School Autonomy and Regression Discontinuity Imbalance

Todd Kawakita¹ and Colin Sullivan²

Abstract

In this research note, we replicate and assess Damon Clark’s (2009) analysis of school autonomy reform in the United Kingdom. Clark implements a regression discontinuity design to estimate the effect of vote-based school autonomy reform on pass rates in secondary school. We show that Clark’s findings are flawed due to his failure to verify the assumptions of the regression discontinuity and reliance on the entire range of data rather than a narrow band near the threshold. Moreover, baseline covariates are severely imbalanced between the treatment and control groups, even at narrow bandwidths around the cutoff. We question Clark’s use of the regression discontinuity design in estimating the impact of school autonomy.

I. Introduction: School Autonomy in the United Kingdom

School independence has become a salient issue in education reform in recent years. In the United States, reformers have touted the benefits of charter schools in part because of their autonomy in curriculum and staffing decisions.³ Supporters further argue that local school control leverages knowledge of a community’s needs, and that traditional schools are forced to use resources inefficiently. In the opposing camp, reformers have called for stricter oversight to ensure accountability for performance.

A 1988 education reform in the United Kingdom made it possible for secondary school districts to hold a vote for “Grant Maintained (GM) status” - independence from regional authority. These elections for school autonomy were held at the discretion of the school board and were executed by secret postal ballot by parents of current students. A majority of voting parents was needed to obtain GM status. Schools that failed to meet the majority requirement were permitted to hold later ballots at two-year intervals.

Damon Clark provides the first evidence of the success of grant maintained school autonomy in his 2009 article The Performance and Competitive Effects of School Autonomy.⁴ Clark considers the set of school districts that held first time votes for autonomy between 1992 and 1997. Comparing the treatment group (those who attain GM status) and control group (those who do not) is not a straightforward matter: the schools that attain autonomy might differ systematically from those schools that do not in student achievement and composition prior to reform, or unobserved factors such as parental support for autonomy, 

¹Harvard University Graduate School of Education  
²Harvard University Kennedy School of Government  
parental involvement in their children’s education, or trust in the school administration. However, since school districts theoretically have no control over the outcome of the vote immediately around 50 percent, schools that barely fail to attain the majority vote should be a reasonable control group for those just above the cutoff. Exploiting this 50 percent vote threshold, Clark uses a regression discontinuity design to estimate the impact of school autonomy and concludes that freedom from local school district authority resulted in significant improvements in 10th grade pass rates.

Regression discontinuity (RD) is commonly used to estimate the effect of treatment that is assigned on the basis of a continuous variable exceeding a given threshold. As long as subjects do not have precise control over the assignment variable, “a consequence of this is that the variation in treatment near the threshold is randomized as though from a randomized experiment.” Randomization around the threshold yields two comparable groups, a treatment and a control, which are balanced in their observed and unobserved characteristics. In order to verify this, Lee and Lemieux recommend that “[as] in a randomized experiment, the distribution of observed baseline covariates should not change discontinuously at the threshold.” The researcher must also check the balance of observables between the groups as he includes observations further from the threshold, as these observations induce bias in the estimates. Selecting a bandwidth large enough to estimate the treatment effect yet narrow enough that the two groups are balanced is key to the success of the RD design.

Using Clark’s data and code, we review Clark’s findings and check the balance of several covariates near the 50 percent vote cutoff. We find that rather than use a narrow balanced bandwidth, Clark includes the entire range of observations in his analysis. Moreover, many of the covariates have significant discontinuities and imbalance in their distributions, particularly in the area proximal to the cutoff. These findings are evidence of serious bias in Clark’s estimates, and further question the validity of RD designs in electoral data.

II. Covariate Imbalance

The idealized regression discontinuity design estimates the treatment effect by a simple comparison of observations immediately above and below the cutoff. To include more observations and obtain more precise estimates, however, researchers generally must choose a range around the cutoff, known as the bandwidth. The bandwidth must encompass enough observations for a reasonably precise estimate, yet not bias the estimate with observations far from the cutoff. A small bandwidth reduces bias but produces noisy estimates in the treatment effect. In his analysis, Clark uses the maximal bandwidth, including all available data within 50 percentage points to the left of the cutoff and within 50 percentage points to the right of the cutoff. Throughout our discussion of bandwidth, we assume that the bounds of the included range are symmetrical about the cutoff, although

---

6Lee & Lemieux, 14.
7Note that if observations far from the cutoff did not induce bias, a simple OLS would be sufficient to estimate the treatment effect.
8A maximal or 100 percent bandwidth indicates that Clark uses all data within 50 percentage points to the left of the cutoff and within 50 percentage points to the right of the cutoff.
able data to estimate the treatment effect. We assess this choice by examining multivariate imbalance and running local regression models at varying bandwidths.

In Figure 1 we examine imbalance at variable bandwidths using $L_1$, a measure of multivariate imbalance bounded between 0 and 1 with a larger fraction indicating greater imbalance.\textsuperscript{9} Imbalance is lowest at bandwidths between 10 and 20 percent - wider bandwidths invite bias from the systematic differences between winning and losing schools. In light of this, Clark’s choice of maximal bandwidth clearly leads to biased estimates. The higher imbalance at bandwidths below 10 percent also indicates differences between the small groups directly above and below the discontinuity, undermining the fundamental assumption of the regression discontinuity design altogether. Moreover, the presence of imbalance in several of the observed covariates suggests the possibility of more unobserved differences.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{multivariateimbalance.png}
\caption{Multivariate Imbalance and Bandwidth} Lower $L_1$ indicates better balance, suggesting that bandwidths between 10 and 20 percent produce the most balanced.
\end{figure}

In Figure 2 we plot the difference in means for vote winners and losers at different bandwidths for base year observables: pass rates, FSM eligibility, and number of pupils. The differences in means would be zero in a perfectly balanced sample. The sample means of winning schools are substantially different from losing schools at bandwidths above 35 to the size of the treatment and control groups may not be equal.

Figure 2: **Treatment and Control Groups Base Year Differences in Means** Positive values indicate that schools above the cutoff have higher means than schools below the cutoff. The differences in means in base year covariates display clear disparities above the 35 to 50 percent bandwidths. Larger bandwidths induce bias in Clark’s estimates.

50 percent: winning schools have higher pass rates and lower FSM rates, and they begin to have fewer students. Therefore schools with landslide elections - vote shares less than 25 percent or more than 75 - should not be used to estimate the treatment effect.

**III. Conclusion**

The regression discontinuity design can be extremely useful for analyzing non-random treatment as in a randomized experiment. The requirements are not particularly exigent: assignment to treatment must be determined by an assignment variable exceeding a given threshold, and subjects must not exercise precise control over the assignment variable. These two requirements are enough to produce “as-good-as-randomized” data around the cutoff. Assuming that these requirements hold true for school autonomy ballots, Damon Clark concludes that autonomy had a significant positive effect on those schools whose vote barely
passed, compared to those schools where autonomy barely failed.

As in randomization, however, it is necessary to verify that the groups are in fact comparable by testing balance. The successful regression discontinuity design will select the bandwidth that produces balanced control and treatment groups, and verify that all observable variables are continuous around the cutoff. Imbalance immediately near the cutoff or a bandwidth that fails to exclude observations far from the cutoff invalidate the regression discontinuity design.

Using multivariate and univariate tests of imbalance, we find multivariate imbalance to be high near the cutoff, and lowest at a bandwidth of about 10 percent, or in marginal elections with vote share between 45 percent and 55 percent. As bandwidth increases, univariate imbalance increases as one might expect: schools gaining autonomy tend to have higher test scores and lower FSM eligibility rates. Clark uses all the data, rather than selecting a narrower bandwidth, thereby comparing imbalanced groups. He does not control for many of these observable imbalances, and they suggest other unobserved imbalances as well. These results raise further questions about the use of regression discontinuity designs in election data, and emphasize the importance of verifying the assumptions of these models.

References


