Online Appendix: Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff

Joshua D. Angrist
MIT Department of Economics and NBER

Miikka Rokkanen
Columbia University Department of Economics

December 2014

This appendix contains additional material to accompany our paper, "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff". Section 1 presents proofs of Theorems 1 and 2 in the main paper. Section 2 discusses the theoretical basis for parametric extrapolation in an RD design, illustrating it with additional results for 10th grade Math. Section 3 discusses an alternative to the conditional independence test discussed in the main paper in Section 3.1. The test here is based on a comparison of RD and CIA estimates of the effects at the cutoff.

1 Proofs

Proof of Theorem 1

We continue to assume that GCIA and other LATE assumptions hold. Given these assumptions, Theorem 3.1 in Abadie (2003) implies that for any measurable function, $g(y_i, W_i, x_i)$, we have

$$E\left[g\left(y_i, W_i, x_i\right) \mid x_i, W_{1i} > W_{0i}\right] = \frac{1}{P\left[W_{1i} > W_{0i} \mid x_i\right]} E\left[\kappa\left(W_i, D_i, x_i\right) g\left(y_i, W_i, x_i\right) \mid x_i\right] \quad (1)$$

where

$$\kappa\left(W_i, D_i, x_i\right) = 1 - \frac{W_i\left(1 - D_i\right)}{1 - P\left[D_i = 1 \mid x_i\right]} - \frac{\left(1 - W_i\right) D_i}{P\left[D_i = 1 \mid x_i\right]}$$

and

$$E\left[g\left(Y_{Wi}, x_i\right) \mid x_i, W_{1i} > W_{0i}\right] = \frac{1}{P\left[W_{1i} > W_{0i} \mid x_i\right]} E\left[\kappa_W\left(W_i, D_i, x_i\right) g\left(y_i, x_i\right) \mid x_i\right],$$

where $W \in \{0,1\}$ and

$$\begin{aligned}
\kappa_0\left(W_i, D_i, x_i\right) &= \left(1 - W_i\right) \frac{P\left[D_i = 1 \mid x_i\right] - D_i}{\left(1 - P\left[D_i = 1 \mid x_i\right]\right) P\left[D_i = 1 \mid x_i\right]} \\
\kappa_1\left(W_i, D_i, x_i\right) &= W_i \frac{D_i - P\left[D_i = 1 \mid x_i\right]}{\left(1 - P\left[D_i = 1 \mid x_i\right]\right) P\left[D_i = 1 \mid x_i\right]}.
\end{aligned}$$

Using the GCIA, we can simplify as follows:

$$\begin{aligned}
& E\left[g\left(Y_{Wi}, x_i\right) \mid W_{1i} > W_{0i}, 0 < r_i \leq c\right] \\
&= E\left\{E\left[g\left(Y_{Wi}, x_i\right) \mid x_i, W_{1i} > W_{0i}\right] \mid W_{1i} > W_{0i}, 0 < r_i \leq c\right\} \\
&= \int \frac{1}{P\left[W_{1i} > W_{0i} \mid x_i\right]} E\left[\kappa_W\left(W_i, D_i, x_i\right) g\left(y_i, x_i\right) \mid X\right] dP\left[x_i \mid W_{1i} > W_{0i}, 0 < r_i \leq c\right] \\
&= \frac{1}{P\left[W_{1i} > W_{0i} \mid 0 < r_i \leq c\right]} \int E\left[\kappa_W\left(W_i, D_i, x_i\right) g\left(y, x_i\right) \mid x_i\right] \frac{P\left[0 < r_i \leq c \mid x_i\right]}{P\left[0 < r_i \leq c\right]} dP\left[x_i\right] \\
&= \frac{1}{P\left[W_{1i} > W_{0i} \mid 0 < r_i \leq c\right]} E\left[\kappa_W\left(W_i, D_i, x_i\right) \frac{P\left[0 < r_i \leq c \mid x_i\right]}{P\left[0 < r_i \leq c\right]} g\left(y_i, x_i\right)\right]. \qquad (2)
\end{aligned}$$

This implies that LATE can be written:

$$\begin{aligned}
& E\left[Y_{1i} - Y_{0i} \mid W_{1i} > W_{0i}, 0 < r_i \leq c\right] \\
&= E\left[Y_{1i} \mid W_{1i} > W_{0i}, 0 < r_i \leq c\right] - E\left[Y_{0i} \mid W_{1i} > W_{0i}, 0 < r_i \leq c\right] \\
&= \frac{1}{P\left[W_{1i} > W_{0i} \mid 0 < r_i \leq c\right]} E\left[\psi\left(D_i, x_i\right) \frac{P\left[0 < r_i \leq c \mid x_i\right]}{P\left[0 < r_i \leq c\right]} y_i\right]
\end{aligned}$$

where

$$\begin{aligned}
\psi\left(D_i, x_i\right) &= \kappa_1\left(W_i, D_i, x_i\right) - \kappa_0\left(W_i, D_i, x_i\right) \\
&= \frac{D_i - P\left[D_i = 1 \mid x_i\right]}{\left(1 - P\left[D_i = 1 \mid x_i\right]\right) P\left[D_i = 1 \mid x_i\right]}.
\end{aligned}$$

Finally, by setting $g\left(y_i, W_i, x_i\right) = 1$ in equation (1) we get:

$$P\left[W_{1i} > W_{0i} \mid x_i\right] = E\left[\kappa\left(W_i, D_i, x_i\right) \mid x_i\right].$$

Using the same steps as in equation (2), the GCIA implies:

$$\begin{aligned}
P\left[W_{1i} > W_{0i} \mid 0 < r_i \leq c\right] &= E\left\{P\left[W_{1i} > W_{0i} \mid x_i\right] \mid 0 < r_i \leq c\right\} \\
&= E\left[\kappa\left(W_i, D_i, x_i\right) \frac{P\left[0 < r_i \leq c \mid x_i\right]}{P\left[0 < x_i \leq c\right]}\right].
\end{aligned}$$

Proof of Theorem 2

Theorem 1 in Angrist and Imbens (1995) implies:

$$E[y_i \mid D_i = 1, x_i] - E[y_i \mid D_i = 0, x_i] = \sum_j P[w_{1i} \geq j > w_{0i} \mid x_i] \, E[Y_{ji} - Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}, x_i]$$

$$E[w_i \mid D_i = 1, x_i] - E[w_i \mid D_i = 0, x_i] = \sum_j P[w_{1i} \geq j > w_{0i} \mid x_i] \, .$$

Given the GCIA, we have:

$$\begin{aligned}
&E\{E[y_i \mid D_i = 1, x_i] - E[y_i \mid D_i = 0, x_i] \mid 0 < r_i \leq c\} \\
&= \sum_j \int P[w_{1i} \geq j > w_{0i} \mid x_i] \, E[Y_{ji} - Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}, x_i] \, dP[x_i \mid 0 < r_i \leq c] \\
&= \sum_j \int P[w_{1i} \geq j > w_{0i} \mid x_i, 0 \leq r_i \leq c] \, E[Y_{ji} - Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}, x_i] \, dP[x_i \mid 0 < r_i \leq c] \\
&= \sum_j P[w_{1i} \geq j > w_{0i} \mid 0 < r_i \leq c] \\
&\quad \times \int E[Y_{ji} - Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}, x_i] \, dP[x_i \mid w_{1i} \geq j > w_{0i}, 0 < r_i \leq c] \\
&= \sum_j P[w_{1i} \geq j > w_{0i} \mid 0 < r_i \leq c] \, E[Y_{ji} - Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}, 0 < r_i \leq c] \, .
\end{aligned}$$

The GCIA can similarly be shown to imply:

$$\begin{aligned}
&E[E[w_i \mid D_i = 1, x_i] - E[w_i \mid D_i = 0, x_i] \mid 0 < r_i \leq c] \\
&= \sum_j P[w_{1i} \geq j > w_{0i} \mid 0 < r_i \leq c] \, .
\end{aligned}$$

Combining these results, the ACR can be written:

$$\begin{aligned}
&\frac{E\{E[y_i \mid D_i = 1, x_i] - E[y_i \mid D_i = 0, x_i] \mid 0 < r_i \leq c\}}{E\{E[w_i \mid D_i = 1, x_i] - E[w_i \mid D_i = 0, x_i] \mid 0 < r_i \leq c\}} \\
&= \sum_j \nu_{jc} E[Y_{ji} - Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}, 0 < r_i \leq c]
\end{aligned}$$

where

$$\nu_{ijc} = \frac{P[w_{1i} \geq j > w_{0i} \mid 0 < r_i \leq c]}{\sum_\ell P[w_{1i} \geq \ell > w_{0i} \mid 0 < r_i \leq c]} .$$

2 Parametric Extrapolation

The running variable is the star covariate in any RD scene, but the role played by the running variable is distinct from that played by covariates in matching and regression-control strategies. In the latter, we look to comparisons of treated and non-treated observations *conditional* on covariates to eliminate omitted variables bias. However, in an RD design, there is *no* value of the running variable at which both treatment and control subjects are observed. Nonparametric identification comes from infinitesimal changes in covariate values across the RD cutoff. As a practical matter, however, nonparametric inference procedures compare applicants with covariate values in a small - though not infinitesimal - neighborhood to the left of the cutoff with applicants whose covariate values put them in a small neighborhood to the right. This empirical comparison requires some extrapolation, however modest. Identification of causal effects away from the cutoff requires a more substantial extrapolative leap.

In the main paper we consider a parametric estimating equation

$$y_i = \sum_t \alpha_t w_{it} + \sum_j \beta_j p_{ij} + \sum_\ell \delta_\ell d_{i\ell} + (1 - D_i) f_0 (r_i) + D_i f_1 (r_i) + \rho D_i + \eta_i \tag{3}$$

where the effects of the running variable are controlled by a pair of 3th-order polynomials that differ on either side of the cutoff, specifically:

$$f_j (r_i) = \pi_{1j} r_i + \pi_{2j} r_i^2 + ... + \pi_{pj} r_i^p; \quad j = 0, 1. \tag{4}$$

In a parametric setup such as described by equations (3) and (4), extrapolation is easy though not necessarily credible. For any distance, c, we have

$$\rho (c) \equiv E [Y_{1i} - Y_{0i} \mid r_i = c] = \rho + \pi_1^* c + \pi_2^* c^2 + ... + \pi_p^* c^p, \tag{5}$$

where $\pi_1^* = \pi_{11} - \pi_{10}$, and so on. The notation in equation (5) masks the extrapolation challenge inherent in identification away from the cutoff: potential outcomes in the treated state are observed for $r_i = c > 0$, but the value of $E [Y_{0i} \mid r_i = c]$ for positive c is never seen. It seems natural to use observations to the left of the cutoff in an effort to pin down functional form, and then extrapolate this to impute $E [Y_{0i} \mid r_i = c]$. With enough data, and sufficiently well-behaved conditional mean functions, $E [Y_{0i} \mid r_i = c]$ is identified for all values of c, including those never seen in the data. It's easy to see, however, why this approach may not generate robust or convincing findings.

The unsatisfying nature of parametric extrapolation emerges in Figures 1a and 1b. These figures show observed and imputed counterfactual 10th grade math scores for 7th and 9th grade applicants. Specifically, the figures plot nonparametric estimates of the observed conditional mean function $E [Y_{0i} \mid r_i = c]$ for O'Bryant applicants to the left of the cutoff, along with imputed $E [Y_{1i} \mid r_i = c]$ to the left. Similarly, for BLS applicants, the figures plot nonparametric estimates of observed $E [Y_{1i} \mid r_i = c]$ for applicants to the right of the cutoff, along with imputed $E [Y_{0i} \mid r_i = c]$ to the right. The imputations use linear, quadratic, and cubic specifications for $f_j(r_i)$. These models generate a wide range of estimates, especially as distance from the cutoff grows. For instance, the estimated effect of BLS attendance to the right of the cutoff for 9th grade applicants changes sign when the polynomial goes from second to third degree. This variability seems unsurprising and consistent with Campbell and Stanley (1963)'s observation that, "at each greater degree of extrapolation, the number of plausible rival hypotheses becomes greater." On the other hand, given that $f_0(r_i)$ looks reasonably linear for $r_i < 0$ and $f_1(r_i)$ looks reasonably linear for $r_i > 0$, we might have hoped for results consistent with those from linear models, even when the specification allows something more elaborate.

Table 1, which reports the estimates and standard errors from the models used to construct the fitted values plotted in Figure 1, shows that part of the problem uncovered in the figure is imprecision. Estimates constructed with $p = 3$ are too noisy to be useful at $c = +/-5$ or higher. Models setting $p = 2$ generate more precise estimates than when $p = 3$, though still fairly imprecise for $c \geq 10$. On the other hand, for very modest extrapolation ($c = 1$), a reasonably consistent picture emerges. Like RD estimates at the cutoff, this slight extrapolation generates small positive estimates at O'Bryant and small negative effects at BLS for both 7th and 9th grade applicants, though few of these estimates are significantly different from zero.[1]

Using Derivatives Instead

Dong and Lewbel (2012) propose an alternative to parametric extrapolation based on the insight that the derivatives of conditional mean functions are nonparametrically identified at the cutoff (a similar idea appears in Section 3.3.2 of DiNardo and Lee, 2011). First-order derivative-based extrapolation exploits the fact that

$$f_j(c) \approx f_j(0) + f_j'(0)\,c. \tag{6}$$

This approximation can be implemented using a nonparametric estimate of $f_j'(0)$.

The components of equation (6) are estimated consistently by fitting linear models to $f_j(r_i)$ in a neighborhood of the cutoff, using a data-driven bandwidth and slope terms that vary across the cutoff. In the main paper we consider a nonparametric estimating equation

$$y_i = \sum_t \alpha_t w_{it} + \sum_j \beta_j p_{ij} + \sum_\ell \delta_\ell d_{i\ell} + \gamma_0 (1 - D_i) r_i + \gamma_1 D_i r_i + \rho D_i + \eta_i \tag{7}$$

[1] Paralleling Figure 1, the estimates in Table 1 are from models omitting controls for test year, application year and application preferences. Estimates from models with these controls differ little from those reported in the table.

Given this specification, the effect of an offer at cutoff value c can be approximated as

$$\rho(c) \approx \rho + \gamma^* c, \tag{8}$$

where $\gamma^* = \gamma_1 - \gamma_0$. The innovation in this procedure relative to LLR estimation of equation (7) is in the interpretation of the interaction term, γ^*. Instead of a bias-reducing nuisance parameter, γ^* is seen in this context as identifying a derivative that facilitates extrapolation. As a practical matter, the picture that emerges from derivative-based extrapolation of exam school effects is similar to that shown in Figure 1.

3 Alternative Conditional Independence Test

The RD design coupled with the conditional independence assumption provides also another testable implication that can be used to guide our choice of the conditioning vector, x_i. This is based on the observation that under CIA the only difference between an RD estimand and a matching-style estimand is in the way they weight the covariate-specific average treatment effects. Specifically, under the CIA, the RD estimand is:

$$\lim_{\epsilon \downarrow 0} E[y_i \mid r_i = +\epsilon] - \lim_{\epsilon \downarrow 0} E[y_i \mid r_i = -\epsilon] \quad = \quad E\{E[Y_{1i} - Y_{0i} \mid x_i] \mid r_i = 0\} \tag{9}$$

In other words, the RD estimand weights the covariate-specific average treatment effects using the distribution of x_i at the cutoff. This implies that under the CIA the RD estimand should equal a matching-style estimand that uses similarly the distribution of x_i at the cutoff to weight the covariate-specific average treatment effects:

$$E\{E[y_i \mid x_i, D_i = 1] - E[y_i \mid x_i, D_i = 0] \mid r_i = 0\} \quad = \quad E\{E[Y_{1i} - Y_{0i} \mid x_i] \mid r_i = 0\}. \tag{10}$$

This observation motivates us to test the difference between RD and CIA estimates of the effects at the cutoff. The RD estimates we use are based on the non-parametric estimator described in the main paper. For the CIA estimates we first estimate the covariate-specific average treatment effects using the linear reweighting and propensity score estimators described in the main paper. We then weight these estimates as follows:

$$E\left\{E[Y_{1i} - Y_{0i} \mid x_i] \times \frac{P[r_i=0|x_i]}{P[r_i=0]}\right\}$$

We approximate $\frac{P[r_i=0|x_i]}{P[r_i=0]}$ by $\frac{P[-5<r_i\leq 5|x_i]}{P[-5<r_i\leq 5]}$ and estimate the numerator using a logit model with the same specifiation as our propensity score model.

Table 2 reports the t-test statistics for the difference between the RD and CIA estimates of the effects at the cutoffs (the standard errors are computed using nonparameric bootstrap with 500 replications). These results tell the same story as the CIA tests shown in the main paper. For 7th grade applicants all of the RD and CIA estimates of the effects at the cutoffs are significantly different from each other, suggesting failure of CIA. At the same time, none of the RD and CIA estimates for 7th graders are significantly different from each other, providing additional support for the validity of CIA.

Table 1: Parametric Extrapolation Estimates for 10th Grade Math

	O'Bryant				Latin School			
	c = -1	c = -5	c = -10	c = -15	c = 1	c = 5	c = 10	c = 15
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: 7th Grade Applicants				
Linear	0.041	0.061	0.085	0.110	-0.076**	-0.051	-0.021	0.010
	(0.052)	(0.057)	(0.072)	(0.093)	(0.035)	(0.040)	(0.049)	(0.061)
	1832	1832	1832	1832	1854	1854	1854	1854
Quadratic	0.063	0.204	0.391*	0.588	-0.056	-0.111	-0.152	-0.161
	(0.075)	(0.125)	(0.237)	(0.384)	(0.051)	(0.088)	(0.162)	(0.261)
	1832	1832	1832	1832	1854	1854	1854	1854
Cubic	0.034	0.167	0.247	0.266	-0.050	-0.096	-0.106	-0.065
	(0.110)	(0.336)	(0.921)	(1.927)	(0.073)	(0.220)	(0.589)	(1.215)
	1832	1832	1832	1832	1854	1854	1854	1854
				Panel B: 9th Grade Applicants				
Linear	0.088	0.083	0.077	0.071	-0.090	0.079	0.291***	0.502***
	(0.057)	(0.059)	(0.070)	(0.088)	(0.065)	(0.063)	(0.108)	(0.168)
	1559	1559	1559	1559	606	606	606	606
Quadratic	0.170**	0.264**	0.427*	0.639*	-0.147*	-0.106	0.078	0.409
	(0.085)	(0.133)	(0.237)	(0.372)	(0.088)	(0.142)	(0.303)	(0.713)
	1559	1559	1559	1559	606	606	606	606
Cubic	0.143	0.069	-0.059	-0.355	-0.061	0.196	0.996	3.094
	(0.119)	(0.327)	(0.851)	(1.735)	(0.118)	(0.338)	(0.910)	(2.543)
	1559	1559	1559	1559	606	606	606	606

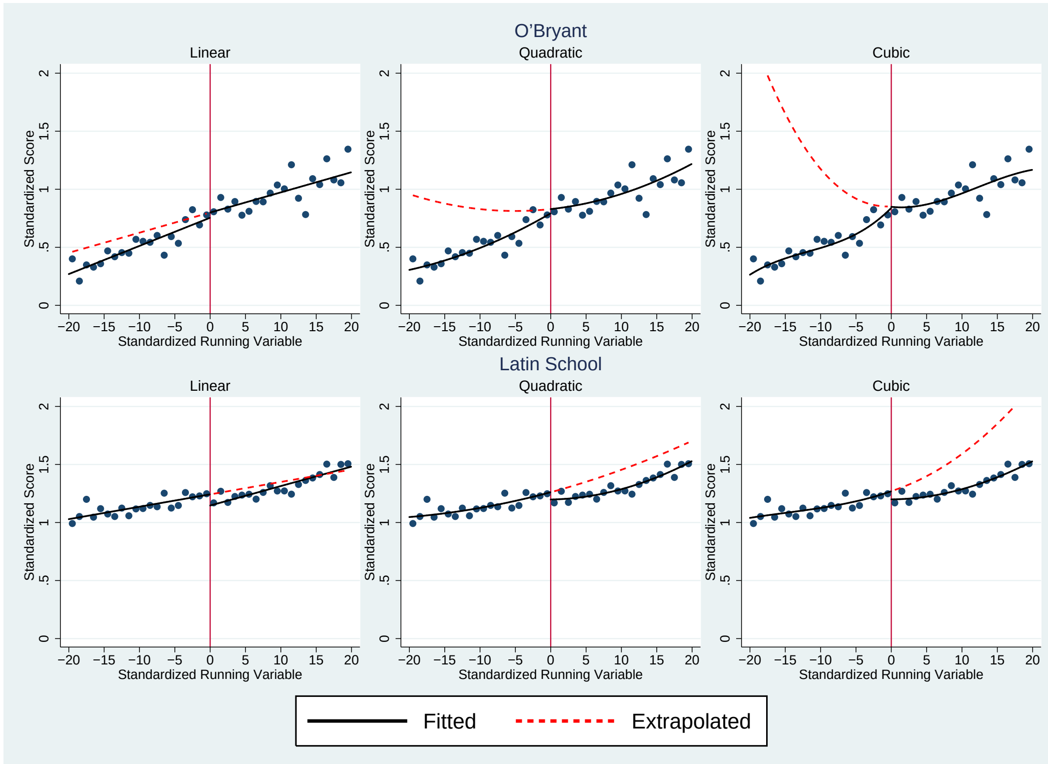
Notes: This table reports estimates of effects on 10th grade Math scores away from the RD cutoff at points indicated in the column heading. Columns 1-4 report estimates of the effect of O'Bryant attendance on unqualified O'Bryant applicants. Columns 5=8 report the effects of BLS attendance on qualified BLS applicants. The estimates are based on first, second, and third order polynomials, as inidcated in rows of the table. Robust standard errors are shown in parentheses.

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

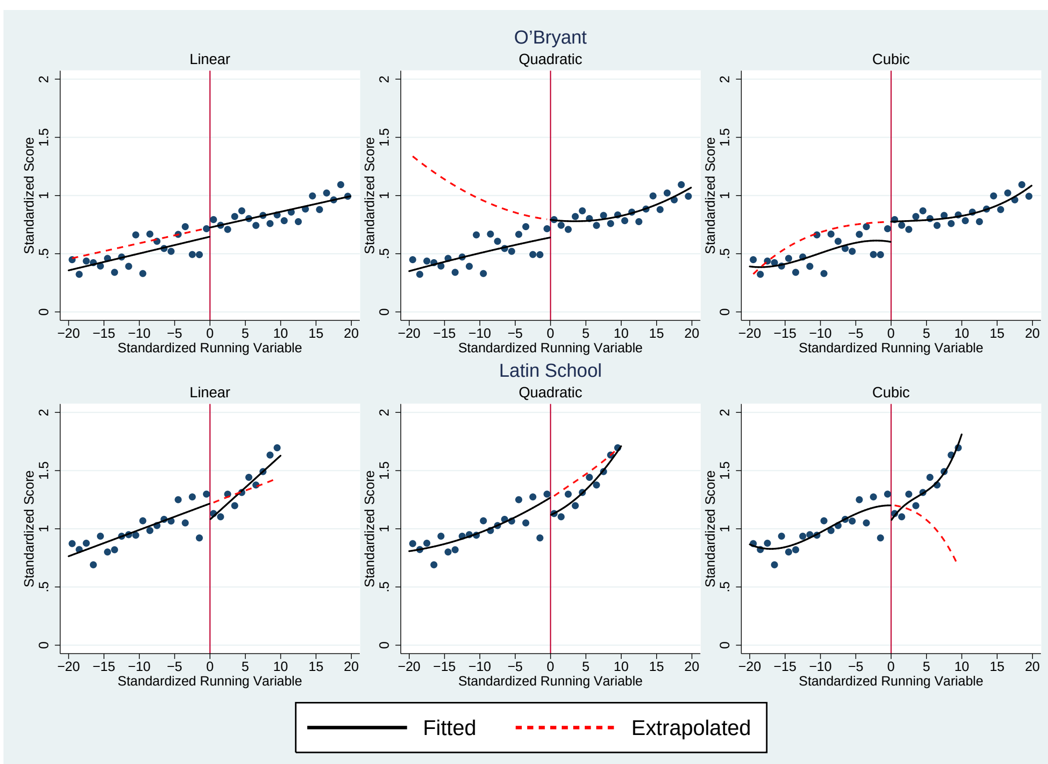
Table 2: Alternative Conditional Independence Based on a Comparison of RD and CIA Estimates

	Linear Reweighting		Propensity Score Weighting	
	O'Bryant	Latin School	O'Bryant	Latin School
	(1)	(2)	(3)	(4)
		Panel A. 7th Grade Applicants		
Math	5.005	4.698	5.548	2.669
ELA	2.670	3.264	2.625	3.020
		Panel B. 9th Grade Applicants		
Math	-0.349	0.867	-0.303	1.041
ELA	0.511	0.684	1.047	0.609

Notes: This table reports t-test statistics for the difference between RD and CIA estimates of the effect of exam school offers on MCAS scores at the admissions cutoffs. The RD estimates use the nonparametric RD estimator described in the text. The CIA estimates use the linear reweighting estimator described in the text. Standard errors for the difference between the estimates were computed using nonparametric bootstrap with 500 replications.



(a) 7th Grade Applicants



(b) 9th Grade Applicants

Figure 1: Parametric Extrapolation at O'Bryant and Boston Latin School for 10th Grade Math. O'Bryant extrapolation is for $E[Y_{1i}|r_i = c]$; BLS extrapolation is for $E[Y_{0i}|r_i = c]$.

References

ABADIE, A. (2003): "Semiparametric Instrumental Variables Estimation of Treatment Response Models," *Journal of Econometrics*, 113(2), 231–263.

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

CAMPBELL, D. T., AND J. STANLEY (1963): *Experimental and Quasi-experimental Designs for Research*. Rand McNally.

DINARDO, J., AND D. S. LEE (2011): *Program Evaluation and Research Designs*chap. 5, pp. 463–536, Handbook of Labor Economics 4. Elsevier.

DONG, Y., AND A. LEWBEL (2012): "Identifying the Effects of Changing the Policy Threshold in Regression Discontinuity Models," Unpublished manuscript.