# Test Results and Income

Presented By: Matt Eding

# Introduction - School Funding


CA Dept. of Education

Are test results indicative of local income?

Update funding formulas to bridge achievement gap.

“Property tax revenues … often reinforces inequity …. State funding … is rarely successful at overcoming these differences.” (Gartner, Jess. *How are Public Schools Funded?*, 2017)

# Strategy

Collect data from California Assessment of Student Performance & Progress.

Use zip codes for median household income.

Feature engineer data for income prediction.

Leverage results to inform education budgeting.

# Methodology - Data Sources

## Test Score Data:

CA school data from 2018 Smarter Balanced testing

caaspp

California Assessment of Student Performance and Progress

## Income Data:

Median household income by zip code

United States Zip Codes.org

# Methodology - Process

**Automate data retrieval**

**Cross Validation (80% train) with four models:**

- Ordinary Least Squares
- Ridge
- LASSO
- Polynomial

**Feature engineer maximal model efficacy**

## Tools Used

- Python
- Pandas
- Matplotlib
- Seaborn
- Selenium
- Patsy
- SciKit-Learn
- Statsmodels

# Findings

Chose OLS over other models because it yielded only marginally worse results.

Predictions are 31.7% off from the true median income on average.

Almost all features are statistically significant.

29.6% of the variance of true median income is accounted for.

# Conclusion

While my findings are more robust than naive approaches, they should not constitute much towards overhaul of the funding schematics.

The variability may accidentally penalize some schools that truly need the extra funding and vice versa.

# Future Work

- Test and income data from prior years
- Investigate prediction of home value
- Data from other states
- More feature engineering

# Thank You!

https://github.com/MattEding/ProjectLuther

# Appendix - Test Summary Pt. 1

| | | | |
|---|---|---|---|
| **Dep. Variable:** | y | **R-squared:** | 0.296 |
| **Model:** | OLS | **Adj. R-squared:** | 0.296 |
| **Method:** | Least Squares | **F-statistic:** | 598.0 |
| **Date:** | Fri, 25 Jan 2019 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 08:44:48 | **Log-Likelihood:** | -3066.1 |
| **No. Observations:** | 11386 | **AIC:** | 6150. |
| **Df Residuals:** | 11377 | **BIC:** | 6216. |
| **Df Model:** | 8 | | |
| **Covariance Type:** | nonrobust | | |

**y:** np.log(Median_Income)

**RMSE Score:** 0.31713498883040286

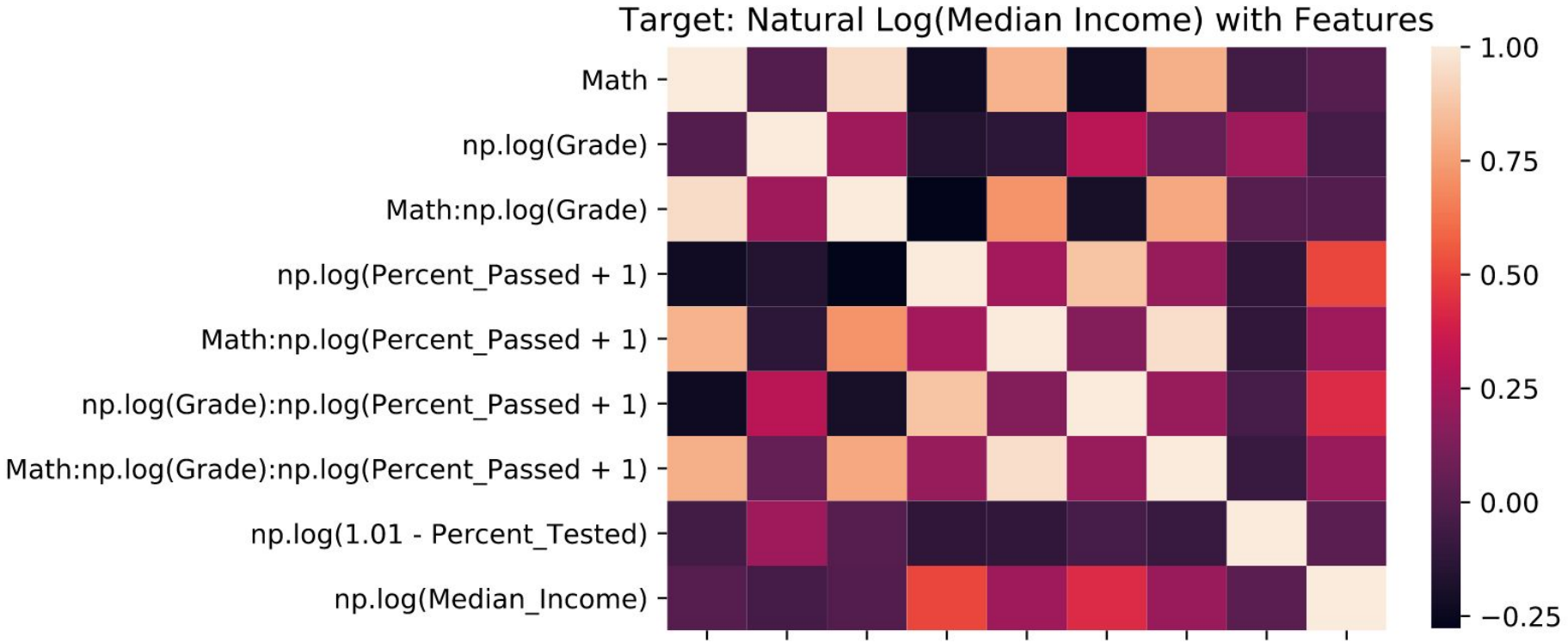
| | | | |
|---|---|---|---|
| **Omnibus:** | 84.035 | **Durbin-Watson:** | 1.994 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 92.251 |
| **Skew:** | -0.173 | **Prob(JB):** | 9.29e-21 |
| **Kurtosis:** | 3.274 | **Cond. No.** | 367. |

# Appendix - Test Summary Pt. 2

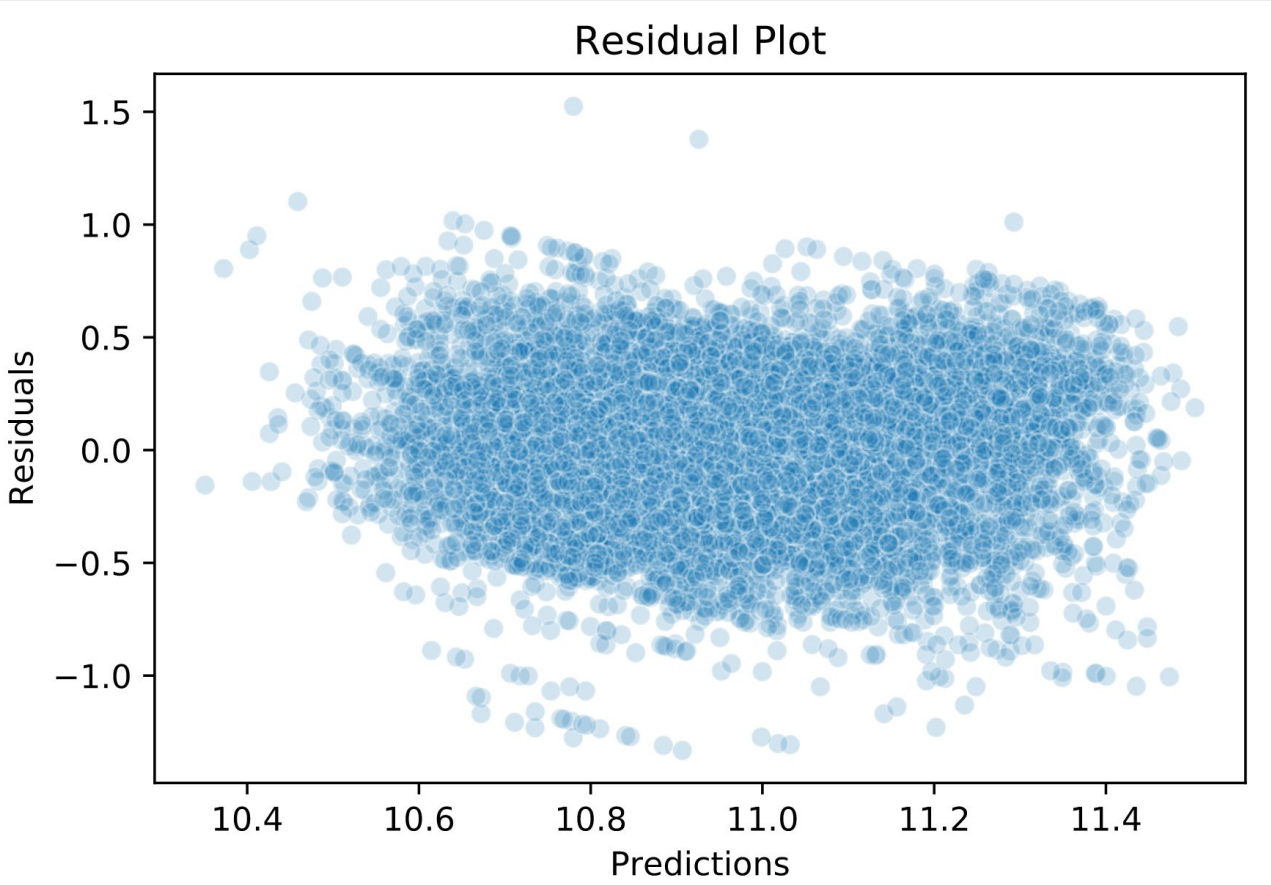
| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 10.2510 | 0.055 | 187.108 | 0.000 | 10.144 | 10.358 |
| **x1** | 0.1460 | 0.066 | 2.228 | 0.026 | 0.018 | 0.274 |
| **x2** | 0.1848 | 0.030 | 6.160 | 0.000 | 0.126 | 0.244 |
| **x3** | -0.0073 | 0.038 | -0.193 | 0.847 | -0.081 | 0.067 |
| **x4** | 2.5263 | 0.127 | 19.915 | 0.000 | 2.278 | 2.775 |
| **x5** | -0.7331 | 0.171 | -4.280 | 0.000 | -1.069 | -0.397 |
| **x6** | -0.6639 | 0.074 | -8.946 | 0.000 | -0.809 | -0.518 |
| **x7** | 0.3942 | 0.101 | 3.921 | 0.000 | 0.197 | 0.591 |
| **x8** | 0.0370 | 0.004 | 9.592 | 0.000 | 0.029 | 0.045 |

**X1:** Math
**X2:** np.log(Grade)
**X3:** Math:np.log(Grade)
**X4:** np.log(Percent_Passed + 1)
**X5:** Math:np.log(Percent_Passed + 1)
**X6:** np.log(Grade):np.log(Percent_Passed + 1)
**X7:** Math:np.log(Grade):np.log(Percent_Passed + 1)
**X8:** np.log(1.01 - Percent_Tested)

# Appendix - Heatmap



# Appendix - Residuals



# Appendix - CV Means / 5 Folds / 80% Train

linreg_mse_mean,
0.10067587790392145,

lasso_mse_optimal,
0.10093148393675706,

ridge_mse_optimal
0.100675691532643O2

linreg_adj_r2_mean,
0.2869029336781394,

lasso_adj_r2_optimal,
0.2851035847982211,

ridge_adj_r2_optimal
0.28690427102712207

# Appendix - CV polynomial degrees

In [151]:
poly_mse_mean

Out[151]:
{0: 0.14139573391162893,
 1: 0.10067587790392145,
 2: 0.09742838086516811,
 3: 0.09673969713674237}

In [152]:
poly_adj_r2_mean

Out[152]:
{0: -0.0005043993899957577,
 1: 0.2868245576350533,
 2: 0.307081431468514,
 3: 0.3027521797426715 5}

# Appendix: Naive Simple Linear Regression

## Median Income ~ Percent Met & Above

| | | | |
|---|---|---|---|
| **Dep. Variable:** | y | **R-squared:** | 0.267 |
| **Model:** | OLS | **Adj. R-squared:** | 0.267 |
| **Method:** | Least Squares | **F-statistic:** | 2.078e+04 |
| **Date:** | Fri, 25 Jan 2019 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 06:38:17 | **Log-Likelihood:** | -16281. |
| **No. Observations:** | 56928 | **AIC:** | 3.257e+04 |
| **Df Residuals:** | 56926 | **BIC:** | 3.258e+04 |
| **Df Model:** | 1 | | |
| **Covariance Type:** | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 10.5833 | 0.003 | 3584.754 | 0.000 | 10.577 | 10.589 |
| **x1** | 0.8856 | 0.006 | 144.169 | 0.000 | 0.874 | 0.898 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 319.540 | **Durbin-Watson:** | 0.313 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 355.348 |
| **Skew:** | -0.143 | **Prob(JB):** | 6.87e-78 |
| **Kurtosis:** | 3.260 | **Cond. No.** | 5.42 |

## Median Income ~ Median Scale Score

| | | | |
|---|---|---|---|
| **Dep. Variable:** | y | **R-squared:** | 0.142 |
| **Model:** | OLS | **Adj. R-squared:** | 0.142 |
| **Method:** | Least Squares | **F-statistic:** | 9449. |
| **Date:** | Fri, 25 Jan 2019 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 08:37:45 | **Log-Likelihood:** | -20769. |
| **No. Observations:** | 56928 | **AIC:** | 4.154e+04 |
| **Df Residuals:** | 56926 | **BIC:** | 4.156e+04 |
| **Df Model:** | 1 | | |
| **Covariance Type:** | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 5.5859 | 0.055 | 100.965 | 0.000 | 5.477 | 5.694 |
| **x1** | 0.0022 | 2.22e-05 | 97.204 | 0.000 | 0.002 | 0.002 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 28.927 | **Durbin-Watson:** | 0.203 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 29.264 |
| **Skew:** | -0.048 | **Prob(JB):** | 4.42e-07 |
| **Kurtosis:** | 3.056 | **Cond. No.** | 9.43e+04 |

# Appendix: OLS without Feature Engineering Pt. 1

| | | | |
|---|---|---|---|
| **Dep. Variable:** | Median_Income | **R-squared:** | 0.319 |
| **Model:** | OLS | **Adj. R-squared:** | 0.319 |
| **Method:** | Least Squares | **F-statistic:** | 2669. |
| **Date:** | Fri, 25 Jan 2019 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 09:10:27 | **Log-Likelihood:** | -6.4510e+05 |
| **No. Observations:** | 56928 | **AIC:** | 1.290e+06 |
| **Df Residuals:** | 56917 | **BIC:** | 1.290e+06 |
| **Df Model:** | 10 | | |
| **Covariance Type:** | nonrobust | | |

| | | | |
|---|---|---|---|
| **Omnibus:** | 8407.110 | **Durbin-Watson:** | 0.301 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 19727.684 |
| **Skew:** | 0.858 | **Prob(JB):** | 0.00 |
| **Kurtosis:** | 5.319 | **Cond. No.** | 1.09e+06 |

# Appendix: OLS without Feature Engineering Pt. 2

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **Test_Year** | -78.1326 | 4.809 | -16.246 | 0.000 | -87.559 | -68.706 |
| **Total_Tested_At_Entity_Level** | 201.6205 | 43.434 | 4.642 | 0.000 | 116.490 | 286.751 |
| **Total_Tested_with_Scores** | -202.8167 | 43.493 | -4.663 | 0.000 | -288.063 | -117.570 |
| **Grade** | -815.4109 | 94.785 | -8.603 | 0.000 | -1001.190 | -629.632 |
| **Test_Id** | 5379.5464 | 182.553 | 29.468 | 0.000 | 5021.742 | 5737.351 |
| **CAASPP_Reported_Enrollment** | 240.0734 | 12.930 | 18.567 | 0.000 | 214.730 | 265.416 |
| **Students_Tested** | -228.7238 | 13.214 | -17.310 | 0.000 | -254.622 | -202.825 |
| **Mean_Scale_Score** | 40.0326 | 3.893 | 10.284 | 0.000 | 32.403 | 47.662 |
| **Percentage_Standard_Met_and_Above** | 5.337e+04 | 978.592 | 54.534 | 0.000 | 5.14e+04 | 5.53e+04 |
| **Type_Id** | -2307.5353 | 107.803 | -21.405 | 0.000 | -2518.829 | -2096.242 |
| **Zip** | 1.1686 | 0.048 | 24.314 | 0.000 | 1.074 | 1.263 |