graduating in different fields of education in Norway by instrumenting for graduation with the initial (random) admission offer and measuring labor market returns eight years after the initial application. In such a setup, re-enrolling in another field year(s) after the initial application results in deferred graduation, which reduces labor market experience (as labor market returns are measured eight years after the initial application regardless of the actual graduation date). In the likely case that the length of the labor market experience affects labor market returns, the estimates in Kirkeboen et al. (2016) can be interpreted as returns to fields of study only if the cost of

[17] If Croatian students re-enroll, they also lose their national-level tuition waiver (otherwise covered by the Ministry of Science and Education), which is constant (homogenous) across programs.

foregoing labor market experience is not field-specific and if the probabilities of non-compliance do not depend on the initial assignment. We show that the latter is not the case in Croatia, using data on the Croatian student-college matching market from 2012 to 2018, where both the probability of non-compliance (re-enrollment) and the probability of re-taking the national exam (re-application) do depend on the initial assignment.

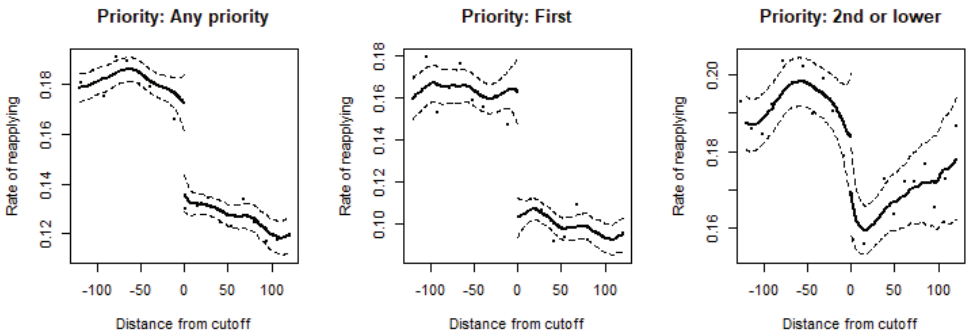
It is reasonable to assume that in the school-college matching market framework, non-compliance with the initial assignment comes at a cost. Not only does it likely imply deferred graduation, but, as demonstrated in the case of Croatia, it also often results in retaking the national exam which is, potentially, also costly (in terms of the cognitive costs of preparation).

The bounding method developed in this paper can be applied in other empirical settings where non-compliance costs arise. For example, when programs are offered through randomized list and applicants can apply to several lotteries (de Chaisemartin and Behaghel (2020)), or in college applications without matching markets (see e.g. Zimmerman, 2014, Goodman et al. (2017), Goodman et al. (2020) and Kozakowski (2020)).

Therefore, our analysis suggests that RDD based IV studies relying on centralized student-school matching markets should first test whether the probabilities of non-compliance with treatment assignment depend on the assignment. If treatment assignment does affect the probability of non-compliance, and if the homogenous costs assumption is not rejected, we suggest employing sharp LATE bounds.

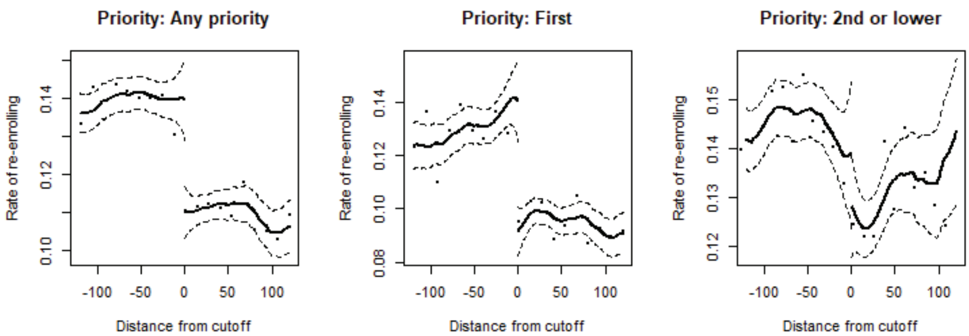
5. Appendix - Tables and Figures

Figure 3: Re-application probability at the initial-application admission cutoff



Notes: The graphs show re-application probabilities, defined using a two-year window following on the initial-application year, around the admission cutoff in the initial application year. The bandwidths used for the local polynomials correspond to optimal bandwidths according to Imbens and Kalyanaraman (2012). The three graphs show cases where the cutoff school (the school where an applicant was near the school admission cutoff at the initial application) was anywhere on the student's ranked choice list, when it was the student's first priority, and when it was his lower-ranked priority, respectively.

Figure 4: Re-enrollment probability at the initial-application admission cutoff



Notes: The graphs show re-enrollment probabilities, defined using a two-year window following on the initial-application year, around the admission cutoffs in the initial application year. The bandwidths used for the local polynomials correspond to optimal bandwidths according to Imbens and Kalyanaraman (2012). The three graphs show cases where the cutoff school (the school where an applicant was near the school admission cutoff at the initial application) was anywhere on the student's ranked choice list, when it was the student's first priority, and when it was his lower-ranked priority, respectively.

Table 1: Summary statistics

	All data (1)	RDD estimation sample (2)
Number of programs	620	620
Number of universities	49	49
Number of applicants	101,484	22,383
Number of applications	571,354	80,702
Average admission score	634.53	619.19
	(122.76)	(143.98)
Average GPA	4.01	3.96
	(0.62)	(0.58)
Fraction male	0.47	0.45
Average re-applying rate	0.13	0.16
	(0.33)	(0.36)
Average re-enrolling rate	0.10	0.13
	(0.31)	(0.34)

Notes: The table presents summary statistics calculated for the entire administrative dataset and for the RDD estimation sample (based on the bandwidth of 60 admission score points corresponding to 0.5 of standard deviations, calculated on the ranking lists reported 5 days before the final admission deadline). Standard errors are in the parentheses. The first panel shows the number of programs, universities, applicants, and applications. The second panel shows the average admission score calculated over all applicant-program pairs and the average GPA and fraction male calculated over all applicants. The third panel shows the rates of re-applying and re-enrolling (within a two-year window after the initial-application year) calculated over applicant-program pairs.

Table 2: Probability of re-application and re-enrollment

	Cutoff program : Any priority			Cutoff program : 1^{st} priority			Cutoff program : 2^{nd} or lower priority		
	Any program	Same program	Different program	Any program	Same program	Different program	Any program	Same program	Different program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A - Probability of re-applying									
Admission offer	-0.062^{***}	-0.036^{***}	-0.025^{***}	-0.079^{***}	-0.050^{***}	-0.028^{***}	-0.043^{***}	-0.018^{***}	-0.025^{**}
	(0.008)	(0.003)	(0.007)	(0.009)	(0.004)	(0.009)	(0.013)	(0.003)	(0.013)
Observations	59,495	59,495	59,495	28,966	28,966	28,966	30,529	30,529	30,529
Baseline	0.183	0.036	0.148	0.178	0.054	0.123	0.192	0.016	0.175
	(0.005)	(0.002)	(0.005)	(0.007)	(0.003)	(0.006)	(0.007)	(0.002)	(0.007)
Panel B - Probability of re-enrolling									
Admission offer	-0.045^{***}	-0.029^{***}	-0.016^{**}	-0.064^{***}	-0.041^{***}	-0.023^{***}	-0.024^{**}	-0.013^{***}	-0.011
	(0.007)	(0.002)	(0.007)	(0.009)	(0.004)	(0.008)	(0.012)	(0.003)	(0.011)
Observations	59,495	59,495	58,216	28,966	28,966	28,345	30,529	30,529	29,871
Baseline	0.146	0.030	0.116	0.153	0.045	0.108	0.140	0.014	0.126
	(0.004)	(0.002)	(0.004)	(0.006)	(0.003)	(0.006)	(0.006)	(0.002)	(0.006)
Program FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents RDD estimates corresponding to equation (4) of the effect of students' marginal admission to a program (in the initial application year) on the probability of re-application (Panel A) and re-enrollment (Panel B) in the following two years. The first three columns of the table consider marginal admissions to all programs, the middle three columns focus on marginal admissions to programs ranked as top priority on students' school choice lists, and the last three columns focus on lower-ranked programs from students' ranked choice lists. For each of these specifications, we consider separately re-applying/re-enrollment to any program, to the 'cutoff program', i.e. the program where in their initial-application year they were near the program's admission score cutoff, and to a program other than the cutoff program. All specifications use bandwidths calculated according to the Imbens and Kalyanaraman (2012) optimal bandwidth procedure. All specifications control for a local quadratic polynomial in students' admission score centered around program admission cutoffs. Application year fixed effects and program fixed effects are also used in all specifications. A triangular kernel is used to give more weight to observations close to the cutoffs.
*p-value<0.1 **p-value<0.05 ***p-value<0.01.

References

Abdulkadiroglu, A., Angrist, J., and Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, pages 137–196.

Altmejd, A., Fernandez, A. B., Drlje, M., Kovac, D., and Neilson, C. (2019). Siblings Effects on University and Major Choice: Evidence from Chile and Croatia. *Princeton University Industrial Relations Section Working Paper Series N633r*.

Angrist, J. D. and Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings. *The Quarterly Journal of Economics*, 106:979–1014.

de Chaisemartin, C. and Behaghel, L. (2020). Estimating the Effect of Treatments Allocated by Randomized Waiting Lists. *Econometrica*, 88:1453–1477.

Dustan, A. (2018). Family networks and school choice. *Journal of Development Economics*.

Fernandez, A. B. (2019). Should I Stay or Should I go? Neighbor's Effects on University Enrollment. *LSE Working paper*.

Goodman, J., Hurwitz, M., and Smith, J. (2017). Access to 4-Year Public Colleges and Degree Completion. *Journal of Labor Economics*, 35.

Goodman, J., Hurwitz, M., and Smith, J. (2020). The Economic Impact of Access to Public Four-Year Colleges. *NBER Working Paper No. w27177*, 35.

Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2014). Are Some Degrees Worth More than Others? Evidence From College Admission Cutoffs in Chile. *NBER Working Paper Series*.

Heckman, J. (1997). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *Journal of Human Resources*, (32):441–462.

Imbens, G. and Kalyanaraman, K. (2012). Breaking Ties: Regression Discontinuity Design Meets Market Design. *The Review of Economic Studies*, 79.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62:467–475.

Jones, D. (2015). The Economics of Exclusion Restrictions in IV Models. *NBER Working Paper*.

Kaufmann, K. M., Messner, M., and Solis, A. (2013). Returns to Elite Higher Education in the Marriage Market: Evidence from Chile. *NBER Working Paper*.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of Study, Earnings and Self-selection. *Quarterly Journal of Econometrics*, 131:1057–1111.

Kozakowski, W. (2020). Are Public Four-year Colleges Engines for Mobility? Evidence from Statewide Admissions Thresholds. *Working Paper*.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, (76):1071–1102.

Lucas, A. M. and Mbiti, I. M. (2014). Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya. *American Economic Journal: Applied Economics*, pages 234–263.

Zimmerman, S. D. (2014). The Returns to College Admissions for Academically Marginal Students. *Journal of Labor Economics*, 32.

Abstrakt

V několika nedávných studiích se odhadují kauzální dopady různých typů vzdělání pomocí dat z centrálních systémů, které přiřazují studenty ke školám na základě tzv. algoritmu odložených přijetí. Tyto systémy totiž umožňují porovnávat uchazeče o studium s podobným skóre (numerickým výsledkem testů) v přijímacím řízení, kteří jsou nebo nejsou přijati ke studiu na danou školu na základě velmi malých, kvazi-náhodných rozdílů v tomto skóre. Tato literatura pak předpokládá, že tito marginální studenti se liší pouze v přiřazení ke škole a využívá skóre jako instrumentální proměnné pro přijetí do programu nebo absolvování daného vzdělávacího programu. V tomto článku upozorňujeme, že nedodržení přiřazení vytvořeného v těchto systémech většinou odpovídá situaci, kdy se daný uchazeč dostane na danou školu, kam se na základě původního přiřazení marginálně nedostal, o rok později. Takové nedodržení původního přiřazení vytváří náklady, které narušují restrikci na výlučnost (exclusion restriction), tj. klíčový předpoklad LATE teorému, což vede ke zkresleným odhadům v případě, že se instrumentuje absolvování daného programu, tj. intervence, která probíhá po té, co jsou náklady nedodržení přiřazení realizovány. S pomocí dat z národního párovacího systému v Chorvatsku ukazujeme empirickou relevanci tohoto problému a nabízíme metodu inspirovanou prací Lee (2009), která odhaduje meze kauzálních efektů na základě předpokladu, že náklady nedodržení nejsou provázány s původním přiřazením.

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the CAS, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

Published by
Charles University, Center for Economic Research and Graduate Education (CERGE)
and
Economics Institute of the CAS, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

Editor: Byeongju Jeong

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-493-9 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-582-9 (Národohospodářský ústav AV ČR, v. v. i.)