IE5202 Applied Forecasting Techniques
Chong Woon Kiat
A0209349X
Project 1: US Presidential Primary Election

1. Data Exploration

Raw data file consists of 52 independent variables, X_i, preliminary checking is done to select predictors to be used for modelling:

(i) Correlation of dependent variable, Y and all independent variable, X is studied to select variables with correlation to the model.
(ii) Correlation of all independent variables, X are studied to avoid model with high collinearity.

From Figure 1(a), 1(b) and Figure 2, thirteen variables that have correlations with ‘Hilary Percent’ and are not strongly correlated with each other are chosen to be the pool of predictors.

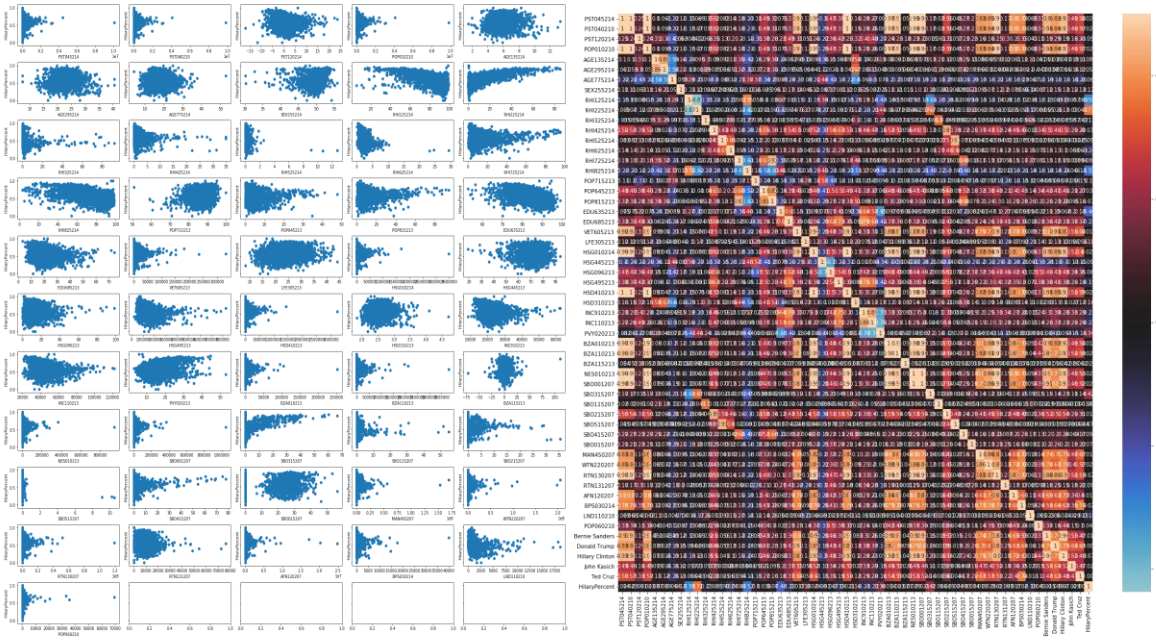


Figure 1: (a) Relationship between Hilary Percent and 52 independent variables, (b) correlation plot of 52 independent variables.

Table 1: Predictors list

Predictors	Description
C(state_abbreviation)	State
PST120214	Population, percent change - April 1, 2010 to July 1, 2014
SEX255214	Female persons, percent, 2014
RHI125214	White alone, percent, 2014
RHI225214	Black or African American alone, percent, 2014
POP815213	Language other than English spoken at home, pct age 5+, 2009-2013
EDU635213	High school graduate or higher, percent of persons age 25+, 2009-2013
EDU685213	Bachelor’s degree or higher, percent of persons age 25+, 2009-2013
HSG495213	Median value of owner-occupied housing units, 2009-2013
INC910213	Per capita money income in past 12 months (2013 dollars), 2009-2013
INC110213	Median household income, 2009-2013
PVY020213	Persons below poverty level, percent, 2009-2013
SBO315207	Black-owned firms, percent, 2007

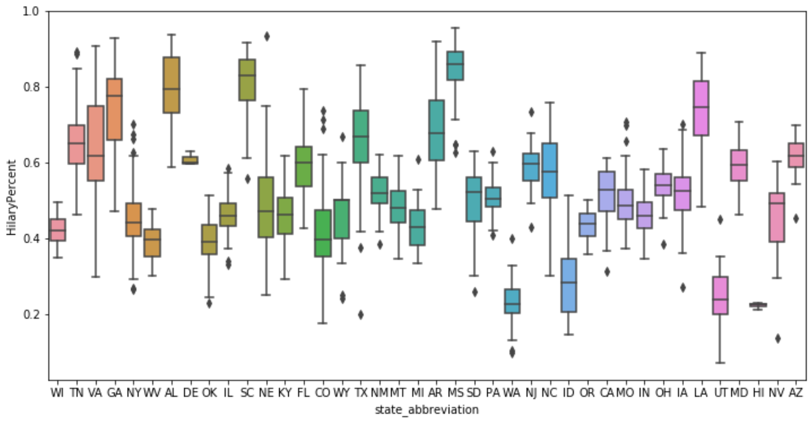


Figure 2: Boxplot of Hilary Percent by state

2. Simple Linear Regression

2.1 Model Description

Without any interaction term, all combination of the 13 variables, with a maximum of five, are tested to obtain a linear regression of equation $Y = \beta_1X_1 + \beta_2X_2 + \beta_2X_2 + \beta_2X_2 + \beta_2X_2$.

Metrics like R^2, adjusted R^2, AIC, BIC and Mallow's C_p of all 2379 combinations are evaluated to choose the best model.

All metrics suggest the same result where five independent variables in Table 2 make the best model. Transformation is done on the variables after evaluating the residual plot.

Table 2: Best Model's Coefficients

Predictors	Description	Transformation
C(state_abbreviation)	State	-
RHI225214	Black or African American alone, percent, 2014	SQRT(RHI225214)
POP815213	Language other than English spoken at home, pct age 5+, 2009-2013	-
EDU685213	Bachelor's degree or higher, percent of persons age 25+, 2009-2013	-
INC910213	Per capita money income in past 12 months (2013 dollars), 2009-2013	LOG(INC910213)

Table 3: Best Model's OLS Regression Results Post Transformation

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           HilaryPercent   R-squared:                       0.829
Model:                             OLS   Adj. R-squared:                  0.826
Method:                  Least Squares   F-statistic:                     276.5
Date:                 Sat, 28 Sep 2019   Prob (F-statistic):               0.00
Time:                         18:19:59   Log-Likelihood:                 3254.5
No. Observations:                 2488   AIC:                            -6421.
Df Residuals:                     2444   BIC:                            -6165.
Df Model:                           43
Covariance Type:             nonrobust
===============================================================================================
                                  coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
Intercept                      -0.3470      0.107     -3.237      0.001      -0.557      -0.137
C(state_abbreviation)[T.AR]    -0.0577      0.012     -4.829      0.000      -0.081      -0.034
C(state_abbreviation)[T.AZ]    -0.0870      0.022     -4.008      0.000      -0.130      -0.044
C(state_abbreviation)[T.CA]    -0.1740      0.014    -12.560      0.000      -0.201      -0.147
C(state_abbreviation)[T.CO]    -0.2191      0.013    -16.728      0.000      -0.245      -0.193
C(state_abbreviation)[T.DE]    -0.1769      0.039     -4.515      0.000      -0.254      -0.100
C(state_abbreviation)[T.FL]    -0.1701      0.013    -13.377      0.000      -0.195      -0.145
C(state_abbreviation)[T.GA]    -0.0587      0.010     -5.701      0.000      -0.079      -0.038
```

```
C(state_abbreviation)[T.HI]    -0.4293      0.039    -10.890      0.000       -0.507      -0.352
C(state_abbreviation)[T.IA]    -0.1333      0.012    -11.158      0.000       -0.157      -0.110
C(state_abbreviation)[T.ID]    -0.3407      0.014    -23.844      0.000       -0.369      -0.313
C(state_abbreviation)[T.IL]    -0.2217      0.012    -19.010      0.000       -0.245      -0.199
C(state_abbreviation)[T.IN]    -0.2081      0.012    -17.543      0.000       -0.231      -0.185
C(state_abbreviation)[T.KY]    -0.2130      0.011    -18.965      0.000       -0.235      -0.191
C(state_abbreviation)[T.LA]    -0.0986      0.013     -7.702      0.000       -0.124      -0.073
C(state_abbreviation)[T.MD]    -0.1626      0.018     -9.283      0.000       -0.197      -0.128
C(state_abbreviation)[T.MI]    -0.2343      0.012    -19.162      0.000       -0.258      -0.210
C(state_abbreviation)[T.MO]    -0.1771      0.011    -15.592      0.000       -0.199      -0.155
C(state_abbreviation)[T.MS]     0.0018      0.012      0.155      0.877       -0.021       0.025
C(state_abbreviation)[T.MT]    -0.1395      0.014    -10.334      0.000       -0.166      -0.113
C(state_abbreviation)[T.NC]    -0.1846      0.011    -16.441      0.000       -0.207      -0.163
C(state_abbreviation)[T.NE]    -0.1563      0.012    -12.740      0.000       -0.180      -0.132
C(state_abbreviation)[T.NJ]    -0.1423      0.019     -7.641      0.000       -0.179      -0.106
C(state_abbreviation)[T.NM]    -0.1709      0.016    -10.717      0.000       -0.202      -0.140
C(state_abbreviation)[T.NV]    -0.2546      0.020    -13.016      0.000       -0.293      -0.216
C(state_abbreviation)[T.NY]    -0.2417      0.013    -18.933      0.000       -0.267      -0.217
C(state_abbreviation)[T.OH]    -0.1426      0.012    -12.120      0.000       -0.166      -0.120
C(state_abbreviation)[T.OK]    -0.2870      0.012    -23.621      0.000       -0.311      -0.263
C(state_abbreviation)[T.OR]    -0.2029      0.015    -13.448      0.000       -0.232      -0.173
C(state_abbreviation)[T.PA]    -0.1746      0.013    -13.843      0.000       -0.199      -0.150
C(state_abbreviation)[T.SC]    -0.0366      0.014     -2.648      0.008       -0.064      -0.010
C(state_abbreviation)[T.SD]    -0.1276      0.013     -9.930      0.000       -0.153      -0.102
C(state_abbreviation)[T.TN]    -0.0444      0.012     -3.814      0.000       -0.067      -0.022
C(state_abbreviation)[T.TX]    -0.0734      0.011     -6.867      0.000       -0.094      -0.052
C(state_abbreviation)[T.UT]    -0.3736      0.017    -22.027      0.000       -0.407      -0.340
C(state_abbreviation)[T.VA]    -0.1111      0.011    -10.392      0.000       -0.132      -0.090
C(state_abbreviation)[T.WA]    -0.4228      0.015    -28.564      0.000       -0.452      -0.394
C(state_abbreviation)[T.WI]    -0.2278      0.013    -18.064      0.000       -0.253      -0.203
C(state_abbreviation)[T.WV]    -0.2763      0.014    -20.391      0.000       -0.303      -0.250
C(state_abbreviation)[T.WY]    -0.1893      0.017    -11.045      0.000       -0.223      -0.156
SQRT_RHI225214                  0.0397      0.001     36.625      0.000        0.038       0.042
POP815213                       0.0015      0.000      9.414      0.000        0.001       0.002
EDU685213                      -0.0045      0.000    -17.516      0.000       -0.005      -0.004
LOG_INC910213                   0.1027      0.011      9.384      0.000        0.081       0.124
==============================================================================
Omnibus:                      145.085   Durbin-Watson:                   1.961
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              569.826
Skew:                          -0.110   Prob(JB):                    1.84e-124
Kurtosis:                       5.334   Cond. No.                     2.08e+03
==============================================================================
```

2.2 Result

Regression model post-transformation gave an adjusted R^2 value 0.829, indicating that the model explains 83% of the variability. Large F-statistic value and low p-value also indicate that there is a strong evidence that at 95% significant level to reject the null hypothesis that none of the predictors are significant. Most of the coefficients for predictors also have low p-value close to 0, indicating that the predictor coefficients are non-zero at 95% significant level.

The model says that percentage of vote Hilary received is positively correlated to percentage of African American, language other than English spoken at home and per capita income, but negatively correlated to percentage of persons with Bachelor's degree or above.

Predictor RHI225214, percentage of African American is showing non-uniform residual when plotted against the predictor value (top right plot Table 4). It is rectified by applying square root transformation. Log transformation is also applied to INC910213, per capita income. Partial residual plots (bottom left plot in Table 4) for all the four predictors have linear correlation with the target variable, showing that they are significant and are telling additional information of Y as a predictor.

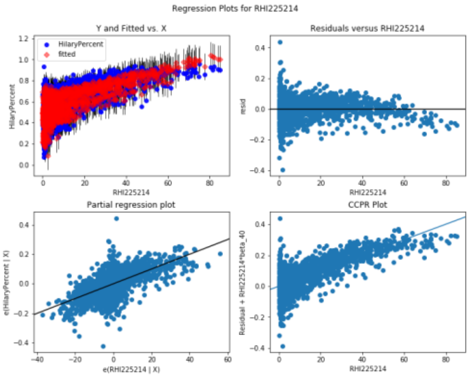
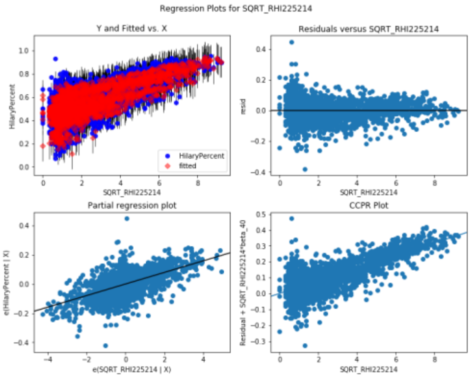
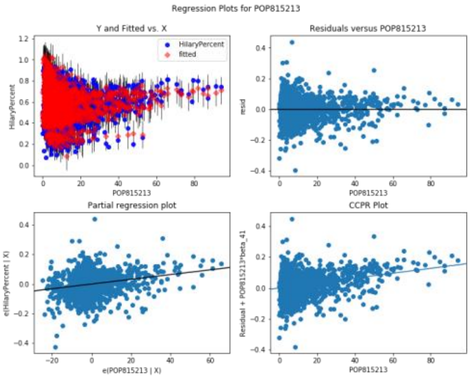
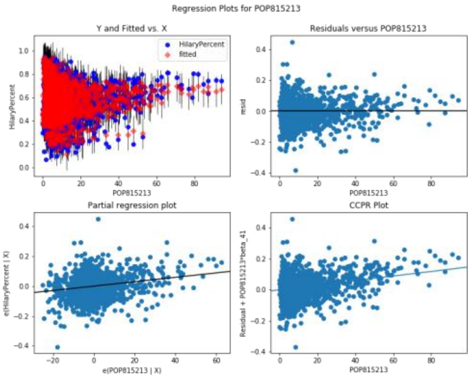
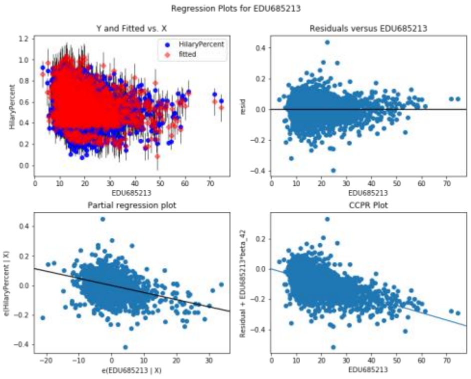
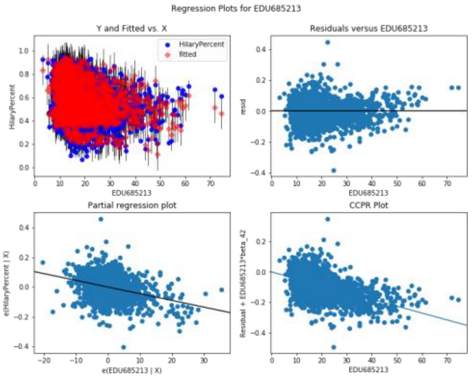
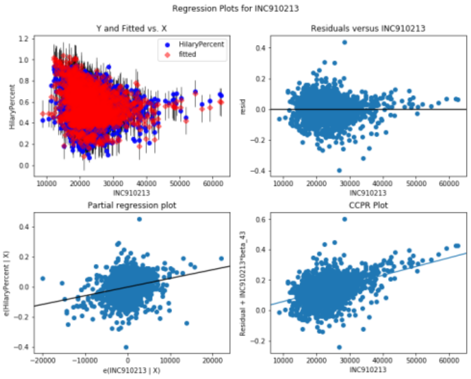
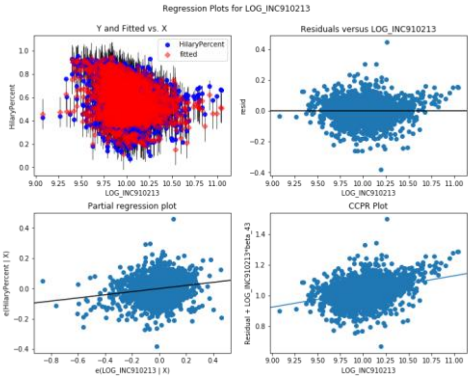
The QQ-plot in Figure 4(a) shows the residual is slightly deviated from normal distribution, with a slightly heavier tails. This indicates that there is some extreme residual unexplainable by the model. Figure 4(b) also shows that residual's variance is constant across fitted values and residual has a zero mean, which suggests a good model.

VIF in Table 5 shows all predictors all low VIF < 5, indicating minimal collinearity between predictors. Even though there slight correlation between education and per capita income (VIF ≈ 2), which is acceptable for the model.

Figure 5 shows that Cook's distance which depends on the size of residual and leverage, is <0.5 for all observations. This indicates that there is no highly influential observation in the model.

20-fold cross validation in Figure 6 shows that MSE of all runs are consistently at around 0.0045 and adjusted R^2 of about 0.827. This means that the model is stable.

Table 4: Regression Results against Predictors (Pre and Post Transformation)

Pre Transformation	Post Transformation
RHI225214  	SQRT(RHI225214) 
POP815213  	POP815213 
EDU685213  	EDU685213 
INC910213  	LOG(INC910213) 

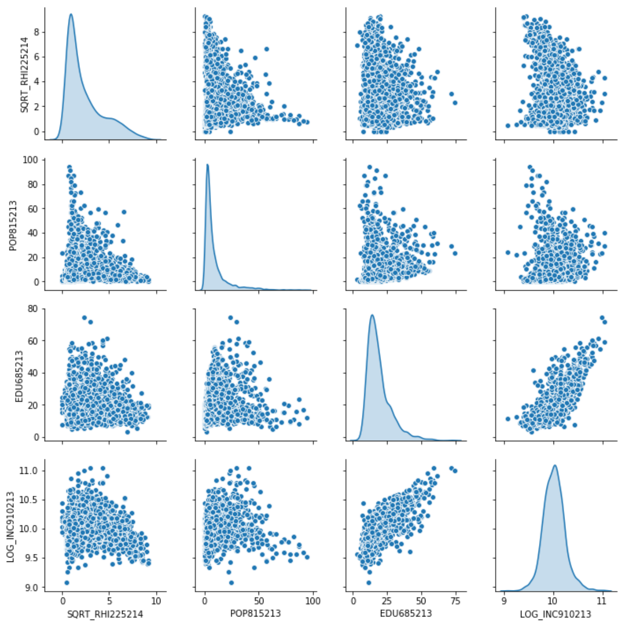


Table 5:

Variable	VIF
C(state_abbreviation)	6562.56
LOG_INC910213	2.27
EDU685213	2.16
RHI225214	2.14
POP815213	1.29

Figure 3: Correlation plot of predictors

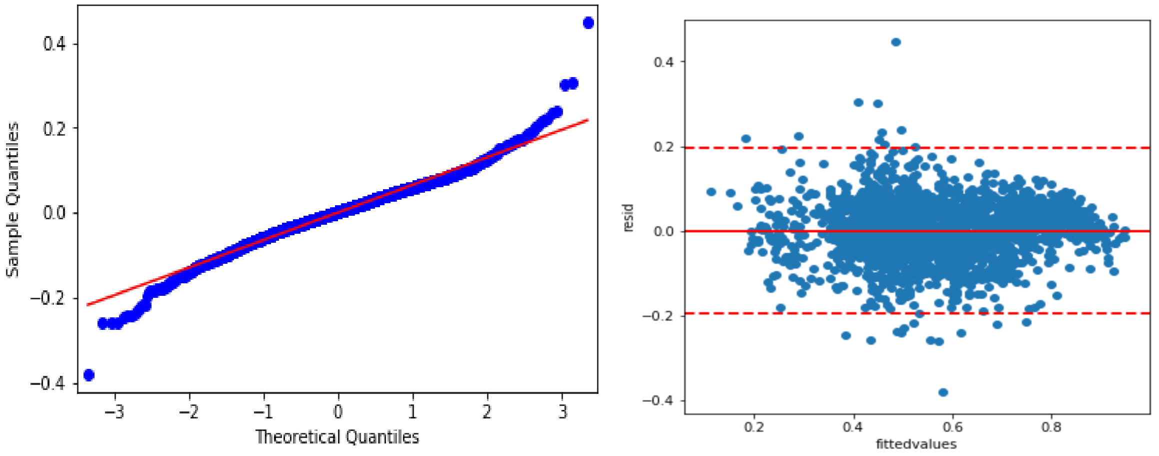


Figure 4: (a) QQ-plot, (b) Residual versus fitted values.

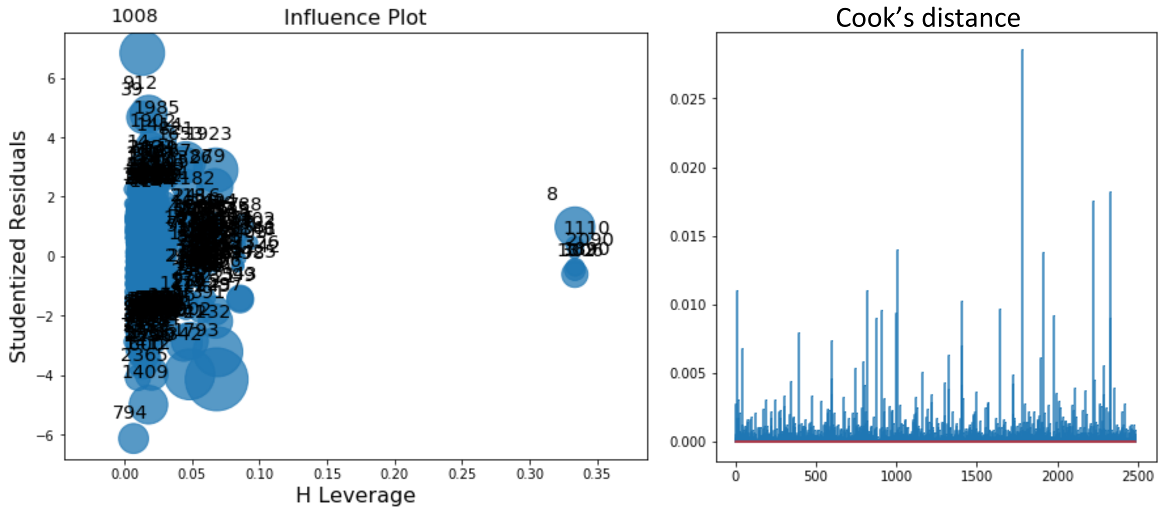


Figure 5: (a) Influence plot, (b) Cook's distance

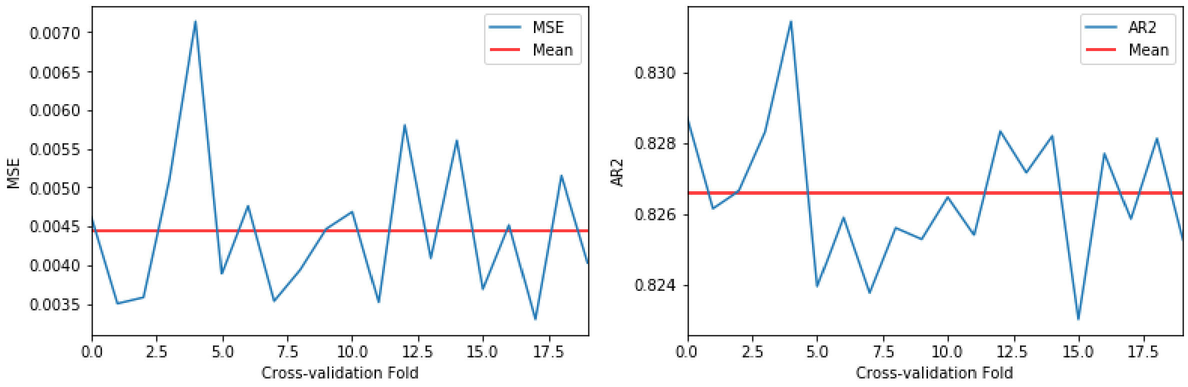


Figure 6: K-Fold Cross Validation Result

3. Complex Regression Model

3.1 Model Description

In this part, interaction term is added on top of the five transformed predictors in part two. Forward selection method is used to determine four best models when criterion is set as AIC, BIC, AR^2, R^2 respectively. The four models are then evaluated using k-fold cross validation to select the best model that have the highest consistency in term of MSE.

Through this method, AIC, AR^2, R^2 suggested the same model and BIC suggested another. The result of predictors are found in Table 6. 30-Fold cross validation shows that both models give an MSE value of about 0.0045 for mean value and 0.0014 for standard deviation. However, BIC model is picked as it explains as much variability as another model, with a smaller number of predictors. Also, the p-values for coefficients of all predictors is significantly closer to 0 as compared to the another model.

Table 6: Predictors obtained via forward selection

AIC	BIC	AR2	R2
EDU685213:LOG_INC910213	EDU685213	EDU685213:LOG_INC910213	EDU685213:LOG_INC910213
POP815213:LOG_INC910213	EDU685213:LOG_INC910213	POP815213:LOG_INC910213	POP815213:LOG_INC910213
SQRT_RHI225214:EDU685213	POP815213	SQRT_RHI225214:EDU685213	SQRT_RHI225214:EDU685213
SQRT_RHI225214:LOG_INC910213	SQRT_RHI225214:POP815213	SQRT_RHI225214:LOG_INC910213	SQRT_RHI225214:LOG_INC910213
state_abbreviation	state_abbreviation	state_abbreviation	state_abbreviation
state_abbreviation:EDU685213	state_abbreviation:SQRT_RHI225214	state_abbreviation:EDU685213	state_abbreviation:EDU685213
state_abbreviation:LOG_INC910213		state_abbreviation:LOG_INC910213	state_abbreviation:LOG_INC910213
state_abbreviation:POP815213		state_abbreviation:POP815213	state_abbreviation:POP815213
state_abbreviation:SQRT_RHI225214		state_abbreviation:SQRT_RHI225214	state_abbreviation:SQRT_RHI225214

Table 7: Best Model's OLS Regression Results (with interaction term)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           HilaryPercent   R-squared:                       0.840
Model:                             OLS   Adj. R-squared:                  0.835
Method:                  Least Squares   F-statistic:                     152.8
Date:                 Sun, 29 Sep 2019   Prob (F-statistic):               0.00
Time:                         10:36:28   Log-Likelihood:                 3324.4
No. Observations:                 2492   AIC:                            -6481.
Df Residuals:                     2408   BIC:                            -5992.
Df Model:                           83
Covariance Type:             nonrobust
=========================================================================================================
                                            coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------
Intercept                                 0.6840      0.021     32.492      0.000       0.643       0.725
state_abbreviation[T.AR]                 -0.0322      0.025     -1.291      0.197      -0.081       0.017
state_abbreviation[T.AZ]                 -0.0801      0.064     -1.248      0.212      -0.206       0.046
```

```
state_abbreviation[T.CA]                 -0.1985      0.031     -6.329      0.000      -0.260      -0.137
state_abbreviation[T.CO]                 -0.2276      0.027     -8.352      0.000      -0.281      -0.174
state_abbreviation[T.DE]                  0.0619      0.256      0.242      0.809      -0.439       0.563
state_abbreviation[T.FL]                 -0.1324      0.037     -3.562      0.000      -0.205      -0.060
state_abbreviation[T.GA]                 -0.1391      0.026     -5.377      0.000      -0.190      -0.088
state_abbreviation[T.HI]                 -0.3517      0.648     -0.543      0.587      -1.621       0.918
state_abbreviation[T.IA]                 -0.0681      0.026     -2.617      0.009      -0.119      -0.017
state_abbreviation[T.ID]                 -0.2672      0.038     -6.941      0.000      -0.343      -0.192
state_abbreviation[T.IL]                 -0.1735      0.024     -7.117      0.000      -0.221      -0.126
state_abbreviation[T.IN]                 -0.1773      0.024     -7.324      0.000      -0.225      -0.130
state_abbreviation[T.KY]                 -0.2018      0.025     -8.155      0.000      -0.250      -0.153
state_abbreviation[T.LA]                 -0.3052      0.047     -6.456      0.000      -0.398      -0.213
state_abbreviation[T.MD]                 -0.1890      0.049     -3.841      0.000      -0.285      -0.093
state_abbreviation[T.MI]                 -0.2060      0.024     -8.446      0.000      -0.254      -0.158
state_abbreviation[T.MO]                 -0.1588      0.023     -6.826      0.000      -0.204      -0.113
state_abbreviation[T.MS]                  0.0020      0.035      0.058      0.954      -0.067       0.071
state_abbreviation[T.MT]                 -0.1482      0.036     -4.114      0.000      -0.219      -0.078
state_abbreviation[T.NC]                 -0.1830      0.026     -6.970      0.000      -0.235      -0.132
state_abbreviation[T.NE]                 -0.0699      0.025     -2.841      0.005      -0.118      -0.022
state_abbreviation[T.NJ]                 -0.2130      0.048     -4.447      0.000      -0.307      -0.119
state_abbreviation[T.NM]                 -0.1347      0.044     -3.097      0.002      -0.220      -0.049
state_abbreviation[T.NV]                 -0.3821      0.052     -7.313      0.000      -0.485      -0.280
state_abbreviation[T.NY]                 -0.2816      0.029     -9.701      0.000      -0.339      -0.225
state_abbreviation[T.OH]                 -0.1249      0.025     -5.011      0.000      -0.174      -0.076
state_abbreviation[T.OK]                 -0.2890      0.028    -10.418      0.000      -0.343      -0.235
state_abbreviation[T.OR]                 -0.2030      0.037     -5.431      0.000      -0.276      -0.130
state_abbreviation[T.PA]                 -0.1616      0.027     -6.051      0.000      -0.214      -0.109
state_abbreviation[T.SC]                 -0.1266      0.052     -2.454      0.014      -0.228      -0.025
state_abbreviation[T.SD]                 -0.1444      0.031     -4.588      0.000      -0.206      -0.083
state_abbreviation[T.TN]                 -0.0560      0.025     -2.286      0.022      -0.104      -0.008
state_abbreviation[T.TX]                 -0.0934      0.024     -3.898      0.000      -0.140      -0.046
state_abbreviation[T.UT]                 -0.2591      0.054     -4.838      0.000      -0.364      -0.154
state_abbreviation[T.VA]                 -0.1759      0.025     -7.092      0.000      -0.225      -0.127
state_abbreviation[T.WA]                 -0.4017      0.034    -11.859      0.000      -0.468      -0.335
state_abbreviation[T.WI]                 -0.1986      0.027     -7.240      0.000      -0.252      -0.145
state_abbreviation[T.WV]                 -0.2679      0.029     -9.393      0.000      -0.324      -0.212
state_abbreviation[T.WY]                 -0.1653      0.069     -2.407      0.016      -0.300      -0.031
state_abbreviation[AL]:SQRT_RHI225214     0.0397      0.004     10.126      0.000       0.032       0.047
state_abbreviation[AR]:SQRT_RHI225214     0.0311      0.003      9.131      0.000       0.024       0.038
state_abbreviation[AZ]:SQRT_RHI225214     0.0315      0.041      0.772      0.440      -0.048       0.111
state_abbreviation[CA]:SQRT_RHI225214     0.0511      0.013      3.929      0.000       0.026       0.077
state_abbreviation[CO]:SQRT_RHI225214     0.0355      0.013      2.790      0.005       0.011       0.060
state_abbreviation[DE]:SQRT_RHI225214    -0.0105      0.056     -0.189      0.850      -0.120       0.099
state_abbreviation[FL]:SQRT_RHI225214     0.0313      0.008      3.908      0.000       0.016       0.047
state_abbreviation[GA]:SQRT_RHI225214     0.0563      0.003     18.682      0.000       0.050       0.062
state_abbreviation[HI]:SQRT_RHI225214    -0.0475      0.709     -0.067      0.947      -1.438       1.343
state_abbreviation[IA]:SQRT_RHI225214    -0.0175      0.013     -1.344      0.179      -0.043       0.008
state_abbreviation[ID]:SQRT_RHI225214    -0.0765      0.042     -1.815      0.070      -0.159       0.006
state_abbreviation[IL]:SQRT_RHI225214     0.0146      0.006      2.417      0.016       0.003       0.026
state_abbreviation[IN]:SQRT_RHI225214     0.0185      0.007      2.542      0.011       0.004       0.033
state_abbreviation[KY]:SQRT_RHI225214     0.0293      0.007      4.090      0.000       0.015       0.043
state_abbreviation[LA]:SQRT_RHI225214     0.0794      0.008     10.300      0.000       0.064       0.095
state_abbreviation[MD]:SQRT_RHI225214     0.0438      0.011      4.071      0.000       0.023       0.065
state_abbreviation[MI]:SQRT_RHI225214     0.0211      0.007      3.129      0.002       0.008       0.034
state_abbreviation[MO]:SQRT_RHI225214     0.0262      0.005      4.900      0.000       0.016       0.037
state_abbreviation[MS]:SQRT_RHI225214     0.0393      0.004      8.914      0.000       0.031       0.048
state_abbreviation[MT]:SQRT_RHI225214     0.0545      0.045      1.206      0.228      -0.034       0.143
state_abbreviation[NC]:SQRT_RHI225214     0.0404      0.004     11.151      0.000       0.033       0.047
state_abbreviation[NE]:SQRT_RHI225214    -0.0480      0.014     -3.466      0.001      -0.075      -0.021
state_abbreviation[NJ]:SQRT_RHI225214     0.0585      0.013      4.583      0.000       0.033       0.083
state_abbreviation[NM]:SQRT_RHI225214     0.0082      0.027      0.299      0.765      -0.046       0.062
state_abbreviation[NV]:SQRT_RHI225214     0.1206      0.027      4.387      0.000       0.067       0.174
state_abbreviation[NY]:SQRT_RHI225214     0.0575      0.008      7.101      0.000       0.042       0.073
state_abbreviation[OH]:SQRT_RHI225214     0.0298      0.007      4.525      0.000       0.017       0.043
state_abbreviation[OK]:SQRT_RHI225214     0.0415      0.009      4.434      0.000       0.023       0.060
state_abbreviation[OR]:SQRT_RHI225214     0.0382      0.033      1.170      0.242      -0.026       0.102
state_abbreviation[PA]:SQRT_RHI225214     0.0337      0.008      4.375      0.000       0.019       0.049
state_abbreviation[SC]:SQRT_RHI225214     0.0552      0.008      7.238      0.000       0.040       0.070
state_abbreviation[SD]:SQRT_RHI225214     0.0571      0.027      2.139      0.033       0.005       0.109
state_abbreviation[TN]:SQRT_RHI225214     0.0430      0.005      8.676      0.000       0.033       0.053
state_abbreviation[TX]:SQRT_RHI225214     0.0497      0.004     11.093      0.000       0.041       0.058
state_abbreviation[UT]:SQRT_RHI225214    -0.1091      0.060     -1.817      0.069      -0.227       0.009
state_abbreviation[VA]:SQRT_RHI225214     0.0556      0.003     17.743      0.000       0.049       0.062
state_abbreviation[WA]:SQRT_RHI225214     0.0199      0.021      0.956      0.339      -0.021       0.061
state_abbreviation[WI]:SQRT_RHI225214     0.0116      0.016      0.745      0.456      -0.019       0.042
state_abbreviation[WV]:SQRT_RHI225214     0.0306      0.013      2.405      0.016       0.006       0.056
state_abbreviation[WY]:SQRT_RHI225214    -0.0019      0.061     -0.032      0.975      -0.121       0.117
EDU685213                                -0.0509      0.004    -13.398      0.000      -0.058      -0.043
EDU685213:LOG_INC910213                   0.0046      0.000     12.837      0.000       0.004       0.005
POP815213                                 0.0020      0.000      7.426      0.000       0.001       0.003
SQRT_RHI225214:POP815213                 -0.0004      0.000     -3.076      0.002      -0.001      -0.000
==============================================================================
Omnibus:                      322.060   Durbin-Watson:                   1.963
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3747.995
Skew:                           0.029   Prob(JB):                         0.00
Kurtosis:                       9.008   Cond. No.                     1.61e+05
==============================================================================
```

3.2 Result

The new regression model has an adjusted R^2 of 0.835, 1% higher than simple regression model of 0.826, indicating the variability being able to be explained by the model has improved o 83.5%. The p-values for all predictors' coefficient is <0.05 which indicates that they significantly different than zero at 95% significant level.

The model says that percentage of vote received by Hilary is correlated to

1. Percentage of African American (slope is affected by state)
2. Percentage of persons with degree (slope is affected by per capita income)
3. Language other than English spoken at home (slope is affected by percentage of African American)

QQ-plot in Figure 7(a) shows that the residual normality is slightly violated as the residual is again having fatter tail, due to some variability unexplainable by the model. Residual versus fitted value plot in Figure 7(b) shows that residual's variance is consistent across fitted values and residual has zero mean.

Average MSE for 30-fold cross validation is at 0.0046 with standard deviation of 0.0015, indicating that the model is stable and not over fitted.

Partial regression plots for all predictors are also found in Figure 9, showing that the additional interaction terms are significant in the model and are linear to the target variable. There is also no critical non-constant variance observed for residual against the predictors.

Cook's distance plot in Figure 10 show that generally all observations in dataset has normal influence except for observation 8 which has high Cook's D value, indicating that the influence for this observation is high. Thorough understanding on the observation is needed to decide if it can be discarded. For the time being, it will be kept in the current model.

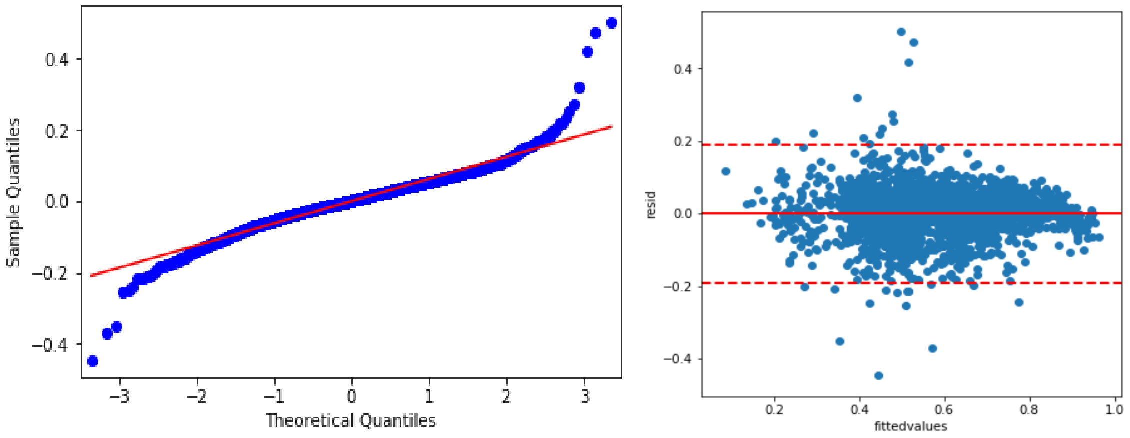


Figure 7: (a) QQ-plot, (b) Residual versus fitted values.

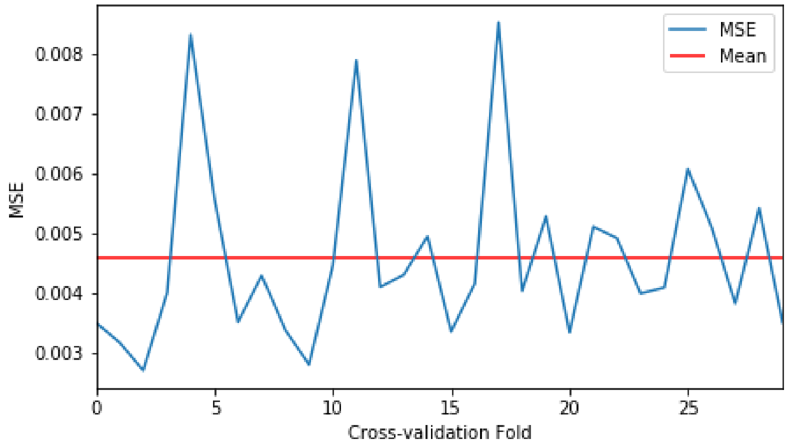


Figure 8: MSE for 30-Fold cross validation

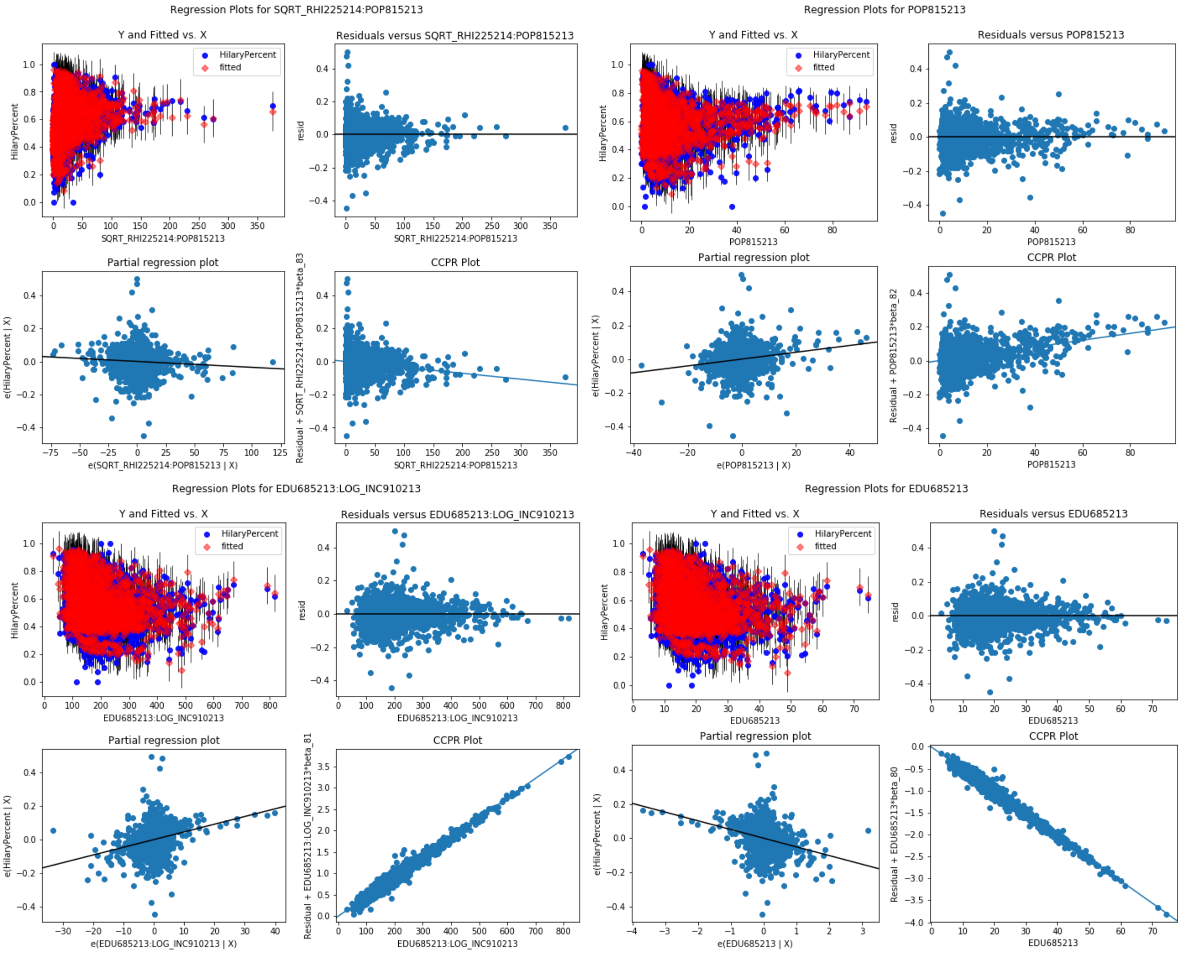


Figure 9: Regression Results against Predictors

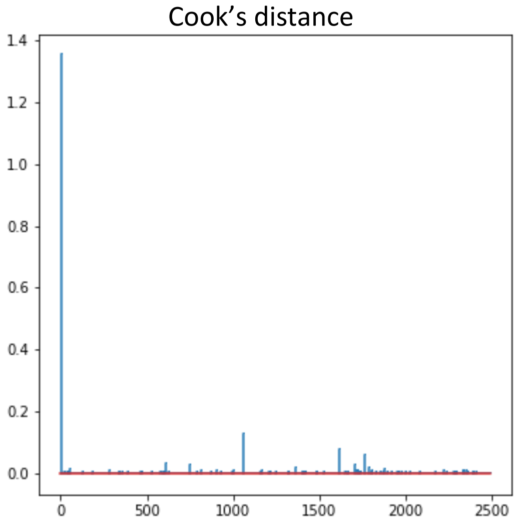


Figure 10: Cook's distance plot