Beating the Odds Statistical Model Technical Report
Prepared for Georgia Governor's Office of Student Achievement by

Douglas Lee Lauen, PhD MPP
Associate Professor of Public Policy
University of North Carolina at Chapel Hill

Purpose

The purpose of this technical report is to examine the Beating the Odds statistical model used to adjust Georgia's CCRPI scores for the mix of non-malleable factors outside of the control of school leaders. This is an important purpose given that research shows that factors such as family poverty and social background have large effects on student performance, and thus, school performance and the accountability metrics based upon these sources. This report summarizes the results of an analysis of a *non-linear* model using the standard deviation of the *forecast* to create confidence intervals (CIs), comparing the results of this proposed model to the baseline linear specification, which uses the standard deviation of the *prediction*.

Schools in Sample and Distribution of Predictors

The descriptive statistics and much of the results in this section are based on the 2017 BTO data file. There are 1,151 elementary schools, 461 middle schools, and 408 high schools in the calculations. There are 125 schools with non-traditional grade spans included.[1] There are 30 non-traditional schools in the BTO dataset. In 2017, 77 charter schools are included in the BTO file (11 conversions, 44 start-ups, and 22 state-authorized).

The outcome of the prediction equations is Single Score, with a mean and median of about 71.5 and a standard deviation of 11.4. The outcome is normally distributed with few scores below 50 and few above 90. Table 1 and Figures 1-4, below, show the descriptive statistics and distribution of the predictors. White, Black, and Hispanic median percentages are about 40, 32 and 9, respectively. Very few schools have large Multi-racial or Native American percentages. The distribution of Hispanic students is skewed with many high outliers (mean is 14, median is 8.6). Most schools have a low percentage of Asian students, but there are many outliers on this variable as well (schools with high Asian percentages).

Percent of students with disabilities (SWD) has a mean of about 12 and a median of 11.6. There are five outliers[2] with much higher percentages between 55 and 100 percent: Tapestry Public Charter School, Margaret Harris Comprehensive School, Atlanta School for the Deaf, Georgia Academy for the Blind, and Georgia School for the Deaf.[3] The last four have 100% SWD. Schools do not vary much on their percentages of female students, but there are some schools

[1] Traditional grade spans are elementary school (K-5), middle school (6-8), and high school (9-12).
[2] Outliers are observations more than 1.5*Interquartile Range (IQR) beyond the nearest quartile (the 75th or 25th percentile).
[3] GOSA does not include the state special schools (Atlanta School for the Deaf, Georgia Academy for the Blind, and Georgia School for the Deaf) in BTO.

that have 100% males or females and others with high percentages of one or the other gender. Percent English Language Learner (ELL) has some high outliers – schools with high ELL populations. The median Direct Certification rate is about 34, but this percentage varies a great deal from about 2 to 79, with no outliers.

The school churn rate (number of exits and entries from the October FTE count date to May 1 divided by number of students present at October FTE count date) has a mean of 20 and a median of 15. The middle 50% of the observations lie between 11 and 23 (between the 25^{th} and 75^{th} percentiles), but there are many outliers. 50 schools have churn rates above 100. All but four are non-traditional schools. The mean and median enrollment count is 776 and 659, respectively. The 25^{th} and 75^{th} percentiles of enrollment are 478 and 902, with some outliers. Several high schools in Gwinnett County have large enrollments (between 3,000 and 4,000). The Georgia Cyber Academy has an enrollment of about 14,000, putting it about 24 standard deviations above the median.

Recommended Changes to Beating the Odds Adjustment

The goal of the BTO adjustment is to level the playing field for schools with social and other disadvantages. Schools with such populations are less likely to have high test score levels due to the strong relationship between family poverty and prior test score levels, to take just one example.[4] Therefore, family economic background should be considered in determining whether a school BTO because this factor is beyond the control of school staff. In general, BTO attempts to adjust predictions for each school for the student body characteristics that are not within the control of educators. Because BTO hinges on both predictors and functional form considerations, these factors have weighed heavily in the following recommendations:

1. Use the standard deviation of the forecast rather than the standard deviation of the prediction.
2. Include squared and cubic terms to allow for non-linear relationships between predictors and the outcome.
3. Stratify prediction models by school size.
4. Include a binary predictor for non-traditional schools.
5. Use percent Direct Certification rather than percent free/reduced-price lunch as a socio-economic status predictor.
6. Use percent Female as a predictor.
7. Omit percent Native American as a predictor.

What follows is a discussion of each recommendation. At the end of this document is an estimation equation showing the variables to be included as predictors.

This report uses full-time equivalent (FTE) variables when available. It does not use percent of students who are academically gifted as approximately five percent of the schools are missing this variable. I considered whether or not to also stratify models by grade span. There was not sufficient evidence to justify such a change. Virtually none of the coefficients varied by grade

[4] See, for example: Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The future of children*, 55-71.

span (see table 8; note the largely non-significant interaction terms of grade span coded E, H, M, and O, for "other"). An additional consideration is that elementary, middle, and high schools vary in size and the recommendation to stratify by school size will partially address the concern that predictions may differ by grade span.

<u>(1) Use the standard deviation of the forecast rather than the standard deviation of the prediction.</u>

There are two definitions for the general term "prediction": the predicted value and the forecast. The former is the average value of the outcome for given values of the covariates. The latter is the expected value of the outcome for a given set of covariates. The point estimates of regression slopes are the same, but the variance of the forecast is larger than the variance of the prediction. The variance of both the prediction and the forecast depends on a measure of overall model fit called root mean squared error (RMSE, also known as the standard error of the regression) and the "leverage" of a particular unit of observation. Leverage is higher for units with extreme values on regressors (i.e., those with predictor values far from the mean value of the predictor).[5] For example, suppose the mean of a predictor is 50. A school with a value of 90 on the predictor has higher leverage than a school with a value of 51. Leverage for unit *j* is defined as:

$$h_j = x_j(\boldsymbol{X}'\boldsymbol{X})^{-1}x_j'$$

where $\boldsymbol{X}$ is a *n x k* matrix of all the values of all of the covariates, $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is the inverse of the $(\boldsymbol{X}'\boldsymbol{X})$ matrix, x_j is a *1 x k* row vector of covariate values for unit *j*, and x_j' is the *k x 1* transpose of the row vector of covariate values for unit *j*.

The standard error for the *prediction* (STDP) is defined as:

$$s_{p_j} = RMSE\sqrt{h_j}$$

where *RMSE* is root mean squared error:

$$RMSE = \sqrt{\frac{\sum(Y_j - \hat{Y}_j)^2}{N - k}}$$

RMSE is a measure of the average distance of actual and predicted/forecast scores in the model as a whole. Poorly fit models have higher RMSE than well fit models. For example, suppose a bivariate model is specified as a linear model, but the true relationship is actually non-linear. The RMSE of a properly specified non-linear model will be lower than the RMSE of the mis-specified linear model.

[5] This section draws heavily from Baum, C. F. (2006). *An introduction to modern econometrics using Stata*. Stata press.

The standard error for the *forecast* (STDF) is defined as:

$$s_{f_j} = RMSE\sqrt{1 + h_j}$$

In this application, leverage for most units is quite small (M=.016; SD=.037 for the preferred specification, all school sizes combined), so the standard error of the prediction, and thus the prediction confidence interval for most units, is quite small as well. Since leverage is small, the standard error of the forecast is approximately equal to $RMSE\sqrt{1} = RMSE$.

To illustrate these concepts, figures 15 and 16 display scatterplots of leverage values from bivariate regressions involving the outcome and one predictor (the one shown on the x-axis). As shown in the figures, leverage is minimized at the mean of the predictor and is particularly high for outlying values. Figure 17 shows that the CIs created from STDP are quite narrow and vary by the value of the predictor, whereas Figure 18 shows that the CIs created from STDF are wider and more consistent across different values of the predictors. As careful observers will notice, Figures 16 and 17 also show that linear specifications produce poor predictions (above 100 and below 0), which suggests that a non-linear model is necessary (see recommendation 2).

If the BTO adjustment continues to use STDP, then leverage will remain the main determinant of the confidence interval, confidence intervals will be much narrower than under STDF, and RMSE will not matter much in the calculation of confidence intervals. This makes little sense because confidence in predictions should be a function of overall model fit, which as we will see below varies by school size. I recommend a threshold of 1*STDF, which will leave about 14-15% of schools as Beating the Odds.

The main consequence of this decision will be that more schools will be above their prediction but below the top confidence interval (CI). This is because under the status quo alternative (STDP), the CI is narrow, so very few schools are above prediction and below the top CI.

Stata, the software used to compute the BTO model, does not permit robust standard errors with STDF, so the proposed specification recommends not using the robust option. This should not raise concerns given that the standard errors with and without the robust option are virtually identical.

(2) Include squared and cubic terms to allow for non-linear relationships between predictors and the outcome.

I recommend including squared and cubed terms of all continuous covariates in the prediction model. A linear model can produce poor predictions if the true relationship between a predictor or set of predictors and the outcome is actually non-linear.

To test for non-linearities, I fit a model with squared and cubed terms for all continuous predictors. Table 2 shows this model in column 1 and the baseline linear specification in column 2. The cubic model fits the data better (R^2 is .689 versus .619 for the linear model; F-test of joint significance of additional predictors is significant at $p<.001$). Better model fit improves the

STDF, which narrows the CI. Allowing for non-linear relationships also allows curves to bend to fit schools with outlying values, which is particularly important because many predictors have outlying values.

Due to multicollinearity of squared and cubed terms, it can be difficult to assess the necessity of including them on the basis of statistical testing of individual coefficients. To address this complication, I conducted a test of joint significance of each set of squared and cubed terms to determine whether together they explained additional variation in the outcome. For percent Female, Asian, Multi-racial, Direct Certification, ELL, SWD and Churn, I reject the null of no additional explained variation at $p<.05$. I accept the null of no additional explained variation for percent Black and Hispanic at $p<.05$. So, in general, there is evidence of non-linear relationships between the predictors and the outcome.

Another approach to assessing non-linearity is to examine plots of model predictions by values of covariates in the model holding other predictors constant at their actual values. Figures 5-13 show these prediction plots for all the continuous variables in the model (from a model that pools across school size for convenience). For example, most schools have about 50% female (see Figure 2), but those that deviate from this norm generally experience an advantage in percentage Female, at least within a range (Figure 5). At the extreme low and high values, percent Female is negatively predictive of single score. As discussed above, percent Black and percent Hispanic are negative and approximately linear predictors, except at the extremes where they tail off in their influence (Figures 7 and 8). Percent Direct Certification appears approximately linear (Figure 10) even though the F-test rejects the null of no additional variation explained by the higher order terms. I recommend retaining the higher order terms because it is also a very strong predictor (see range of variation in linear prediction on the y-axis for this predictor compared to a weaker predictor such as percent Black). The strength of this relationship could affect schools between the linear and non-linear prediction. Churn rate is also a strong negative predictor with many outliers and strong evidence of non-linearity.

To see the impact of non-linearity on prediction, consider the case of percent ELL in Figure 14. The red line plots a cubic relationship and the blue line plots a linear relationship. Between about 20 and 80 percent ELL, a linear model would make it easier to BTO, while at values higher than about 80 percent ELL, a linear model would make it harder to BTO. This is because the cubic fit predicts higher values than the linear model within the first range and lower values above 80 percent ELL. In short, because percent ELL is non-linearly related to the outcome, predictions for some schools will differ.

Examination of margins plots and the range of variation of each predictor suggests that the following negatively predict single score: percent Black, Hispanic, Direct Certification, Churn, and SWD. Percent Female and percent Asian positively predict the outcome. Percent ELL and Multi-racial are not strong positive or negative predictors for most values of these predictors, but both negatively predict the outcome at higher values of these predictors.

For reference, I also include the results using 2016 data for the pooled model for the non-linear and linear specifications. Measures of model fit are superior for the non-linear specification relative to the linear one. F-tests of higher order terms produce largely the same conclusions – ample evidence of non-linearity in all predictors except percent Black and Multi-racial.

For the purpose of this technical report, I have retained squared and cubed terms whether they are jointly or individually significant or not. Extraneous terms will harm the precision of coefficient estimates, so it may be preferable to fine-tune the specification to remove these terms each year the model is run.

(3) Stratify prediction models by school size.

The primary reason to stratify prediction models by school size is that the reliability of the outcome will be worse for smaller schools than for larger schools. It will be more difficult for any model to predict a noisy outcome than it will be for a model to predict a less noisy one. If this is true, then model fit and the CI of the prediction will vary by school size. I find evidence supporting these hypotheses in the data. Table 4 shows non-linear models stratified by school size. For the purposes of this report, I coded small as 0-500, medium as 501-1000, and large as 1001 or more. Using fixed numbers rather than comparative numbers, such as terciles, ensures the categories do not fluctuate each year. These categories also ensure sufficient sample size in each. About 28%, 52%, and 20% of schools are small, medium, and large, respectively. Specifically, these models were separately fit to these mutually exclusive and exhaustive school size classifications, which allows the predictors to flexibly vary across the size classifications. Model fit is better for larger schools than smaller schools. RMSE is 7.8 for the smallest schools and 4.9 for the largest schools. Other things equal, this will make it more difficult for a small school to exceed the BTO threshold. This is appropriate if we have more uncertainty about the prediction for these types of schools. A wider confidence interval is appropriate for smaller schools if it is more difficult to determine the distance between actual and predicted due to noise in the outcome.

Table 5 shows that under the status quo linear specification with STDF, about 18.5% of small schools would Beat the Odds, whereas only 11% of larger schools would do so. Under a non-linear model with STDF, the percent of larger schools above the threshold would be 13% and the percent of small schools above the threshold would fall to about 15%.

For reference, I include descriptive statistics on the outcome, predicted scores, errors of forecast, confidence intervals and Beating the Odds variables in Table 6. Note that there are some actual and predicted scores above 100 in 2017. There is no evidence that high scoring schools are disadvantaged by the BTO adjustment. In fact, schools with actual single scores above 95 are more likely to exceed the BTO threshold than for the sample as a whole.

As shown in Table 7, the percentage of schools that would Beat the Odds in 2017 with a model adopting all of these recommendations would be 14.4%. About 37% of schools would have observed scores above predicted, but would be below the top CI. About 34% of schools would have observed scores less than predicted, but above the lower CI. And 15% would have observed less than predicted and below the lower CI.

(4) Include a binary predictor for non-traditional schools.

This accounts for the special populations of these schools not captured by the other predictors. In the pooled model shown in Table 2, it is a highly significant predictor with a coefficient of about 9 points (t=-4.54, p<.000), which adds to explained variance in the outcome (F-test comparing model fit with and without this variable is significant at p<.001). It is also a significant predictor in Table 4 among schools in the smallest size category.

(5) Use percent Direct Certification rather than percent free/reduced-price lunch as a socio-economic status predictor.

With the implementation of the Community Eligibility Provision, percent economic disadvantaged (as measured by free/reduced-price lunch) is becoming a less reliable measure of school poverty. As can be seen in Figure 19, the FRL percentage is 100% for schools that vary a great deal on percent Direct Certification, which is a measure of the percentage of students who automatically qualify for food assistance due to TANF, SNAP, and other federal programs. In addition, there are some schools that have quite low percent FRL and yet higher than expected percent Direct Certification. There is also quite a wide range of variability in percent Direct Certification for schools that do and do not participate in the Community Eligibility Program (see Figures 20 and 21). In short, I believe that percent Direct Certification will be a more valid measure of school socio-economic status moving forward.

(6) Use percent Female as a predictor.

Although there is limited range of variation in this predictor, it is a significant predictor of the outcome as is shown in Tables 2, 3, and among small schools in column 1 of Table 4.

(7) Omit percent Native American as a predictor.

Percent Native American is not a significant predictor of the outcome and very few schools have more than three Native American students.

Proposed Model

The proposed model is shown in the equation below. This model should be estimated separately by school size category, as discussed above. Most variables are continuous school-level percentages of student-level characteristics. For example, the variable Female is percentage of students that are female. All continuous variables include linear, quadratic, and cubed terms to allow for non-linearity in the effect of the variable on the outcome. EM, EMH, H, M, and MH are each binary indicators for grade spans (E is reference). Non-traditional is a binary variable coded as 1 if the school is a non-traditional school, as defined by GOSA, and 0 otherwise.

$$\begin{aligned}Y_i = \beta_0 &+ \beta_1 Female + \beta_2 Female^2 + \beta_3 Female^3 + \beta_4 Asian + \beta_5 Asian^2 + \beta_6 Asian^3 \\ &+ \beta_7 Hispanic + \beta_8 Hispanic^2 + \beta_9 Hispanic^3 + \beta_{10} MultiEthnic \\ &+ \beta_{11} MultiEthnic^2 + \beta_{12} MultiEthnic^3 + \beta_{13} DirectCertification \\ &+ \beta_{14} DirectCertification^2 + \beta_{15} DirectCertification^3 + \beta_{16} ELL + \beta_{17} ELL^2 \\ &+ \beta_{18} ELL^3 + \beta_{19} SWD + \beta_{20} SWD^2 + \beta_{21} SWD^3 + \beta_{22} Churn + \beta_{23} Churn^2 \\ &+ \beta_{24} Churn^3 + \beta_{25} EM + \beta_{26} EMH + \beta_{27} H + \beta_{28} M + \beta_{29} MH \\ &+ \beta_{30} NonTraditional + \epsilon_i\end{aligned}$$

Table 1. Descriptive Statistics of Outcome and Predictors, 2017

	count	mean	sd	min	max
singlescore	2145	71.36	11.41	16.40	104.60
pct_female	2145	48.65	3.96	0.00	100.00
pct_asian	2145	3.19	6.46	0.00	66.00
pct_black	2145	39.38	30.78	0.00	100.00
pct_hispanic	2145	13.93	15.59	0.00	97.00
pct_multi	2145	3.47	1.99	0.00	11.00
pct_directcert	2145	34.82	19.00	0.53	85.90
pct_ell	2145	8.23	13.04	0.00	100.00
pct_swd	2145	11.97	5.47	0.00	100.00
pct_churn	2145	20.07	26.69	0.00	504.10
cluster_numeric	2145	2.54	1.77	1.00	6.00
nontrad17	2145	0.01	0.12	0.00	1.00
enrollment_count	2145	776.39	557.03	53.00	14319.00

Table 2. Coefficients and Standard Errors from Pooled Preferred Cubic Specification and Baseline Linear Specification, 2017

	(1) Cubic Model	(2) Baseline Linear Model
pct_female	-0.806407**	
	(0.269766)	
pct_female2	0.020756***	
	(0.005419)	
pct_female3	-0.000129***	
	(0.000034)	
pct_asian	0.499566***	0.302312***
	(0.110306)	(0.029569)
pct_asian2	-0.014532*	
	(0.005731)	
pct_asian3	0.000152*	
	(0.000074)	
pct_black	-0.029924	-0.085993***
	(0.043213)	(0.006979)
pct_black2	-0.001115	
	(0.001108)	
pct_black3	0.000010	

	(0.000008)	
pct_hispanic	0.037058	0.017071
	(0.074747)	(0.021662)
pct_hispanic2	-0.001669	
	(0.002207)	
pct_hispanic3	0.000005	
	(0.000018)	
pct_multi	-0.820705	0.313839***
	(0.531245)	(0.087547)
pct_multi2	0.247328	
	(0.126665)	
pct_multi3	-0.016984	
	(0.008934)	
pct_directcert	-0.780983***	
	(0.075611)	
pct_directcert2	0.012851***	
	(0.002053)	
pct_directcert3	-0.000090***	
	(0.000017)	
pct_ell	-0.054414	-0.060219*
	(0.083035)	(0.027633)
pct_ell2	0.003315	
	(0.002524)	
pct_ell3	-0.000040	
	(0.000021)	
pct_swd	-0.170950	-0.239773***
	(0.150116)	(0.029924)
pct_swd2	0.001618	
	(0.006511)	
pct_swd3	-0.000014	
	(0.000052)	

pct_churn	-0.217007^{***}	-0.104278^{***}
	(0.025266)	(0.006110)
pct_churn2	0.000798^{***}	
	(0.000172)	
pct_churn3	-0.000001^{**}	
	(0.000000)	
2.cluster_numeric	-2.162976^{**}	-1.019787
	(0.742983)	(0.792779)
3.cluster_numeric	-5.527453^{**}	-2.682525
	(1.698361)	(1.843430)
4.cluster_numeric	0.917106	2.364418^{***}
	(0.476640)	(0.500864)
5.cluster_numeric	-1.207266^{**}	-0.498007
	(0.438614)	(0.437679)
6.cluster_numeric	-0.221642	4.694634^{**}
	(1.598262)	(1.536253)
1.nontrad17	-8.894668^{***}	
	(1.957489)	
pct_native		-0.059207
		(0.494667)
pct_ed		-0.176870^{***}
		(0.010707)
enrollment_count		0.000394
		(0.000338)
1.cep_sas		2.329751^{***}
		(0.538761)

_cons	98.083267***	88.794678***
	(5.026255)	(0.823599)
N	2145	2145
R^2	0.689	0.623
rmse	6.413644	7.037039

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Coefficients and Standard Errors from Pooled Preferred Cubic Specification and Baseline Linear Specification, 2016

	(1) Cubic Model	(2) Baseline Linear Model
pct_female	-0.526044*	
	(0.204552)	
pct_female2	0.013561**	
	(0.004536)	
pct_female3	-0.000083**	
	(0.000030)	
pct_asian	0.574006***	0.343769***
	(0.115891)	(0.031288)
pct_asian2	-0.017009**	
	(0.006468)	
pct_asian3	0.000183*	
	(0.000090)	
pct_black	-0.134430**	-0.105137***
	(0.042717)	(0.007137)
pct_black2	0.001285	
	(0.001093)	
pct_black3	-0.000008	
	(0.000008)	
pct_hispanic	0.162986*	0.028269
	(0.073774)	(0.021762)
pct_hispanic2	-0.005254*	
	(0.002196)	

pct_hispanic3	0.000029	
	(0.000018)	
pct_multi	-0.982050	0.421782***
	(0.549127)	(0.090324)
pct_multi2	0.308944*	
	(0.140658)	
pct_multi3	-0.022406*	
	(0.010640)	
pct_directcert	-0.687545***	
	(0.071603)	
pct_directcert2	0.010360***	
	(0.001829)	
pct_directcert3	-0.000067***	
	(0.000014)	
pct_ell	-0.127446	-0.099241***
	(0.082367)	(0.027785)
pct_ell2	0.005163*	
	(0.002516)	
pct_ell3	-0.000057**	
	(0.000021)	
pct_swd	-0.368319**	-0.213318***
	(0.124496)	(0.031735)
pct_swd2	0.012399*	
	(0.005185)	
pct_swd3	-0.000105*	
	(0.000042)	
pct_churn	-0.270865***	-0.133890***
	(0.025646)	(0.006989)
pct_churn2	0.001003***	
	(0.000214)	

pct_churn3	-0.000001**	
	(0.000000)	
2.cluster_numeric	-4.045356***	-2.567769***
	(0.727556)	(0.774229)
3.cluster_numeric	-4.418517*	-2.740298
	(1.766947)	(1.819395)
4.cluster_numeric	0.759579	2.497377***
	(0.478703)	(0.498425)
5.cluster_numeric	-0.673975	0.287758
	(0.431011)	(0.432324)
6.cluster_numeric	-3.324971*	2.313753
	(1.406077)	(1.420596)
1.nontrad17	-9.012829***	
	(1.905652)	
pct_native		0.065082
		(0.494799)
pct_ed		-0.170270***
		(0.010582)
enrollment_count		0.000374
		(0.000344)
1.cep_sas		1.686617**
		(0.522947)
_cons	96.238439***	87.113530***
	(3.704515)	(0.820248)
N	2140	2140
R^2	0.723	0.665
rmse	6.390731	7.002432

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4. Coefficients and Standard Errors from Preferred Cubic Specification, by School Size, 2017

	(1) Cubic, Small (0-500)	(2) Cubic, Medium (501-1000)	(3) Cubic, Large (1001+)
pct_female	-0.797188*	-6.334598	31.156237
	(0.377936)	(9.849751)	(97.484761)
pct_female2	0.019952*	0.116635	-0.639794
	(0.008197)	(0.179372)	(1.967532)
pct_female3	-0.000125*	-0.000675	0.004403
	(0.000052)	(0.001071)	(0.013226)
pct_asian	0.489456	0.468165**	0.246887
	(0.421323)	(0.153145)	(0.160668)
pct_asian2	0.003507	-0.013744	-0.003719
	(0.040295)	(0.008539)	(0.007612)
pct_asian3	-0.000475	0.000129	0.000042
	(0.000946)	(0.000120)	(0.000090)
pct_black	-0.176192	0.027361	0.207453*
	(0.095324)	(0.058893)	(0.082899)
pct_black2	0.001371	-0.001833	-0.006041**
	(0.002470)	(0.001488)	(0.002097)
pct_black3	-0.000002	0.000011	0.000045**
	(0.000017)	(0.000010)	(0.000015)
pct_hispanic	0.302892	-0.171444	-0.084721
	(0.172833)	(0.115996)	(0.117877)
pct_hispanic2	-0.008577	0.004276	0.001294
	(0.005380)	(0.003740)	(0.003500)
pct_hispanic3	0.000054	-0.000047	-0.000009
	(0.000043)	(0.000032)	(0.000030)
pct_multi	-1.575738	-0.084990	-0.814327
	(1.009796)	(0.763610)	(1.348656)
pct_multi2	0.353216	0.049813	0.389629
	(0.248675)	(0.177052)	(0.312379)

pct_multi3	-0.017896	-0.003745	-0.036664
	(0.017750)	(0.012237)	(0.021884)
pct_directcert	-0.539765*	-0.785899***	-1.238770***
	(0.233492)	(0.102495)	(0.187022)
pct_directcert2	0.009111	0.012633***	0.027427***
	(0.005712)	(0.002813)	(0.006981)
pct_directcert3	-0.000071	-0.000092***	-0.000249**
	(0.000042)	(0.000024)	(0.000080)
pct_ell	-0.202281	0.182903	-0.000149
	(0.188663)	(0.118322)	(0.174608)
pct_ell2	0.005385	-0.003851	0.002669
	(0.006057)	(0.004051)	(0.005219)
pct_ell3	-0.000053	0.000029	-0.000025
	(0.000048)	(0.000038)	(0.000048)
pct_swd	-0.364727	-1.349454*	0.216035
	(0.262131)	(0.630946)	(2.412652)
pct_swd2	0.002471	0.096561	-0.014054
	(0.010227)	(0.050237)	(0.222253)
pct_swd3	-0.000002	-0.002093	0.000274
	(0.000080)	(0.001278)	(0.006641)
pct_churn	-0.223264***	-0.230123**	-0.145103
	(0.048136)	(0.080828)	(0.141319)
pct_churn2	0.000992***	0.002198	-0.004694
	(0.000271)	(0.001650)	(0.003236)
pct_churn3	-0.000001**	-0.000007	0.000029
	(0.000000)	(0.000006)	(0.000015)
2.cluster_numeric	-2.231509	-0.957876	-8.343723***
	(1.512025)	(0.946152)	(2.219521)
3.cluster_numeric	-6.361500*	-3.727663	-4.708242
	(3.097787)	(2.729861)	(3.388656)

4.cluster_numeric	-0.578273	1.624192*	1.788176
	(1.265085)	(0.767882)	(1.100944)
5.cluster_numeric	-2.428952*	-0.833826	0.024716
	(1.089636)	(0.558561)	(1.048079)
6.cluster_numeric	-1.117524	1.457334	3.422638
	(2.970554)	(2.172312)	(5.109549)
1.nontrad17	-12.132811***	-6.694226	
	(3.278954)	(5.479516)	
_cons	99.267417***	206.023101	-417.218750
	(7.492707)	(177.426349)	(1613.470175)
N	605	1115	425
R^2	0.582	0.699	0.780
rmse	7.787064	5.956214	4.863994

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. Proportion Beating the Odds Under Preferred, Baseline, and Unconditional Models, by School Size, 2017

	0-500 mean	501-1000 mean	1000+ mean	Total mean
BTO2_pref_f	0.147	0.148	0.129	0.144
BTO2_base_f	0.185	0.131	0.113	0.143

All models use the standard deviation of the forecast to compute BTO.

Table 6. Descriptive Statistics for predictions, errors and Confidence Intervals from Preferred Model, 2017

	count	mean	sd	min	max
singlescore	2145	71.36	11.41	16.40	104.60
predscore_pref_f	2145	71.36	9.58	29.43	101.98
error_pref_f	2145	6.40	1.10	4.90	10.76
ci_low_pref_f	2145	64.96	10.03	19.18	95.67
ci_high_pref_f	2145	77.76	9.25	39.68	108.45
BTO4_pref_f	2145	2.50	0.92	1.00	4.00
BTO2_pref_f	2145	0.14	0.35	0.00	1.00

Predictions and errors from model run separately by school size category.

Table 7. Frequency and Percentages of Schools that Would BTO with CI Categories, Preferred Cubic Model, 2016

	freq	pct	cumpct
obs>pred, above CI	309	14.41	14.41
obs>pred, below CI	787	36.69	51.10
obs<pred, above CI	719	33.52	84.62
obs<pred, below CI	330	15.38	100.00
Total	2145	100.00	

Table 8. Cubic model run separately by school size category including interactions of predictors with school grade span

	(1) Cubic, Small (1-500)	(2) Cubic, Medium (501-1000)	(3) Cubic, Large (1001+)
pct_female	17.184771	-14.755683	-520.503027*
	(51.513711)	(66.852583)	(260.692266)
pct_female # pct_female	-0.431325	0.269118	10.832486*
	(1.068591)	(1.348914)	(5.373014)
pct_female # pct_female # pct_female	0.003501	-0.001589	-0.075063*
	(0.007376)	(0.009063)	(0.036932)
pct_asian	-0.144469	0.369833*	0.038185

	(0.502231)	(0.181027)	(0.368153)
pct_asian # pct_asian	0.052310	-0.011832	0.008074
	(0.046721)	(0.009597)	(0.015153)
pct_asian # pct_asian # pct_asian	-0.001268	0.000113	-0.000099
	(0.001050)	(0.000131)	(0.000165)
pct_black	-0.064598	0.133283	-0.034672
	(0.117591)	(0.079748)	(0.333640)
pct_black # pct_black	-0.000778	-0.004289*	0.005403
	(0.003040)	(0.001987)	(0.009674)
pct_black # pct_black # pct_black	0.000010	0.000024	-0.000053
	(0.000021)	(0.000014)	(0.000074)
pct_hispanic	0.151453	-0.350119	-1.952828**
	(0.285928)	(0.179681)	(0.705914)
pct_hispanic # pct_hispanic	-0.000143	0.008059	0.055545**
	(0.009103)	(0.005442)	(0.019749)
pct_hispanic # pct_hispanic # pct_hispanic	-0.000027	-0.000069	-0.000354*
	(0.000074)	(0.000044)	(0.000154)
pct_multi	-2.066876	0.810122	-1.173499
	(1.287605)	(1.049787)	(4.153271)
pct_multi # pct_multi	0.430275	-0.065390	0.228730
	(0.310404)	(0.226671)	(0.853476)
pct_multi # pct_multi # pct_multi	-0.019583	-0.000040	-0.013376
	(0.021898)	(0.014882)	(0.052846)
direct_crt_perc	-0.494532	-0.806393***	0.160539
	(0.290267)	(0.130376)	(0.659269)
direct_crt_perc # direct_crt_perc	0.006666	0.012245***	-0.031033

	(0.006880)	(0.003569)	(0.027194)
direct_crt_perc # direct_crt_perc # direct_crt_perc	-0.000055	-0.000084**	0.000347
	(0.000050)	(0.000030)	(0.000340)
pct_ell	-0.056679	0.331075*	1.075871
	(0.275544)	(0.165778)	(0.609560)
pct_ell # pct_ell	-0.004618	-0.007653	-0.033684
	(0.010063)	(0.005612)	(0.018640)
pct_ell # pct_ell # pct_ell	0.000064	0.000060	0.000211
	(0.000095)	(0.000050)	(0.000162)
pct_swd	-1.798153	0.367870	-14.008448
	(1.440431)	(0.963101)	(9.405437)
pct_swd # pct_swd	0.087228	-0.026578	1.109412
	(0.109298)	(0.073962)	(0.780769)
pct_swd # pct_swd # pct_swd	-0.001022	0.000647	-0.028217
	(0.002627)	(0.001792)	(0.021035)
churnrate	-0.248864	-0.668189**	-2.607558
	(0.438958)	(0.246675)	(2.136801)
churnrate # churnrate	0.005341	0.022181*	0.131480
	(0.016341)	(0.009716)	(0.133293)
churnrate # churnrate # churnrate	-0.000045	-0.000236*	-0.002275
	(0.000182)	(0.000113)	(0.002619)
H	171.146234	-712.940652	-9959.665627*
	(831.551817)	(1351.046557)	(4849.196979)
M	797.047476	-2030.871686	817.642404
	(833.028284)	(1347.680472)	(565.288378)
O	209.095477	-438.164729	2088.834803
	(826.606924)	(1235.345670)	(3860.206372)

H # pct_female	-15.517535	39.305072	612.374050*
	(51.873674)	(79.880189)	(297.654208)
M # pct_female	-48.930733	123.887913	-35.060009
	(51.807969)	(82.017693)	(23.525215)
O # pct_female	-17.731863	23.936988	-97.387256
	(51.517775)	(73.171314)	(163.907175)
H # pct_female # pct_female	0.409348	-0.728386	-12.631135*
	(1.075507)	(1.570229)	(6.108241)
M # pct_female # pct_female	0.964122	-2.488236	0.363735
	(1.072653)	(1.662908)	(0.242563)
O # pct_female # pct_female	0.441988	-0.430585	1.082584
	(1.068769)	(1.447565)	(1.774825)
H # pct_female # pct_female # pct_female	-0.003390	0.004504	0.086798*
	(0.007417)	(0.010293)	(0.041799)
M # pct_female # pct_female # pct_female	-0.006266	0.016621	0.000000
	(0.007392)	(0.011228)	(.)
O # pct_female # pct_female # pct_female	-0.003548	0.002542	0.000000
	(0.007377)	(0.009561)	(.)
H # pct_asian	5.086246	1.291065	-0.015265
	(6.215903)	(1.866774)	(0.495645)
M # pct_asian	-0.447156	0.307049	0.334119
	(3.543009)	(0.619089)	(0.513838)
O # pct_asian	1.938092	0.399828	-76.577095
	(2.167009)	(2.527721)	(216.162702)
H # pct_asian # pct_asian	-2.802181	-0.536486	-0.006271
	(5.513300)	(0.697403)	(0.025735)

M # pct_asian # pct_asian	2.038226	-0.021453	-0.006128
	(2.302469)	(0.079315)	(0.023654)
O # pct_asian # pct_asian	0.017394	-0.001708	22.366619
	(0.260548)	(0.701334)	(74.677884)
H # pct_asian # pct_asian # pct_asian	0.457973	0.011835	0.000064
	(1.139425)	(0.014992)	(0.000353)
M # pct_asian # pct_asian # pct_asian	-0.337466	0.000657	-0.000030
	(0.314145)	(0.002755)	(0.000305)
O # pct_asian # pct_asian # pct_asian	-0.003205	0.000288	-1.838899
	(0.007823)	(0.034107)	(6.691548)
H # pct_black	-0.240616	-0.183618	0.212491
	(0.315943)	(0.225183)	(0.349948)
M # pct_black	-0.376384	-0.104408	0.045953
	(0.283542)	(0.146393)	(0.386428)
O # pct_black	-0.286192	-0.447327	12.802397
	(0.395295)	(0.345260)	(8.819129)
H # pct_black # pct_black	0.004566	0.001550	-0.011584
	(0.007842)	(0.005705)	(0.010043)
M # pct_black # pct_black	0.009291	0.003564	-0.006730
	(0.007432)	(0.003704)	(0.010958)
O # pct_black # pct_black	0.005452	0.006767	-0.359483
	(0.009814)	(0.009047)	(0.237457)
H # pct_black # pct_black # pct_black	-0.000014	0.000008	0.000104
	(0.000055)	(0.000041)	(0.000076)
M # pct_black #	-0.000078	-0.000028	0.000060

pct_black # pct_black			
	(0.000054)	(0.000026)	(0.000084)
O # pct_black # pct_black # pct_black	-0.000030	-0.000016	0.002670
	(0.000069)	(0.000065)	(0.001710)
H # pct_hispanic	-0.079740	0.068294	1.980313**
	(0.745883)	(0.825577)	(0.727068)
M # pct_hispanic	-1.094572	0.274916	1.637024*
	(0.731066)	(0.324146)	(0.805319)
O # pct_hispanic	1.907439	0.157463	1.255761
	(1.151325)	(1.008151)	(0.819475)
H # pct_hispanic # pct_hispanic	0.020395	0.005370	-0.060216**
	(0.043376)	(0.056394)	(0.020471)
M # pct_hispanic # pct_hispanic	0.077506	-0.010069	-0.049111*
	(0.045030)	(0.011057)	(0.022662)
O # pct_hispanic # pct_hispanic	-0.151532*	0.020296	0.000000
	(0.066476)	(0.044621)	(.)
H # pct_hispanic # pct_hispanic # pct_hispanic	-0.000384	-0.000102	0.000404*
	(0.000724)	(0.001049)	(0.000162)
M # pct_hispanic # pct_hispanic # pct_hispanic	-0.001275	0.000095	0.000308
	(0.000690)	(0.000103)	(0.000183)
O # pct_hispanic # pct_hispanic # pct_hispanic	0.002616*	-0.000307	0.000000
	(0.001078)	(0.000516)	(.)
H # pct_multi	4.774384	5.607791	2.824821
	(3.040976)	(4.708761)	(4.881761)

M # pct_multi	1.609037 (3.259127)	-4.603760* (2.186431)	0.887538 (5.959109)
O # pct_multi	-2.197179 (6.040065)	0.677529 (3.464790)	0.000000 (.)
H # pct_multi # pct_multi	-1.044490 (0.767143)	-2.155133 (1.483653)	-0.439970 (1.167756)
M # pct_multi # pct_multi	-0.404401 (0.886096)	0.569062 (0.538684)	0.111120 (1.470992)
O # pct_multi # pct_multi	1.548169 (2.183857)	-0.001282 (0.915795)	0.000000 (.)
H # pct_multi # pct_multi # pct_multi	0.056104 (0.052761)	0.229110 (0.139868)	0.025257 (0.091406)
M # pct_multi # pct_multi # pct_multi	0.024080 (0.069304)	-0.012731 (0.040062)	-0.017045 (0.117275)
O # pct_multi # pct_multi # pct_multi	-0.270367 (0.234903)	0.014761 (0.070203)	0.000000 (.)
H # direct_crt_perc	0.263345 (1.281285)	-0.357375 (0.920065)	-1.565639* (0.707559)
M # direct_crt_perc	0.374737 (0.995523)	-0.151698 (0.323920)	-0.744761 (0.889990)
O # direct_crt_perc	0.320446 (1.133357)	0.896775 (0.592666)	0.000000 (.)
H # direct_crt_perc # direct_crt_perc	-0.013434 (0.034381)	0.010468 (0.023745)	0.065719* (0.028808)
M # direct_crt_perc # direct_crt_perc	-0.007150 (0.025031)	0.004920 (0.008342)	0.040372 (0.035348)

O # direct_crt_perc # direct_crt_perc	-0.006080	-0.017674	0.000000
	(0.031280)	(0.017872)	(.)
H # direct_crt_perc # direct_crt_perc # direct_crt_perc	0.000131	-0.000085	-0.000654
	(0.000285)	(0.000190)	(0.000357)
M # direct_crt_perc # direct_crt_perc # direct_crt_perc	0.000079	-0.000044	-0.000495
	(0.000192)	(0.000066)	(0.000434)
O # direct_crt_perc # direct_crt_perc # direct_crt_perc	0.000063	0.000115	0.000000
	(0.000258)	(0.000158)	(.)
H # pct_ell	-0.883170	2.117180	-0.750179
	(1.562376)	(2.657248)	(0.794858)
M # pct_ell	-1.298977	0.076528	-1.238676
	(1.463770)	(0.647553)	(0.918011)
O # pct_ell	-1.172206	-0.928666	0.000000
	(1.843609)	(1.127165)	(.)
H # pct_ell # pct_ell	-0.005400	-0.618370	0.037337
	(0.207609)	(1.026737)	(0.040369)
M # pct_ell # pct_ell	0.039855	-0.036534	0.038427
	(0.176411)	(0.072093)	(0.037437)
O # pct_ell # pct_ell	0.348371	0.014901	0.000000
	(0.237735)	(0.062932)	(.)
H # pct_ell # pct_ell # pct_ell	0.002728	0.042351	-0.000319
	(0.009356)	(0.088515)	(0.000598)
M # pct_ell # pct_ell # pct_ell	0.004847	0.001552	-0.000217
	(0.005831)	(0.002254)	(0.000468)

O # pct_ell # pct_ell # pct_ell	-0.016905	-0.000101	0.000000
	(0.008611)	(0.000941)	(.)
H # pct_swd	1.455538	3.578319	13.013940
	(2.244992)	(4.455497)	(10.270075)
M # pct_swd	3.779289	-3.482473	0.659773
	(1.998196)	(1.869992)	(16.502354)
O # pct_swd	2.222230	-7.682096	0.000000
	(1.855979)	(4.186300)	(.)
H # pct_swd # pct_swd	-0.099379	-0.417435	-1.004448
	(0.156826)	(0.372382)	(0.864604)
M # pct_swd # pct_swd	-0.178411	0.247584	0.056312
	(0.122926)	(0.145551)	(1.430214)
O # pct_swd # pct_swd	-0.106609	0.961953	0.000000
	(0.113082)	(0.586571)	(.)
H # pct_swd # pct_swd # pct_swd	0.001124	0.014615	0.024543
	(0.003315)	(0.010075)	(0.023666)
M # pct_swd # pct_swd # pct_swd	0.001686	-0.005465	-0.003783
	(0.002663)	(0.003635)	(0.040300)
O # pct_swd # pct_swd # pct_swd	0.001156	-0.041264	0.000000
	(0.002634)	(0.024925)	(.)
H # churnrate	-0.153007	0.047204	2.251883
	(0.452115)	(0.573717)	(2.143904)
M # churnrate	-0.099318	0.495351	1.755268
	(0.592654)	(0.586994)	(2.540272)
O # churnrate	-0.607938	2.122152	0.000000
	(0.645704)	(1.192818)	(.)
H # churnrate # churnrate	-0.003299	-0.014631	-0.135051
	(0.016351)	(0.016046)	(0.133338)

M # churnrate # churnrate	-0.002494	-0.023980	-0.076880
	(0.018261)	(0.025021)	(0.152044)
O # churnrate # churnrate	0.006499	-0.122286	0.000000
	(0.019081)	(0.083208)	(.)
H # churnrate # churnrate # churnrate	0.000042	0.000213	0.002305
	(0.000182)	(0.000121)	(0.002620)
M # churnrate # churnrate # churnrate	0.000039	0.000266	0.001310
	(0.000183)	(0.000323)	(0.002885)
O # churnrate # churnrate # churnrate	0.000012	0.001835	0.000000
	(0.000184)	(0.001662)	(.)
nontrad17=1	-4.250046	1.056318	
	(5.650156)	(7.320768)	
nontrad17=0			0.000000
			(.)
Constant	-118.958291	353.921560	8493.009683*
	(826.241439)	(1103.366353)	(4226.707903)
Observations	605	1115	425
R^2	0.680	0.723	0.821
rmse	7.333564	5.928575	4.767177

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

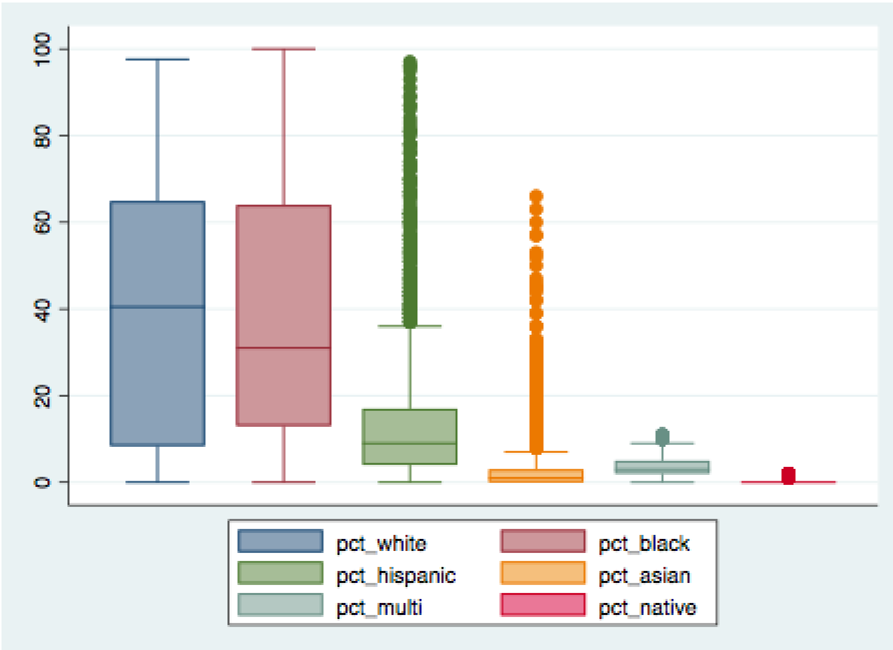

Figure 1. Distribution of Selected School Predictors I

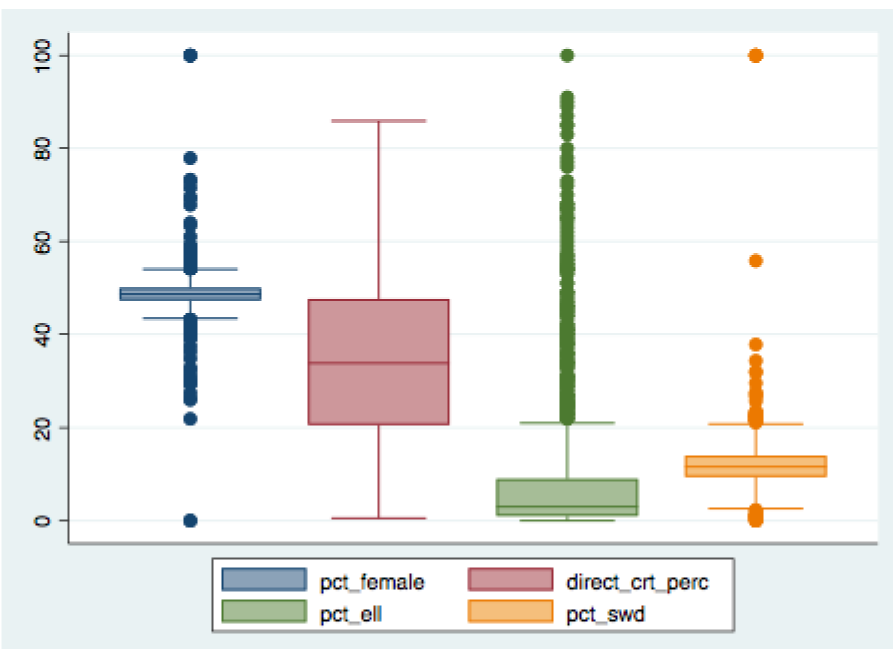

Figure 2. Distribution of Selected School Predictors II

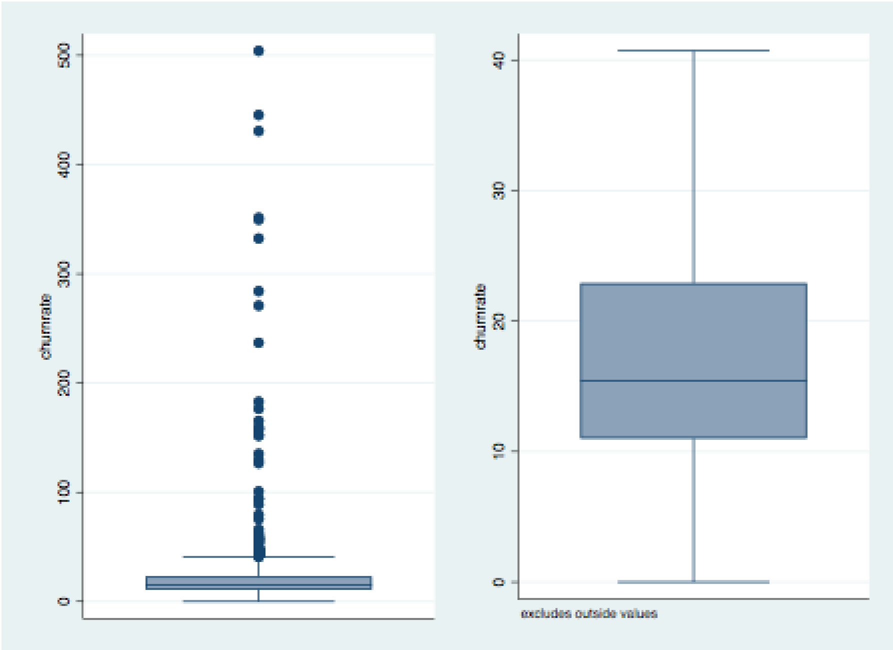

Figure 3. Distribution of Churn Rate with and without outliers

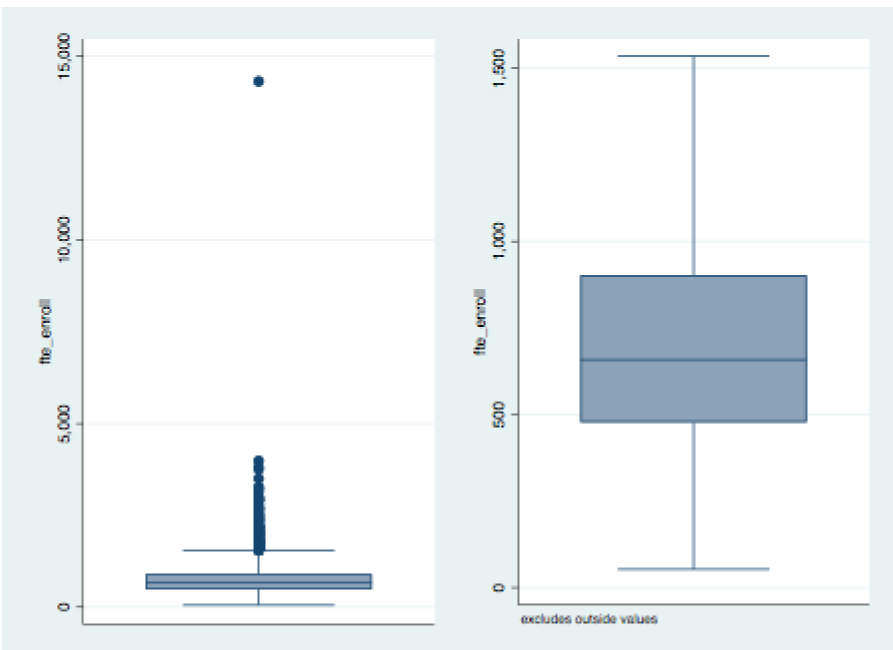

Figure 4. Distribution of Enrollment with and without outliers

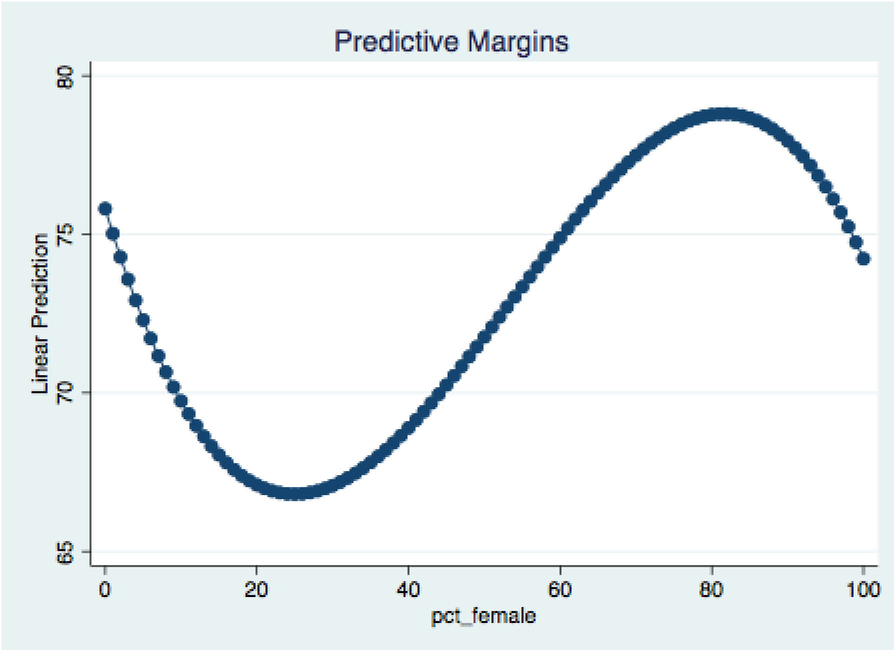

Figure 5. Margins Plot of Percent Female Holding Other Predictors Constant

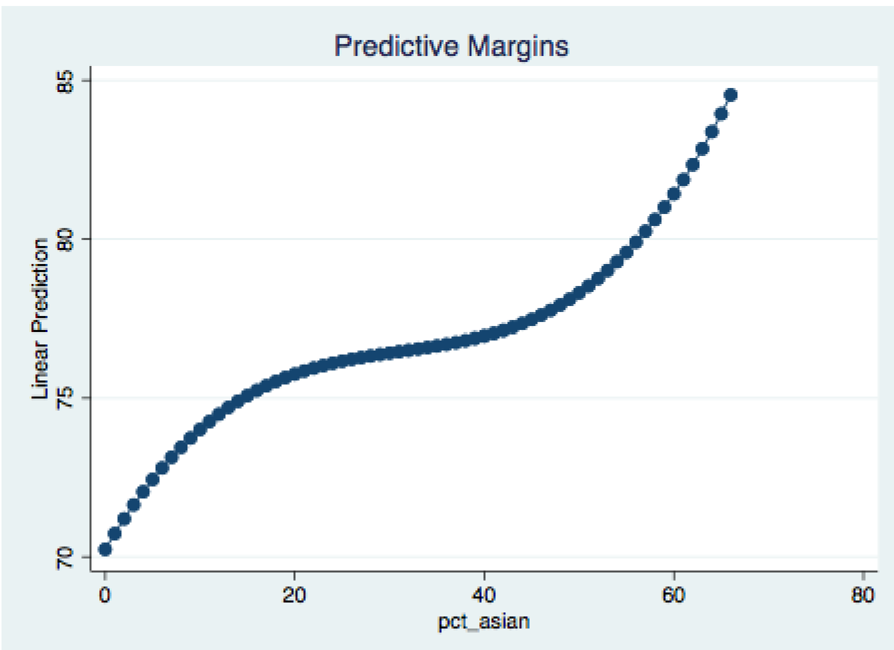

Figure 6. Margins Plot of Percent Asian Holding Other Predictors Constant

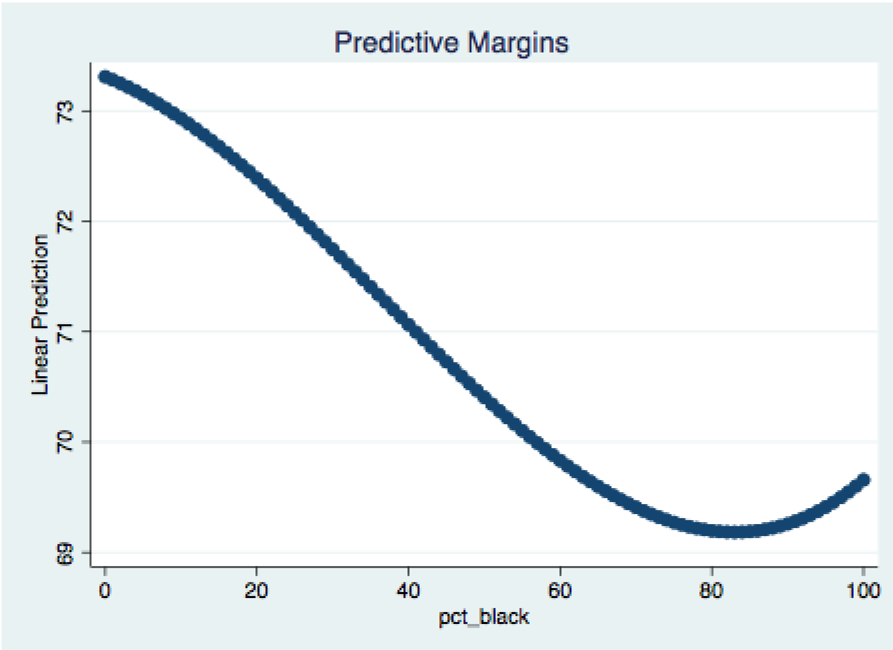

Figure 7. Margins Plot of Percent Black Holding Other Predictors Constant

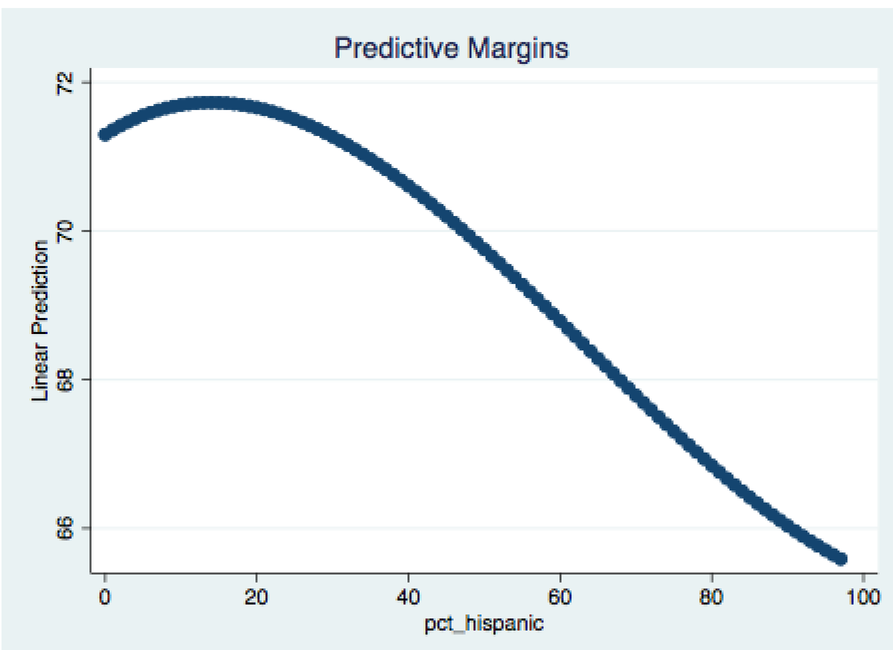

Figure 8. Margins Plot of Percent Hispanic Holding Other Predictors Constant

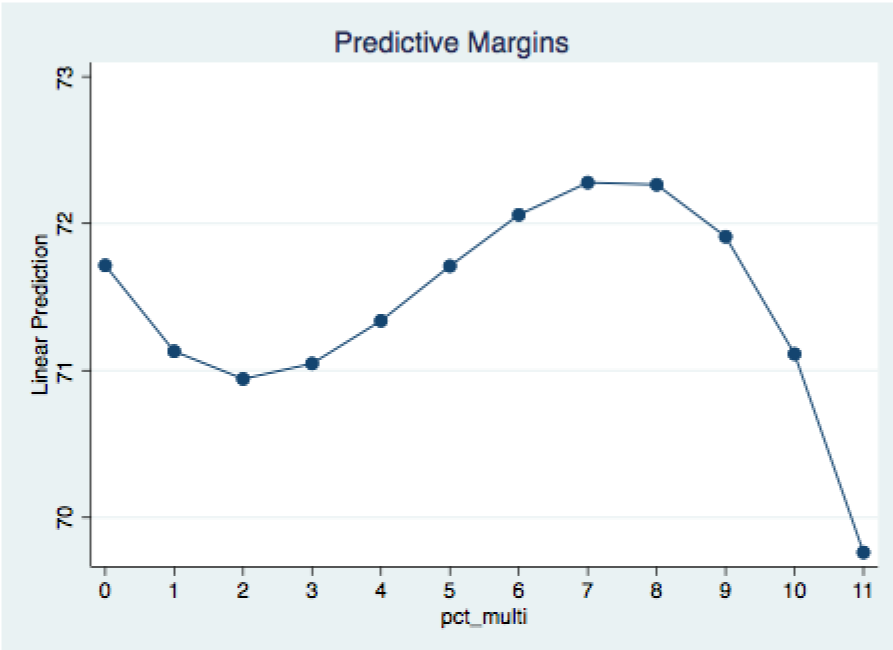

Figure 9. Margins Plot of Percent Multi Ethnic Holding Other Predictors Constant

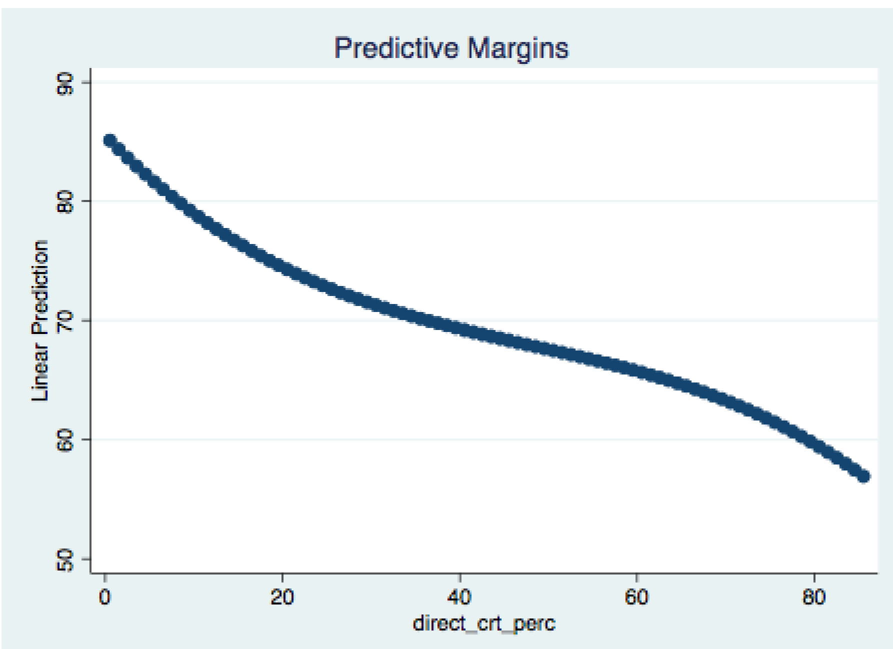

Figure 10. Margins Plot of Percent Direct Certification Holding Other Predictors Constant

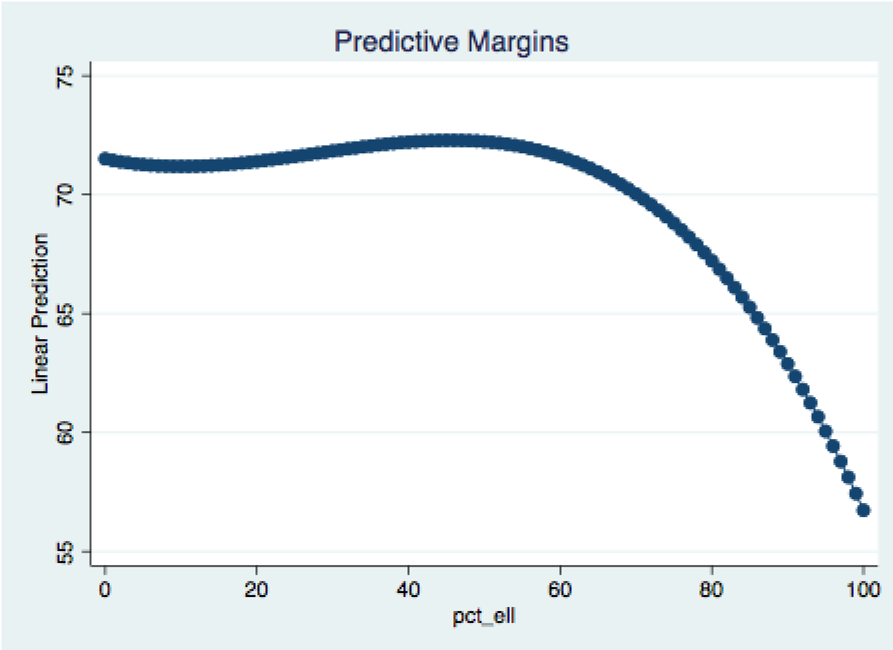

Figure 11. Margins Plot of Percent ELL Holding Other Predictors Constant

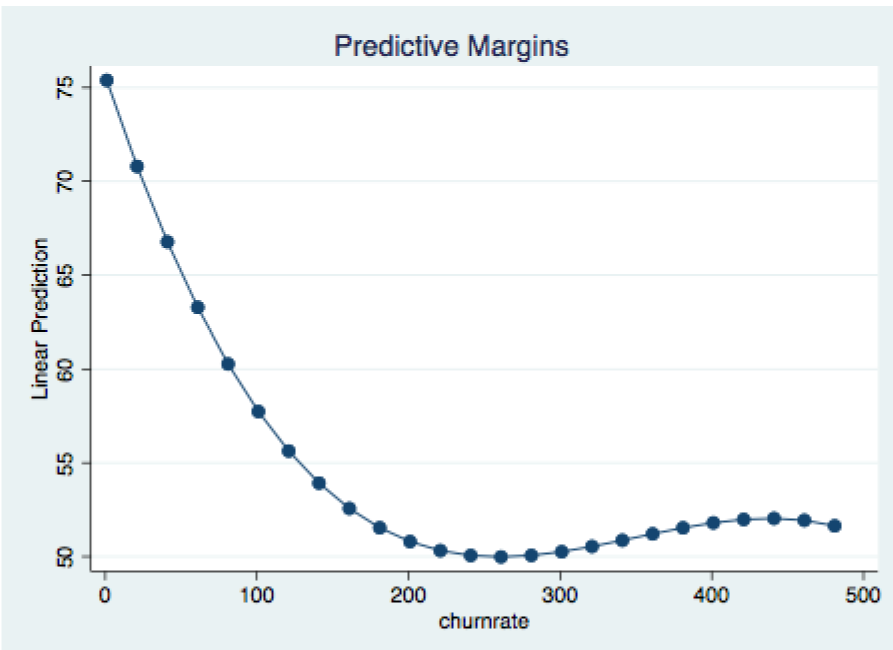

Figure 12. Margins Plot of Percent Churn Holding Other Predictors Constant

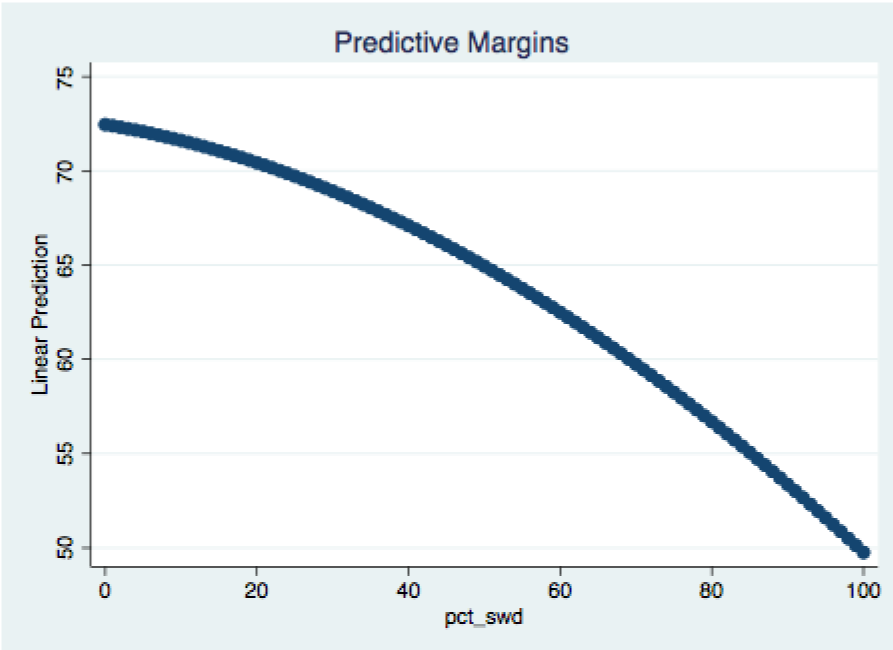

Figure 13. Margins Plot of Percent SWD Holding Other Predictors Constant

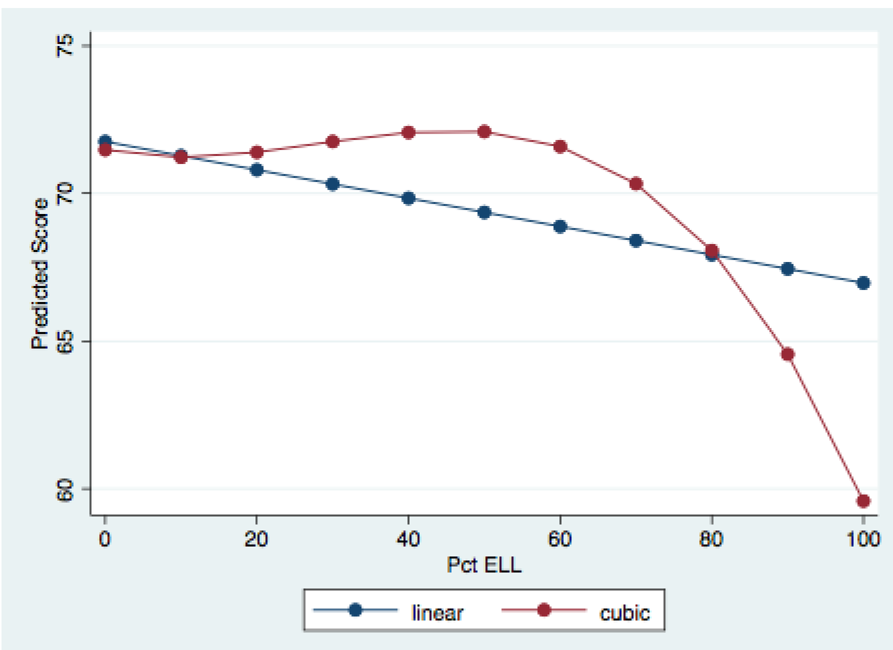

Figure 14. Margins Plot Comparison for Percent ELL from Cubic (red) and Linear (blue) Specifications

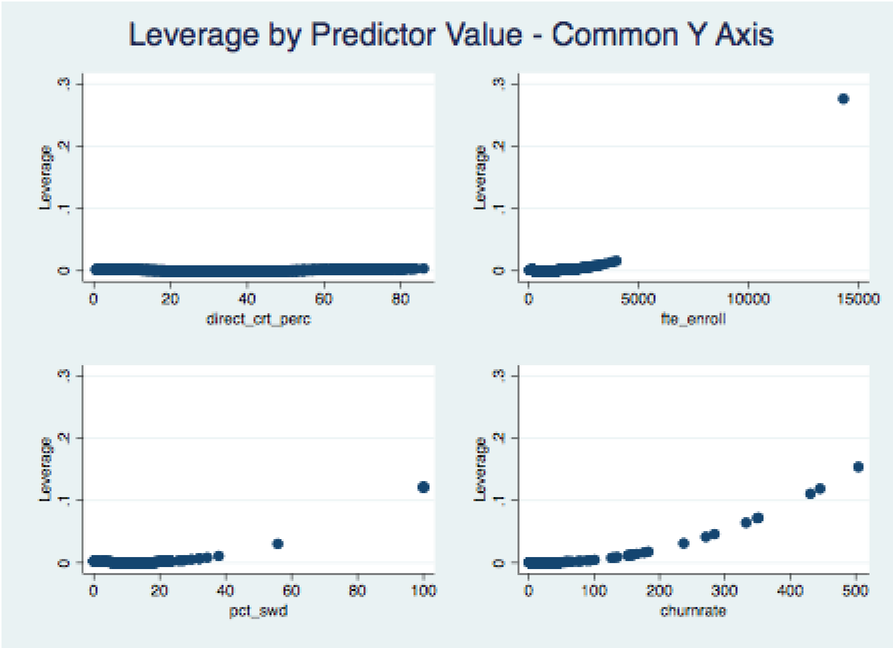

Figure 15. Leverage plots for selected predictors, common Y Axis

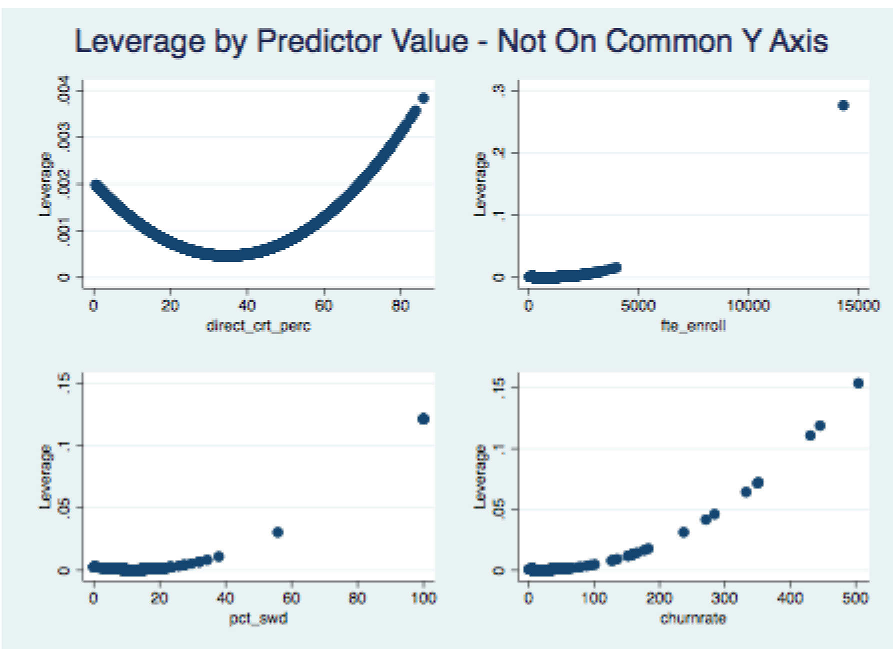

Figure 16. Leverage plots for selected predictors, not on a common Y Axis

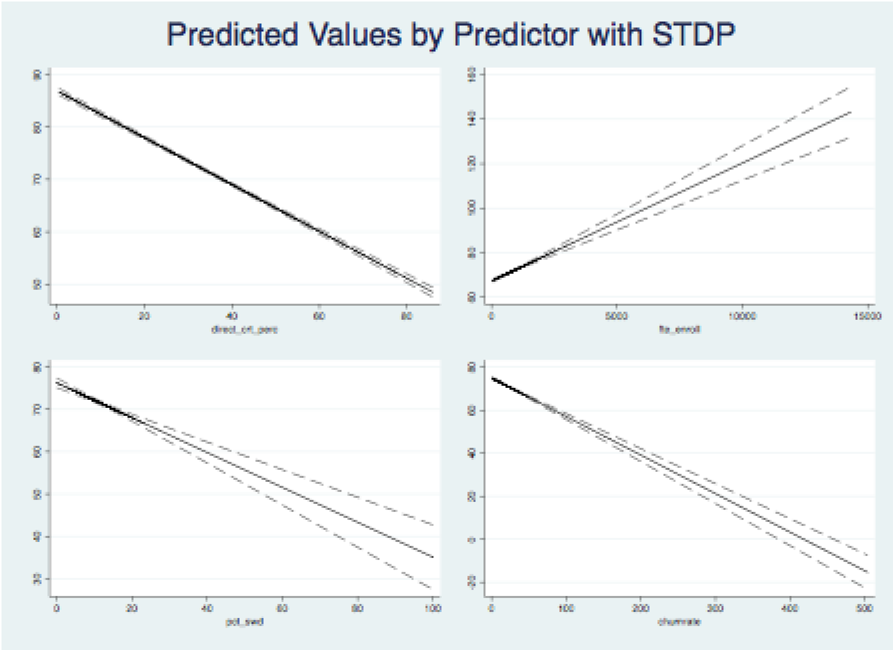

Figure 17. Predictions and Confidence Intervals for Selected Predictors, Standard Deviation of the Prediction

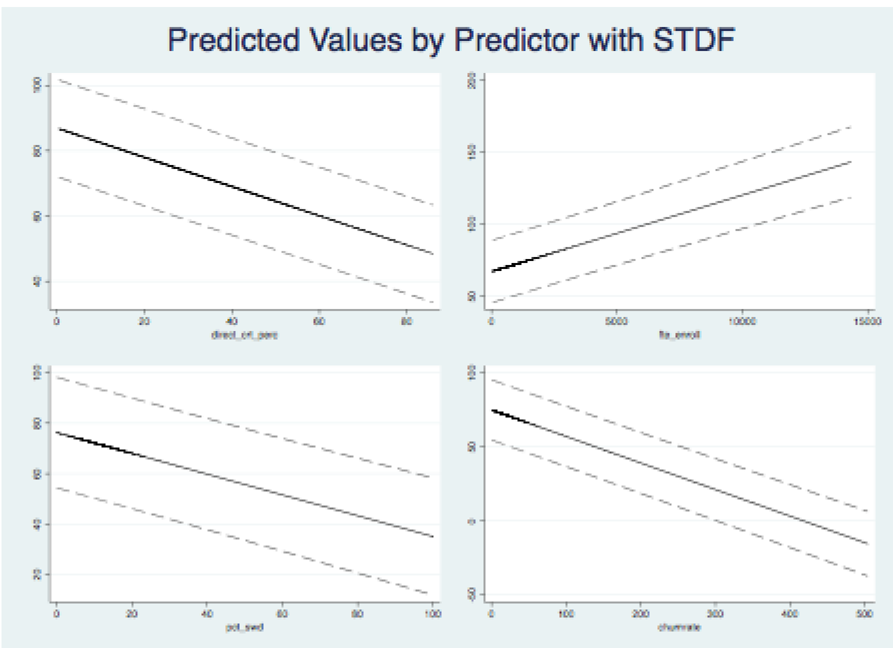

Figure 18. Predictions and Confidence Intervals for Selected Predictors, Standard Deviation of the Forecast

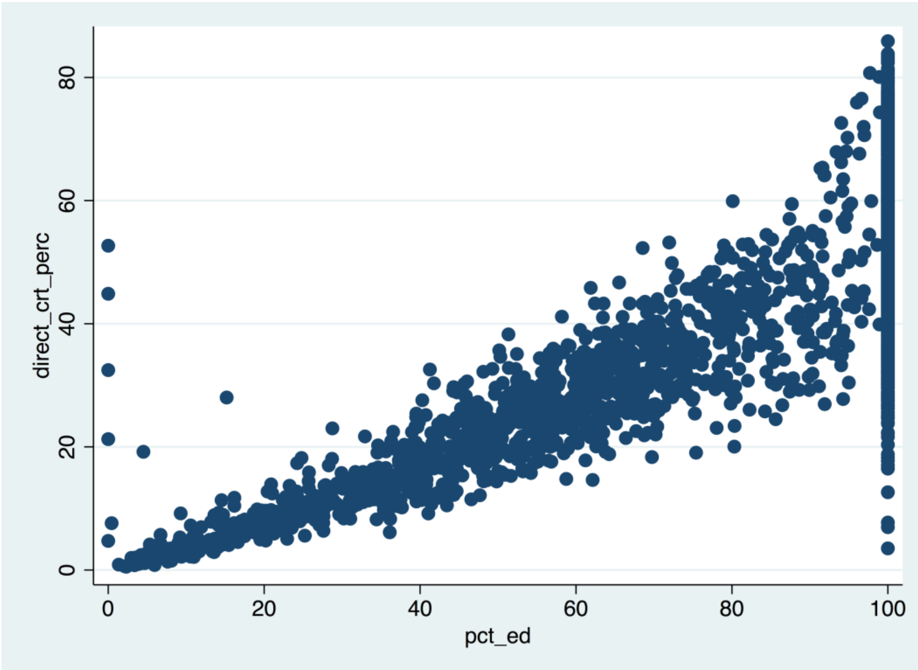

Figure 19. Scatterplot of Percent Economic Disadvantage and Percent Direct Certification

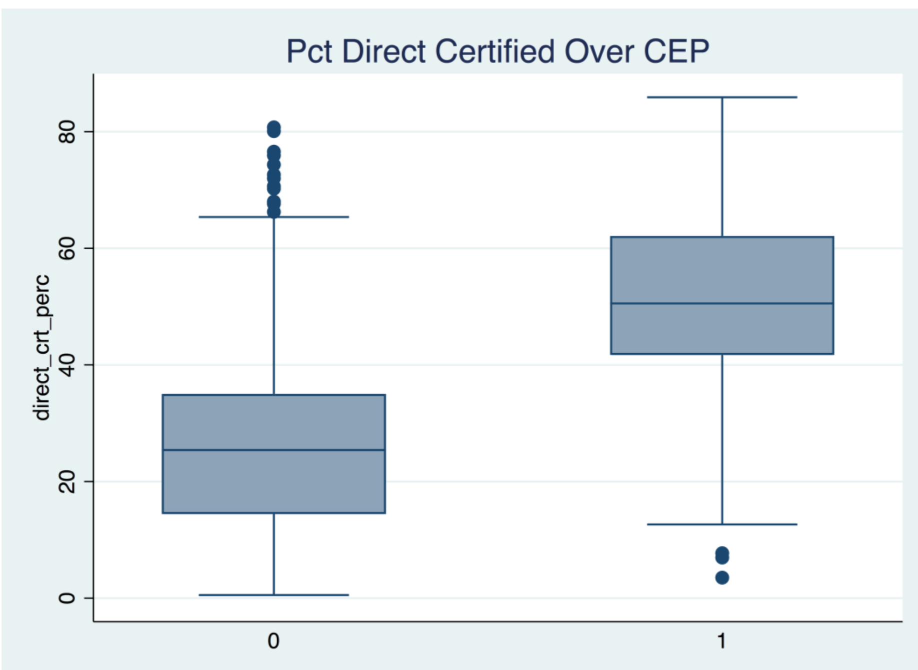

Figure 20. Box Plots of Percent Direct Certification by CEP Status

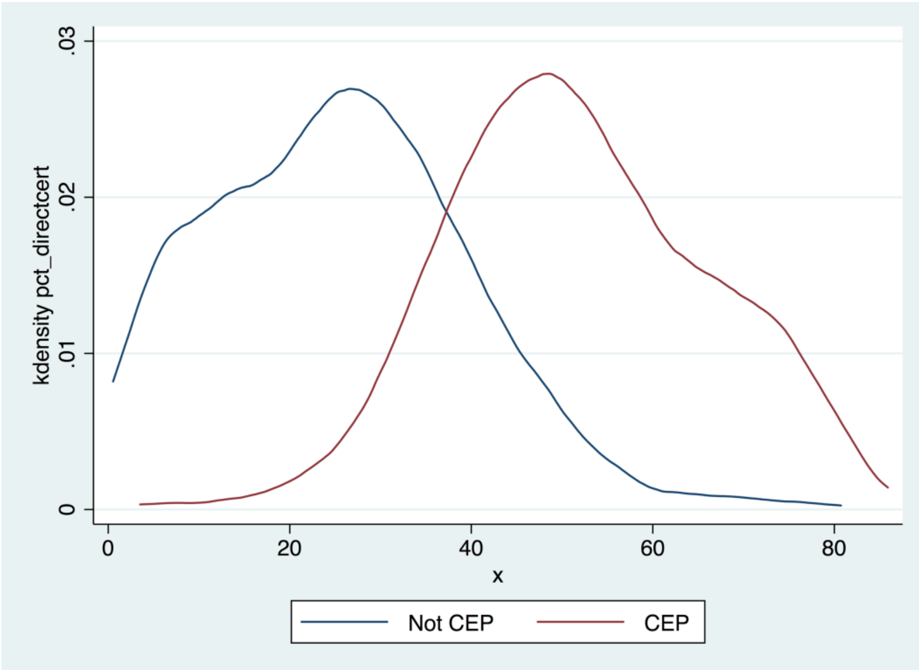

Figure 21. Kernel Density Plots of Percent Direct Certification by CEP Status