A Statistical Analysis of Ohio School District Ratings

Statistics 36-707: Applied Regression
Carnegie Mellon University

December 12, 2001

Abstract

Since 1997, the Ohio Department of Education has graded school districts based on the extent the which they meet 27 standards, most of which are percentages of students passing the state proficiency tests. A low grade has several practical consequences and can even lead the state to take partial control of the district. In this study, I look for district and student characteristics that help predict the number of standards a district will meet. I do this by studying and fitting a generalized linear model to 1999–2000 ratings data. I conclude that the number of standards met is most strongly influenced by the economic well-being of the student population, rather than by academic practices of the district.

1 Introduction

In 1997, the Ohio General Assembly passed Senate Bill 55, which was intended to make school districts accountable for the academic performance of their students. This law set forth 27 standards that school districts were expected to meet, and it required that districts be given annual ratings based on how many of the standards they actually met.

Most of the 27 standards are minimum passage rates for state proficiency tests, which are given in the fourth, sixth, ninth[1], and twelfth grades. Each test has five sections (citizenship, mathematics, reading, writing, and science), and so five standards are associated with each test. The remaining two standards relate to attendance and graduation rates. The exact criteria are these:

- Standards 1–5: 75% of students passing each portion of fourth-grade test
- Standards 6–10: 75% of students passing each portion of sixth-grade test
- Standards 11–15: 75% of students passing each portion of ninth-grade test by the end of ninth grade
- Standards 16–20: 85% of students passing each portion of ninth-grade test by the end of tenth grade
- Standards 21–25: 60% of students passing each portion of twelfth-grade test
- Standard 26: 93% attendance rate
- Standard 27: 90% graduation rate

The ratings are then assigned as follows:

[1] By 2003, the ninth grade test will be replaced by the "Ohio Graduation Test," a more rigorous test given in the tenth grade.

- Effective: meets 26 or more standards
- Continuous Improvement: meets 14 to 25 standards
- Academic Watch: meets 9 to 13 standards
- Academic Emergency: meets 8 or fewer standards

The ratings have several practical consequences, in addition to their effect on districts' reputations. First, any district not receiving the highest rating must submit an improvement plan to the state. Second, any district receiving one of the two lowest ratings must establish a citizen committee to suggest further ways to improve. Third, any district that does not improve sufficiently after two years must undergo a site evaluation, which can lead the state to appoint a committee to help direct the district's improvement efforts.

Given the practical significance of these ratings, it is important to know what factors influence the number of standards a district meets. If the influential factors are within a district's control, then knowing them will help administrators design the best policies. If the influential factors are outside a district's control, then the state's approach—which assumes that these ratings measure the quality of a district's academic and operational practices—will be called into question.

In the present study, I look for a relationship between the characteristics of a district and the number of standards it meets. I do this by analyzing and fitting a generalized linear model to a data set containing 1999–2000 information on 607 Ohio school districts.

2 Data

2.1 Overview

The data set has one observation for each of the 607 Ohio school districts, and it has columns for performance and district information. The performance information consists of two columns for each standard—one for the district's

percentage of disabled students
district stability rate
percentage of economically disadvantaged students
median income
pupil-teacher ratio
percentage of courses taught by appropriately certified teachers
percentage of core courses taught by appropriately certified teachers
staff attendance rate
local revenue per pupil
state revenue per pupil
federal revenue per pupil
administrative expenditures per pupil
building operation expenditures per pupil
staff support expenditures per pupil
pupil support expenditures per pupil
instructional expenditures per pupil
total expenditures per pupil

Table 1: Potential explanatory variables in the data set

actual numerical score (e.g., the actual percentage of students passing the reading section of the fourth grade test) and one for an indicator variable of whether the district met that standard—as well as columns recording the total number of standards met and the overall rating.

The district information includes characteristics of the school system, such as the pupil-teacher ratio, and characteristics of the student body, such as the percentage of economically disadvantaged students. Table 1 lists these potential explanatory variables.

There are no missing values among the performance variables but 8 among the explanatory variables. These data are missing because school districts did not report the information to the ODE; since there are so few missing values and since the missing values are evenly distributed among the four ratings, observations with missing values will be dropped from all subsequent analyses. The working data set thus has 599 observations.

Academic Emergency	11%
Academic Watch	22%
Continuous Improvement	62%
Effective	5%

Table 2: Distribution of overall ratings

Minimum	0
First Quartile	12
Median	16
Mean	16
Third Quartile	20
Maximum	27

Table 3: Summary of number of standards met

2.2 Ratings Data

Table 2 shows the distribution of overall ratings. Most districts fall into the "Continuous Improvement" category, but more are in "Academic Watch" or "Academic Emergency" than "Effective," which includes only 5 percent of the districts.

The number of standards met provides more detailed information about a district's overall performance. Table 3 summarizes this variable numerically, and Figure 1 summarizes it graphically. Its distribution is roughly bell-shaped, with a mode between 15 and 20 (which qualifies for "Continuous Improvement").

This measure of overall performance thus behaves in a fairly regular way (much like the stereotypical class curve); small proportions of districts receive very low and very high scores, while most districts fall somewhere in the middle.

An examination of the correlations among the 27 numerical sub-measures reveals that all are highly correlated (with most correlation coefficients above 0.60) and that the percentage of students passing a given portion of a given test tends to be most highly correlated with other passing percentages for

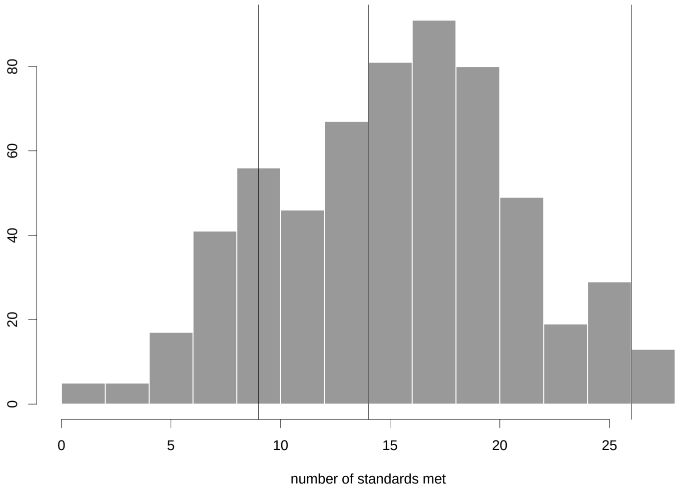

Figure 1: Histogram of number of standards met, with vertical lines marking the cutoffs for different ratings

the same test. For instance, the percentage of students passing the citizenship section of the fourth-grade test is most highly correlated (0.88) with the the percentage of students passing the science section of the same test. This suggests that a given passing percentage provides some information about all of the other passing percentages and much information about other passing percentages within the same test. All of the sub-measures are highly correlated with the total number of standards met, with most correlation coefficients close to 0.80.

2.3 Explanatory Variables

In an examination of the correlations among the numerical sub-measures and the potential explanatory variables, several variables stand out as having relatively high (greater than 0.40) correlations with most performance measures. The student disadvantaged rate, which measures economic deprivation, has a fairly high (around 0.60) negative correlation with the overall measure (total

Abbreviation	Full Name
disadv	percentage of economically disadvantaged students
medinc	median income
ratio	pupil-teacher ratio
cert	percentage of courses taught by approrpriately certified teachers
attend	staff attendance rate
locrev	local revenue per pupil
statrev	state revenue per pupil
expend	total expenditures per pupil

Table 4: Key to variable name abbreviations in the figures

number of standards met) and with most sub-measures. The median income has a similarly high positive correlation with the performance measures, and the state revenue per pupil has a moderate (around 0.45) negative correlation with the performance measures.

I have chosen several other substantively interesting variables to add to the analysis: the pupil-teacher ratio, the percentage of courses taught by appropriately certified teachers, the staff attendance rate, the local revenue per pupil, and the total expenditures per pupil. These choices are highly subjective, and an expert in the educational literature could probably make better ones.

Figures 2 and 3 summarize graphically the relationships between district performance and the eight explanatory variables mentioned above. The figures show, for each explanatory variable, a scatterplot between the variable and the number of standards met, and side-by-side boxplots of the variable split by overall rating. Figure 4 shows all possible scatterplots among the explanatory variables. Table 4 lists the abbreviated variable names used in these plots.

These figures show that the four variables describing academic practices—pupil-teacher ratio, percentage of courses taught by appropriately certified teachers, staff attendance rate, and total expenditures per pupil—have no visible relationship with district performance. The variables that have visible relationships with performance are the ones that describe the economic status of the district's students and residents.

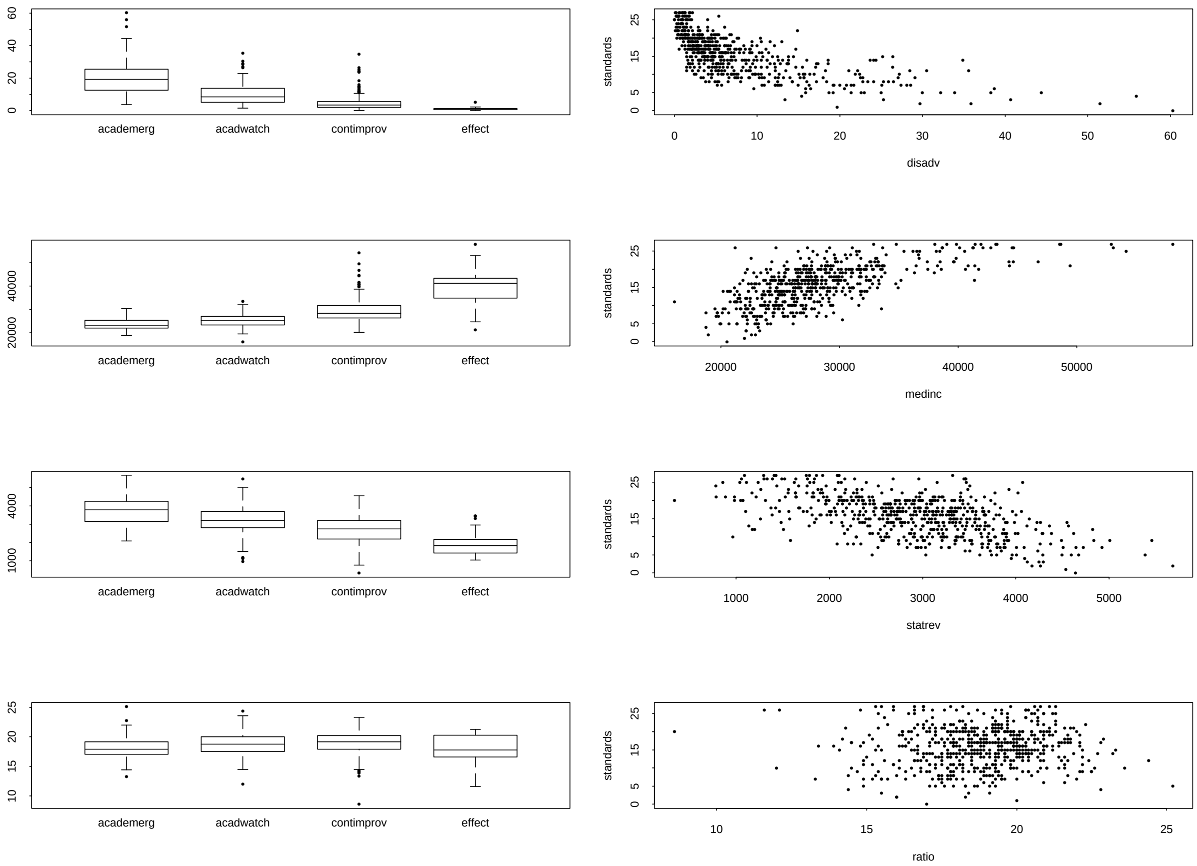

Figure 2: Graphical summaries of the relationships between the explanatory variables and school district performance

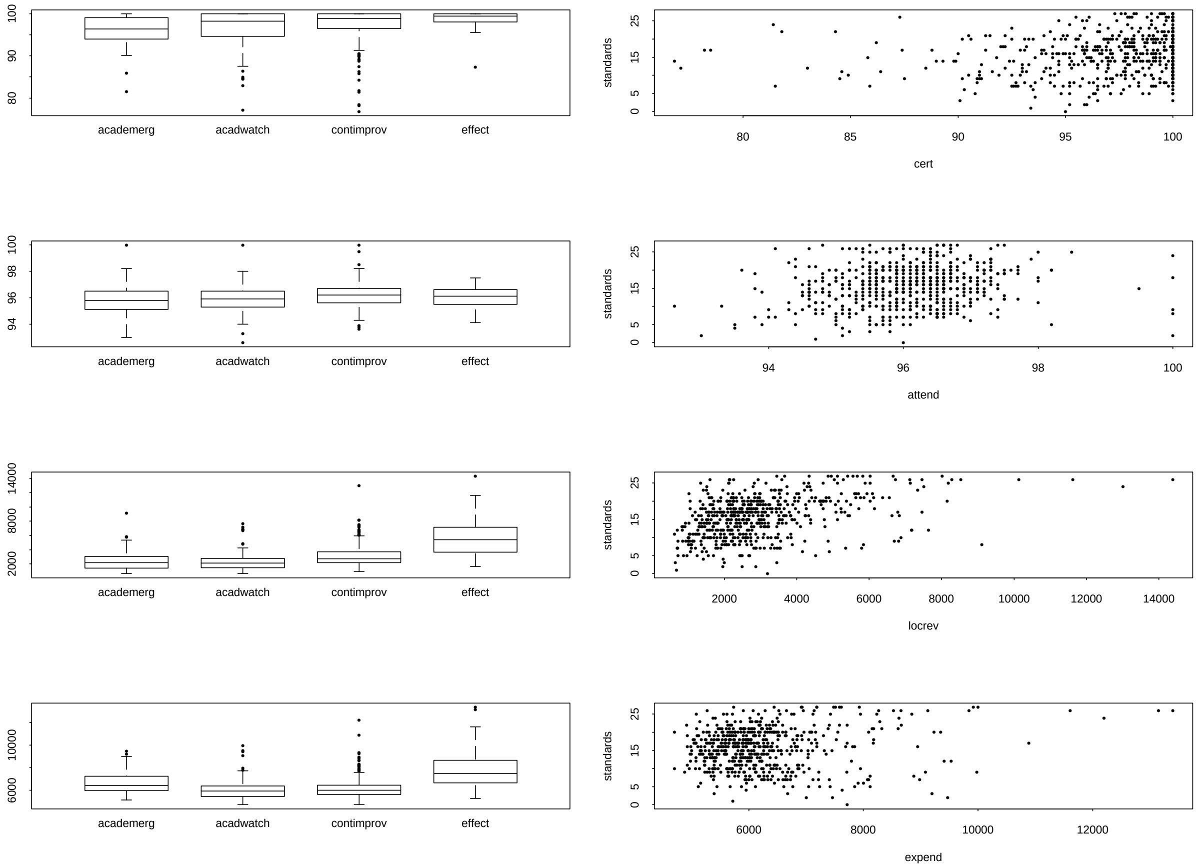

Figure 3: Graphical summaries of the relationships between the explanatory variables and school district performance

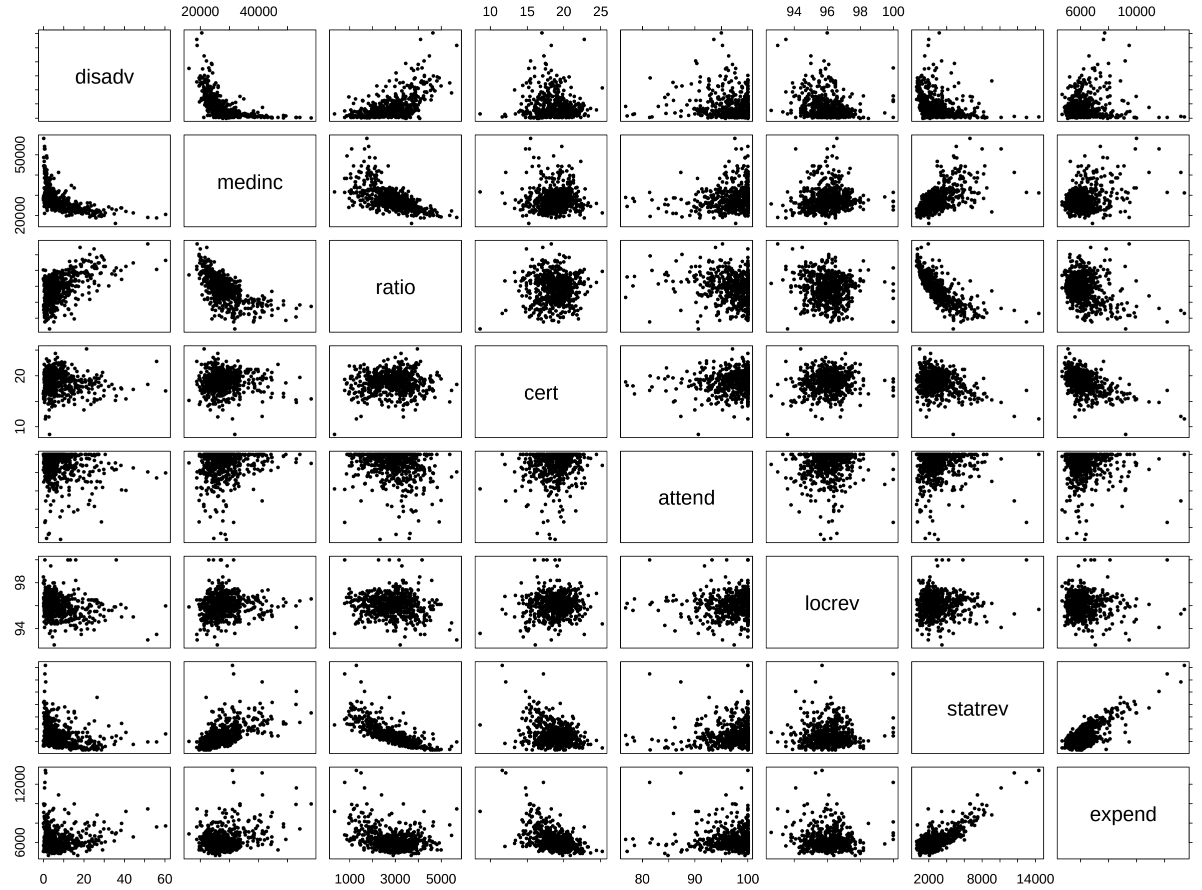

Figure 4: Scatterplots among the explanatory variables

Among the explanatory variables, local revenue per pupil is strongly associated with both state revenue per pupil and total expenditures per pupil. The relationship with state revenue is negative (with a correlation of -0.59), which makes sense because Ohio gives more aid to districts that receive less local revenue. The relationship with total expenditures per pupil is positive (with a correlation of 0.55), which also makes sense: The more income a district receives, the more it will spend.

The exploratory analysis leaves one puzzle. The most likely explanation for the strong negative relationship between performance and state revenue is local revenue: Since state aid is inversely tied to local revenue and since local revenue comes from property taxes, high state revenue signals low property values; therefore, state revenue acts as a proxy for the wealth of a district's residents. However, this explanation implies a strong positive relationship between performance and local revenue, and both the scatterplot (Figure 3) and the correlation (0.23) show that this relationship is quite weak.

3 Model

I took the number of standards met as my response variable, since it is a more detailed measure of overall performance than the rating. Since there are 27 standards altogether, this variable is a count between 0 and 27, and thus I modeled it as a binomial random variable with 27 trials and a probability p of success on each trial. I modeled this probability as a function of the explanatory variables, adopting the framework of a generalized linear model (McCullagh and Nelder, 1989). In this framework, I specified that p is some function of a linear combination of the variables; that is,

$$p = f(\textstyle\sum_{j=1}^{k} \beta_j x_j)$$

Now, p is bounded by 0 and 1, but $\sum \beta_j x_j$ can range all over the real line; therefore, f must map the real line to the unit interval. The two most common transformations are logit, defined by

$$f(\textstyle\sum_{j=1}^{k} \beta_j x_j) = \frac{\exp(\sum_{j=1}^{k} \beta_j x_j)}{1+\exp(\sum_{j=1}^{k} \beta_j x_j)}$$

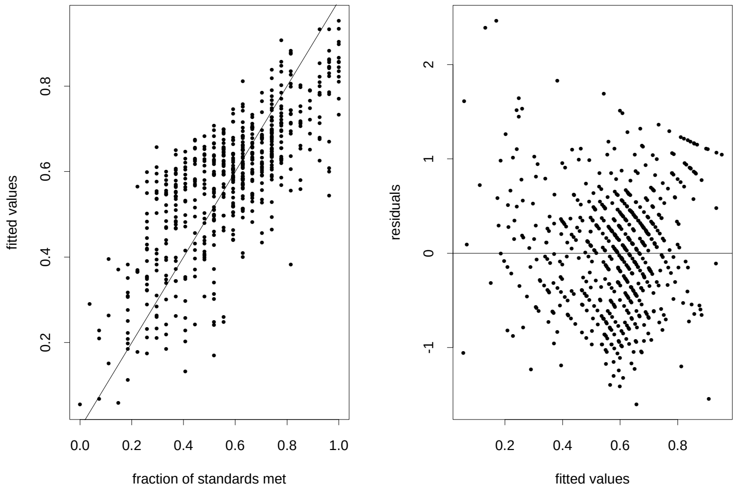

Figure 5: Diagnostic plots for the full logit model

and probit, defined by

$$f(\sum_{j=1}^{k}\beta_j x_j) = \Phi^{-1}(\sum_{j=1}^{k}\beta_j x_j)$$

where Φ is the standard normal cumulative distribution function. Because I had no theoretical reason to prefer one transformation, I tried fitting both logit and probit models.

I initially fit models with all eight of the explanatory variables. The fit was quite good in both cases, as Figures 5 and 6 show: The fitted values tracked the true values fairly closely, and the residuals looked randomly distributed around 0. Furthermore, the logit and probit models produced almost identical fitted values, as Figure 7 shows. Appendix A contains the full logit results, and Appendix B contains the full probit results.

I next fit models with a reduced set of explanatory variables. Since the exploratory data analysis showed no relationships between the response variable and pupil-teacher ratio, percentage of courses taught by appropriately certified teachers, or staff attendance rate, I dropped those variables. Furthermore, since the analysis showed strong collinearity between local revenue

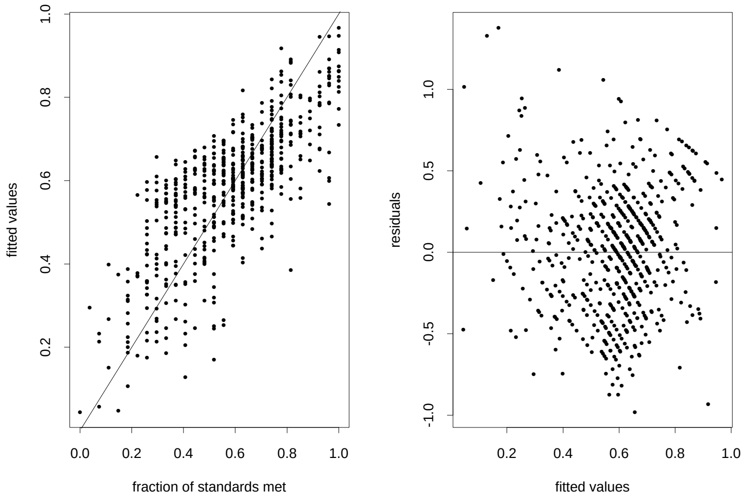

Figure 6: Diagnostic plots for the full probit model

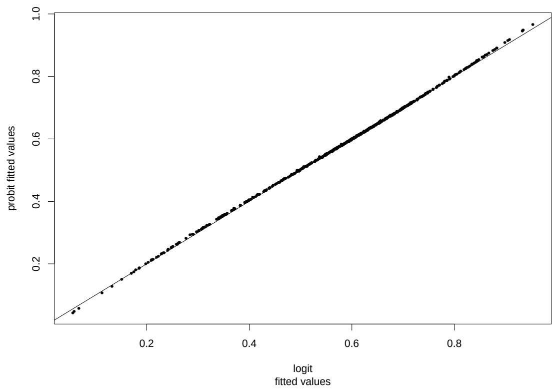

Figure 7: A plot of logit fitted values against probit fitted values for the full models

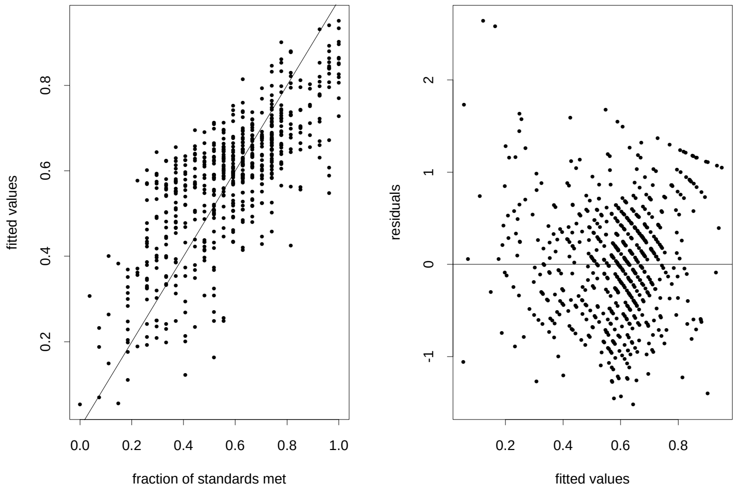

Figure 8: Diagnostic plots for the reduced logit model

per pupil and both state revenue per pupil and total expenditures per pupil, I dropped the latter two. Local revenue seems more substantively relevant than state revenue, and it is more strongly associated with performance than total expenditures. Therefore, the reduced model included only percentage of disadvantaged students, median income, and local revenue per pupil.

The reduced models fit about as well as the full ones; Figures 8 and 9 show that the fitted values are relatively close to the true values and that the residuals seem appropriately dispersed around 0. Again, the logit and probit models gave very similar results, as Figure 10 shows. Appendix C contains the full logit results, and Appendix D contains the full probit results.

Since the extra explanatory variables contributed nothing to the model fit, I chose the reduced models; and since the logit and probit models fit equally well, analytic simplicity led me to choose the logit model. Therefore, the final model can be summarized as follows, where i indexes observations and j indexes explanatory variables:

$$y_i \overset{\text{indep}}{\sim} B(27, p_i) \qquad i \in \{1, \ldots, 599\}$$

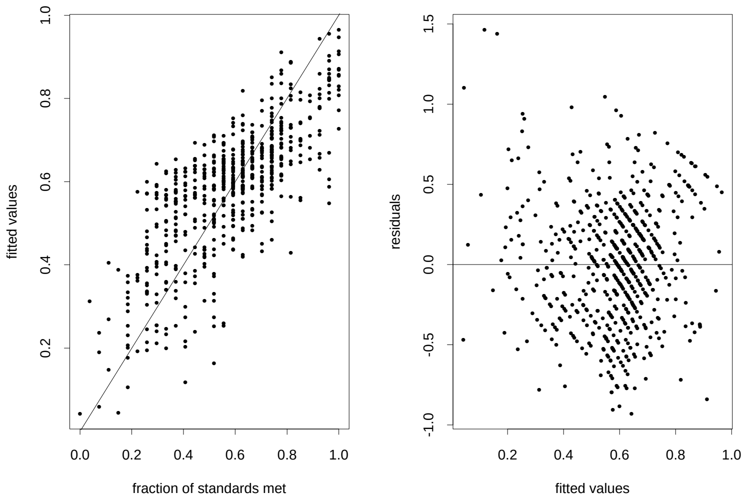

Figure 9: Diagnostic plots for the reduced probit model

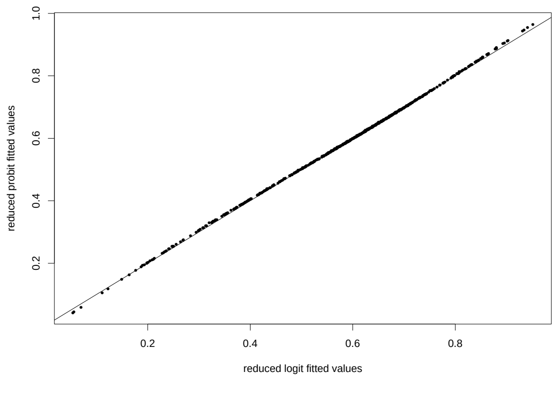

Figure 10: A plot of logit fitted values against probit fitted values for the reduced models

Starting From Average Values Baseline $p = 0.596$		
Variable	Initial Value	New p After Increase
percentage disadvantaged	7.5	0.584
median income	28,160	0.613
local revenue per pupil	3,002	0.608

Table 5: Effect on predicted p of increasing each explanatory variable, holding the others constant; percentage disadvantaged is increased by 1; median income and local revenue per pupil are increased by $1,000

$$p_i = \frac{\exp\left(\sum_{j=1}^{4} \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{4} \beta_j x_{ij}\right)}$$

$$\begin{aligned} y &= \text{number of standards met} \\ x_1 &= 1(\text{intercept}) \\ x_2 &= \text{percentage of economically disadvantaged students} \\ x_3 &= \text{median income} \\ x_4 &= \text{local revenue per pupil} \end{aligned}$$

4 Results

Tables 5, 6, and 7 help interpret the results in substantive terms. The effect of a change in an explanatory variable depends on the initial levels of the variables, and so Table 5 starts with the variables' average values, Table 6 starts with their minimum values, and Table 7 starts with their maximum values. The tables show the change in predicted p that results from an increase in each explanatory variable, holding the other variables constant.

For the average school district, increasing the percentage of economically disadvantaged students by one point would decrease the predicted probability of meeting a standard by about 0.01; increasing the median income by $1,000 would increase the predicted probability by about 0.02; and increasing the

Starting From Minimum Values Baseline $p = 0.449$		
Variable	Initial Value	New p After Increase
percentage disadvantaged	0.0	0.437
median income	16,080	0.467
local revenue per pupil	624	0.462

Table 6: Effect on predicted p of increasing each explanatory variable, holding the others constant; percentage disadvantaged is increased by 1; median income and local revenue per pupil are increased by \$1,000

Starting From Maximum Values Baseline $p = 0.601$		
Variable	Initial Value	New p After Increase
percentage disadvantaged	60.3	0.589
income median	58,100	0.618
local revenue per pupil	14,390	0.613

Table 7: Effect on predicted p of increasing each explanatory variable, holding the others constant; percentage disadvantaged is increased by 1; median income and local revenue per pupil are increased by \$1,000

local revenue per pupil by $1,000 would increase the predicted probability by about 0.01. Although these changes seem minor, going from the lowest observed values to the highest observed values boosts the predicted probability from 0.45 to 0.60, a substantial increase.

5 Discussion

5.1 Findings

In this study, I looked for the characteristics of a school district or of its students that would help predict how many state standards the district meets. Exploratory data analysis suggested that attainment of these standards depends mainly on the economic well-being of the students, rather than on academic or operational practices of the district. Formal model fitting supported this finding and quantified the impact of the most influential factors: the percentage of economically disadvantaged students, the median income, and the local revenue per pupil (which measures property values in the district). Each of these factors affects the predicted probability of meeting a standard by about 0.01.

This has two main implications for policy.

1. Improving students' performance—as measured by the state proficiency tests—may require policies that improve their families' economic well-being.

2. Proficiency test passage rates may not measure the quality of personnel and practices in a school district.

5.2 Study Limitations and Future Research

Response Variable. The data set in this study contains a single measure of a school district's performance: attainment of the state's 27 performance standards, which are based mainly on proficiency test passage rates. This is, however, an imperfect measure of the true object of interest: the quality of the district's personnel and practices. The results of the present study suggest that this variable is a particularly flawed measure. It would be worthwhile to study other measures of district quality, such as SAT scores or expert evaluations. One could even form a composite of several such measures.

Explanatory Variables. Although this data set contains a fair number of variables, it leaves many factors—such as teacher, principal, and superintendent quality; characteristics of the board of education; characteristics of the curriculum; location in a rural or urban area—unmeasured. One can imagine stories in which such unmeasured variables alter the relationships suggested by this study; for instance, the economic well-being of a district's residents may have some relation to the quality of teachers it can attract, and this may be producing the apparent relationship between district performance and median income. Making more definitive causal statements requires a richer set of explanatory variables.

Model. This study assumed that for a given district, p was the same for all standards; that is, the probability of meeting standard 1 was the same as the probability of meeting standard 27. However, different explanatory variables might influence the probability of meeting different standards, and so it would be interesting to fit a model in which the coefficients varied by standard as well as by district.

6 Acknowledgements

In writing this paper, I drew on technical material learned in Statistics 36-711, taught by Professor Brian Junker. I also benefited from private conversations with Professor Junker and from his review of an early draft.

References

[1] McCullagh, P. and Nelder, N.A. (1989) *Generalized Linear Models.* (2nd ed.) New York: Chapman and Hall.

[2] `http://www.ode.state.oh.us/pa` (web site of the Ohio Department of Education)

A Full Logit

```
Coefficients:
                  Value Std. Error     t value
(Intercept) -5.540642965 9.53567680 -0.58104349
       V412 -0.048319098 0.01700759 -2.84103061
      V414b  0.067963362 0.02756902  2.46520771
      V426b -0.084748836 0.24396134 -0.34738633
       V416  0.003137699 0.05414533  0.05794957
       V418  0.008088455 0.02406419  0.33611997
       V422  0.036058919 0.09558289  0.37725286
      V424b -0.003475609 0.18667352 -0.01861865
      V440b  0.045537600 0.22201737  0.20510828

    Null Deviance: 116.5709 on 598 degrees of freedom

Residual Deviance: 51.46384 on 590 degrees of freedom

Number of Fisher Scoring Iterations: 4

Correlation of Coefficients:
      (Intercept)       V412      V414b      V426b       V416       V418
 V412 -0.1714153
V414b -0.1043068   0.4626409
V426b -0.0263787  -0.1746801  0.0987179
 V416 -0.0448455  -0.1125805 -0.1705724  0.0165003
 V418 -0.2231271   0.0622783  0.0090566  0.1107231 -0.0771397
 V422 -0.9516483   0.1678987  0.0461814 -0.0498862 -0.0779878 -0.0236547
V424b  0.0666724  -0.0038360 -0.1208105  0.7919928 -0.0398690  0.0579278
V440b -0.1054052  -0.2088384 -0.0726467 -0.5510281  0.3339864 -0.0300878

           V422      V424b
 V412
V414b
V426b
 V416
 V418
 V422
V424b -0.0656812
```

```
V440b  0.0303150 -0.8267383
sink()
```

B Full Probit

```
Coefficients:
                   Value Std. Error     t value
(Intercept) -3.518142469 5.83858349 -0.60256781
       V412 -0.029183201 0.01005114 -2.90347176
      V414b  0.041227228 0.01605766  2.56744969
      V426b -0.052525888 0.14782960 -0.35531375
       V416  0.002719756 0.03312557  0.08210441
       V418  0.004856665 0.01476867  0.32884909
       V422  0.023346514 0.05841747  0.39964953
      V424b -0.001677470 0.11242146 -0.01492126
      V440b  0.029627188 0.13426336  0.22066473

    Null Deviance: 116.5709 on 598 degrees of freedom

Residual Deviance: 51.29184 on 590 degrees of freedom

Number of Fisher Scoring Iterations: 4

Correlation of Coefficients:
      (Intercept)       V412      V414b      V426b       V416       V418
 V412 -0.1701732
V414b -0.1032543   0.4534606
V426b -0.0268773  -0.1793675  0.1114996
 V416 -0.0496307  -0.1146066 -0.1744689  0.0138337
 V418 -0.2269715   0.0643497  0.0019545  0.1108221 -0.0732471
 V422 -0.9519370   0.1703082  0.0517773 -0.0491659 -0.0739228 -0.0205040
V424b  0.0680366   0.0070338 -0.1089277  0.7892189 -0.0445312  0.0589510
V440b -0.1091253  -0.2206869 -0.0876824 -0.5512043  0.3377997 -0.0275374

            V422      V424b
 V412
V414b
```

```
V426b
 V416
 V418
 V422
V424b -0.0678407
V440b  0.0349327 -0.8297325
```

C Reduced Logit

```
Coefficients:
                  Value Std. Error    t value
(Intercept) -1.35548226 0.73947331 -1.8330374
       V412 -0.05108766 0.01519234 -3.3627242
      V414b  0.06907416 0.02698395  2.5598239
      V424b  0.04824947 0.06781341  0.7115033

    Null Deviance: 116.5709 on 598 degrees of freedom

Residual Deviance: 51.87959 on 595 degrees of freedom

Number of Fisher Scoring Iterations: 4

Correlation of Coefficients:
      (Intercept)       V412      V414b
 V412 -0.6649558
V414b -0.9527110   0.5467443
V424b  0.2725753  -0.1304965 -0.5130341
```

D Reduced Probit

```
Coefficients:
                  Value  Std. Error    t value
(Intercept) -0.82276706 0.430561320 -1.9109173
       V412 -0.03086123 0.008875921 -3.4769608
      V414b  0.04197227 0.015676806  2.6773484
```

```
      V424b  0.03054910 0.040612549  0.7522084

    Null Deviance: 116.5709 on 598 degrees of freedom

Residual Deviance: 51.73061 on 595 degrees of freedom

Number of Fisher Scoring Iterations: 4

Correlation of Coefficients:
      (Intercept)       V412      V414b
 V412 -0.6652008
V414b -0.9489666   0.5425075
V424b  0.2717147  -0.1234818 -0.5220465
```