Assignment 3: Teacher Performance Bonus Program

Andrew Heiss (working with Jessica Lewis, Michelle Welch, and Jennifer Shen)

PubPol 604 • October 15, 2013

All the code and logs for this assignment are available at https://github.com/andrewheiss/Causal-inference-code.

1. Program compliance

According to the program rules, any school that achieves positive average expected growth should receive a $750 per teacher bonus. However, it appears that the program was not administered perfectly in 2005 and 2006—some schools that did not meet the standard received bonuses, while others that did meet the standard did not (see Table 1; $\chi^2 = 2{,}101.9$, $p < 0.001$). Fortunately the rates of noncompliance are quite low—there was a 98% chance of getting a bonus if the school met the standard.

Table 1: Compliance with bonus program in 2006

	Did not receive bonus	Received bonus
Negative growth	1033 (99.0%)	23 (1.9%)
Positive growth	10 (1.0%)	1170 (98.1%)

2. Local linear regression discontinuity

Given that there is minimal noncompliance, we can use nonparametric fuzzy regression discontinuity to determine if receiving a bonus in 2005 improves teacher performance in 2006. The Wald estimate and bootstrapped standard errors (over 100 draws) for the optimal local linear bandwidth of 0.061 is included in Table 2, along with estimates for half and twice that bandwidth. As seen in the results, the local average treatment effect has the smallest standard errors and is thus ostensibly the most accurate of the three bandwidth estimates (which isn't surprising, given that the Imbens-Kalyanaraman estimator is optimized for the data). However, even with precise estimates, there is no significant discontinuity in treatment effect at any bandwidth. Receiving a bonus appears to have no effect on school performance in the following year.

Table 2: Wald estimates for various local linear bandwidth levels

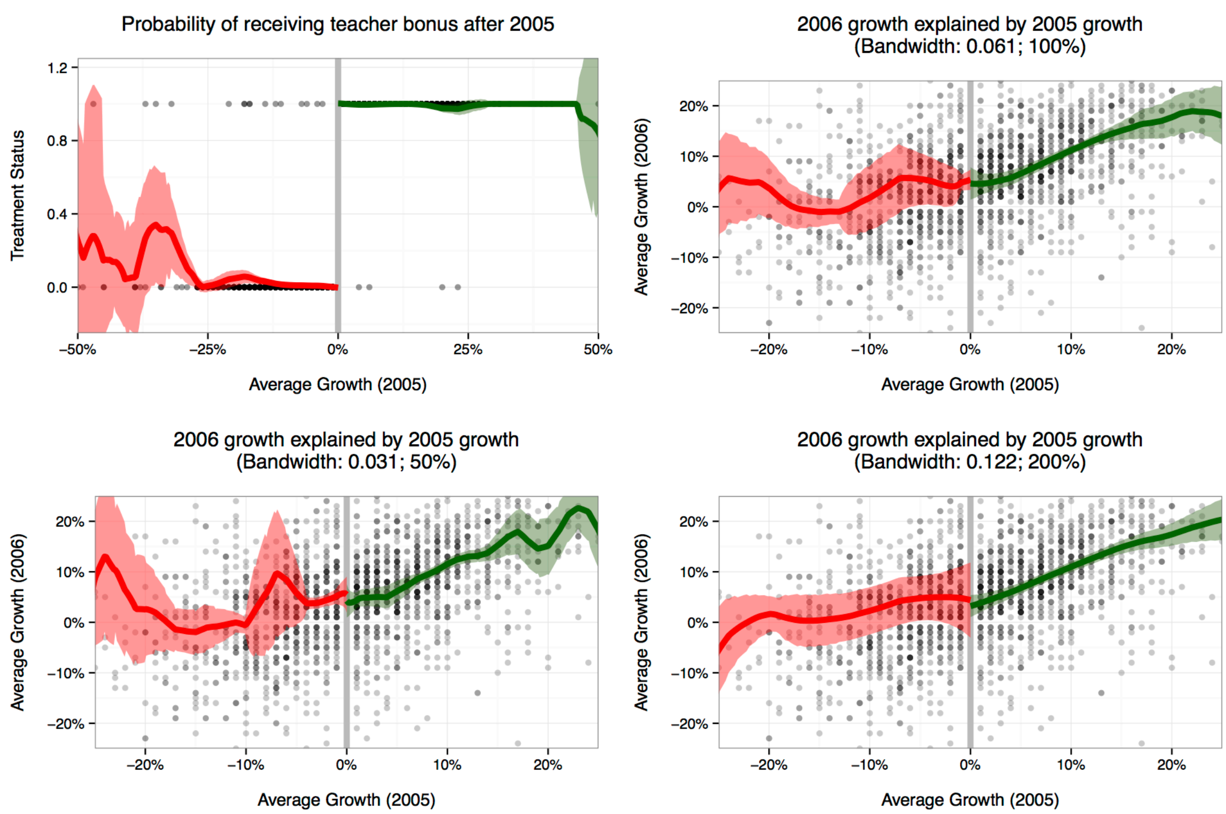
	Bandwidth	Estimate	Bootstrap Std Err.	z	P>z
Local Average Treatment Effect	0.06096	-.0101625	.0177568	-0.57	0.567
Half bandwidth	0.03048	-.0136931	.0228536	-0.60	0.549
Double bandwidth	0.12192	-.0172546	.0112783	-1.53	0.126

3. Visualizing the nonparametric discontinuity

Part A: Discontinuity in treatment

One prerequisite for regression discontinuity is that there must be a discontinuity in the treatment around an arbitrary cutoff. In the case of the bonus program, this means that there should be a change in the probability of receiving a bonus above and below 0% growth. The table in Question 1 showed this was true, and the top left panel of Figure 1 demonstrates this discontinuity visually. There is a clear jump in the probability of receiving a bonus when average growth crosses the 0% threshold.

Figure 1: Nonparametric discontinuities



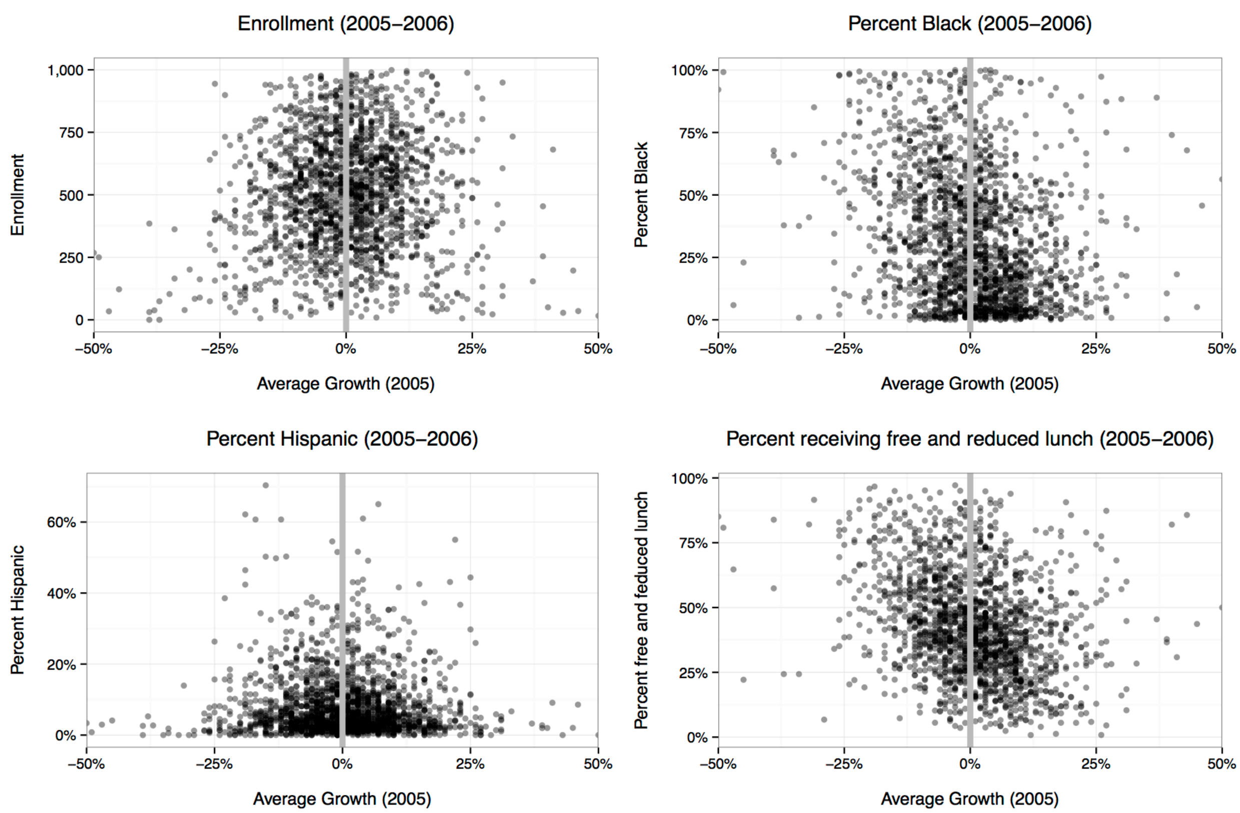
Part B: Discontinuity in outcome

Though there is a clear discontinuity in treatment, there does not appear to be any discontinuity in outcome (see the remaining panels of Figure 1), verifying what was found numerically in Question 2. That is, school performance in 2005 does not significantly increase or decrease performance in 2006 near the 0% threshold—a hypothetical school that received teacher bonuses for achieving just over 0% growth performed no better than a hypothetical school that achieved just under 0% and did not receive bonuses. This is true at all three bandwidths.

Part C: Pretreatment covariates

Another assumption for regression discontinuity is that pretreatment covariates do not show a discontinuity at the treatment threshold—that is, other relevant factors are continuous across the cutoff. In the case of this bonus program, student characteristics should be relatively constant across the threshold in order to isolate the effect of the bonus on school performance. Figure 2 shows scatterplots of average growth in relation to school enrollment and the proportion of black, Hispanic, and low socioeconomic status (SES) students students in each school. As is evident from the plots, there appears to be no discontinuity in any of the characteristics. Enrollment and percent Hispanic appear relatively uniform across all levels of average growth, while percent black and percent low SES have a negative relationship—schools with more black students or poor students tend to have lower growth—but that relationship lacks a discontinuity at zero.

Figure 2: Pretreatment characteristics



4. Parametric discontinuity

Instead of fitting nonparametric local linear models, we can estimate parametric regression models to approximate possible discontinuities in the outcome variable. An advantage to this approach is that covariates can be added to the model, allowing for better isolation of the intent-to-treat (ITT) effect. However, there are at least two disadvantages to this approach: (1) accounting for noncompliance, or calculating the treatment-on-the-treated (TOT) effect requires a more complicated two-stage instrumental variable regression, and (2) the estimated curves do not generally fit the data as well as nonparametric approximations.

One example of parametric estimation is to use use the following cubic model, controlling for pretreatment characteristics.[1] The results of this model are presented in column 1 of Table 3.

$$\text{2006 growth} = \beta_0 + \beta_1 \text{2005 growth} + \beta_2 \text{2005 growth}^2 + \beta_3 \text{2005 growth}^3 + \beta_4 \text{Reward given} + \beta_5 \text{Controls}$$

According to this model, after accounting for the potentially cubic nature of the data and controlling for pretreatment school characteristics, the receipt of a bonus does have a significant effect on 2006 performance—average growth is 12% higher in schools that receive teacher bonuses ($t = 3.547$, $p < 0.001$). The magenta line in Figure 3 demonstrates this effect visually. The line on the lefthand side of the threshold represents the model's fitted values where no reward was given, while the line on the righthand side of the threshold represents fitted values where the reward was given (all other modal controls are held at their median values). There is a clear discontinuity at the threshold.

However, the model clearly does not fit the data well—it overpredicts growth in 2006 for negative values of 2005 growth and underpredicts 2006 growth for positive values of growth in 2005. Additionally, this parametric model only shows the ITT effects of the program and does not account for noncompliance, which is a problem given that the majority of noncompliance occurs near the threshold (some schools that just barely missed the cutoff received funding, and vice versa). Thus, while the model technically shows that the reward has a significant effect on 2006 performance, the finding is likely spurious.

---

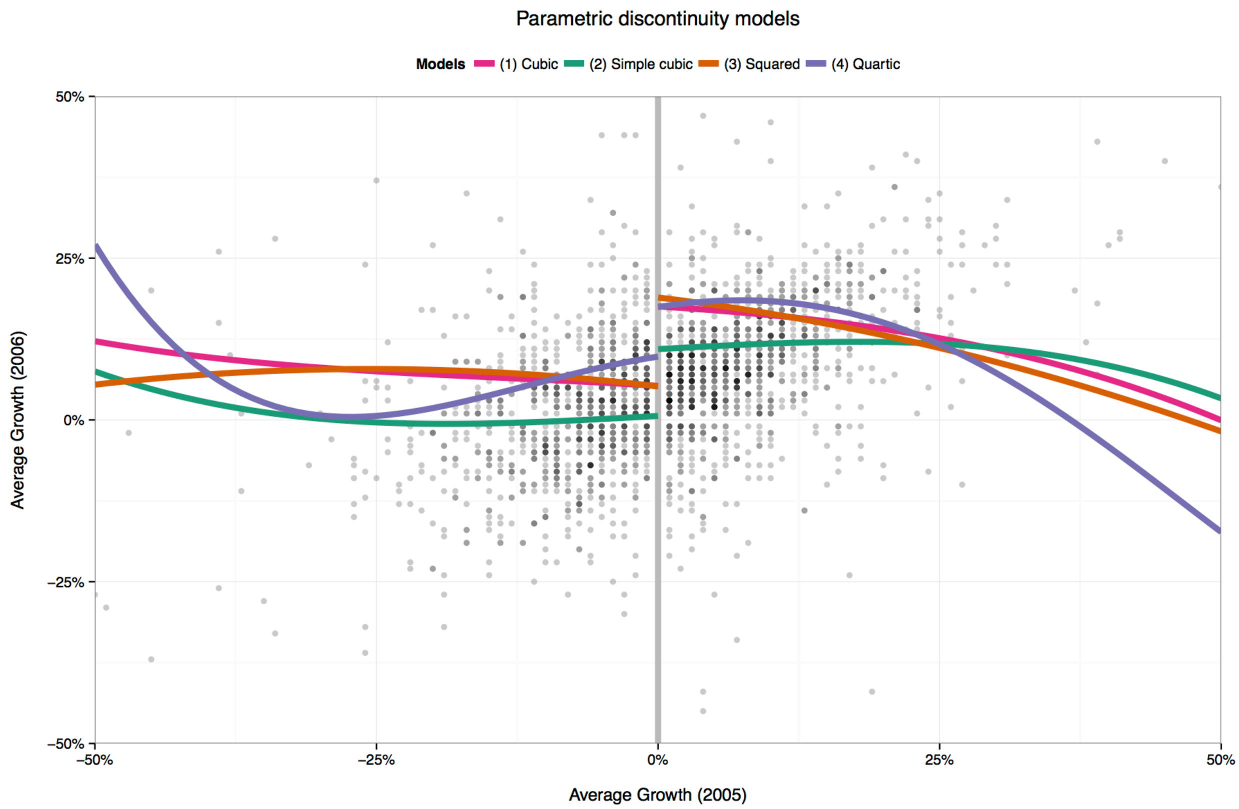
[1] Controls = enrollment, percent black, percent Hispanic, percent low SES, and indicators for elementary, middle, and high school.

Table 3: Parametric models of discontinuity

	Dependent variable:				
	Average growth (2006)				
	(1)	(2)	(3)	(4)	(5)
Average growth (2005)	−0.110	0.097	−0.210	0.265	0.563***
	(0.169)	(0.138)	(0.127)	(0.203)	(0.078)
Average growth $(2005)^2$	−0.214	−0.015	−0.409	−1.476**	0.524*
	(0.341)	(0.279)	(0.264)	(0.511)	(0.214)
Average growth $(2005)^3$	−0.542	−0.967		−3.143**	−0.950
	(0.604)	(0.523)		(0.990)	(1.048)
Average growth $(2005)^4$				4.508***	
				(1.361)	
Percent black	−0.037		−0.035	−0.031	−0.068***
	(0.059)		(0.059)	(0.059)	(0.017)
Percent Hispanic	0.076		0.077	0.065	0.069
	(0.141)		(0.141)	(0.140)	(0.036)
Percent low SES	−0.141		−0.147	−0.122	−0.086***
	(0.086)		(0.086)	(0.086)	(0.025)
Enrollment	−0.0001		−0.0001	−0.0001	−0.00001
	(0.0001)		(0.0001)	(0.0001)	(0.00002)
Elementary school	−0.037		−0.036	−0.041	
	(0.033)		(0.033)	(0.033)	
Middle school	−0.031		−0.030	−0.037	
	(0.041)		(0.041)	(0.041)	
High school	0.025		0.030	0.032	
	(0.049)		(0.049)	(0.049)	
Received bonus	0.123***	0.103***	0.137***	0.077*	−0.012
	(0.035)	(0.029)	(0.031)	(0.037)	(0.012)
Constant	0.123*	0.006	0.123*	0.154**	0.107***
	(0.055)	(0.018)	(0.055)	(0.056)	(0.016)
Observations	1,660	2,059	1,660	1,660	945
R^2	0.029	0.022	0.029	0.036	0.361
Adjusted R^2	0.023	0.020	0.023	0.029	0.355
Residual Std. Error	0.455 (df = 1648)	0.425 (df = 2054)	0.455 (df = 1649)	0.454 (df = 1647)	0.093 (df = 936)
F Statistic	4.501*** (df = 11; 1648)	11.297*** (df = 4; 2054)	4.872*** (df = 10; 1649)	5.066*** (df = 12; 1647)	66.078*** (df = 8; 936)

*p<0.05; **p<0.01; ***p<0.001

Figure 3: Parametric models of discontinuity (controls held at median values)



5. Parametric specification checks

Parts A–C

Modifications of the parametric model reveal similar findings. Columns 2–4 of Table 3 show the results for the following models:

$$(2)\ \text{2006 growth} = \beta_0 + \beta_1 \text{2005 growth} + \beta_2 \text{2005 growth}^2 + \beta_3 \text{2005 growth}^3 + \beta_4 \text{Reward given}$$

$$(3)\ \text{2006 growth} = \beta_0 + \beta_1 \text{2005 growth} + \beta_2 \text{2005 growth}^2 + \beta_3 \text{Reward given} + \beta_4 \text{Controls}$$

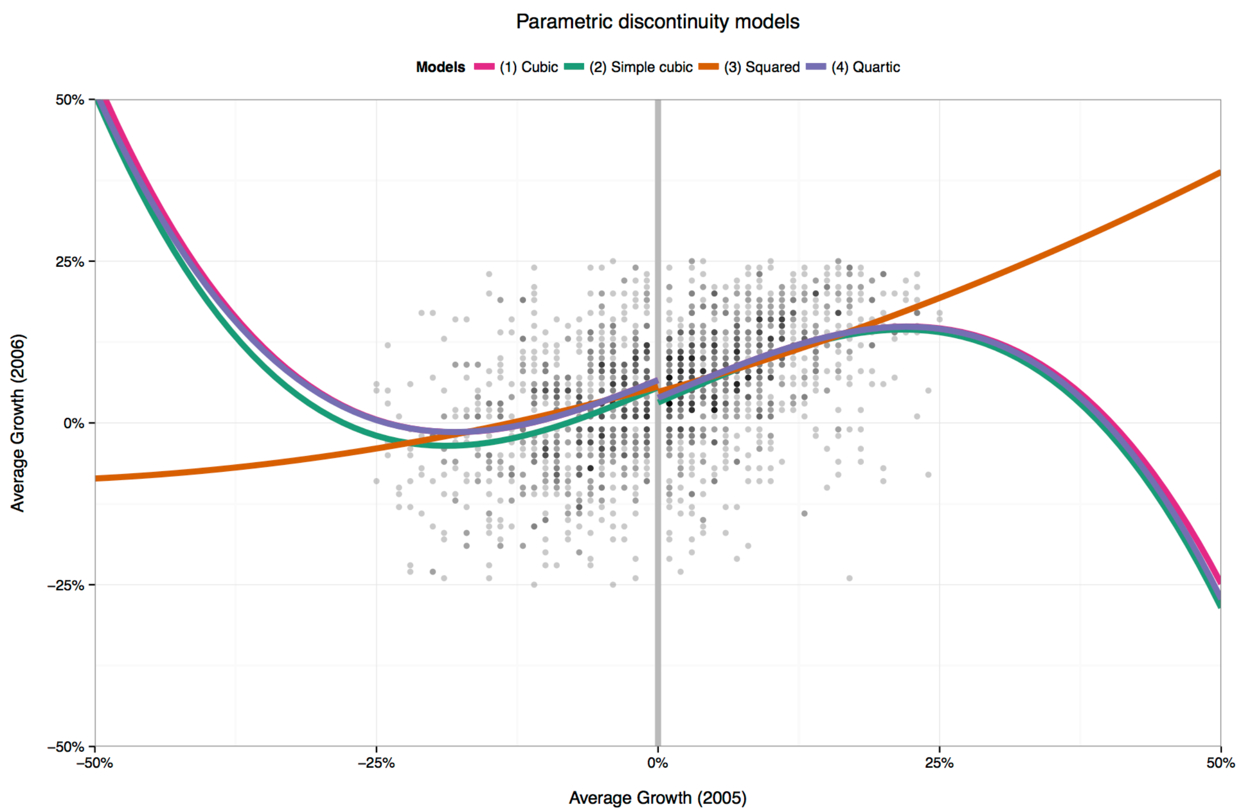
$$(4)\ \text{2006 growth} = \beta_0 + \beta_1 \text{2005 growth} + \beta_2 \text{2005 growth}^2 + \beta_3 \text{2005 growth}^3 + \beta_4 \text{2005 growth}^4 + \beta_5 \text{Reward given} + \beta_6 \text{Controls}$$

The effect of the reward is significant in each model, with improvements in 2006 performance between 8–13%. However, like the model in Question 4, none of these modified parametric models fit the data very well (see Figure 3), and therefore do not give very accurate estimates of the program's effect.

Extra checks for parts A–C (just for fun)

None of the parametric models fit well because of outliers far from the threshold. Since regression discontinuity deals primarily with data near the threshold, we can potentially improve these models' fit by weighting data near the cutoff more heavily than data far away. One brute force method of doing this is dropping observations that had growth greater than 25% in either 2005 or 2006. The results of this approach are presented in Table 4 in the appendix, and can be seen visually in Figure 4. As is evident in the plot, the models have a much more accurate fit, with tiny (yet significant) negative discontinuities. Using this approach, the reward appears to lower a school's performance by around 2% the following year. This counterintuitive result is most likely the result of throwing away so much data and shows another important caveat to parametric discontinuity estimation—getting a good fit requires either (1) excellent model specification (which is rare) or (2) torturing the model or data (which leads to incorrect or uninterpretable results).

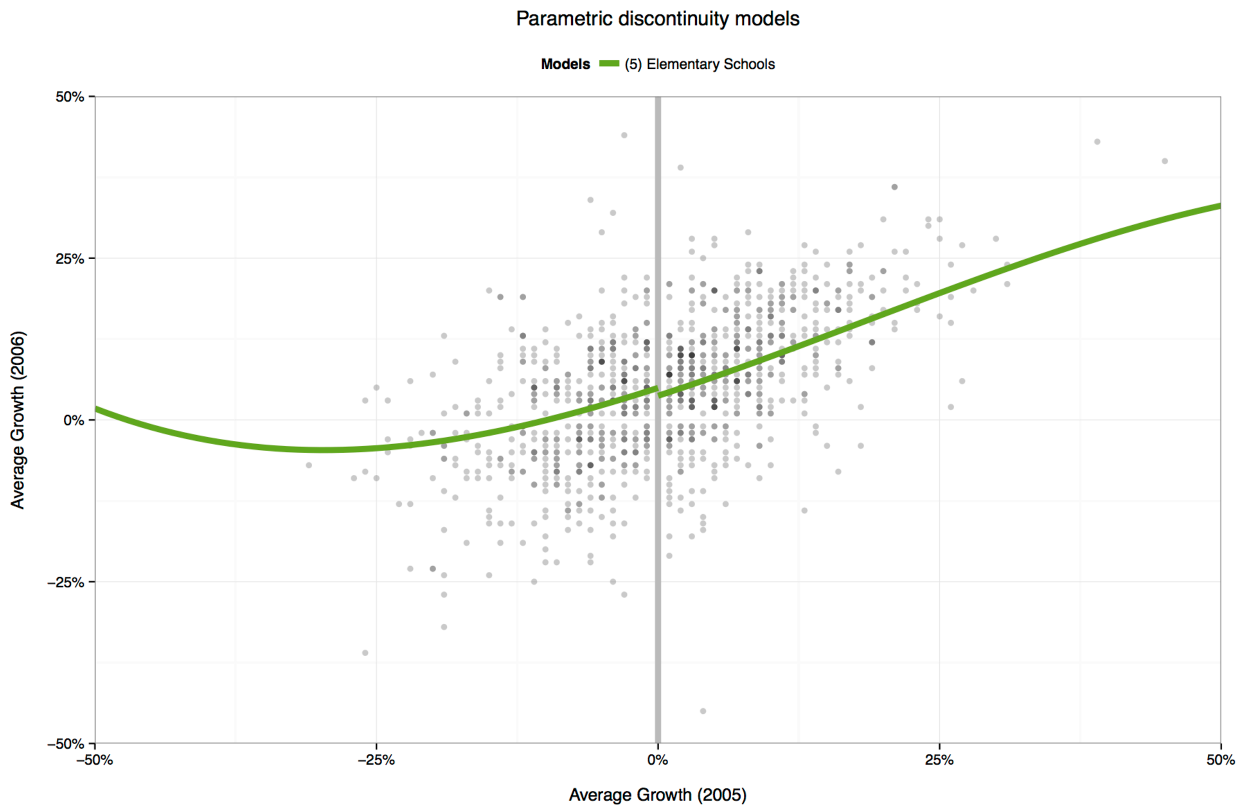
Figure 4: Parametric models, data limited to values between -25% and 25%



Part D

Another method for checking the model's accuracy is limiting the model to a smaller subset of data. The majority of elementary schools had growth between -25% and 25%, and this lack of significant outliers allows us to more naturally fit the data without actually dropping data. Column 5 of Table 3 is the same model as column 1, but limited to elementary schools. Using this subset of data, teacher bonuses have no significant effect on school improvement in 2006 ($t = -1.025$, $p = 0.306$). Figure 5 demonstrates this visually—the model's predicted values fit the original data much more accurately than any of the previous models, and there is no significant discontinuity at the threshold. Teacher bonuses did nothing for elementary school performance.

Figure 5: Parametric model, data limited to elementary schools



6. Policy implications

Thus, given the findings of both the nonparametric and parametric regression discontinuity models, I conclude that the teacher bonus program has no effect on school performance. Though the parametric models seem to indicate a strong program effect, the models' incomplete fit bias that effect upward. The nonparametric and elementary-school-only models give a much more accurate conclusion, indicating the lack of any effect.

Accordingly, I would not endorse the expansion of this program. However, I would not necessarily feel comfortable eliminating the program either. While the program has no effect on growth (which is ostensibly its primary goal), it may have other important effects on teacher morale, student engagement, or other elements of education not captured by the expected growth measure. Given that the 0% average growth threshold lends itself nicely as an arbitrary cutoff, I suggest using it to measure the effect of bonuses on other dependent variables (like morale) to see if the program helps. If there is no effect elsewhere, I would endorse phasing the program out. If there is an effect, though, I would recommend continuing the program.

Appendix

Table 4: Parametric models of discontinuity, data limited to -0.25 ≤ x ≤ 0.25

	Dependent variable:				
	Average growth (2006)				
	(1)	(2)	(3)	(4)	(5)
Average growth (2005)	0.702***	0.747***	0.482***	0.702***	0.730***
	(0.077)	(0.072)	(0.038)	(0.077)	(0.103)
Average growth $(2005)^2$	0.368*	0.272	0.395*	0.386	0.719**
	(0.182)	(0.171)	(0.183)	(0.478)	(0.245)
Average growth $(2005)^3$	−5.844***	−6.066***		−5.848**	−4.983*
	(1.772)	(1.681)		(1.776)	(2.397)
Average growth $(2005)^4$				−0.442	
				(10.829)	
Percent black	−0.058***		−0.057***	−0.058***	−0.069***
	(0.011)		(0.011)	(0.011)	(0.015)
Percent Hispanic	0.040		0.039	0.040	0.060
	(0.026)		(0.026)	(0.026)	(0.031)
Percent low SES	−0.051**		−0.053**	−0.051**	−0.059**
	(0.017)		(0.017)	(0.017)	(0.022)
Enrollment	0.00002		0.00002	0.00002	−0.00001
	(0.00001)		(0.00001)	(0.00001)	(0.00002)
Elementary school	0.006		0.006	0.006	
	(0.006)		(0.006)	(0.006)	
Middle school	0.012		0.012	0.012	
	(0.008)		(0.008)	(0.008)	
High school	0.015		0.015	0.015	
	(0.010)		(0.010)	(0.010)	
Received bonus	−0.027**	−0.024**	−0.008	−0.027**	−0.029*
	(0.009)	(0.009)	(0.007)	(0.009)	(0.012)
Constant	0.077***	0.055***	0.068***	0.076***	0.097***
	(0.011)	(0.005)	(0.011)	(0.011)	(0.015)
Observations	1,521	1,874	1,521	1,521	900
R^2	0.325	0.243	0.320	0.325	0.369
Adjusted R^2	0.320	0.241	0.316	0.320	0.364
Residual Std. Error	0.080 (df = 1509)	0.086 (df = 1869)	0.080 (df = 1510)	0.080 (df = 1508)	0.081 (df = 891)
F Statistic	66.056*** (df = 11; 1509)	149.603*** (df = 4; 1869)	71.109*** (df = 10; 1510)	60.511*** (df = 12; 1508)	65.183*** (df = 8; 891)

*p<0.05; **p<0.01; ***p<0.001