



Columbia University

*Department of Economics
Discussion Paper Series*

**Exam Schools, Ability, and the Effects of Affirmative Action:
Latent Factor Extrapolation in the Regression Discontinuity
Design**

Miikka Rokkanen

Discussion Paper No.: 1415-03

*Department of Economics
Columbia University
New York, NY 10027*

March 2015

Exam Schools, Ability, and the Effects of Affirmative Action: Latent Factor Extrapolation in the Regression Discontinuity Design*

Miikka Rokkanen[†]

March 26, 2015

Abstract

Selective school admissions give rise to a Regression Discontinuity (RD) design that non-parametrically identifies causal effects for marginal applicants. Without stronger assumptions nothing can be said about causal effects for inframarginal applicants. Estimates of causal effects for inframarginal applicants are valuable for many policy questions, such as affirmative action, that substantially alter admissions cutoffs. This paper develops a latent factor-based approach to RD extrapolation that is then used to estimate effects of Boston exam schools away from admissions cutoffs. Achievement gains from Boston exam schools are larger for applicants with lower English and Math abilities. I also use the model to predict the effects of introducing either minority or socioeconomic preferences in exam school admissions. Affirmative action has modest average effects on achievement, while increasing the achievement of the applicants who gain access to exam schools as a result.

Keywords: Affirmative Action, Extrapolation, Latent Factor, Regression Discontinuity Design, Selective School

*I would like to thank Isaiah Andrews, Joshua Angrist, David Autor, Alex Bartik, Erich Battistin, Stephane Bonhomme, Victor Chernozhukov, Clement de Chaisemartin, Yingying Dong, Kirill Evdokimov, Benjamin Feigenberg, Eliza Forsythe, Jerry Hausman, Guido Imbens, Patrick Kline, Liisa Laine, Bradley Larsen, Anna Mikusheva, Conrad Miller, Yusuke Narita, Whitney Newey, Juan Passadore, Parag Pathak, Adam Sacarny, Matti Sarvimäki, Annalisa Scognamiglio, Henry Swift, Roope Uusitalo, Chris Walters and Heidi Williams as well as the seminar participants at Boston University, Columbia University, Dartmouth College, Federal Reserve Bank of New York, Helsinki Center of Economic Research, INSEAD, Institute for International Economic Studies, London School of Economics, Massachusetts Institute of Technology, Michigan State University, Norwegian School of Economics, Teachers College, UC San Diego, University of Warwick, University of Western Ontario, and Uppsala University for their helpful comments and suggestions. I would also like to thank Kamal Chavda and the Boston Public Schools for graciously sharing data.

[†]Columbia University, Department of Economics, 420 West 118th Street, New York, NY 10027. Email: mr3454@columbia.edu.

1 Introduction

Regression Discontinuity (RD) methods identify treatment effects for individuals at the cutoff value determining treatment assignment under relatively mild assumptions. Without stronger assumptions, however, nothing can be said about treatment effects for individuals away from the cutoff. Such effects may be valuable for predicting the effects of policies that change treatment assignments of a broader group. An important example of this are affirmative action policies that change cutoffs substantially.

Motivated by affirmative action considerations, this paper develops a strategy for the identification and estimation of causal effects for inframarginal applicants to Boston’s selective high schools, known as exam schools. The exam schools, spanning grades 7-12, are seen as the flagship of the Boston Public Schools (BPS) system. They offer higher-achieving peers and an advanced curriculum. Admissions to these schools are based on Grade Point Average (GPA) and the Independent School Entrance Exam (ISEE). The RD design generated by exam school admissions nonparametrically identifies causal effects of exam school attendance for marginal applicants at admissions cutoffs. Abdulkadiroglu, Angrist, and Pathak (2014) use this strategy and find little evidence of effects for these applicants.¹ Other applicants, however, may benefit or suffer as a consequence of exam school attendance.

Treatment effects away from RD cutoffs are especially important for discussions of affirmative action at exam schools. Boston exam schools have played an important role in the history of attempts to ameliorate racial imbalances in Boston. In 1974 a federal court ruling introduced the use of minority preferences in Boston exam school admissions as part of a city-wide desegregation plan. Court challenges later led Boston to drop racial preferences. Similarly, Chicago switched from minority to socioeconomic preferences in exam school admissions following a federal court ruling in 2009.²

This paper develops a latent factor-based approach to the identification and estimation of treatment effects away from the cutoff. I assume that the source of omitted variables bias in an RD design can be modeled using latent factors. The running variable is one of a number of noisy measures of these factors. Assuming other noisy measures are available, causal effects for all values

¹Dobbie and Fryer (2014) find similar results in an RD study of New York City exam schools.

²The use of affirmative action in exam school admissions is a highly contentious issue also in New York City where a federal complaint was filed in 2012 against the purely achievement-based exam school admissions because of the disproportionately low minority shares at these schools.

of the running variable are nonparametrically identified.³ In related work on the same problem, Angrist and Rokkanen (forthcoming) postulate a strong conditional independence assumption that identifies causal effects away from RD cutoffs. The framework developed here relies on weaker assumptions and is likely to find wider application.⁴

I use this framework to estimate causal effects of exam school attendance for the full population of applicants. These estimates suggest that the achievement gains from exam school attendance are larger among applicants with lower baseline measures of ability. I use these estimates to simulate effects of introducing either minority or socioeconomic preferences in Boston exam school admissions. The reforms change the admissions cutoffs faced by different applicant groups and affect the exam school assignment of 27-35% of applicants. The simulations suggest that the reforms boost achievement among applicants. These effects are largely driven by achievement gains experienced by lower-achieving applicants who gain access to exam schools as a result.

In developing the latent factor-based approach to RD extrapolation I build on the literatures on measurement error models (Hu and Schennach, 2008) and nonparametric instrumental variable models (Newey and Powell, 2003). Latent factor models have a long tradition in economics (Aigner, Hsiao, Kapteyn, and Wansbeek, 1984). In the program evaluation literature they have been used, for instance, to identify the joint distribution of potential outcomes (Aakvik, Heckman, and Vytlacil, 2005), time-varying treatment effects (Cooley Fruehwirth, Navarro, and Takahashi, 2014), and distributional treatment effects (Bonhomme and Sauder, 2011).

The educational consequences of affirmative action have mostly been studied in post-secondary schools with a focus on application and enrollment margins. Several papers have studied affirmative action bans in California and Texas as well as the introduction of the Texas 10% plan (Long, 2004; Card and Krueger, 2005; Dickson, 2006; Andrews, Ranchhod, and Sathy, 2010; Cortes, 2010; Antonovics and Backes, 2013, 2014). Howell (2010) uses a structural model to simulate the effects of a nation-wide elimination of affirmative action in college admissions, and Hinrichs (2012) studies the effects of various affirmative action bans around the United States. Only a few studies have looked at the effects of affirmative action in selective school admissions on later outcomes (Arcidiacono, 2005; Rothstein and Yoon, 2008; Bertrand, Hanna, and Mullainathan, 2010; Francis and Tannuri-Pianto, 2012; Kapor, 2015).

³This is similar to ideas put forth by Lee (2008), Lee and Lemieux (2010), DiNardo and Lee (2011), and Bloom (2012). However, this is the first paper that discusses how this framework can be used in RD extrapolation.

⁴For other approaches to RD extrapolation, see Angrist and Pischke (2009), Jackson (2010), DiNardo and Lee (2011), Wing and Cook (2013), Bargain and Doorley (2013), Bertanha and Imbens (2014), and Dong and Lewbel (forthcoming).

The rest of the paper is organized as follows. The next section outlines the econometric framework in a sharp RD design. Section 3 discusses an extension to a fuzzy RD design. Section 4 discusses identification and estimation of the latent factor model in the Boston exam schools setting. Section 5 reports the estimation results. Section 6 uses the estimates to simulate effects of affirmative action. Section 7 concludes.

2 Latent Factor Extrapolation in a Sharp RD Design

2.1 Framework

In a sharp RD design a binary treatment $D \in \{0, 1\}$ is a deterministic function of an observed, continuous running variable $R \in \mathcal{R}$ and a known cutoff c :

$$D = 1(R \geq c).$$

Individuals with running values less than c do not receive the treatment whereas individuals with running variable values above c receive the treatment. Each individual is associated with two potential outcomes: $Y(0)$ is the outcome of an individual if she does not receive the treatment ($D = 0$), and $Y(1)$ is the outcome of an individual if she receives the treatment ($D = 1$). The observed outcome of an individual is

$$Y = (1 - D) \times Y(0) + D \times Y(1).$$

I ignore additional covariates to simplify notation. Everything that follows can be generalized to allow for additional covariates by conditioning on them throughout.

Sharp RD allows one to nonparametrically identify the Average Treatment Effect (ATE) at the cutoff, $E[Y(1) - Y(0) | R = c]$, under relatively mild conditions, listed in Assumption A. Under these assumptions, the ATE is given by the discontinuity in $E[Y | R]$ at the cutoff, as shown in Lemma 1 (Hahn, Todd, and van der Klaauw, 2001).

Assumption A.

1. $f_R(r) > 0$ in a neighborhood around c .
2. $E[Y(0) | R = r]$ and $E[Y(1) | R = r]$ are continuous in r at c for all $y \in \mathcal{Y}$.
3. $E[|Y(0)| | R = c], E[|Y(1)| | R = c] < \infty$.

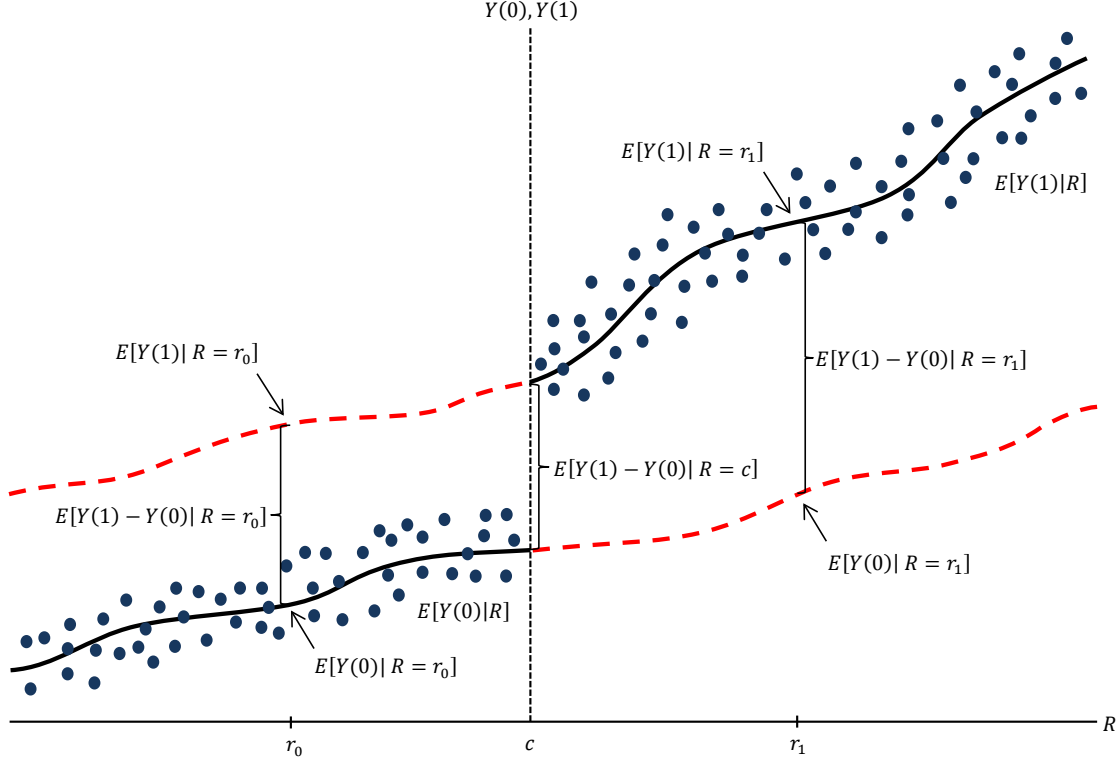


Figure 1: Extrapolation Problem in a Sharp Regression Discontinuity Design

Lemma 1. (Hahn, Todd, and van der Klaauw, 2001) Suppose Assumption A holds. Then

$$E[Y(1) - Y(0) | R = c] = \lim_{\delta \downarrow 0} \{E[Y | R = c + \delta] - E[Y | R = c - \delta]\}.$$

The drawback of sharp RD is that it does not allow one to say anything about the ATE away from the cutoff, $E[Y(1) - Y(0) | R = r]$ for $r \neq c$. Figure 1 illustrates this extrapolation problem. To the left of the cutoff we observe $Y(0)$ whereas to the right of the cutoff we observe $Y(1)$. The relevant counterfactual outcomes are instead unobservable. Suppose we wanted to know the ATE at $R = r_0$ to the left of the cutoff, $E[Y(1) - Y(0) | R = r_0]$. We observe $E[Y(0) | R = r_0]$, but the counterfactual $E[Y(1) | R = r_0]$ is unobservable. Similarly, suppose we wanted to know the ATE at $R = r_1$ to the right of the cutoff, $E[Y(1) - Y(0) | R = r_1]$. We observe $E[Y(1) | R = r_1]$, but the counterfactual $E[Y(0) | R = r_1]$ is unobservable. Therefore, in order to identify these ATEs, one is forced to extrapolate from one side of the cutoff to the other.

This paper develops a latent factor-based solution to the extrapolation problem. I consider a

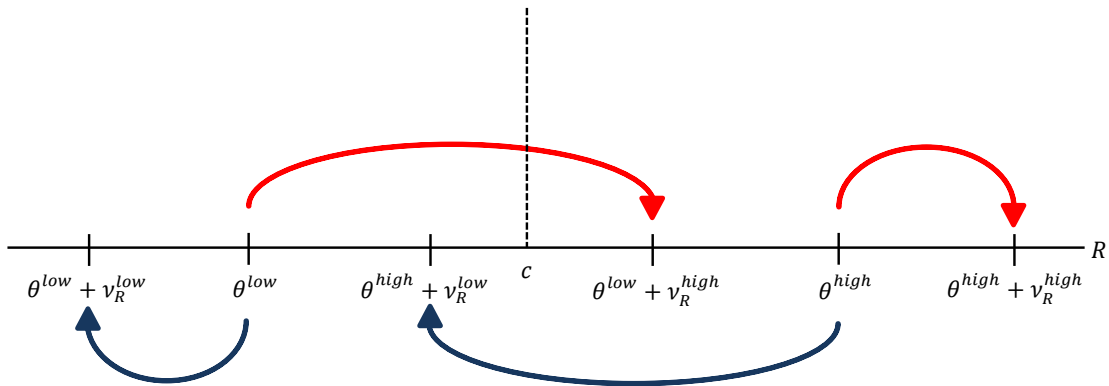


Figure 2: Treatment Assignment in the Latent Factor Framework

setting in which R is a function of a latent factor θ and a disturbance ν_R :

$$R = g_R(\theta, \nu_R)$$

where g_R is an unknown function, and both θ and ν_R are unobservable and potentially multidimensional. One example of such setting is selective school admissions where R is an entrance exam score that can be thought of as a noisy measure of an applicant's academic ability θ .

Figure 2 illustrates the latent factor framework when both θ and ν_R are scalars and $R = \theta + \nu_R$. Consider two types of individuals with low and high levels of θ , $\theta^{low} < c$ and $\theta^{high} > c$. If there was no noise in R , individuals with $\theta = \theta^{low}$ would not receive the treatment whereas individuals with $\theta = \theta^{high}$ would receive the treatment. However, because of the noise in R , some of the individuals with $\theta = \theta^{low}$ end up to the right of the cutoff, and some of the individuals with $\theta = \theta^{high}$ to the left of the cutoff. This means that both types of individuals are observed with and without the treatment. In the selective school admissions example this means that applicants with the same level of academic ability end up on different sides of the admissions cutoff because of, say, having a good or a bad day when taking the entrance exam.

I assume that the potential outcomes $Y(0)$ and $Y(1)$ are conditionally independent of R given θ , as stated in Assumption B. In the selective school admissions example this means that while $Y(0)$ and $Y(1)$ may depend on an applicant's academic ability, they do not depend on the noise in the entrance exam score. Assumption B implies that any dependence of $(Y(0), Y(1))$ on R is solely due to the dependence of $(Y(0), Y(1))$ on θ and the dependence of R on θ . Lemma 2 highlights the key implication of Assumption B: the ATE at $R = r$, $E[Y(1) - Y(0) | R = r]$, depends on the

latent conditional ATE given θ , $E[Y(1) - Y(0) | \theta]$, and the conditional distribution of θ given R , $f_{\theta|R}$. Therefore, the identification of the ATE away from the cutoff depends on one's ability to identify these two objects.

Assumption B. $(Y(0), Y(1)) \perp\!\!\!\perp R | \theta$.

Lemma 2. *Suppose that Assumption B holds. Then*

$$E[Y(1) - Y(0) | R = r] = E\{E[Y(1) - Y(0) | \theta] | R = r\}$$

for all $r \in \mathcal{R}$.

Consider again the identification of the ATE for individuals with $R = r_0$ to the left of the cutoff and for individuals with $R = r_1$ to the right of the cutoff, illustrated in Figure 1. As discussed above, the sharp RD design allows one to observe $E[Y(0) | R = r_0]$ and $E[Y(1) | R = r_1]$, and the extrapolation problem arises from the unobservability of $E[Y(1) | R = r_0]$ and $E[Y(0) | R = r_1]$. Suppose the conditional expectation functions of $Y(0)$ and $Y(1)$ given θ , $E[Y(0) | \theta]$ and $E[Y(1) | \theta]$ are known. In addition, suppose the conditional distributions of θ given $R = r_0$ and $R = r_1$, $f_{\theta|R}(\theta | r_0)$ and $f_{\theta|R}(\theta | r_1)$, are known. Then, under Assumption B, the unobserved counterfactuals $E[Y(1) | R = r_0]$ and $E[Y(0) | R = r_1]$ are given by

$$\begin{aligned} E[Y(1) | R = r_0] &= E\{E[Y(1) | \theta] | R = r_0\} \\ E[Y(0) | R = r_1] &= E\{E[Y(0) | \theta] | R = r_1\}. \end{aligned}$$

There is only one remaining issue: how does one identify $E[Y(0) | \theta]$, $E[Y(1) | \theta]$, and $f_{\theta|R}$? If θ was observable, these objects could be identified following the covariate-based approach developed by Angrist and Rokkanen (forthcoming). However, here the unobservability of θ requires an alternative approach. To achieve identification, I rely on the availability of multiple noisy measures of θ . I focus on a setting where θ is one-dimensional for simplicity. Extension to a setting where θ is multidimensional is discussed in the end of this section.

I assume that the data contains three noisy measures of θ , denoted by M_1 , M_2 , and M_3 :

$$\begin{aligned} M_1 &= g_{M_1}(\theta, \nu_{M_1}) \\ M_2 &= g_{M_2}(\theta, \nu_{M_2}) \\ M_3 &= g_{M_3}(\theta, \nu_{M_3}) \end{aligned}$$

where g_{M_1} , g_{M_2} , and g_{M_3} are unknown functions, and ν_{M_1} , ν_{M_2} , and ν_{M_3} are potentially multi-dimensional disturbances. I require θ , M_1 , and M_2 to be continuous but allow M_3 to be either continuous or discrete; even a binary M_3 is sufficient. I denote the supports of θ , M_1 , M_2 , and M_3 by Θ , \mathcal{M}_1 , \mathcal{M}_2 , and \mathcal{M}_3 . I will occasionally also use the notation $M = (M_1, M_2, M_3)$ and $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{M}_3$.

I focus on a setting in which R is a deterministic function of a subset of M . It is possible to allow for a more general setting where the relationship between R and M is stochastic as long as this additional noise is not related to the potential outcomes. In the selective school admissions example, one could think of M_1 as the entrance exam score and M_2 and M_3 as two (pre-application) baseline test scores.

2.2 Parametric Example

To provide a benchmark for the discussion about nonparametric identification of the latent factor model, I start by considering the identification of a simple parametric example.

The measurement model takes the following form:

$$\begin{aligned} M_1 &= \theta + \nu_{M_1}, k = 1, 2, 3 \\ M_2 &= \mu_{M_2} + \lambda_{M_2}\theta + \nu_{M_2} \\ M_3 &= \mu_{M_3} + \lambda_{M_3}\theta + \nu_{M_3} \end{aligned}$$

where

$$\begin{bmatrix} \theta \\ \nu_{M_1} \\ \nu_{M_2} \\ \nu_{M_3} \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_\theta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\theta^2 & 0 & 0 & 0 \\ 0 & \sigma_{\nu_{M_1}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\nu_{M_2}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\nu_{M_3}}^2 \end{bmatrix} \right).$$

The intercept and slope in the first measurement equation are normalized to 0 and 1 without loss of generality in order to pin down the location and scale of θ .

The latent outcome model takes the following form:

$$\begin{aligned} E[Y(0) | \theta] &= \alpha_0 + \beta_0 \theta \\ E[Y(1) | \theta] &= \alpha_1 + \beta_1 \theta. \end{aligned}$$

I assume that the potential outcomes $Y(0)$ and $Y(1)$ are conditionally independent of M given θ . Formally,

$$(Y(0), Y(1)) \perp\!\!\!\perp M | \theta.$$

This implies that the measurements are related to the potential outcomes only through θ . They are used as instruments to identify the relationships between $Y(0)$ and θ and $Y(1)$ and θ .

The parameters of the measurement model can be identified from the means, variances, covariances of the joint distribution of M_1 , M_2 , and M_3 by noticing that

$$\begin{aligned} E[M_1] &= \mu_\theta \\ E[M_k] &= \mu_{M_k} + \lambda_{M_k} \mu_\theta, \quad k = 2, 3 \\ \text{Var}[M_1] &= \sigma_\theta^2 + \sigma_{\nu_{M_1}}^2 \\ \text{Var}[M_k] &= \lambda_{M_k}^2 \sigma_\theta^2 + \sigma_{\nu_{M_k}}^2, \quad k = 2, 3 \\ \text{Cov}[M_2, M_3] &= \lambda_{M_2} \lambda_{M_3} \sigma_\theta^2 \\ \text{Cov}[M_1, M_k] &= \lambda_{M_k} \sigma_\theta^2, \quad k = 2, 3. \end{aligned}$$

This allows one to derive closed-form solutions for all of the parameters.

The parameters of the latent outcome model can be identified from the conditional expectation of Y given M and D and the conditional expectation of θ given M and D by noticing that

$$\begin{aligned} E[Y | M = m^0, D = 0] &= \alpha_0 + \beta_0 E[\theta | M = m^0, D = 0] \\ E[Y | M = m^1, D = 1] &= \alpha_1 + \beta_1 E[\theta | M = m^1, D = 1] \end{aligned}$$

for all $m^0 \in \mathcal{M}^0$ and $m^1 \in \mathcal{M}^1$ where \mathcal{M}^d denotes the conditional support of M given $D = d$. This

allows one to identify the parameters of the latent outcome model by essentially running regressions of Y on predicted θ (given M) using data only to the left ($D = 0$) or to the right ($D = 1$) of the cutoff.

Finally, $E[Y(0) | R = r]$ and $E[Y(1) | R = r]$ and $E[Y(1) - Y(0) | R = r]$ are given by

$$\begin{aligned} E[Y(0) | R = r] &= \alpha_0 + \beta_0 E[\theta | R = r] \\ E[Y(1) | R = r] &= \alpha_1 + \beta_1 E[\theta | R = r] \\ E[Y(1) - Y(0) | R = r] &= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) E[\theta | R = r] \end{aligned}$$

where $E[\theta | R] = E[E[\theta | M] | R]$ for all $r \in \mathcal{R}$.

2.3 Nonparametric Identification

In this section I discuss nonparametric identification in the latent factor framework. I start by discussing the identification of the measurement model that takes the form

$$\begin{aligned} M_1 &= g_{M_1}(\theta, \nu_{M_1}) \\ M_2 &= g_{M_2}(\theta, \nu_{M_2}) \\ M_3 &= g_{M_3}(\theta, \nu_{M_3}). \end{aligned}$$

There are several approaches available in the literature for identifying the underlying measurement model from the joint distribution of noisy measurements (see, for instance, Chen, Hong, and Nekipelov (2011) for a review). These approaches differ in how they deal with the tradeoff between (a) how many restrictions are imposed on the measurement model and (b) how much is required from the data. I focus on a particular approach by Hu and Schennach (2008) as it allows one to be relatively agnostic about the measurement model.

Assumption C lists the conditions under which the measurement model is nonparametrically identified. (Hu and Schennach, 2008; Cunha, Heckman, and Schennach, 2010).

Assumption C.

1. $f_{\theta, M}(\theta, m)$ is bounded with respect to the product measure of the Lebesgue measure on $\Theta \times \mathcal{M}_1 \times \mathcal{M}_2$ and some dominating measure μ on \mathcal{M}_3 . All the corresponding marginal and conditional densities are also bounded.
2. M_1 , M_2 , and M_3 are jointly independent conditional on θ .

3. For all $\theta', \theta'' \in \Theta$, $f_{M_3|\theta}(m_3 | \theta')$ and $f_{M_3|\theta}(m_3 | \theta'')$ differ over a set of strictly positive probability whenever $\theta' \neq \theta''$.
4. There exists a known functional H such that $H[f_{M_1|\theta}(\cdot | \theta)] = \theta$ for all $\theta \in \Theta$.
5. $f_{\theta|M_1}(\theta | m_1)$ and $f_{M_1|M_2}(m_1 | m_2)$ form (boundedly) complete families of distributions indexed by $m_1 \in \mathcal{M}_1$ and $m_2 \in \mathcal{M}_2$.

Assumption C.1 requires θ , M_1 , and M_2 to be continuous but allows M_3 to be either continuous or discrete. It also restricts the joint, marginal and conditional densities of θ , M_1 , M_2 , and M_3 to be bounded. Assumption C.2 requires ν_{M_1} , ν_{M_2} , and ν_{M_3} to be jointly independent conditional on θ but allows for arbitrary dependence between θ and these disturbances, such as heteroscedasticity. Assumption C.3 requires the conditional distribution of M_3 given θ to vary as a function of θ . This assumption can be satisfied, for instance, by assuming strict monotonicity of the conditional expectation of M_3 given θ . Assumption C.4 imposes a normalization on the conditional distribution of M_1 given θ in order to pin down the location and scale of θ . This normalization can be achieved by, for instance, requiring the conditional mean, mode or median of M_1 to be equal to θ .

Assumption C.5 requires that the conditional distributions $f_{\theta|M_1}$ and $f_{M_1|M_2}$ are either complete or boundedly complete.⁵ The concept of completeness, originally introduced in statistics by Lehmann and Scheffe (1950, 1955), arises regularly in econometrics, for instance, as a necessary condition for the identification of nonparametric instrumental variable models (Newey and Powell, 2003). It can be seen as an infinite-dimensional generalization of a full rank condition. Loosely speaking, this assumption requires that $f_{\theta|M_1}$ and $f_{M_1|M_2}$ vary sufficiently as functions of M_1 and M_2 .

Theorem 1 states the identification result by Hu and Schennach (2008). This result relies on an eigenvalue-eigenfunction decomposition of the integral equation relating the distribution of the measurements to the latent distributions characterizing the measurement model. Having identified these latent distributions, one can then construct $f_{\theta,M}$ and consequently $f_{\theta|R}$.

Theorem 1. *Suppose Assumption C holds. Then the equation*

$$f_{M_1, M_3 | M_2}(m_1, m_3 | m_2) = \int_{\Theta} f_{M_1|\theta}(m_1 | \theta) f_{M_3|\theta}(m_3 | \theta) f_{\theta|M_2}(\theta | m_2) d\theta$$

⁵ Let X and Z denote generic random variables with supports \mathcal{X} and \mathcal{Z} . $f_{X|Z}(x | z)$ is said to form a (boundedly) complete family of distributions indexed by $z \in \mathcal{Z}$ if for all measurable (bounded) real functions h such that $E[h(X)] < \infty$, $E[h(X) | Z] = 0$ a.s. implies $h(X) = 0$ a.s. (Lehmann and Romano, 2005).

for all $m_1 \in \mathcal{M}_1$, $m_2 \in \mathcal{M}_2$ and $m_3 \in \mathcal{M}_3$ admits unique solutions for $f_{M_1|\theta}$, $f_{M_3|\theta}$, and $f_{\theta|M_2}$.

Let us now turn to the identification of the latent outcome model. This can be achieved by relating the variation in $E[Y | M, D]$ to the variation in $f_{\theta|M, D}$ to the left ($D = 0$) and to the right ($D = 1$) of the cutoff. This is analogous to the identification of nonparametric instrumental variable models (Newey and Powell, 2003). Assumption D lists the conditions under which $E[Y(0) | \theta]$ and $E[Y(1) | \theta]$ are nonparametrically identified for all $\theta \in \Theta$.

Assumption D.

1. $(Y(0), Y(1)) \perp\!\!\!\perp M | \theta$.
2. $0 < P[D = 1 | \theta] < 1$ a.s.
3. $f_{\theta|M, D}(\theta | m^0, 0)$ and $f_{\theta|M, D}(\theta | m^1, 1)$ form (boundedly) complete families of distributions indexed by $m^0 \in \mathcal{M}^0$ and $m^1 \in \mathcal{M}^1$.

Assumption D.1 requires that $Y(0)$ and $Y(1)$ are related to M only through θ . This can be thought of as an exclusion restriction. Assumption D.2 requires that the subset of M entering R are sufficiently noisy measures of θ so that an individual can always end up on either side of the cutoff. This can be thought of as a common support condition. Assumption D.3 imposes again a (bounded) completeness condition. This can be thought of as an infinite-dimensional first stage condition.

Theorem 2 states the identification result. The conditional independence assumption allows one to write down the integral equations given in the theorem. Under the (bounded) completeness assumption, $E[Y(0) | \theta]$ and $E[Y(1) | \theta]$ are unique solutions to these integral equations. Finally, the common support assumption ensures that both $E[Y(0) | \theta]$ and $E[Y(1) | \theta]$ are determined for all $\theta \in \Theta$.

Theorem 2. *Suppose Assumption D holds. Then the equations*

$$\begin{aligned} E[Y | M = m^0, D = 0] &= E\{E[Y(0) | \theta] | M = m^0, D = 0\} \\ E[Y | M = m^1, D = 1] &= E\{E[Y(1) | \theta] | M = m^1, D = 1\} \end{aligned}$$

for all $m^0 \in \mathcal{M}^0$ and $m^1 \in \mathcal{M}^1$ admit unique solutions for (bounded) $E[Y(0) | \theta]$ and $E[Y(1) | \theta]$ for all $\theta \in \Theta$.

All of the results discussed in this section can be easily extended to a setting with multidimensional θ . This only requires that one reinterprets θ , M_1 , and M_2 as vectors of the same dimension.

M_3 can instead still be one-dimensional. Therefore, the data requirements become stricter because of the additional measures needed for identification, but otherwise all of the results go through.

3 Extension to a Fuzzy RD Design

Section 2 focused on a sharp RD design where treatment received is fully determined by the running variable. In this section I discuss an extension to a fuzzy RD design in which treatment received is only partly determined by the running variable. The fuzziness of the RD design has no implications on the identification of the measurement model. It only causes modifications to the identification of the latent outcome model.

In a fuzzy RD design treatment assignment $Z \in \{0, 1\}$ is a deterministic function of the running variable R and the cutoff c :

$$Z = 1(R \geq c)$$

However, there is imperfect compliance with the treatment assignment: some of the individuals assigned to treatment ($Z = 1$) end up not receiving the treatment ($D = 0$) whereas some of the individuals not assigned to treatment ($Z = 0$) end up receiving the treatment ($D = 1$). Because of this, the probability of receiving the treatment jumps at the cutoff but by less than 1:

$$\lim_{\delta \downarrow 0} P[D = 1 | R = r + \delta] > \lim_{\delta \downarrow 0} P[D = 1 | R = r - \delta].$$

In addition to the potential outcomes $Y(0)$ and $Y(1)$, each individual is associated with two potential treatment status: $D(0)$ is the treatment status of an individual if she is not assigned to the treatment ($Z = 0$), and $D(1)$ is the treatment status of an individual if she is assigned to the treatment ($Z = 1$). The observed outcome and treatment status are

$$\begin{aligned} Y &= (1 - D) \times Y(0) + D \times Y(1) \\ D &= (1 - Z) \times D(0) + Z \times D(1). \end{aligned}$$

Assumption E lists conditions under which the fuzzy RD design allows one to nonparametrically

identify the Local Average Treatment Effect (LATE) for the compliers at the cutoff,

$$E[Y(1) - Y(0) \mid D(1) > D(0), R = c].$$

The compliers are individuals who are induced to receive the treatment by being assigned the treatment. Under these assumptions, the LATE is given by the ratio of the discontinuities in $E[Y \mid R]$ and $E[D \mid R]$ at the cutoff, as shown in Lemma 3 (Hahn, Todd, and van der Klaauw, 2001).

Assumption E.

1. $f_R(r) > 0$ in a neighborhood around c .
2. $P[D(0) = d_0, D(1) = d_1 \mid R = r]$ are continuous in r at c for $d_0, d_1 = 0, 1$.
3. $E[Y(0) \mid D(0) = d_0, D(1) = d_1, R = r]$ and $E[Y(1) \mid D(0) = d_0, D(1) = d_1, R = r]$ are continuous in r at c for $d_0, d_1 = 0, 1$.
4. $P[D(1) \geq D(0) \mid R = c] = 1$ and $P[D(1) > D(0) \mid R = c] > 0$
5. $E[Y(0) \mid D(1) > D(0), R = c], E[Y(1) \mid D(1) > D(0), R = c] < \infty$.

Lemma 3. (Hahn, Todd, and van der Klaauw, 2001) Suppose Assumption E holds. Then

$$E[Y(1) - Y(0) \mid D(1) > D(0), R = c] = \lim_{\delta \downarrow 0} \frac{E[Y \mid R = c + \delta] - E[Y \mid R = c - \delta]}{E[D \mid R = c + \delta] - E[D \mid R = c - \delta]}.$$

Assumption F lists conditions under which the LATE for compliers at $R = r$,

$$E[Y(1) - Y(0) \mid D(1) > D(0), R = r],$$

is identified in the latent factor framework. Assumption F.1 requires that the potential outcomes $Y(0)$ and $Y(1)$ and the potential treatment status $D(0)$ and $D(1)$ are jointly independent of R conditional on the latent factor θ . Assumption F.2 requires that being assigned the treatment can only make an individual more likely to receive the treatment for all $\theta \in \Theta$. Assumption F.3 imposes this relationship to be strict at least for some individual for all $\theta \in \Theta$.

Assumption F.

1. $(Y(0), Y(1), D(0), D(1)) \perp\!\!\!\perp R \mid \theta$.
2. $P[D(1) \geq D(0) \mid \theta] = 1$ a.s.
3. $P[D(1) > D(0) \mid \theta] > 0$ a.s.

The identification result is stated in Lemma 4. Under Assumption F, the LATE for compliers at $R = r$ is given by the ratio of the reduced form effect of treatment assignment on the outcome and the first stage effect of treatment assignment on treatment received at $R = r$.

Lemma 4. *Suppose Assumption F holds. Then*

$$E[Y(1) - Y(0) \mid D(1) > D(0), R = r] = \frac{E\{E[Y(D(1)) - Y(D(0)) \mid \theta] \mid R = r\}}{E\{E[D(1) - D(0) \mid \theta] \mid R = r\}}$$

for all $r \in \mathcal{R}$.

Assumption G lists the conditions under which the latent reduced form effect of treatment assignment on the outcome, $E[Y(D(1)) - Y(D(0)) \mid \theta]$, and the latent first stage effect of treatment assignment on the probability of receiving the treatment, $E[D(1) - D(0) \mid \theta]$, are nonparametrically identified from the observed conditional expectation functions $E[Y \mid M, Z]$ and $E[D \mid M, Z]$. Assumption G.1 that the noise in M is unrelated to $(Y(0), Y(1), D(0), D(1))$. Assumption G.2 repeats the common support assumption from Assumption D whereas Assumption G.3 is analogous to the (bounded) completeness condition in Assumption D.

Assumption G.

1. $(Y(0), Y(1), D(0), D(1)) \perp\!\!\!\perp M \mid \theta$.
2. $0 < P[D = 1 \mid \theta] < 1$ a.s.
3. $f_{\theta|M,Z}(\theta \mid m^0, 0)$ and $f_{\theta|M,Z}(\theta \mid m^1, 1)$ form (boundedly) complete families of distributions indexed by $m^0 \in \mathcal{M}^0$ and $m^1 \in \mathcal{M}^1$.

Theorem 3 states the identification result. Together with Lemma 4 this result can be used to nonparametrically identify the LATE for compliers at any point in the running variable distribution. The proof of Theorem 3 is analogous to the proof of Theorem 2.

Theorem 3. *Suppose Assumption G holds. Then the equations*

$$\begin{aligned} E[Y \mid M = m^0, D = 0] &= E\{E[Y(D(0)) \mid \theta] \mid M = m^0, D = 0\} \\ E[Y \mid M = m^1, D = 1] &= E\{E[Y(D(1)) \mid \theta] \mid M = m^1, D = 1\} \\ E[D \mid M = m^0, D = 0] &= E\{E[D(0) \mid \theta] \mid M = m^0, D = 0\} \\ E[D \mid M = m^1, D = 1] &= E\{E[D(1) \mid \theta] \mid M = m^1, D = 1\} \end{aligned}$$

for all $m_0 \in \mathcal{M}_0$ and $m_1 \in \mathcal{M}_1$ admit unique solutions for (bounded) $E[Y(D(0)) | \theta]$, $E[Y(D(1)) | \theta]$, $E[D(0) | \theta]$, and $E[D(1) | \theta]$ for all $\theta \in \Theta$.

4 Boston Exam Schools

4.1 Setting

The BPS system includes three selective high schools, known as exam schools, that span grades 7-12: Boston Latin School, Boston Latin Academy, and John D. O’Bryant High School of Mathematics and Science. These schools are seen as the flagship of the BPS system, and they are regularly among the best high schools in Boston as measured by, for instance, test scores and graduation rates. Because of their selective admissions process, the exam schools enroll a higher-achieving student body than traditional BPS schools. In addition, the teaching staff at these schools is more experienced and qualified, and they offer a rich array of college preparatory classes and extracurricular activities.

The exam schools admit new students for grades 7 and 9 (O’Bryant also admits some students for grade 10). Only Boston residents are allowed to apply. Each applicant submits a preference ordering of the exam schools she is applying to. Admissions decisions are based on the applicant’s GPA in English and Math from the previous school year and the fall term of the ongoing school year as well as the applicant’s scores on the ISEE administered during the fall term of the ongoing school year. The ISEE is an entrance exam used by several selective schools in the United States. It consists of five sections: Reading Comprehension, Verbal Reasoning, Mathematics Achievement, Quantitative Reasoning, and a 30-minute essay. Exam school admissions only use the first four sections of the ISEE.

Each applicant receives an offer from at most one exam school. There is no waitlist practice. The offers are assigned using the student-proposing Deferred Acceptance (DA) algorithm (Gale and Shapley, 1962). The algorithm takes as inputs each exam school’s predetermined capacity, each applicant’s preferences over the exam schools, and the exam schools’ rankings of the applicants based on a weighted average of their standardized GPA and ISEE scores. These rankings differ slightly across the exam schools.

The DA algorithm produces exam school-specific admissions cutoffs that are given by the lowest rank among applicants admitted to a given exam school. Since the applicants receive an offer from at most one exam school, there is no direct link between an applicant’s rank and offer status.

However, as in Abdulkadiroglu, Angrist, and Pathak (2014), it is possible to construct for each exam school a sharp sample that consists of applicants who receive an offer if and only if their rank is above the admissions cutoff. Appendix B describes in detail the DA algorithm and the construction of the sharp samples.

4.2 Data

The data for this paper comes from the following three files provided by the BPS: (1) an exam school application file, (2) a BPS registration and demographic file, and (3) an Massachusetts Comprehensive Assessment System (MCAS) file. In addition, I use the students' home addresses to merge the BPS data with Census tract-level information from the American Community Survey (ACS) 5-year summary file for 2006-2011.

The exam school application file consists of the records for all exam school applications in 1995-2009. It provides information on each applicant's application year and grade, application preferences, GPA in English and Math, ISEE scores, exam school-specific ranks, and admissions decision. This allows me to reproduce the exam school-specific admissions cutoffs. I transform the exam school-specific ranks into percentiles, ranging from 0 to 100, within application year and grade and center them to be 0 at the admissions cutoff. These exam school-specific running variables give an applicant's distance from the admissions cutoff in percentile units. I standardize the ISEE scores and GPA to have a mean of 0 and a standard deviation of 1 in the applicant population within each year and grade.

The BPS registration and demographic file consists of the records for all BPS students in 1996-2012. It provides information on each student's home address, school, grade, gender, race, limited English proficiency (LEP) status, bilingual status, special education (SPED) status, and free or reduced price lunch (FRLP) status.

The MCAS file consists of the records for all MCAS tests taken by BPS students in 1997-2008. It provides information on 4th, 7th, and 10th grade MCAS scores in English, and 4th, 8th, and 10th grade MCAS scores in Math (in the case of retakes I only consider the first time a student took the test). I construct middle school and high school MCAS composite scores that are given by the average of the MCAS scores in 7th grade English and 8th grade Math and the average of the MCAS scores in 10th grade English and 10th grade Math. I standardize the 4th grade MCAS scores in English and Math as well as the middle school and high school MCAS composite scores to have a mean of 0 and standard deviation of 1 in the BPS population within each year and grade.

I use the ACS 5-year summary file for 2006-2011 to obtain information on the median family income, percent of households occupied by the owner, percent of families headed by a single parent, percent of households where a language other than English is spoken, distribution of educational attainment, and number of school-aged children in each Census tract in Boston. I use this information to divide the Census tracts into socioeconomic tiers as described in Section 6.1.

I restrict the sample to 7th grade applicants in 2000-2004. Most students enter the exam schools in 7th grade, and their exposure to the exam school treatment is the longest. This is also the applicant group for which the covariate-based RD extrapolation approach by Angrist and Rokkanen (forthcoming) fails. The restriction to 2000-2004 allows me to observe both 4th grade MCAS scores and middle/high school MCAS composite scores for the applicants. I exclude applicants from outside BPS as they are more likely to remain outside BPS and thus not have follow-up information in the data.

Table 1 reports descriptive statistics. Column (1) includes all BPS students enrolled in 6th grade in 2000-2004. Column (2) includes the subset of students who apply to exam schools. Columns (3)-(6) divide the applicants based on their exam school assignment. Exam school applicants are a highly selected group of students, with higher 4th grade MCAS scores and lower shares of blacks and Hispanics, LEP, and SPED. Similarly, there is considerable selection even within exam school applicants by their exam school assignment, with applicants admitted to a more selective exam school having higher 4th grade MCAS scores and lower shares of blacks and Hispanics, LEP, and FRPL.

4.3 Identification and Estimation

Let $Z \in \{0, 1, 2, 3\}$ denote the exam school assignment of an applicant where 0 stands for no offer, 1 for O’Bryant, 2 for Latin Academy and 3 for Latin School. Let $S \in \{0, 1, 2, 3\}$ denote the enrollment decision of an applicant in the fall following exam school application where 0 stands for traditional BPS, 1 for O’Bryant, 2 for Latin Academy, and 3 for Latin School. Let R_1 , R_2 , and R_3 denote the running variables for O’Bryant, Latin Academy, and Latin School.

The exam school assignment of an applicant is a deterministic function of her running variables and application preferences, denoted by P :

$$Z = g_Z(R_1, R_2, R_3, P)$$

Table 1: Descriptive Statistics for Boston Public School Students and Exam School Applicants

			Exam School Assignment			
	All	All	No		Latin	Latin
	BPS	Applicants	Offer	O'Bryant	Academy	School
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.489	0.545	0.516	0.579	0.581	0.577
Black	0.516	0.399	0.523	0.396	0.259	0.123
Hispanic	0.265	0.189	0.223	0.196	0.180	0.081
FRPL	0.755	0.749	0.822	0.788	0.716	0.499
LEP	0.116	0.073	0.109	0.064	0.033	0.004
Bilingual	0.315	0.387	0.353	0.420	0.451	0.412
SPED	0.227	0.043	0.073	0.009	0.006	0.009
English 4	0.000	0.749	0.251	0.870	1.212	1.858
Math 4	0.000	0.776	0.206	0.870	1.275	2.114
N	21,094	5,179	2,791	755	790	843

Notes: This table reports descriptive statistics for 2000-2004. The All BPS column includes all 6th grade students in Boston Public Schools in who do not have missing covariate or 4th grade MCAS information. The All Applicants column includes the subset of students who apply to Boston exam schools. The Exam School Assignment columns include the subsets of applicants who receive an offer from a given exam school.

The running variables are deterministic functions of the applicant's scores in the Reading Comprehension, Verbal Reasoning, Mathematics Achievement, and Quantitative Reasoning sections of the ISEE, denoted by M_2^E , M_3^E , M_2^M , and M_3^M , as well as her GPA in English and Math, denoted by G :

$$R_s = g_{R_s}(M_2^E, M_3^E, M_2^M, M_3^M, G), s = 1, 2, 3$$

In addition, the data contains 4th grade MCAS scores in English and Math, denoted by M_1^E and M_1^M .

I treat the 4th grade MCAS score in English and the scores in the Reading Comprehension and Verbal Reasoning sections of the ISEE as noisy measures of an applicant's English ability, denoted by θ_E :

$$M_k^E = g_{M_k^E}(\theta_E, \nu_{M_k^E}), k = 1, 2, 3$$

I treat the 4th grade MCAS score in Math and the scores in the Mathematics Achievement and Quantitative Reasoning sections of the ISEE as noisy measures of an applicant's Math ability,

Table 2: Correlations between the ISEE Scores and 4th Grade MCAS Scores

	ISEE				MCAS	
	Reading (1)	Verbal (2)	Math (3)	Quantitative (4)	English 4 (5)	Math 4 (6)
<i>Panel A: ISEE</i>						
Reading	1	0.735	0.631	0.621	0.670	0.581
Verbal	0.735	1	0.619	0.617	0.655	0.587
Math	0.631	0.619	1	0.845	0.598	0.740
Quantitative	0.621	0.617	0.845	1	0.570	0.718
<i>Panel B: MCAS</i>						
English 4	0.670	0.655	0.598	0.570	1	0.713
Math 4	0.581	0.587	0.740	0.718	0.713	1.000

Notes: This table reports correlations between the ISEE scores and 4th grade MCAS scores. The sample size is 5,179.

denoted by θ_M :

$$M_k^M = g_{M_k^M}(\theta_M, \nu_{M_k^M}), k = 1, 2, 3$$

This generates a two-dimensional latent factor model with $\theta = (\theta_E, \theta_M)$ and $M_k = (M_k^E, M_k^M)$, $k = 1, 2, 3$, using the notation of Section 2.

Table 2 reports the correlations between the ISEE scores and 4th grade MCAS scores. The correlations between the test scores are strong but far from perfect. Consistent with the latent factor structure specified above, test scores measuring English ability are more highly correlated with each other than with test scores measuring Math ability, and vice versa. There is also a time-pattern among test scores measuring a given ability: the ISEE scores measuring the same ability are more highly correlated with each other than with the 4th grade MCAS score measuring the same ability.

Let $Y(s)$, $s = 0, 1, 2, 3$, denote potential achievement under different enrollment decisions, and let $S(z)$, $z = 0, 1, 2, 3$, denote potential enrollment decisions under different exam school assignments. I assume that the potential outcomes and enrollment decisions are jointly independent of the test scores conditional on English and Math abilities and a set of covariates, denoted by X :

$$\left(\{Y(s)\}_{s=0}^3, \{S(z)\}_{z=0}^3 \right) \perp\!\!\!\perp M \mid \theta, X$$

The covariates included in X are GPA, application preferences, application year, race, gender,

SES tier, and indicators for FRPL, LEP, SPED, and being bilingual. I also assume that there is sufficient noise in the ISEE scores so that conditional on the covariates it is possible to observe an applicant with a given level of English and Math ability under any exam school assignment:

$$0 < P[Z = z \mid \theta, X] < 1 \text{ a.s.}, z = 0, 1, 2, 3$$

These two assumptions are enough to identify causal effects of exam school assignments on enrollment and achievement. I make two additional assumptions to identify causal effects of enrollment at a given exam school as opposed to traditional BPS for the compliers who enroll at the exam school if they receive an offer and at traditional BPS if they receive no offer. First, I assume that receiving an offer from exam school s as opposed to no offer induces at least some applicants to enroll at exam school s instead of traditional BPS. Second, I assume that this is the only way in which receiving an offer from exam school s as opposed to no offer affects the enrollment decision of an applicant. Formally, these first stage and monotonicity assumptions are given by

$$\begin{aligned} P[S(s) = s, S(0) = 0 \mid \theta, X] &> 0, s = 1, 2, 3 \\ P[S(s) = s', S(0) = s'' \mid \theta, X] &= 0, s' \neq s, s'' \neq 0, s' \neq s'' \end{aligned}$$

The four assumptions made here generalize the fuzzy RD of Section 3 to a setting with multiple treatments.

I approximate the conditional joint distribution of English and Math ability by a bivariate normal distribution:

$$\begin{bmatrix} \theta_E \\ \theta_M \end{bmatrix} \mid X \sim N \left(\begin{bmatrix} \mu'_{\theta_E} X \\ \mu'_{\theta_M} X \end{bmatrix}, \begin{bmatrix} \sigma_{\theta_E}^2 & \sigma_{\theta_E \theta_M} \\ \sigma_{\theta_E \theta_M} & \sigma_{\theta_M}^2 \end{bmatrix} \right).$$

To ensure a valid variance-covariance matrix I use the parametrization

$$\begin{bmatrix} \sigma_{\theta_E}^2 & \sigma_{\theta_E \theta_M} \\ \sigma_{\theta_E \theta_M} & \sigma_{\theta_M}^2 \end{bmatrix} = \begin{bmatrix} \omega_{11} & 0 \\ \omega_{21} & \omega_{22} \end{bmatrix} \begin{bmatrix} \omega_{11} & \omega_{21} \\ 0 & \omega_{22} \end{bmatrix}.$$

I also approximate the conditional distributions of the test scores using by normal distributions:

$$\begin{aligned} M_k^E | \theta, X &\sim N\left(\mu'_{M_k^E} X + \lambda_{M_k^E} \theta_E, \exp\left(\gamma_{M_k^E} + \delta_{M_k^E} \theta_E\right)^2\right), k = 1, 2, 3 \\ M_k^M | \theta, X &\sim N\left(\mu'_{M_k^M} X + \lambda_{M_k^M} \theta_M, \exp\left(\gamma_{M_k^M} + \delta_{M_k^M} \theta_M\right)^2\right), k = 1, 2, 3 \end{aligned}$$

I normalize $\mu_{M_1^E} = \mu_{M_1^M} = 0$ and $\lambda_{M_1^E} = \lambda_{M_1^M} = 1$ to pin down the location and scale of θ_E and θ_M .

Let $D_s(z) = 1(S(z) = s)$, $s = 0, 1, 2, 3$, denote indicators for potential enrollment decisions under different exam school assignments, and let $Y(S(z))$ denote potential achievement under different exam school assignments. I approximate the conditional expectations of $D_s(z)$ and $Y(S(z))$ using the linear models

$$\begin{aligned} E[D_s(z) | \theta, X] &= \alpha'_{D_s(z)} X + \beta_{D_s(z)}^E \theta_E + \beta_{D_s(z)}^M \theta_M \\ E[Y(S(z)) | \theta, X] &= \alpha'_{Y(S(z))} X + \beta_{Y(S(z))}^E \theta_E + \beta_{Y(S(z))}^M \theta_M \end{aligned}$$

where $z = 0, 1, 2, 3$.

Identification of the measurement and latent outcome models specified above follows from the nonparametric identification results of Sections 2 and 3. However, I illustrate this in Appendix C by providing moment equations that identify these particular parametric models.

I estimate the parameters of the measurement model using Maximum Simulated Likelihood (MSL). I use 500 random draws from $f_{\theta|X}$, to evaluate the integral in $f_{M|X}$. For a given observation, $f_{M|X}$ takes the form

$$\begin{aligned} &f_{M|X}(m | X; \mu, \lambda, \gamma, \delta, \omega) \\ &= \int \left[\prod_{k=1}^3 f_{M_k^E|\theta, X}(m_k^E | \theta, X; \mu, \lambda, \gamma, \delta) f_{M_k^M|\theta, X}(m_k^M | \theta, X; \mu, \lambda, \gamma, \delta) \right] \\ &\quad \times f_{\theta|X}(\theta | X; \mu, \omega) d\theta \end{aligned}$$

where $f_{M_k^E|\theta, X}$, $f_{M_k^M|\theta, X}$, $k = 1, 2, 3$, and $f_{\theta|X}$ are as specified above.

I estimate the parameters of the latent outcome model using the Method of Simulated Moments

(MSM) based on the moment equations

$$\begin{aligned} E[D_s | M, X, Z] &= \alpha'_{D_s(Z)} X + \beta_{D_s(Z)}^E E[\theta_E | M, X, Z] + \beta_{D_s(Z)}^M E[\theta_M | M, X, Z] \\ E[Y | M, X, Z] &= \alpha'_{Y(S(z))} X + \beta_{Y(S(z))}^E E[\theta_E | M, X, Z] + \beta_{Y(S(z))}^M E[\theta_M | M, X, Z] \end{aligned}$$

for $Z = 0, 1, 2, 3$. The conditional expectations $E[\theta_E | M, X, Z]$ and $E[\theta_M | M, X, Z]$ are computed using the MSL estimates of the measurement model and 500 random draws from $f_{\theta|X}$. The weighting matrix in the MSM procedure is based on the number of observations in the (M, X, Z) cells.

I compute standard errors based on nonparametric 5-step bootstrap using 500 replications (Andrews, 2002). For each bootstrap sample I re-estimate the measurement model using the original estimates as initial values and stop the MSL procedure after five iterations. I then re-estimate the latent outcome models using these MSL estimates. This provides a computationally attractive approach to taking into account the sampling uncertainty coming from both steps of the estimation procedure.

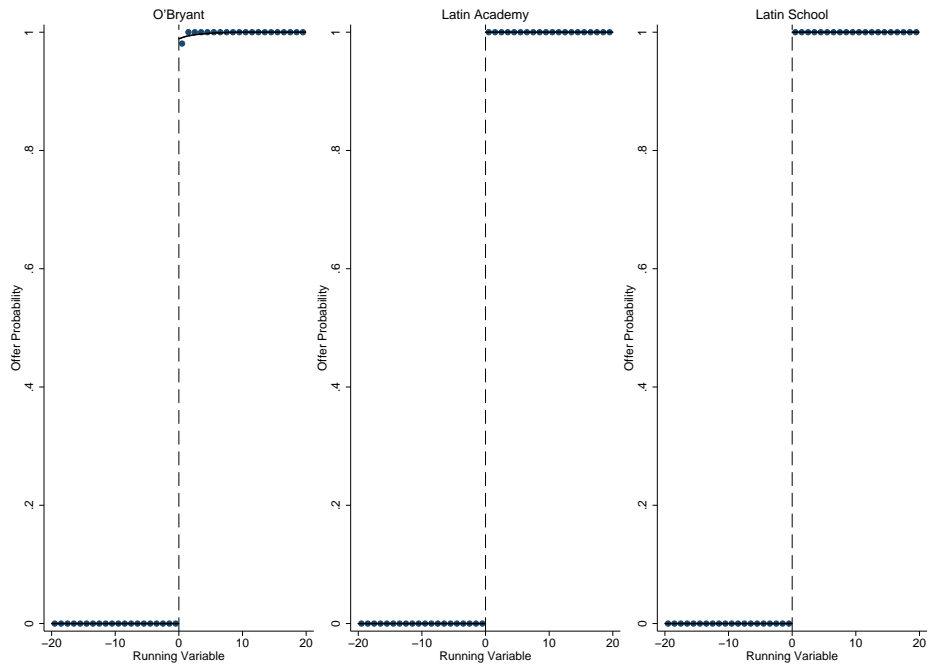
5 Extrapolation Results

5.1 Effects at the Admissions Cutoffs

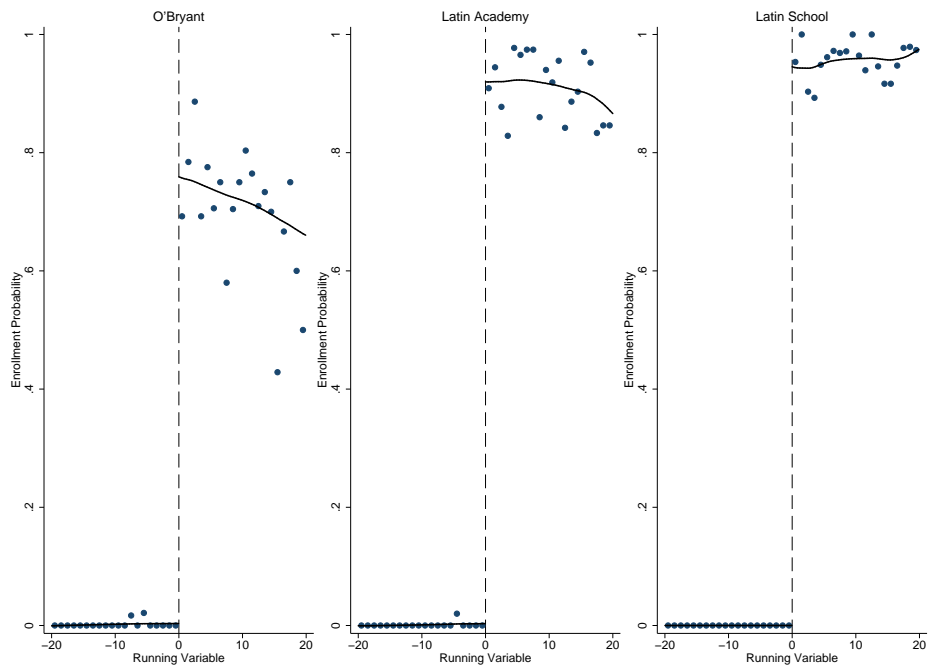
To provide a benchmark for the extrapolations, I begin with RD estimates of causal effects of exam school attendance for marginal applicants at the admissions cutoffs. Figure 3 plots the probability of receiving an offer from and the probability of enrolling at a given exam school as functions of the running variables. The figures are limited to windows of ± 20 around the admissions cutoffs and focus only on the sharp samples. The blue dots show bin averages in windows of width 1. The black solid lines show fits from local linear regressions (LLR) that use the edge kernel and the bandwidth algorithm by Imbens and Kalyanaraman (2012). Figure 4 repeats the exercise for average middle school and high school MCAS composite scores.

Table 3 reports the LLR-based first stage, reduced form, and LATE estimates corresponding to Figures 3 and 4. The first stage and reduced form models are

$$\begin{aligned} D_s &= \alpha_{FS} + \beta_{FS} Z_s + \gamma_{FS} R_s + \delta_{FS} Z_s \times R_s + X' \pi_{FS} + \eta \\ Y &= \alpha_{RF} + \beta_{RF} Z_s + \gamma_{RF} R_s + \delta_{RF} Z_s \times R_s + X' \pi_{RF} + \epsilon \end{aligned}$$

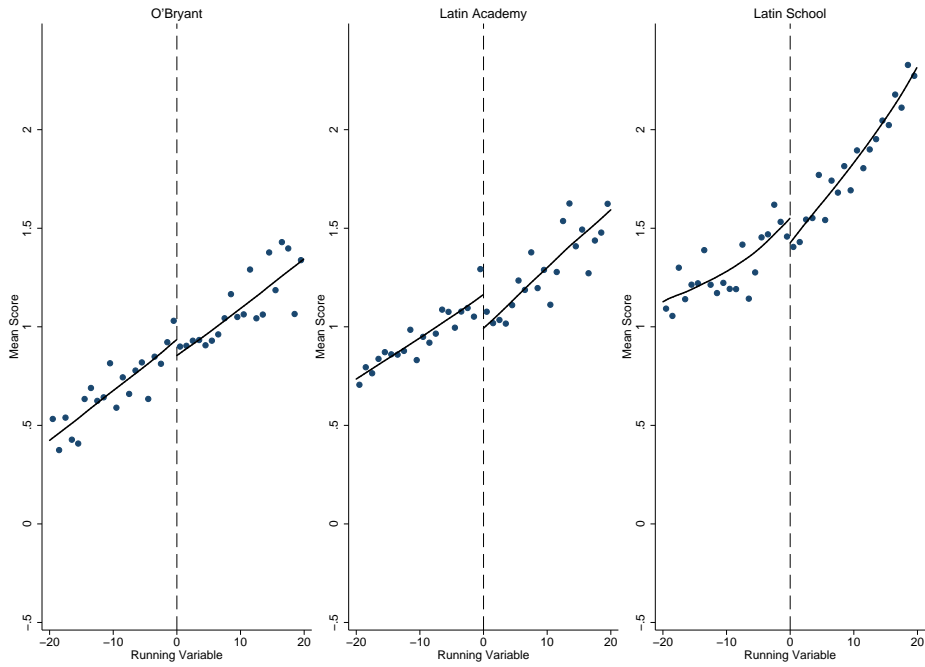


(a) Exam School Offer

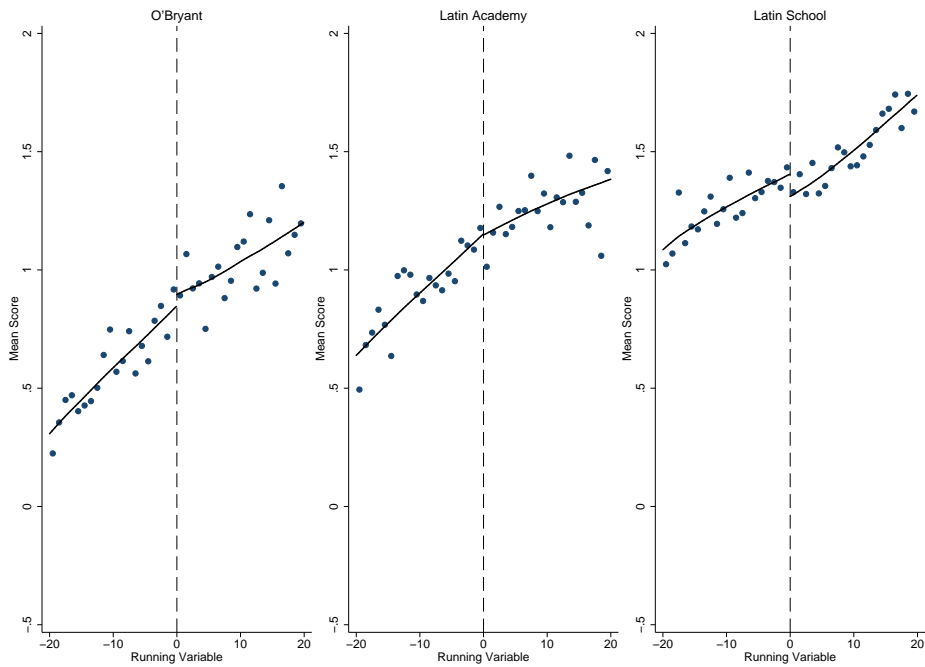


(b) Exam School Enrollment

Figure 3: Relationship between Exam School Offer and Enrollment and the Running Variables



(a) Middle School MCAS Composite Score



(b) High School MCAS Composite Score

Figure 4: Relationship between Middle School and High School MCAS Composite Scores and the Running Variables

where D_s is an indicator for enrollment at exam school s in the following school year, Y is the outcome of interest, Z_s is an indicator for being at or above the admissions cutoff for exam school s , R_s is running variable for exam school s , and X is a vector containing indicators for application years and application preferences. The first stage and reduced form effects are given by β_{FS} and β_{RF} , and the LATE for compliers at the admissions cutoff is given by $\frac{\beta_{RF}}{\beta_{FS}}$. Columns 1-3 report estimates for all applicants in the sharp samples. Columns 4-6 and columns 7-9 report estimates for applicants whose average 4th grade MCAS scores fall below and above the within-year median.

Figure 3a confirms the deterministic nature of exam school offers in the sharp samples: the probability of receiving an offer from a given exam school jump from 0 to 1 at the admissions cutoff. However, as can be seen from Figure 3b and the first stage estimates in Table 3, not all applicants receiving an offer from a given exam school choose to enroll there. The enrollment first stages are nevertheless large in magnitude, ranging from around 0.7-0.8 for O’Bryant to around 0.9-1.0 for Latin Academy and Latin School.

According to Figure 4a and the reduced form estimates in Panel A of Table 3, access to Latin Academy reduces the average middle school MCAS composite score of marginal applicants by 0.181σ . The corresponding estimates for O’Bryant and Latin School are also negative but not statistically significant. These negative effects are concentrated among applicants with high 4th grade MCAS scores. The estimates for applicants with low 4th grade MCAS scores are instead positive (although not statistically significant), with the exception of Latin School. The LATE estimates are similar to the reduced form estimates because of the high enrollment first stages.

According to Figure 4b and the reduced form estimates in Panel B of Table 3, there is no statistically significant evidence of effects of access to the exam schools on the average high school MCAS composite score of marginal applicants. There is, however, some evidence of heterogeneity in these effects: access to O’Bryant is estimated to increase the average high school MCAS composite score of marginal applicants with low 4th grade MCAS scores by 0.204σ . The LATE estimates are again similar to the reduced form estimates because of the high enrollment first stages.

It is important to note the incremental nature of the RD estimates reported here. Applicants just below the O’Bryant admissions cutoff do not receive an offer from any exam school. Therefore, the counterfactual for these applicants is traditional BPS. On the other hand, the vast majority of applicants just below the Latin Academy admissions cutoff receive an offer from O’Bryant, and the vast majority of applicants just below the Latin School admissions cutoff receive an offer from Latin Academy. This means that for most of the applicants at the Latin Academy admissions

Table 3: RD Estimates of the First Stage, Reduced Form and Local Average Treatment Effects at the Admissions Cutoffs

	All Applicants			Low 4th Grade MCAS			High 4th Grade MCAS		
	O'Bryant	Latin Academy	Latin School	O'Bryant	Latin Academy	Latin School	O'Bryant	Latin Academy	Latin School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Panel A: Middle School MCAS</i>									
First Stage	0.775*** (0.031)	0.949*** (0.017)	0.962*** (0.017)	0.734*** (0.072)	0.917*** (0.056)	1.000*** (0.000)	0.780*** (0.033)	0.956*** (0.017)	0.961*** (0.017)
Reduced Form	-0.084 (0.060)	-0.181*** (0.057)	-0.104 (0.079)	0.059 (0.127)	0.130 (0.199)	-0.095 (0.214)	-0.122* (0.065)	-0.200*** (0.056)	-0.084 (0.082)
LATE	-0.108 (0.078)	-0.191*** (0.060)	-0.108 (0.082)	0.080 (0.173)	0.142 (0.215)	-0.095 (0.214)	-0.157* (0.083)	-0.209*** (0.059)	-0.088 (0.085)
N	1,934	2,328	1,008	420	246	681	1,348	1,802	921
<i>Panel B: High School MCAS</i>									
First Stage	0.781*** (0.034)	0.955*** (0.018)	0.964*** (0.018)	0.742*** (0.074)	0.944*** (0.055)	1.000*** (0.000)	0.797*** (0.033)	0.942*** (0.022)	0.964*** (0.017)
Reduced Form	0.047 (0.055)	-0.021 (0.044)	-0.086 (0.052)	0.204** (0.099)	-0.029 (0.141)	-0.272 (0.186)	0.017 (0.054)	-0.029 (0.052)	-0.093* (0.053)
LATE	0.060 (0.070)	-0.022 (0.046)	-0.089 (0.054)	0.275** (0.136)	-0.031 (0.150)	-0.272 (0.186)	0.021 (0.068)	-0.030 (0.055)	-0.097* (0.055)
N	1,475	1,999	907	531	567	436	1,129	974	909

Notes: This table reports RD estimates of the effect of an exam school offer on exam school enrollment (First Stage), the effect of an exam school offer on MCAS scores (Reduced Form), and the effects of exam school enrollment on MCAS scores (LATE). The estimates are shown for all applicants as well as separately for applicants whose average 4th grade MCAS scores fall below and above the within-year median. Heteroskedasticity-robust standard errors shown in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

cutoff the relevant counterfactual is O’Bryant whereas at the Latin School admissions cutoff the relevant counterfactual is Latin Academy. Therefore, the above estimates should be interpreted as incremental effects of gaining access or going to a more selective school.

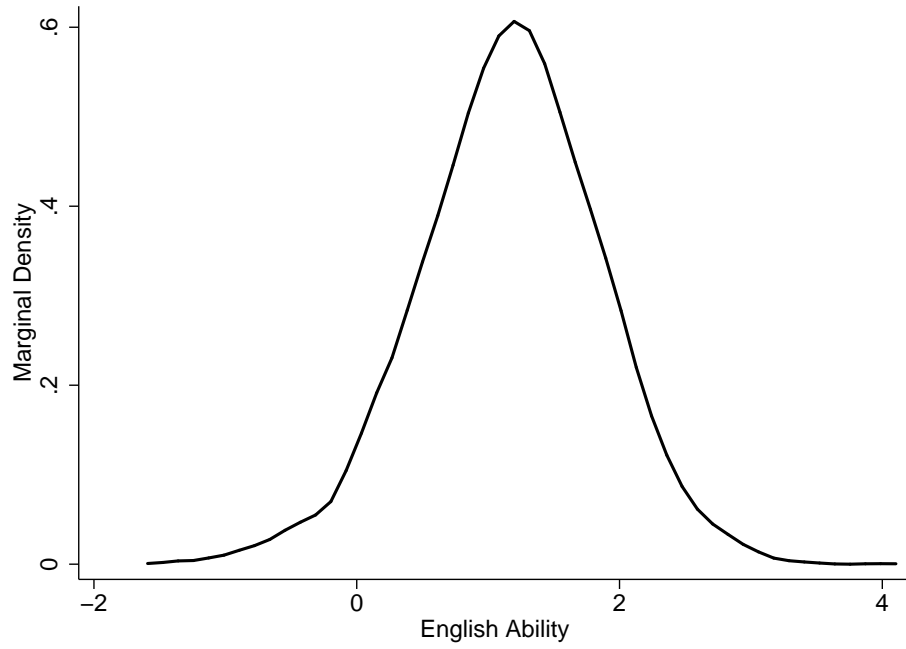
5.2 Estimates of the Latent Factor Model

Before turning to the extrapolation results I briefly discuss the main estimates of the measurement and latent outcome models. Figures 5 and 6 show the marginal distributions and scatterplot of English and Math ability among exam school applicants based on simulations from the estimated measurement model. The mean and standard deviation of English ability are 1.165 and 0.687. The mean and standard deviation of Math ability are 1.121 and 0.831. The correlation between the two abilities is 0.817.

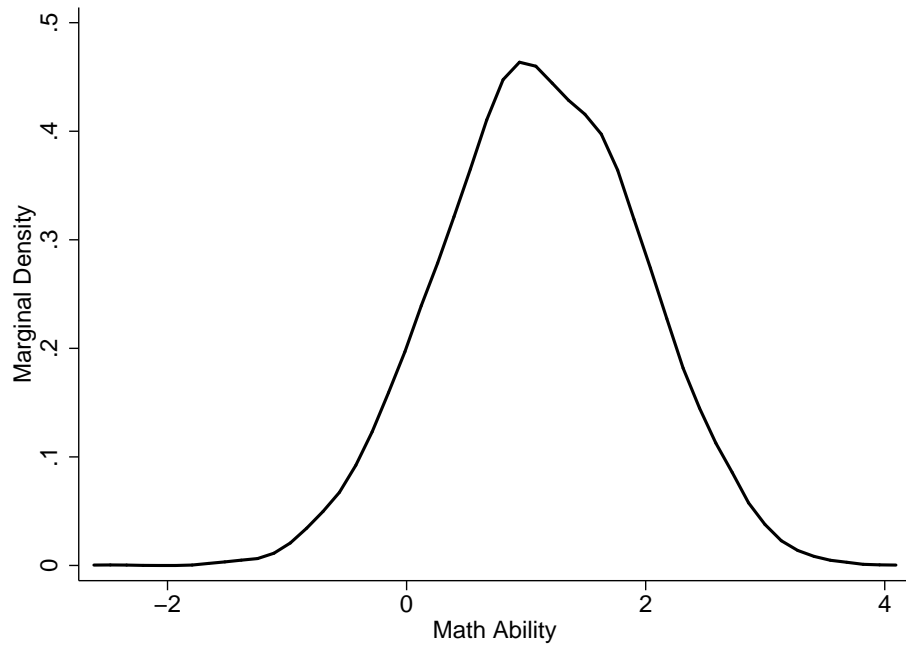
Table 4 reports the estimated factor loadings on the means and (log) standard deviations of ISEE and 4th grade MCAS scores. As mentioned above, the scales of the abilities have been pinned down by normalizing the factor loadings on the means of 4th grade MCAS scores to 1. The estimated factor loadings on the means of ISEE scores are somewhat above 1. The estimated factor loadings on the (log) standard deviations of ISEE scores in Reading Comprehension and Verbal Reasoning and 4th grade MCAS score in Math suggest that the variances of these test scores are increasing in the relevant ability. The estimated factor loadings on the (log) standard deviations of ISEE scores in Mathematical Achievement and Quantitative Reasoning and 4th grade MCAS score in English are small in magnitude and statistically insignificant.

Table 5 reports the estimated factor loadings on enrollment (First Stage) and middle school and high school MCAS composite scores (Reduced Form) under a given exam school assignment. For no exam school offer in column 1 the enrollment outcome is enrollment at traditional BPS. For an offer from a given exam school in columns 2-4 the outcome is enrollment at the exam school in question. With the exception of Latin School, the estimated factor loadings on enrollment tend to be negative, suggesting that applicants with higher ability are less likely to attend a given exam school if they receive an offer (or traditional BPS if they receive no offer). However, most of the estimates are small in magnitude and statistically insignificant.

The estimated factor loadings on middle school and high school MCAS composite scores are positive, large in magnitude, and statistically significant. This is not surprising: applicants with higher English and Math abilities tend to perform, on average, better in middle school and high school English and Math irrespective of their exam school assignment. A more interesting finding



(a) Marginal Distribution of English Ability



(b) Marginal Distribution of Math Ability

Figure 5: Marginal Distributions of English and Math Abilities

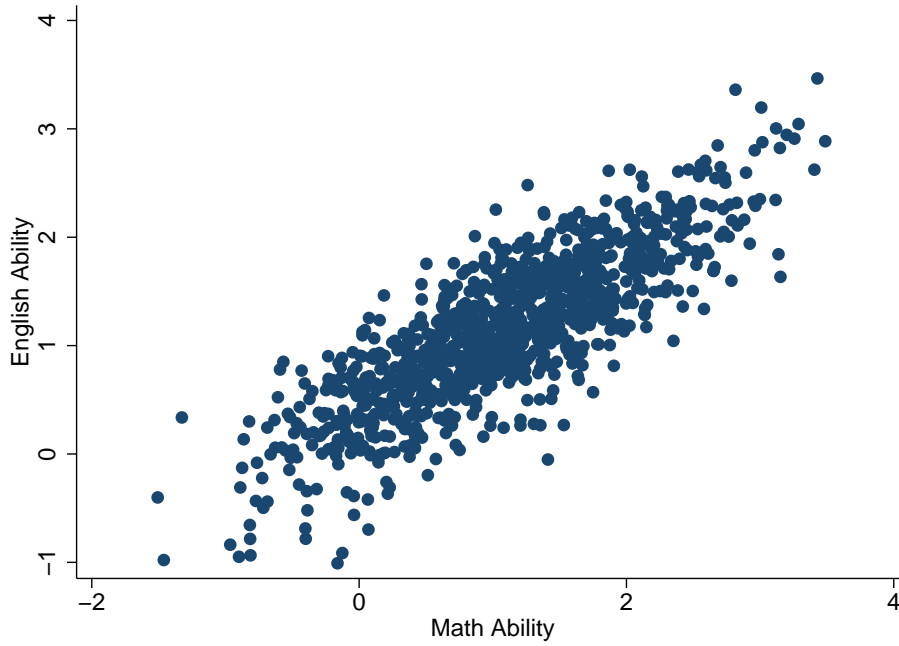


Figure 6: Scatterplot of English and Math Abilities

Table 4: Factor Loadings on the Means and (Log) Standard Deviations of ISEE and 4th Grade MCAS Scores

	ISEE				MCAS	
	Reading (1)	Verbal (2)	Math (3)	Quantitative (4)	English 4 (5)	Math 4 (6)
<i>Panel A: Factor Loading on Mean</i>						
θ_E	1.160*** (0.033)	1.180*** (0.035)			1	
θ_M			1.135*** (0.027)	1.119*** (0.028)		1
<i>Panel B: Factor Loading on (Log) Standard Deviation</i>						
θ_E	0.081*** (0.020)	0.152*** (0.018)			0.016 (0.016)	
θ_M			0.019 (0.018)	-0.013 (0.017)		0.124*** (0.015)

Notes: This table reports the estimated factor loadings on the means and (log) standard deviations of the ISEE and 4th grade MCAS scores. Standard errors based on nonparametric 5-step bootstrap shown in parentheses. The sample size is 5,179.

* significant at 10%, ** significant at 5%, *** significant at 1%

Table 5: Factor Loadings on Enrollment and MCAS Scores Under Given Exam School Assignments

	Middle School MCAS				High School MCAS			
	No Offer	O'Bryant	Latin Academy	Latin School	No Offer	O'Bryant	Latin Academy	Latin School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: First Stage</i>								
θ_E	-0.001 (0.007)	-0.073 (0.069)	-0.107** (0.050)	0.016 (0.011)	-0.001 (0.011)	-0.025 (0.076)	-0.102* (0.053)	0.003 (0.009)
θ_M	-0.008 (0.006)	-0.035 (0.059)	0.011 (0.031)	0.023 (0.015)	-0.012 (0.008)	-0.059 (0.061)	0.012 (0.030)	0.027** (0.012)
<i>Panel B: Reduced Form</i>								
θ_E	0.570*** (0.052)	0.476*** (0.087)	0.190** (0.089)	0.439*** (0.063)	0.446*** (0.062)	0.282*** (0.067)	0.217*** (0.066)	0.239*** (0.046)
θ_M	0.633*** (0.050)	0.601*** (0.067)	0.742*** (0.067)	0.464*** (0.063)	0.604*** (0.056)	0.346*** (0.070)	0.267*** (0.057)	0.158*** (0.044)
N	2,490	690	728	793	1,777	563	625	793

Notes: This table reports the estimated factor loadings on enrollment (First Stage) and MCAS scores (Reduced Form) under a given exam school assignment. First Stage refers to enrollment at a traditional Boston public school in the No Offer column and enrollment at a given exam school in the other columns. Standard errors based on nonparametric 5-step bootstrap shown in parentheses.

* significant at 10%, ** significant at 5%, *** significant at 1%

is that the factor loadings are larger in magnitude under no offer than under an offer from any given exam school. This is especially true for high school MCAS composite scores. This suggests that applicants with lower English and Math abilities gain more in terms of their achievement from access to exam schools.

5.3 Effects Away from the Admissions Cutoffs

Figure 7 plots the latent factor model-based fits and extrapolations for the sharp samples in the RD experiments over the full supports of the running variables. The fits and extrapolations are defined as

$$E [Y (S (z^{act})) | R_s = r], s = 1, 2, 3$$

$$E [Y (S (z^{cf})) | R_s = r], s = 1, 2, 3$$

where the counterfactual assignment z^{cf} is an offer from exam school s for applicants below the admissions cutoff for exam school s and an offer from the next most selective exam school for the

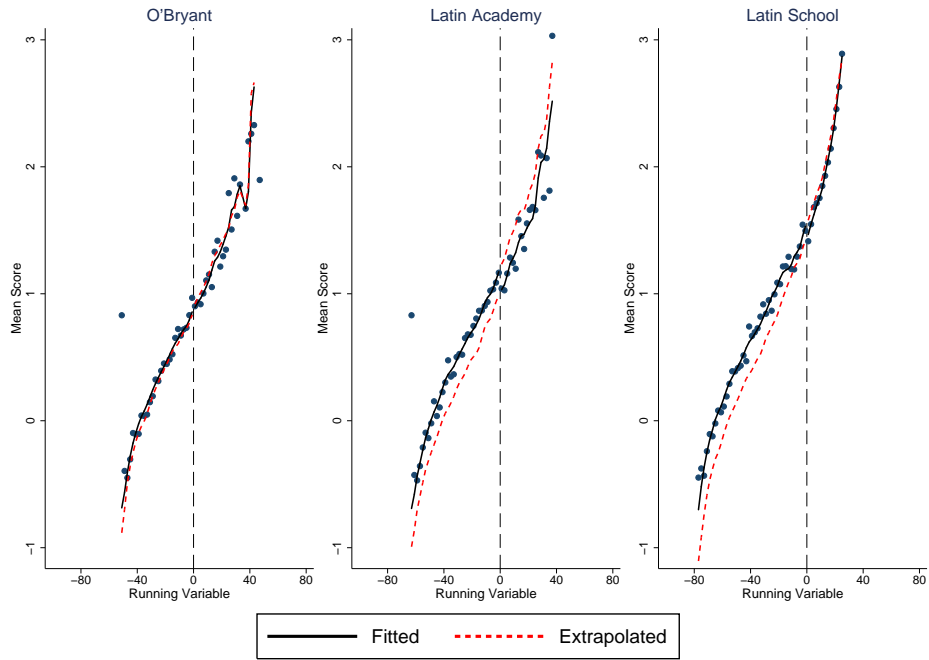
applicants above the admissions cutoff for exam school s (no offer in the case of O’Bryant). The blue dots show bin averages of the observed data in windows of width 1. The black solid lines show the fits, and the dashed red lines show the extrapolations. I smooth the fits and extrapolations with LLR using the edge kernel and a rule of thumb bandwidth (Fan and Gijbels, 1996).

Figure 7a plots the fits and extrapolations for middle school MCAS composite scores. The lack of an effect of access to OBryant found for marginal applicants at the admissions cutoff holds also for inframarginal applicants away from the admissions cutoff. The same is true for the negative effect of access to Latin Academy. The story is more nuanced for Latin School. The lack of an effect of access to Latin School found for marginal applicants at the admissions cutoff holds only for inframarginal applicants above the admissions cutoff. There is instead positive effect for inframarginal applicants above the admissions cutoff.

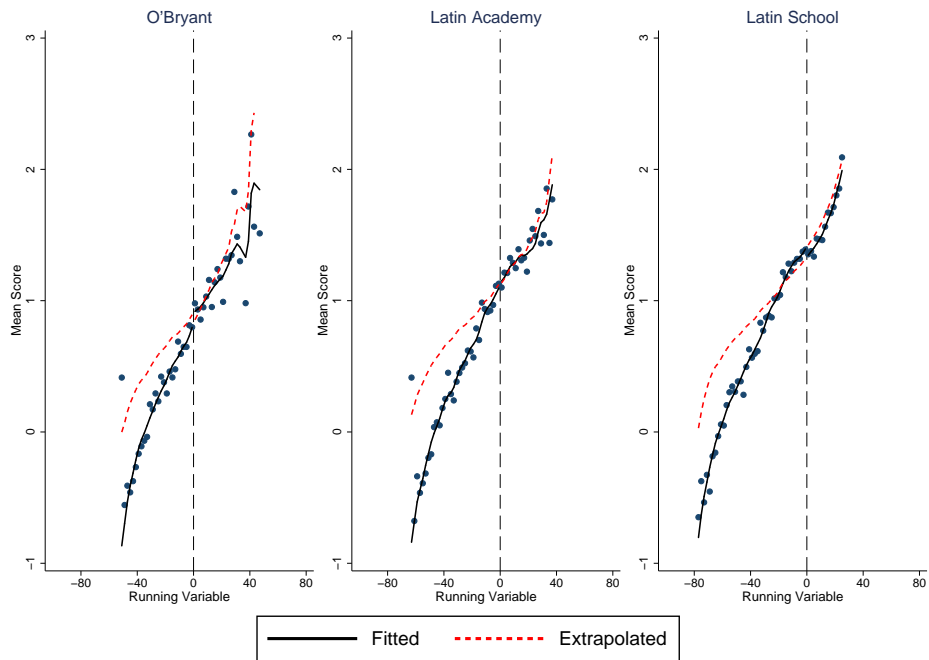
Figure 7b plots the fits and extrapolations for high school MCAS composite scores. For all three exam schools the extrapolations suggest little effect from access to a given exam school for inframarginal applicants above the admissions cutoffs. This is consistent with the findings for marginal applicants at the admissions cutoffs. For inframarginal applicants below the admissions cutoffs the picture arising is instead markedly different. For all three exam schools the extrapolations suggest a positive effect for applicants failing to gain access to the exam school in question.

Table 6 reports the extrapolated first stage, reduced form, and LATE estimates corresponding to Figure 7. Panel A reports estimates for middle school MCAS composite scores and Panel B for high school MCAS composite scores. Columns 1-3 report estimates for all applicants in the sharp samples whereas columns 4-6 and columns 7-9 report estimates separately for applicants who are below and above the admissions cutoffs. The estimates confirm the findings from Figure 7. Overall, the effects of exam school offers are negative for middle school MCAS composite scores and positive for high school MCAS composite scores. Most of these effects are concentrated among the inframarginal applicants below the admissions cutoffs. The LATE estimates are similar to the reduced form estimates because of the high enrollment first stages.

Just like the RD estimates, these extrapolations should be interpreted as incremental effects of an offer from or enrollment at a more selective school as opposed to the comparison between a given exam school versus traditional BPS. Table 7 addresses this question using the full sample of applicants (as opposed to only the sharp samples). I use the latent factor model to construct the effect of an offer from a given exam school versus no exam school offer on enrollment at this school



(a) Middle School MCAS Composite Score



(b) Middle School MCAS Composite Score

Figure 7: Reduced Form Extrapolations in the Exam School-Specific RD Experiments

Table 6: Extrapolated First Stage, Reduced Form, and Local Average Treatment Effects in the Exam School-Specific RD Experiments

	All Applicants			Below Admissions Cutoff			Above Admissions Cutoff		
	O'Bryant (1)	Latin Academy (2)	Latin School (3)	O'Bryant (4)	Latin Academy (5)	Latin School (6)	O'Bryant (7)	Latin Academy (8)	Latin School (9)
<i>Panel A: Middle School MCAS</i>									
First Stage	0.869*** (0.033)	0.971*** (0.013)	0.950*** (0.017)	0.904*** (0.041)	0.992*** (0.015)	0.960*** (0.020)	0.749*** (0.016)	0.891*** (0.018)	0.906*** (0.031)
Reduced Form	-0.047 (0.093)	-0.229** (0.100)	-0.214*** (0.080)	-0.050 (0.118)	-0.234* (0.123)	-0.245** (0.097)	-0.038 (0.034)	-0.211*** (0.061)	-0.080 (0.076)
LATE	-0.054 (0.108)	-0.236** (0.103)	-0.226*** (0.084)	-0.055 (0.132)	-0.236* (0.124)	-0.255** (0.101)	-0.051 (0.045)	-0.237*** (0.069)	-0.088 (0.085)
N	3,029	3,641	4,271	2,339	2,913	3,478	690	728	793
<i>Panel B: High School MCAS</i>									
First Stage	0.858*** (0.038)	0.962*** (0.021)	0.950*** (0.018)	0.892*** (0.050)	0.984*** (0.027)	0.959*** (0.021)	0.758*** (0.017)	0.885*** (0.021)	0.920*** (0.029)
Reduced Form	0.252*** (0.068)	0.279*** (0.075)	0.199*** (0.061)	0.343*** (0.089)	0.362*** (0.092)	0.281*** (0.075)	-0.021 (0.034)	-0.005 (0.046)	-0.091* (0.054)
LATE	0.293*** (0.081)	0.290*** (0.079)	0.209*** (0.064)	0.385*** (0.102)	0.367*** (0.097)	0.293*** (0.079)	-0.028 (0.045)	-0.006 (0.053)	-0.099* (0.059)
N	2,240	2,760	3,340	1,677	2,135	2,601	563	625	739

Notes: This table reports latent factor model based-estimates of the effect of an exam school offer on exam school enrollment (First Stage), the effect of an exam school offer on MCAS scores (Reduced Form), and the effects of exam school enrollment on MCAS scores (LATE) in the RD experiments. The estimates are shown for all applicants as well as separately for applicants whose running variables fall below and above the admissions cutoffs. Standard errors based on nonparametric 5-step bootstrap shown in parentheses.

* significant at 10%, ** significant at 5%, *** significant at 1%

(first stage) and on achievement (reduced form):

$$E [D_s (s) - D_s (0)], s = 1, 2, 3$$

$$E [Y (S (s)) - Y (S (0))], s = 1, 2, 3$$

I use these first stage and reduced form effects to construct the LATE of enrollment at a given exam school versus a traditional BPS for the full population of compliers induced to enroll at the exam school in question as opposed to traditional BPS:

$$E [Y (s) - Y (0) | D_s (s) = 1, D_0 (0) = 1] = \frac{E [Y (S (s)) - Y (S (0))]}{E [D_s (s) - D_s (0)]}, s = 1, 2, 3$$

Panel A reports the estimates for middle school MCAS composite scores and Panel B for high school MCAS composite scores. Columns 1-3 report estimates for all applicants whereas columns 4-6 and columns 7-9 report estimates separately for low-scoring applicants who do not receive an exam school offer and for high-scoring applicants who receive an exam school offer.

The estimates for middle school MCAS composite scores show large negative effects from receiving offers from Latin Academy and Latin School and no effect from receiving an offer from O’Bryant. There is not much evidence of heterogeneity in these effects by admissions status. The estimates for high school MCAS composite scores show instead no effect for the full population of applicants, but there is considerable heterogeneity in these effects by admissions status. For all three schools the estimates show large negative effects for low-scoring applicants who do not receive an exam school offer and large positive effects for high-scoring applicants who receive an exam school offer.

5.4 Placebo Experiments

A concern one might have with the previous results is that they are just an artifact of arbitrary extrapolations away from the admissions cutoffs. To address this concern, I study the performance of the latent factor model using a set of placebo experiments. I split the applicants with exam school assignment $z = 0, 1, 2, 3$ in half based on the within-year median of the running variable.⁶ I re-estimate the latent outcome models to the left and right of the resulting placebo cutoffs and use these estimates to extrapolate away from the placebo cutoffs. All of the applicants to the

⁶For applicant receiving no offer I use the average of their exam school-specific running variables.

Table 7: Extrapolated First Stage, Reduced Form, and Local Average Treatment Effects for Comparisons between a Given Exam School and Traditional Boston Public Schools

	All Applicants			No Exam School Offer			Exam School Offer		
	O'Bryant (1)	Latin Academy (2)	Latin School (3)	O'Bryant (4)	Latin Academy (5)	Latin School (6)	O'Bryant (7)	Latin Academy (8)	Latin School (9)
<i>Panel A: Middle School MCAS</i>									
First Stage	0.767*** (0.017)	0.956*** (0.009)	0.967*** (0.016)	0.901*** (0.040)	0.994*** (0.016)	0.958*** (0.024)	0.749*** (0.016)	0.927*** (0.010)	0.986*** (0.005)
Reduced Form	-0.059 (0.048)	-0.275*** (0.074)	-0.319*** (0.079)	-0.051 (0.117)	-0.246* (0.138)	-0.279** (0.117)	-0.069 (0.074)	-0.307*** (0.048)	-0.363*** (0.051)
LATE	-0.077 (0.063)	-0.288*** (0.078)	-0.330*** (0.082)	-0.057 (0.131)	-0.248* (0.140)	-0.292** (0.122)	-0.092 (0.099)	-0.332*** (0.052)	-0.368*** (0.052)
N		4,701			2,490			2,211	
<i>Panel B: High School MCAS</i>									
First Stage	0.754*** (0.020)	0.948*** (0.013)	0.968*** (0.016)	0.887*** (0.049)	0.987*** (0.030)	0.955*** (0.026)	0.758*** (0.017)	0.925*** (0.010)	0.987*** (0.005)
Reduced Form	0.021 (0.037)	0.049 (0.058)	0.024 (0.064)	0.334*** (0.088)	0.429*** (0.105)	0.428*** (0.094)	-0.268*** (0.066)	-0.300*** (0.048)	-0.348*** (0.052)
LATE	0.027 (0.049)	0.052 (0.062)	0.025 (0.066)	0.376*** (0.102)	0.435*** (0.110)	0.448*** (0.098)	-0.353*** (0.087)	-0.325*** (0.053)	-0.353*** (0.053)
N		3,704			1,777			1,927	

Notes: This table reports latent factor model based-estimates of the effect of receiving an offer from a given exam school versus no offer at all on enrollment at this exam school (First Stage), the effect of receiving an offer from a given exam school versus no offer at all on MCAS scores (Reduced Form), and the effect of enrollment at this exam school versus a traditional Boston public school on MCAS scores (LATE). The estimates are shown for all applicants as well as separately for applicants who do not receive an exam school offer and for applicants who receive an exam school offer. Standard errors based on nonparametric 5-step bootstrap shown in parentheses.

* significant at 10%, ** significant at 5%, *** significant at 1%

left and right of the placebo cutoffs receive the same exam school assignment. Therefore, these extrapolations should show no effects if the identifying assumptions are valid and the empirical specifications provide reasonable approximations to the underlying data generating processes.

Figure 8 plots the fits and extrapolations in the placebo RD experiments.⁷ The corresponding placebo estimates are reported in table 8. The fits and extrapolations lie close to each other, and the corresponding placebo effects are small in magnitude and statistically insignificant. This provides strong evidence supporting the identifying assumptions and empirical specifications.

6 Counterfactual Simulations

6.1 Description of the Admissions Reforms

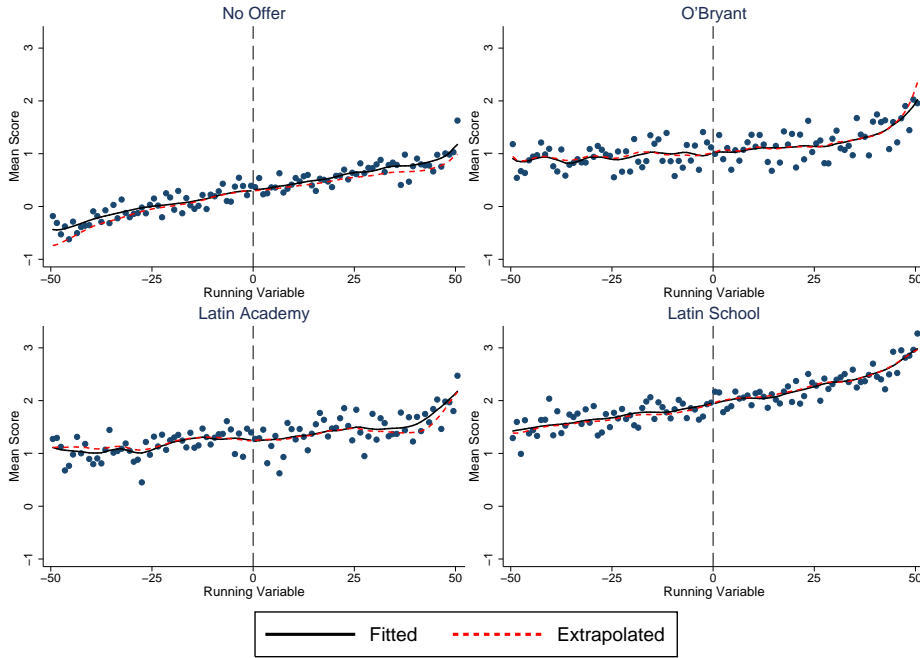
Estimates of exam school effects away from the admissions cutoffs are useful for predicting effects of reforms that change the exam school assignments of inframarginal applicants. A highly contentious example of this is the introduction of affirmative action in the currently purely achievement-based exam school admissions. I use the estimated latent factor model to predict how two particular affirmative action reforms would affect the achievement of exam school applicants.

The first reform considers minority preferences that were in place in Boston exam school admissions in 1975-1998. In this counterfactual admissions process 65% of the exam school seats are first assigned among all applicants. The remaining 35% of the exam school seats are then assigned among black and Hispanic applicants.

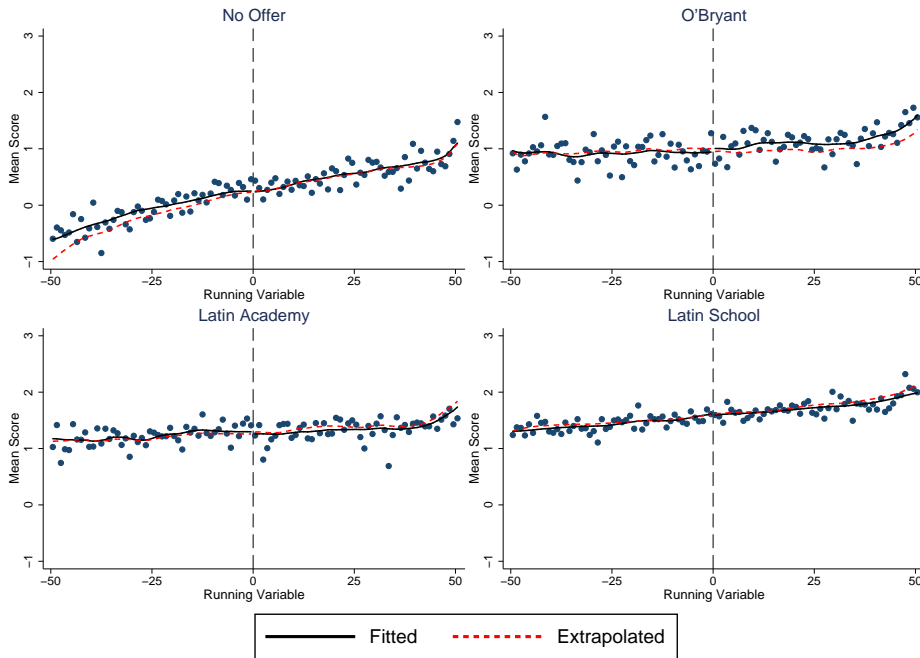
The second reform considers socioeconomic preferences that have been in place in Chicago exam school admissions since 2010. In this counterfactual admissions process 30% of the exam school seats are first assigned among all applicants. The remaining 70% of the exam school seats are then assigned within four socioeconomic tiers.⁸

⁷I transform the running variables into percentile ranks within each year in the placebo RD experiments and re-centered them to be 0 at the placebo cutoff for expositional purposes.

⁸The socioeconomic tiers are constructed by computing for each Census tract in Boston a socioeconomic index that takes into account the following five characteristics: (1) median family income, (2) percent of households occupied by the owner, (3) percent of families headed by a single parent, (4) percent of households where a language other than English is spoken, and (5) an educational attainment score. The educational attainment score is calculated based on the educational attainment distribution among individuals over the age of 25: $educational\ attainment\ score = 0.2 \times (\% \text{ less than high school diploma}) + 0.4 \times (\% \text{ high school diploma}) + 0.6 \times (\% \text{ some college}) + 0.8 \times (\% \text{ bachelors degree}) + 1.0 \times (\% \text{ advanced degree})$. The socioeconomic index for a given Census tract is given by the sum of its percentile ranks in each five characteristics among the Census tracts in Boston (for single-parent and non-English speaking



(a) Middle School MCAS Composite Score



(b) High School MCAS Composite Score

Figure 8: Reduced Form Extrapolations in the Placebo RD Experiments

Table 8: Extrapolated Reduced Form Effects in the Placebo RD Experiments

	No Offer (1)	O'Bryant (2)	Latin Academy (3)	Latin School (4)
<i>Panel A: All Applicants</i>				
Middle School MCAS	0.002 (0.053) 2,490	-0.000 (0.089) 690	0.042 (0.081) 728	-0.023 (0.072) 793
High School MCAS	-0.057 (0.064) 1,777	0.080 (0.068) 563	-0.041 (0.062) 625	-0.008 (0.045) 793
<i>Panel B: Below Placebo Cutoff</i>				
Middle School MCAS	-0.081 (0.082) 1,244	0.015 (0.075) 344	0.034 (0.069) 362	-0.034 (0.067) 395
High School MCAS	-0.133 (0.099) 887	0.027 (0.061) 280	-0.028 (0.055) 311	0.027 (0.049) 368
<i>Panel C: Above Placebo Cutoff</i>				
Middle School MCAS	0.085* (0.048) 1,246	-0.015 (0.134) 346	0.050 (0.127) 366	-0.013 (0.112) 395
High School MCAS	0.020 (0.056) 890	0.133 (0.102) 283	-0.054 (0.097) 314	-0.044 (0.068) 371

Notes: This table reports latent factor model-based estimates of the effects of placebo offers on MCAS scores. The estimates are shown for all applicants and separately for applicants whose running variables fall below and above the placebo admissions cutoffs. Standard errors based on nonparametric 5-step bootstrap shown in parentheses.

* significant at 10%, ** significant at 5%, *** significant at 1%

To study the effects of the two reforms I reassign exam school offers based on the counterfactual admissions processes, considering only applicants in the estimation sample. I then use the estimated latent factor model to predict average middle school and high school MCAS composite scores based on the reassigned exam school offers.⁹

An important feature of both of the reforms is that they cause substantial changes to the admissions cutoffs faced by exam school applicants. This means that predictions of the effects of the reforms based on exam school effects at the admissions cutoffs are likely to be misleading if there is considerable heterogeneity in the effects based on the running variables. Based on the results above, this is the case for Boston exam schools.

As with all counterfactuals, there are other dimensions that may also change as a result. First, the reforms potentially affect the composition of the pool of exam school applicants as some students face a decrease and some students an increase in their ex ante expected probability of being admitted to a given exam school.¹⁰ Second, the reforms lead to changes in the composition of applicants who are admitted to and consequently enroll at the exam schools. These changes may affect the sorting of teachers across schools (Jackson, 2009) and the way teaching is targeted (Duflo, Dupas, and Kremer, 2011). Finally, the changes in student composition may affect achievement directly through peer effects (Epple and Romano, 2011; Sacerdote, 2011).

6.2 Simulation Results

The introduction of minority and socioeconomic preferences have substantial effects on the exam school assignments. They lower the admissions cutoffs faced by minority applicants and applicants from lower socioeconomic tiers and increase the admissions cutoffs faced by non-minority applicants and applicants from higher socioeconomic tiers. Overall, 27 – 35% of the applicants are affected by the reforms. This can be seen from Table 9 that reports the actual and counterfactual exam school assignments.

Table 10 reports descriptive statistics for the exam school applicants based on their counterfactual assignments. The most notable compositional changes caused by the two reforms can be seen

households 1 minus the percentile rank). Each BPS student is assigned a socioeconomic index based on the Census tract they live in and the students are divided into socioeconomic tiers based on the quartiles of the socioeconomic index distribution in the BPS population within each year.

⁹This exercise is closely related to the literature on evaluating the effects of reallocations on the distribution of outcomes (Graham, 2011).

¹⁰Long (2004) and Andrews, Ranchhod, and Sathy (2010) find the college application behavior of high school students in California and Texas to be responsive to affirmative action. However, the evidence on this is somewhat mixed (Card and Krueger, 2005; Antonovics and Backes, 2013).

Table 9: Actual and Counterfactual Assignments under Minority and Socioeconomic Preferences

Counterfactual Assignment	Actual Assignment			
	No Offer (1)	O'Bryant (2)	Latin Academy (3)	Latin School (4)
<i>Panel A: Minority Preferences</i>				
No Offer	2418	221	113	39
O'Bryant	280	129	133	213
Latin Academy	88	389	268	45
Latin School	5	16	276	546
<i>Panel B: Socioeconomic Preferences</i>				
No Offer	2579	159	39	14
O'Bryant	203	319	146	87
Latin Academy	9	106	403	272
Latin School	0	171	202	470

Notes: This table reports the actual assignments and the counterfactual assignments under minority and socioeconomic preferences in the exam school admissions.

among applicants receiving no exam school offer and among applicants receiving an offer from Latin School. The reforms reduce baseline achievement and increase the share of minority applicants in the former group while having the opposite effects for the latter group. There are less marked changes in the composition of applicants receiving offers from O'Bryant and Latin Academy.

Table 11 reports the Average Reassignment Effects (ARE) of the reforms on middle school and high school MCAS composite scores. The ARE is given by the difference in average potential achievement among exam school applicants under the counterfactual and actual assignments:

$$\begin{aligned}
 & E \left[Y^{cf} - Y^{act} \right] \\
 &= \sum_{z'=0}^3 \sum_{z''=0}^3 P \left[Z^{cf} = z', Z^{act} = z'' \right] E \left[Y \left(S \left(z' \right) \right) - Y \left(S \left(z'' \right) \right) \mid Z^{cf} = z', Z^{act} = z'' \right]
 \end{aligned}$$

where Z^{act} and Z^{cf} are an applicant's actual and counterfactual assignments. Panel A reports the ARE estimates for full population of applicants and Panel B for applicants whose assignment is affected by the reform ($Z^{act} \neq Z^{cf}$).

The estimates for middle school MCAS composite scores show no effect of either reform on the average score among the full population of applicants. However, there is substantial heterogeneity in these effects. Introduction of minority preferences is estimated to reduce the average score

Table 10: Composition of Applicants by Counterfactual Assignments under Minority and Socioeconomic Preferences

	No Offer (1)	O'Bryant (2)	Latin Academy (3)	Latin School (4)
<i>Panel A: Minority Preferences</i>				
Female	0.502	0.588	0.629	0.572
Black	0.430	0.440	0.386	0.274
Hispanic	0.182	0.220	0.203	0.172
FRPL	0.810	0.771	0.715	0.556
LEP	0.116	0.030	0.022	0.018
Bilingual	0.386	0.396	0.380	0.391
SPED	0.073	0.009	0.013	0.005
English 4	0.277	0.981	1.215	1.668
Math 4	0.277	0.963	1.289	1.781
<i>Panel B: Socioeconomic Preferences</i>				
Female	0.514	0.572	0.597	0.575
Black	0.499	0.360	0.295	0.203
Hispanic	0.223	0.191	0.143	0.120
FRPL	0.813	0.728	0.673	0.624
LEP	0.107	0.058	0.030	0.017
Bilingual	0.359	0.423	0.391	0.445
SPED	0.073	0.012	0.006	0.006
English 4	0.261	1.067	1.309	1.556
Math 4	0.217	1.146	1.365	1.744
N	2,791	755	790	843

Notes: This table reports descriptive statistics for the exam school applicants by their counterfactual assignment under minority and socioeconomic preferences in the exam school admissions.

Table 11: Average Reassignment Effects of Introducing Minority or Socioeconomic Preferences

	Minority Preferences			Socioeconomic Preferences				
	All Applicants	Minority	Non-Minority	All Applicants	SES Tier 1	SES Tier 2	SES Tier 3	SES Tier 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: All Applicants</i>								
Middle School MCAS	0.001 (0.005) 4,701	-0.028*** (0.007) 2,741	0.043*** (0.010) 1,960	0.007 (0.005) 4,701	-0.032*** (0.011) 948	-0.005 (0.006) 1,113	0.004 (0.007) 1,155	0.041*** (0.011) 1,468
High School MCAS	0.023*** (0.004) 3,704	0.027*** (0.007) 2,086	0.017* (0.009) 1,618	0.015*** (0.005) 3,704	0.024** (0.011) 737	0.011* (0.006) 876	0.006 (0.005) 901	0.018* (0.010) 1,190
<i>Panel B: Affected Applicants</i>								
Middle School MCAS	0.003 (0.014) 1,670	-0.084*** (0.021) 932	0.113*** (0.026) 738	0.024 (0.019) 1,294	-0.092*** (0.032) 328	-0.021 (0.027) 263	0.021 (0.031) 241	0.133*** (0.034) 426
High School MCAS	0.062*** (0.011) 1,367	0.074*** (0.017) 754	0.046** (0.020) 613	0.050*** (0.016) 1,100	0.068** (0.031) 265	0.044* (0.023) 223	0.026 (0.024) 208	0.054* (0.029) 404

Notes: This table reports the latent factor model-based estimates of the effects of minority and socioeconomic preferences on MCAS scores. The estimates are shown for all applicants and separately for the applicant groups who face different admissions cutoffs after the reforms. The estimates are also shown separately for the affected applicants whose exam school assignment is altered by the reforms. Standard errors based on nonparametric 5-step bootstrap shown in parentheses.

* significant at 10%, ** significant at 5%, *** significant at 1%

among minority applicants and increase the average score among non-minority applicants. The introduction of socioeconomic preferences is estimated to reduce the average score among applicants from the lowest socioeconomic tier and to increase the average score among applicants from the highest socioeconomic tier. The estimates for high school MCAS composite scores show instead a positive effect of both reforms on the average score among the full population of applicants. These estimates show less marked heterogeneity, but the effects are somewhat larger for minority applicants and for applicants from lower socioeconomic tiers.

There are two mechanisms at work behind these estimates. First, the reforms lower the admissions cutoffs faced by minority applicants and applicants from lower socioeconomic tiers. This leads to more lower-achieving applicants, who experience achievement gains from exam school attendance, to gain access to exam schools. Second, the reforms increase the admissions cutoff faced by non-minority applicants and applicants from higher socioeconomic tiers. This leads to some of the higher-achieving applicants, who experience achievement losses from exam school attendance, to lose access to exam schools.

7 Conclusions

RD design allows for nonparametric identification and estimation of treatment effects for individuals at the cutoff value determining treatment assignment. However, many policies of interest change treatment assignment of individuals away from the cutoff, making knowledge of treatment effects for these individuals of substantial interest. A highly contentious example of this is affirmative action in selective schools that affects admissions cutoffs faced by different applicant groups.

The contributions of this paper are two-fold. First, I develop a new latent factor-based approach to the identification and estimation of treatment effects away from the cutoff in RD. The approach relies on the assumption that sources of omitted variables bias in an RD design can be modeled using unobserved latent factors. My main result is nonparametric identification of treatment effects for all values of the running variable based on the availability of multiple noisy measures of the latent factors. Second, I use the latent factor framework to estimate causal effects of Boston exam school attendance for the full population of applicants and to simulate effects of introducing either minority or socioeconomic preferences in Boston exam school admissions.

My findings highlight the local nature of RD estimates that show little evidence of causal effects for marginal applicants at admissions cutoffs (Abdulkadiroglu, Angrist, and Pathak, 2014). The

estimates of the latent factor model suggest that achievement gains from exam school attendance are larger among applicants with lower baseline measures of ability. As a result, lower-achieving applicants who currently fail to gain admission to Boston exam schools would experience substantial achievement gains from attending these schools. The simulations predict that the introduction of either minority or socioeconomic preferences in exam school admissions boosts average achievement among applicants. This is largely driven by achievement gains experienced by lower-achieving applicants who gain access to exam schools as a result of the policy change. These findings are of significant policy-relevance given ongoing discussion about the use of affirmative action in exam school admissions.

I focus in this paper on the heterogeneity in causal effects of exam school attendance based on the running variables used in the admissions process. This is a first-order concern when predicting effects of admissions reforms that widely change the exam school assignments of inframarginal applicants. However, as with all counterfactuals, there are other dimension that may change as a result of these reforms. First, affirmative action might lead to changes in the application behavior of students (Long, 2004; Andrews, Ranchhod, and Sathy, 2010). Second, affirmative action causes changes in student composition that may affect the sorting of teachers across schools (Jackson, 2009) and the way teaching is targeted (Duflo, Dupas, and Kremer, 2011). Finally, the changes in student composition may affect achievement directly through peer effects (Epple and Romano, 2011; Sacerdote, 2011). It is possible to model these channels in the latent factor framework, but this is left for future research.

Boston exam schools, as well as other selective schools, are a natural application for latent factor-based RD extrapolation as admissions are based on noisy measures of applicants' latent abilities. However, the approach is likely to prove useful also in other educational settings, such as gifted and talented programs (Bui, Craig, and Imberman, 2014) and remedial education (Jacob and Lefgren, 2004; Matsudaira, 2008). Moreover, the approach is likely to prove useful in health settings where treatment assignment is based on noisy measures of individuals' latent health conditions. Such settings include, for instance, the use of birth weight to assign additional medical care for newborns (Almond, Doyle, Kowalski, and Williams, 2010; Bharadwaj, Loken, and Neilson, 2013). As illustrated by the findings for Boston exam schools, local effects identified by RD do not necessarily represent the effects of policy interest. Latent factor-based RD extrapolation provides a framework for investigating external validity in these and other RD designs.

References

- AAKVIK, A., J. HECKMAN, AND E. VYTLACIL (2005): “Estimating Treatment Effects for Discrete Outcomes When Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs,” *Journal of Econometrics*, 125(1-2), 15–51.
- ABDULKADIROGLU, A., J. ANGRIST, AND P. PATHAK (2014): “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools,” *Econometrica*, 82(1), 137–196.
- AIGNER, D., C. HSIAO, A. KAPTEYN, AND T. WANSBEEK (1984): “Latent Variable Models in Econometrics,” *Handbook of Econometrics* 2, pp. 1321–1393. Elsevier B.V., Amsterdam.
- ALMOND, D., J. DOYLE, A. KOWALSKI, AND H. WILLIAMS (2010): “Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns,” *Quarterly Journal of Economics*, 125(2), 591–634.
- ANDREWS, D. (2002): “Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators,” *Econometrica*, 70(1), 119–162.
- ANDREWS, R., V. RANCHHOD, AND V. SATHY (2010): “Estimating the Responsiveness of College Applications to the Likelihood of Acceptance and Financial Assistance: Evidence from Texas,” *Economics of Education Review*, 29(1), 104–115.
- ANGRIST, J., AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics*. Princeton University Press, Princeton.
- ANGRIST, J., AND M. ROKKANEN (forthcoming): “Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff,” *Journal of the American Statistical Association*.
- ANTONOVICS, K., AND B. BACKES (2013): “Were Minority Students Discouraged from Applying to University of California Campuses after the Affirmative Action Ban?,” *Education Finance and Policy*, 8(2), 208–250.
- (2014): “The Effect of Banning Affirmative Action on College Admissions Policies and Student Quality,” *Journal of Human Resources*, 49(2), 295–322.
- ARCIDIACONO, P. (2005): “Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?,” *Econometrica*, 73(5), 1477–1524.

- BARGAIN, O., AND K. DOORLEY (2013): “Putting Structure on the RD Design: Social Transfers and Youth Inactivity in France,” IZA Discussion Paper 7508.
- BERTANHA, M., AND G. IMBENS (2014): “External Validity in Fuzzy Regression Discontinuity Designs,” NBER Working Paper 20773.
- BERTRAND, M., R. HANNA, AND S. MULLAINATHAN (2010): “Affirmative Action in Education: Evidence from Engineering College Admissions in India,” *Journal of Public Economics*, 94(1-2), 16–29.
- BHARADWAJ, P., K. LOKEN, AND C. NEILSON (2013): “Early Life Health Interventions and Academic Achievement,” *American Economic Review*, 103(5), 1862–1891.
- BLOOM, H. (2012): “Modern Regression Discontinuity Analysis,” *Journal of Research on Educational Effectiveness*, 5(1), 43–82.
- BONHOMME, S., AND U. SAUDER (2011): “Recovering Distributions in Difference-in-Differences Models: A Comparison of Selective and Comprehensive Schooling,” *Review of Economic Studies*, 93(2), 479–494.
- BUI, S., S. CRAIG, AND S. IMBERMAN (2014): “Is Gifted Education a Bright Idea? Assessing the Impacts of Gifted and Talented Programs,” *American Economic Journal: Economic Policy*, 6(3), 30–62.
- CARD, D., AND A. KRUEGER (2005): “Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas,” *Industrial & Labor Relations Review*, 58(3), 416–434.
- CHEN, X., H. HONG, AND D. NEKIPELOV (2011): “Nonlinear Models of Measurement Errors,” *Journal of Economic Literature*, 49(4), 901–937.
- COOLEY FRUEHWIRTH, J., S. NAVARRO, AND Y. TAKAHASHI (2014): “How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects,” Unpublished manuscript.
- CORTES, K. (2010): “Do Bans on Affirmative Action Hurt Minority Students? Evidence from the Texas Top 10% Plan,” *Economics of Education Review*, 29(6), 1110–1124.

- CUNHA, F., J. HECKMAN, AND S. SCHENNACH (2010): “Estimating the Technology of Cognitive and Noncognitive Skill Formulation,” *Econometrica*, 78(3), 883–931.
- DICKSON, L. (2006): “Does Ending Affirmative Action in College Admissions Lower the Percent of Minority Students Applying to College?,” *Economics of Education Review*, 25(1), 109–119.
- DINARDO, J., AND D. LEE (2011): “Program Evaluation and Research Designs,” Handbook of Labor Economics 4A, pp. 465–536. Elsevier B.V., Amsterdam.
- DOBBIE, W., AND R. FRYER (2014): “The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools,” *American Economic Journal: Applied Economics*, 6(3), 58–75.
- DONG, Y., AND A. LEWBEL (forthcoming): “Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models,” *Review of Economics and Statistics*.
- DUFLO, E., P. DUPAS, AND M. KREMER (2011): “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, 101(5), 1739–1774.
- EPPLE, D., AND R. ROMANO (2011): “Peer Effects in Education: A Survey of the Theory and Evidence,” Handbook of Social Economics 1B, pp. 1053–1163. Elsevier B.V., Amsterdam.
- FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.
- FRANCIS, A., AND M. TANNURI-PIANTO (2012): “Using Brazil’s Racial Continuum to Examine the Short-Term Effects of Affirmative Action in Higher Education,” *Journal of Human Resources*, 47(3), 754–784.
- GALE, D., AND L. SHAPLEY (1962): “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, 69(1), 9–15.
- GRAHAM, B. (2011): “Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers,” Handbook of Social Economics 1B, pp. 965–1052. Elsevier B.V., Amsterdam.
- HAHN, J., P. TODD, AND W. VAN DER KLAUW (2001): “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 69(1), 201–209.

- HINRICHS, P. (2012): “The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities,” *Review of Economics and Statistics*, 94(3), 712–722.
- HOWELL, J. (2010): “Assessing the Impact of Eliminating Affirmative Action in Higher Education,” *Journal of Labor Economics*, 28(1), 113–166.
- HU, Y., AND S. SCHENNACH (2008): “Instrumental Variable Treatment of Nonclassical Measurement Error Models,” *Econometrica*, 76(1), 195–216.
- IMBENS, G., AND K. KALYANARAMAN (2012): “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *Review of Economic Studies*, 79(3), 933–959.
- JACKSON, K. (2009): “Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation,” *Journal of Labor Economics*, 27(2), 213–256.
- (2010): “Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago,” *Economic Journal*, 120(549), 1399–1429.
- JACOB, B., AND L. LEFGREN (2004): “Remedial Education and Student Achievement: A Regression-Discontinuity Analysis,” *Review of Economic and Statistics*, 86(1), 226–244.
- KAPOR, A. (2015): “Distributional Effects of Race-Blind Affirmative Action,” Unpublished manuscript.
- LEE, D. (2008): “Randomized Experiments from Non-Random Selection in U.S. House Elections,” *Journal of Econometrics*, 142(2), 675–697.
- LEE, D., AND T. LEMIEUX (2010): “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 48, 281–355.
- LEHMANN, E., AND J. ROMANO (2005): *Testing Statistical Hypothesis*. Springer, New York.
- LEHMANN, E., AND H. SCHEFFE (1950): “Completeness, Similar Regions, and Unbiased Estimation: Part I,” *Sankhya*, 10(4), 305–340.
- (1955): “Completeness, Similar Regions, and Unbiased Estimation: Part II,” *Sankhya*, 15(3), 219–236.

- LONG, M. (2004): “College Applications and the Effects of Affirmative Action,” *Journal of Econometrics*, 121(1-2), 319–342.
- MATSUDAIRA, J. (2008): “Mandatory Summer School and Student Achievement,” *Journal of Econometrics*, 142(8), 829–850.
- NEWHEY, W., AND J. POWELL (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71(5), 1565–1578.
- ROTHSTEIN, J., AND A. YOON (2008): “Affirmative Action in Law School Admissions: What Do Racial Preferences Do?,” *University of Chicago Law Review*, 75(2), 649–714.
- SACERDOTE, B. (2011): “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” *Handbook of Economics of Education* 3, pp. 249–277. Elsevier B.V., Amsterdam.
- WING, C., AND T. COOK (2013): “Strengthening The Regression Discontinuity Design Using Additional Design Elements: A Within-Study Comparison,” *Journal of Policy analysis and Management*, 32(4), 853–877.

Appendix A: Proofs

Proof of Lemma 2 The result follows directly from the Law of Iterated Expectations. ■

Proof of Theorem 1 Assumptions C.1, C.3, C.4, and C.5 correspond to assumptions 1, 3, 4, and 5 in Hu and Schennach (2008) with $y = M_3$, $x = M_1$, $z = M_2$, and $x^* = \theta$ in their notation. Furthermore, as shown in Cunha, Heckman, and Schennach (2010), Assumption C.2 is equivalent to Assumption 2 in Hu and Schennach (2008). The identification of the conditional distributions $f_{M_1|\theta}$, $f_{M_3|\theta}$, and $f_{\theta|M_2}$ then follows from Theorem 1 in Hu and Schennach (2008). ■

Proof of Theorem 2 Assumption D.1 allows one to write down the integral equations

$$\begin{aligned} E[Y | M = m^0, D = 0] &= E\{E[Y(0) | \theta] | M = m^0, D = 0\} \\ E[Y | M = m^1, D = 1] &= E\{E[Y(1) | \theta] | M = m^1, D = 1\}. \end{aligned}$$

The uniqueness of the solutions to these equations follows directly from Assumption D.3. To see this, suppose that in addition to $E[Y(0) | \theta]$ there exists some $\tilde{E}[Y(0) | \theta]$ such that

$$P\{E[Y(0) | \theta] \neq \tilde{E}[Y(0) | \theta]\} > 0$$

also satisfying the above equation for all $m^0 \in M^0$. Thus,

$$E\{E[Y(0) | \theta] - \tilde{E}[Y(0) | \theta] | R = r^0, D = 0\} = 0$$

for all $m^0 \in M^0$. By Assumption D.3, this implies that $E[Y(0) | \theta] - \tilde{E}[Y(0) | \theta] = 0$ for all $m^0 \in M^0$, which leads to a contradiction. An analogous argument can be given for the uniqueness of $E[Y(1) | \theta]$. Finally, Assumption D.2 guarantees that $E[Y(0) | \theta]$ and $E[Y(1) | \theta]$ are determined for all $\theta \in \Theta$. ■

Proof of Lemma 4 Using Assumptions G.1 and G.2, one can write

$$\begin{aligned}
& E \{ E [Y (D (1)) - Y (D (0)) \mid \theta] \mid R = r \} \\
&= \int E [Y (D (1)) - Y (D (0)) \mid \theta] f_{\theta|R} (\theta \mid r) d\theta \\
&= \int E [Y (1) - Y (0) \mid D (1) > D (0) , \theta] P [D (1) > D (0) \mid \theta] f_{\theta|R} (\theta \mid r) d\theta \\
&= \int E [Y (1) - Y (0) \mid D (1) > D (0) , \theta , R = r] \\
&\quad \times P [D (1) > D (0) \mid \theta , R = r] f_{\theta|R} (\theta \mid r) d\theta.
\end{aligned} \tag{1}$$

Furthermore, using the fact that

$$\begin{aligned}
& P [D (1) > D (0) \mid \theta , R = r] f_{\theta|R} (\theta \mid r) \\
&= f_{\theta|R} (\theta , D (1) > D (0) \mid r) \\
&= P [D (1) > D (0) \mid R = r] f_{\theta|R, D(0), D(1)} (\theta \mid r , 0 , 1) ,
\end{aligned}$$

Equation (1) becomes

$$\begin{aligned}
& E \{ E [Y (D (1)) - Y (D (0)) \mid \theta] \mid R = r \} \\
&= P [D (1) > D (0) \mid R = r] \\
&\quad \times \int E [Y (1) - Y (0) \mid D (1) > D (0) , \theta , R = r] f_{\theta|R, D(0), D(1)} (\theta \mid r , 0 , 1) d\theta \\
&= P [D (1) > D (0) \mid R = r] E [Y (1) - Y (0) \mid D (1) > D (0) , R = r] .
\end{aligned} \tag{2}$$

Using similar arguments, one can write

$$\begin{aligned}
& E \{ E [D (1) - D (0) \mid \theta] \mid R = r \} \\
&= \int E [D (1) - D (0) \mid \theta] f_{\theta|R} (\theta \mid r) d\theta \\
&= \int P [D (1) > D (0) \mid \theta] f_{\theta|R} (\theta \mid r) d\theta \\
&= \int P [D (1) > D (0) \mid \theta , R = r] f_{\theta|R} (\theta \mid r) d\theta \\
&= P [D (1) > D (0) \mid R = r] \int f_{\theta|R, D(0), D(1)} (\theta \mid r , 0 , 1) d\theta \\
&= P [D (1) > D (0) \mid R = r] .
\end{aligned} \tag{3}$$

The result then follows from Equations (2) and (3). ■

Proof of Theorem 3 The proof is analogous to the proof of Theorem 2. ■

Appendix B: Deferred Acceptance Algorithm and the Definition of Sharp Samples

The student-proposing DA algorithm assigns exam school offers as follows:

- Round 1: Applicants are considered for a seat in their most preferred exam school. Each exam schools rejects the lowest-ranking applicants in excess of its capacity. The rest of the applicants are provisionally admitted.
- Round $k > 1$: Applicants rejected in Round $k - 1$ are considered for a seat in their next most preferred exam school. Each exam schools considers these applicants together with the provisionally admitted applicants from Round $k - 1$ and rejects the lowest-ranking students in excess of its capacity. The rest of the students are provisionally admitted.

The algorithm terminates once either all applicants are assigned an offer from one of the exam schools or all applicants with no offer are rejected by every exam school in their preference ordering.

This produces an admissions cutoff for each exam school that is given by the lowest rank among applicants admitted to the school. By definition none of the applicants with a ranking below this cutoff are admitted to the school. On the other hand, applicants with ranks at or above this cutoff are admitted to either this school or a more preferred school depending on their position relative to the admissions cutoffs for these schools.

The DA algorithm-based admissions process implies that only a subset of the applicants to a given exam school that clear the admissions cutoff are admitted to the school. There are three ways in which an applicant can be admitted to exam school s given the admissions cutoffs:

1. Exam school s is the applicant's 1st choice, and she clears the admissions cutoff.
2. The applicant does not clear the admissions cutoff for her 1st choice, exam school s is her 2nd choice, and she clears the admissions cutoff.
3. The applicant does not clear the admissions cutoff for her 1st or 2nd choice, exam school s is her 3rd choice, and she clears the admissions cutoff.

It is possible to define for each exam school a sharp sample that consist of applicants who are admitted to this school if and only if they clear the admissions cutoff (Abdulkadiroglu, Angrist, and Pathak, 2014). The sharp sample for exam school s is the union of the following three subsets of applicants:

1. Exam school s is the applicant's 1st choice.
2. The applicant does not clear the admissions cutoff for her 1st choice, and exam school s is her 2nd choice.
3. The applicant does not clear the admissions cutoff for her 1st or 2nd choice, and exam school s is her 3rd choice.

Note that each applicant is included in the sharp sample for at least one exam school (the exam school they listed as their first choice), but an applicant can be included in the sharp sample for more than one exam school. For instance, an applicant who does not clear the admissions cutoff for any of the exam schools is included in the sharp samples for all three schools.

Appendix C: Identification of the Parametric Latent Factor Model

Identification of the Measurement Model

Under the parametric measurement model specified in Section 4 (ignoring additional covariates) the mean and covariances of the measures can be written as

$$\begin{aligned}
 E [M_1^k] &= \mu_{\theta_k} \\
 E [M_2^k] &= \mu_{M_2^k} + \lambda_{M_2^k} \mu_{\theta_k} \\
 E [M_3^k] &= \mu_{M_3^k} + \lambda_{M_3^k} \mu_{\theta_k} \\
 Cov [M_1^k, M_2^k] &= \lambda_{M_2^k} \sigma_{\theta_k}^2 \\
 Cov [M_1^k, M_3^k] &= \lambda_{M_3^k} \sigma_{\theta_k}^2 \\
 Cov [M_2^k, M_3^k] &= \lambda_{M_2^k} \lambda_{M_3^k} \sigma_{\theta_k}^2 \\
 Cov [M_1^E, M_1^M] &= \sigma_{\theta_E \theta_M}
 \end{aligned}$$

for $k = E, M$. From these equations one can solve for μ_{θ_k} , $\sigma_{\theta_k}^2$, $\sigma_{\theta_E \theta_M}$, $\mu_{M_2^k}$ and $\lambda_{M_2^k}$ that are given by

$$\begin{aligned}
 \mu_{\theta_k} &= E [M_1^k] \\
 \lambda_{M_2^k} &= \frac{Cov [M_2^k, M_3^k]}{Cov [M_1^k, M_3^k]} \\
 \lambda_{M_3^k} &= \frac{Cov [M_3^k, M_2^k]}{Cov [M_1^k, M_2^k]} \\
 \mu_{M_2^k} &= E [M_2^k] - \lambda_{M_2^k} \mu_{\theta_k} \\
 \mu_{M_3^k} &= E [M_3^k] - \lambda_{M_3^k} \mu_{\theta_k} \\
 \sigma_{\theta_k}^2 &= \frac{Cov [M_1^k, M_2^k]}{\lambda_{M_2^k}} \\
 \sigma_{\theta_E \theta_M} &= Cov [M_1^E, M_1^M]
 \end{aligned}$$

for $k = E, M$, provided that $Cov [M_1^k, M_2^k]$, $Cov [M_1^k, M_3^k] \neq 0$, $k = E, M$.

Furthermore, the conditional means and covariances of the measures can be written as

$$\begin{aligned}
E \left[M_1^k \mid M_2^k \right] &= E \left[\theta_k \mid M_2^k \right] \\
E \left[M_1^k \mid M_3^k \right] &= E \left[\theta_k \mid M_3^k \right] \\
Cov \left[M_1^k, M_3^k \mid M_2^k \right] &= \lambda_{M_3^k} Var \left[\theta_k \mid M_2^k \right] \\
Cov \left[M_1^k, M_2^k \mid M_3^k \right] &= \lambda_{M_2^k} Var \left[\theta_k \mid M_3^k \right]
\end{aligned}$$

for $k = E, M$. From these equations one can solve for $E \left[\theta_k \mid M_2^k \right]$, $E \left[\theta_k \mid M_3^k \right]$, $Var \left[\theta_k \mid M_2^k \right]$, and $Var \left[\theta_k \mid M_3^k \right]$ that are given by

$$\begin{aligned}
E \left[\theta_k \mid M_2^k \right] &= E \left[M_1^k \mid M_2^k \right] \\
E \left[\theta_k \mid M_3^k \right] &= E \left[M_1^k \mid M_3^k \right] \\
Var \left[\theta_k \mid M_2^k \right] &= \frac{Cov \left[M_1^k, M_3^k \mid M_2^k \right]}{\lambda_{M_3^k}} \\
Var \left[\theta_k \mid M_3^k \right] &= \frac{Cov \left[M_1^k, M_2^k \mid M_3^k \right]}{\lambda_{M_2^k}}
\end{aligned}$$

for $k = E, M$.

Finally, the conditional variances of the measures can be written as

$$\begin{aligned}
Var \left[M_1^k \mid M_2^k \right] &= E \left[Var \left[M_1^k \mid \theta_k \right] \mid M_2^k \right] + Var \left[E \left[M_1^k \mid \theta_k \right] \mid M_2^k \right] \\
&= E \left[\exp \left(2 \left(\gamma_{M_1^k} + \delta_{M_1^k} \right) \theta_k \right) \mid M_2^k \right] + Var \left[\theta_k \mid M_2^k \right] \\
&= \exp \left(2 \left(\gamma_{M_1^k} + E \left[\theta_k \mid M_2^k \right] \delta_{M_1^k} + Var \left[\theta_k \mid M_2^k \right] \delta_{M_1^k}^2 \right) \right) + Var \left[\theta_k \mid M_2^k \right] \\
Var \left[M_2^k \mid M_3^k \right] &= E \left[Var \left[M_2^k \mid \theta_k \right] \mid M_3^k \right] + Var \left[E \left[M_2^k \mid \theta_k \right] \mid M_3^k \right] \\
&= E \left[\exp \left(2 \left(\gamma_{M_2^k} + \delta_{M_2^k} \right) \theta_k \right) \mid M_3^k \right] + \lambda_{M_2^k}^2 Var \left[\theta_k \mid M_3^k \right] \\
&= \exp \left(2 \left(\gamma_{M_2^k} + E \left[\theta_k \mid M_3^k \right] \delta_{M_2^k} + Var \left[\theta_k \mid M_3^k \right] \delta_{M_2^k}^2 \right) \right) + \lambda_{M_2^k}^2 Var \left[\theta_k \mid M_3^k \right] \\
Var \left[M_3^k \mid M_2^k \right] &= E \left[Var \left[M_3^k \mid \theta_k \right] \mid M_2^k \right] + Var \left[E \left[M_3^k \mid \theta_k \right] \mid M_2^k \right] \\
&= E \left[\exp \left(2 \left(\gamma_{M_3^k} + \delta_{M_3^k} \right) \theta_k \right) \mid M_2^k \right] + \lambda_{M_3^k}^2 Var \left[\theta_k \mid M_2^k \right] \\
&= \exp \left(2 \left(\gamma_{M_3^k} + E \left[\theta_k \mid M_2^k \right] \delta_{M_3^k} + Var \left[\theta_k \mid M_2^k \right] \delta_{M_3^k}^2 \right) \right) + \lambda_{M_3^k}^2 Var \left[\theta_k \mid M_2^k \right].
\end{aligned}$$

These can be further modified to

$$\begin{aligned}
\frac{1}{2} \log \left(\text{Var} \left[M_1^k \mid M_2^k \right] - \text{Var} \left[\theta_k \mid M_2^k \right] \right) &= \gamma_{M_1^k} + E \left[\theta_k \mid M_2^k \right] \delta_{M_1^k} + \text{Var} \left[\theta_k \mid M_2^k \right] \delta_{M_1^k}^2 \\
\frac{1}{2} \log \left(\text{Var} \left[M_2^k \mid M_3^k \right] - \lambda_{M_2^k}^2 \text{Var} \left[\theta_k \mid M_3^k \right] \right) &= \gamma_{M_2^k} + E \left[\theta_k \mid M_3^k \right] \delta_{M_2^k} + \text{Var} \left[\theta_k \mid M_3^k \right] \delta_{M_2^k}^2 \\
\frac{1}{2} \log \left(\text{Var} \left[M_3^k \mid M_2^k \right] - \lambda_{M_3^k}^2 \text{Var} \left[\theta_k \mid M_2^k \right] \right) &= \gamma_{M_3^k} + E \left[\theta_k \mid M_2^k \right] \delta_{M_3^k} + \text{Var} \left[\theta_k \mid M_2^k \right] \delta_{M_3^k}^2
\end{aligned}$$

for $k = E, M$.

Thus, the parameters $\gamma_{M_1^k}$, $\delta_{M_1^k}$, $\gamma_{M_2^k}$, $\delta_{M_2^k}$, $\gamma_{M_3^k}$, and $\delta_{M_3^k}$ can be solved from

$$\begin{aligned}
&\begin{bmatrix} \frac{1}{2} \log \left(\text{Var} \left[M_1^k \mid M_2^k = m_1 \right] - \text{Var} \left[\theta_k \mid M_2^k = m_1 \right] \right) \\ \frac{1}{2} \log \left(\text{Var} \left[M_1^k \mid M_2^k = m_2 \right] - \text{Var} \left[\theta_k \mid M_2^k = m_2 \right] \right) \end{bmatrix} \\
&= \begin{bmatrix} \gamma_{M_1^k} + E \left[\theta_k \mid M_2^k = m_1 \right] \delta_{M_1^k} + \text{Var} \left[\theta_k \mid M_2^k = m_1 \right] \delta_{M_1^k}^2 \\ \gamma_{M_1^k} + E \left[\theta_k \mid M_2^k = m_2 \right] \delta_{M_1^k} + \text{Var} \left[\theta_k \mid M_2^k = m_2 \right] \delta_{M_1^k}^2 \end{bmatrix} \\
&\begin{bmatrix} \frac{1}{2} \log \left(\text{Var} \left[M_2^k \mid M_3^k = m_1 \right] - \lambda_{M_2^k}^2 \text{Var} \left[\theta_k \mid M_3^k = m_1 \right] \right) \\ \frac{1}{2} \log \left(\text{Var} \left[M_2^k \mid M_3^k = m_2 \right] - \lambda_{M_2^k}^2 \text{Var} \left[\theta_k \mid M_3^k = m_2 \right] \right) \end{bmatrix} \\
&= \begin{bmatrix} \gamma_{M_2^k} + E \left[\theta_k \mid M_3^k = m_1 \right] \delta_{M_2^k} + \text{Var} \left[\theta_k \mid M_3^k = m_1 \right] \delta_{M_2^k}^2 \\ \gamma_{M_2^k} + E \left[\theta_k \mid M_3^k = m_2 \right] \delta_{M_2^k} + \text{Var} \left[\theta_k \mid M_3^k = m_2 \right] \delta_{M_2^k}^2 \end{bmatrix} \\
&\begin{bmatrix} \frac{1}{2} \log \left(\text{Var} \left[M_3^k \mid M_2^k = m_1 \right] - \lambda_{M_3^k}^2 \text{Var} \left[\theta_k \mid M_2^k = m_1 \right] \right) \\ \frac{1}{2} \log \left(\text{Var} \left[M_3^k \mid M_2^k = m_2 \right] - \lambda_{M_3^k}^2 \text{Var} \left[\theta_k \mid M_2^k = m_2 \right] \right) \end{bmatrix} \\
&= \begin{bmatrix} \gamma_{M_3^k} + E \left[\theta_k \mid M_2^k = m_1 \right] \delta_{M_3^k} + \text{Var} \left[\theta_k \mid M_2^k = m_1 \right] \delta_{M_3^k}^2 \\ \gamma_{M_3^k} + E \left[\theta_k \mid M_2^k = m_2 \right] \delta_{M_3^k} + \text{Var} \left[\theta_k \mid M_2^k = m_2 \right] \delta_{M_3^k}^2 \end{bmatrix},
\end{aligned}$$

provided that the matrices

$$\begin{aligned}
&\begin{bmatrix} 1 & E \left[\theta_k \mid M_2^k = m_1 \right] + 2\text{Var} \left[\theta_k \mid M_2^k = m_1 \right] \delta_{M_1^k} \\ 1 & E \left[\theta_k \mid M_2^k = m_2 \right] + 2\text{Var} \left[\theta_k \mid M_2^k = m_2 \right] \delta_{M_1^k} \end{bmatrix} \\
&\begin{bmatrix} 1 & E \left[\theta_k \mid M_3^k = m_1 \right] + 2\text{Var} \left[\theta_k \mid M_3^k = m_1 \right] \delta_{M_2^k} \\ 1 & E \left[\theta_k \mid M_3^k = m_2 \right] + 2\text{Var} \left[\theta_k \mid M_3^k = m_2 \right] \delta_{M_2^k} \end{bmatrix} \\
&\begin{bmatrix} 1 & E \left[\theta_k \mid M_2^k = m_1 \right] + 2\text{Var} \left[\theta_k \mid M_2^k = m_1 \right] \delta_{M_3^k} \\ 1 & E \left[\theta_k \mid M_2^k = m_2 \right] + 2\text{Var} \left[\theta_k \mid M_2^k = m_2 \right] \delta_{M_3^k} \end{bmatrix}
\end{aligned}$$

are of full rank.

Identification of the Latent Outcome Models

Under the parametric latent outcome model specified in Section 4 (ignoring additional covariates) the conditional expectation of an outcome Y can be written as

$$\begin{aligned} E[Y | M, Z] &= E[Y(S(Z)) | M, Z] \\ &= \alpha_{Y(S(Z))} + \beta_{Y(S(Z))}^E E[\theta_E | M, Z] + \beta_{Y(S(Z))}^M E[\theta_M | M, Z] \end{aligned}$$

for $Z = 0, 1, 2, 3$. Thus, the parameters $\alpha_{Y(S(Z))}$, $\beta_{Y(S(Z))}^E$, and $\beta_{Y(S(Z))}^M$ can be solved from

$$\begin{aligned} \begin{bmatrix} E[Y | M = m_1^Z, Z] \\ E[Y | M = m_2^Z, Z] \\ E[Y | M = m_3^Z, Z] \end{bmatrix} &= \begin{bmatrix} 1 & E[\theta_E | M = m_1^Z, Z] & E[\theta_M | M = m_1^Z, Z] \\ 1 & E[\theta_E | M = m_2^Z, Z] & E[\theta_M | M = m_2^Z, Z] \\ 1 & E[\theta_E | M = m_3^Z, Z] & E[\theta_M | M = m_3^Z, Z] \end{bmatrix} \begin{bmatrix} \alpha_{Y(S(z))} \\ \beta_{Y(S(z))}^E \\ \beta_{Y(S(z))}^M \end{bmatrix} \\ \Rightarrow \begin{bmatrix} \alpha_{Y(S(Z))} \\ \beta_{Y(S(Z))}^E \\ \beta_{Y(S(Z))}^M \end{bmatrix} &= \begin{bmatrix} 1 & E[\theta_E | M = m_1^Z, Z] & E[\theta_M | M = m_1^Z, Z] \\ 1 & E[\theta_E | M = m_2^Z, Z] & E[\theta_M | M = m_2^Z, Z] \\ 1 & E[\theta_E | M = m_3^Z, Z] & E[\theta_M | M = m_3^Z, Z] \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} E[Y | M = m_1^Z, Z] \\ E[Y | M = m_2^Z, Z] \\ E[Y | M = m_3^Z, Z] \end{bmatrix} \end{aligned}$$

where $m_1^Z, m_2^Z, m_3^Z \in \mathcal{M}^Z$, provided that the matrix

$$\begin{bmatrix} 1 & E[\theta_E | M = m_1^Z, Z] & E[\theta_M | M = m_1^Z, Z] \\ 1 & E[\theta_E | M = m_2^Z, Z] & E[\theta_M | M = m_2^Z, Z] \\ 1 & E[\theta_E | M = m_3^Z, Z] & E[\theta_M | M = m_3^Z, Z] \end{bmatrix}$$

is of full rank.

The identification of $\alpha_{D_s(Z)}$, $\beta_{D_s(Z)}^E$, and $\beta_{D_s(Z)}^M$, $Z = 0, 1, 2, 3$, is analogous.