DatExplorationAssignment

erik campbell

2022-08-08

Research Question: Among colleges that predominantly grant bachelor's degrees, did the release of Score Card shift student interest to high-earning colleges relative to low-earnings ones?

Information about the data used & key use to answer research question:

Google search trends data with a range from March 1, 2013 through March 1, 2016 measuring keyword search by week. Trend indices are comparable only to themselves, thus indices were standardized in order utilized as an independent variable in regression models going forward.

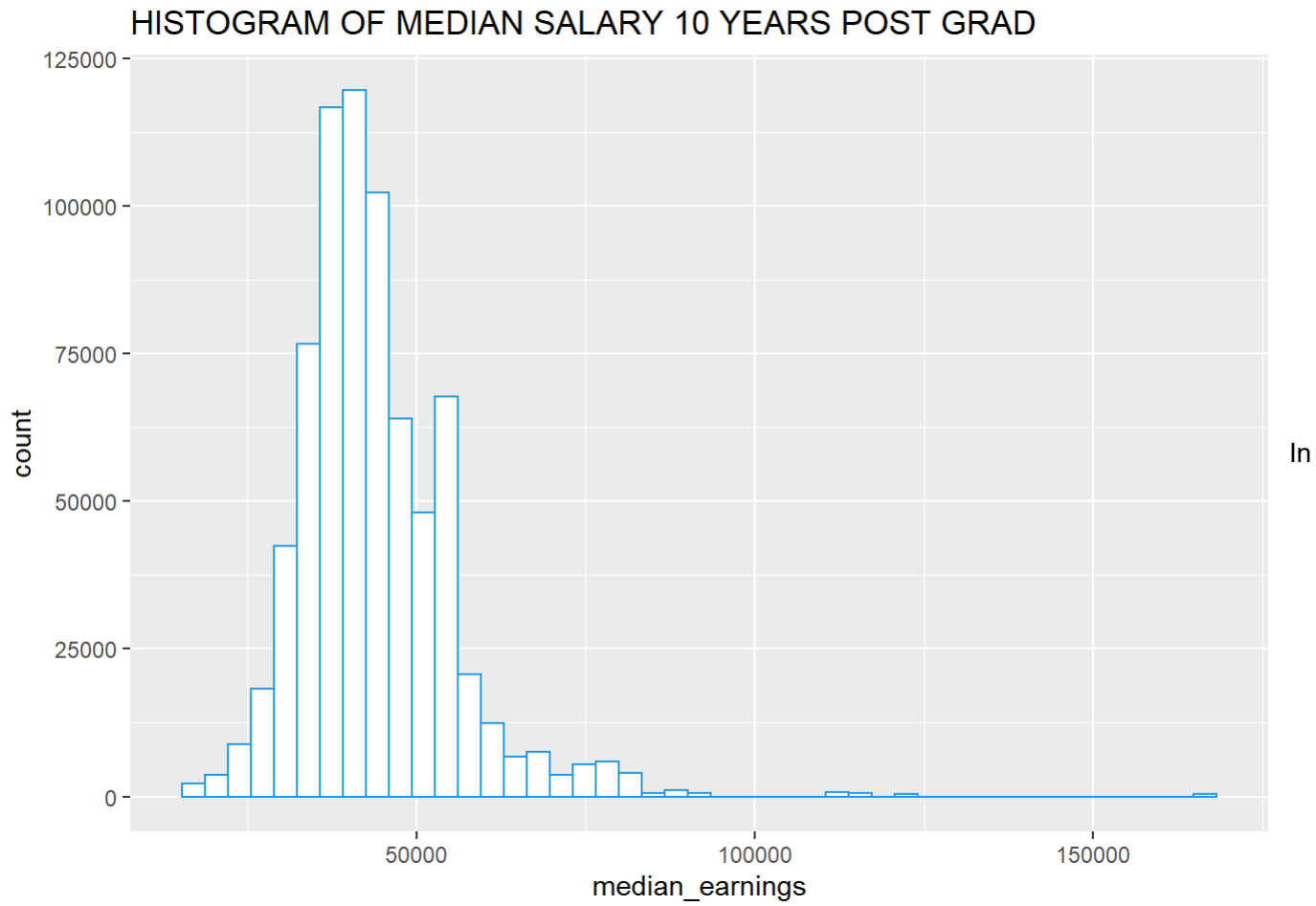
Scorecard was executed on September 1, 2015 and contains a multitude of variables and metrics about schools. The data contains a variable of interest called median earnings and is used a dependent variable in regression models going forward. Details further describing the formulation and use of median earnings is described below in data wrangling.

Id_name_link.csv is a file is a file used to assist in linking Scorecard data to Google trends data. It was used to identify distinct university names.

Data Wrangling & Analysis Foundation:

The Google trends data was aggregated for the 3 years at the weekly level, grouped by school name and keyword and the indices were standardized and used as an independent variable in models. Then, Google trends data was inner-joined with Id_name_link file and then inner-joined with Scorecard data and na omit was applied and observations equaled 1,210,902 with 130 variables. I pulled a working data set that included variables of interest, filtered for schools identified as primarily awarding bachelor's degrees, filtered out any values for median earnings containing "Null" and "Privacy Suppressed," and transformed other variables of interest. The working data set contained 741,598 observations and 16 variables.

Using the working data set, a histogram of median earning over 10 years was evaluated and demonstrated a slight skewness to the right with a couple outlier values.



In order to differentiate between high earning schools vs. low earning schools and turn median earnings into a binary variable, the median earnings of graduates after 10 years was calculated at $42,000. For perspective, the mean of median earnings after 10 years was $43,799 and had a standard deviation was $11,607. I added high earning school as a binary variable to my working data set. My analysis moving forward will have the logic that schools with median earnings for alumni 10 years after graduation above $42k will be considered as 'high earning schools' while those below $42k will be considered as 'low earning schools.'

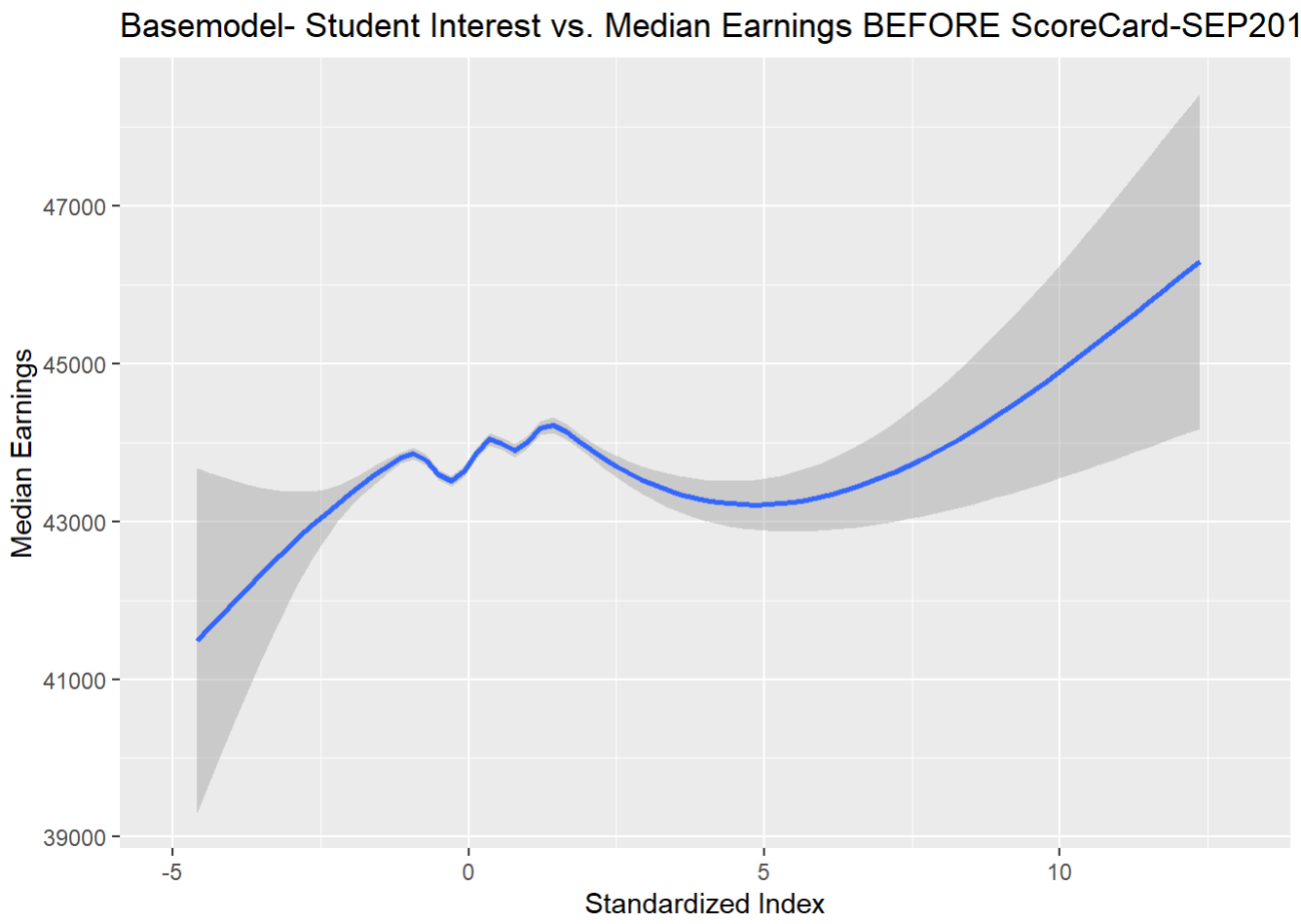
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16700   36600   42000   43799   49400  166200
```

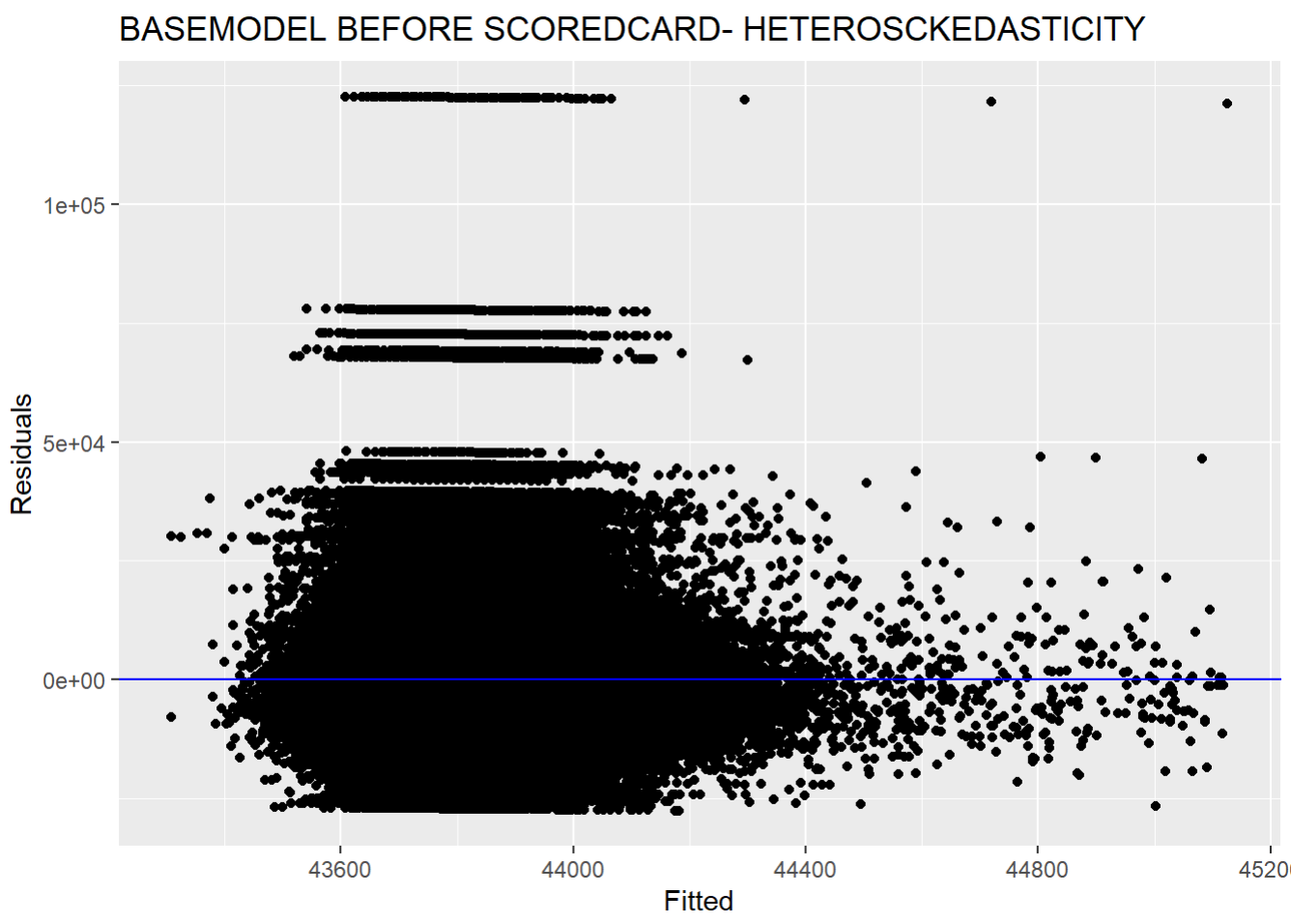
```
## Adding missing grouping variables: `keyword`
## Adding missing grouping variables: `keyword`
```

I reviewed Working data in a vtable to check data types and for missing values and constructed a casual diagram ("../figures/") with the variables of interest. I decided to not use average sat, percentage of student receiving a federal loan and median debt after 10 years due to concerned with colliders, washing out too much variation, and missing values in some variables. Instead, I decided to pare-down variables and focus on high_earning, university type and geography variables in regressions. Scorecard was implemented in September 2015 and our Google trends data precedes scorecard implementation; and in order to evaluate the effect of release of scorecard on student interest ("standardized_index") in high earning schools ("high_earning_school"), the working data set was filtered for before/after scorecard release and split into two data sets. I created baseline regressions evaluating

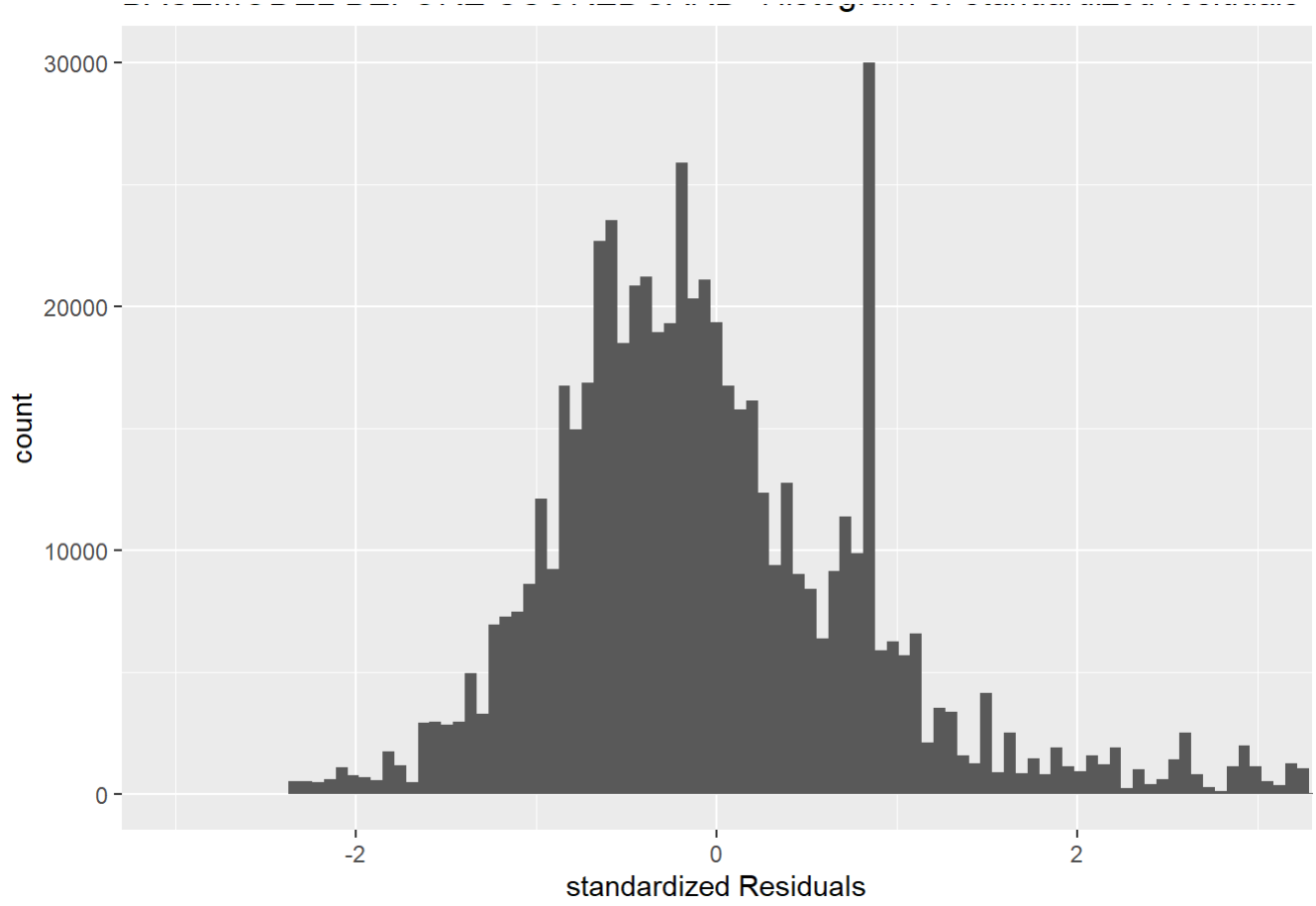
before and after September 1, 2015 by regressing standardized_index on median earnings with the intent of graphically evaluating linearity/non-linearity, heteroskedasticity and homogeneity, and the standard residuals. I decided to add heteroskedasticity robust standard errors for future regressions.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

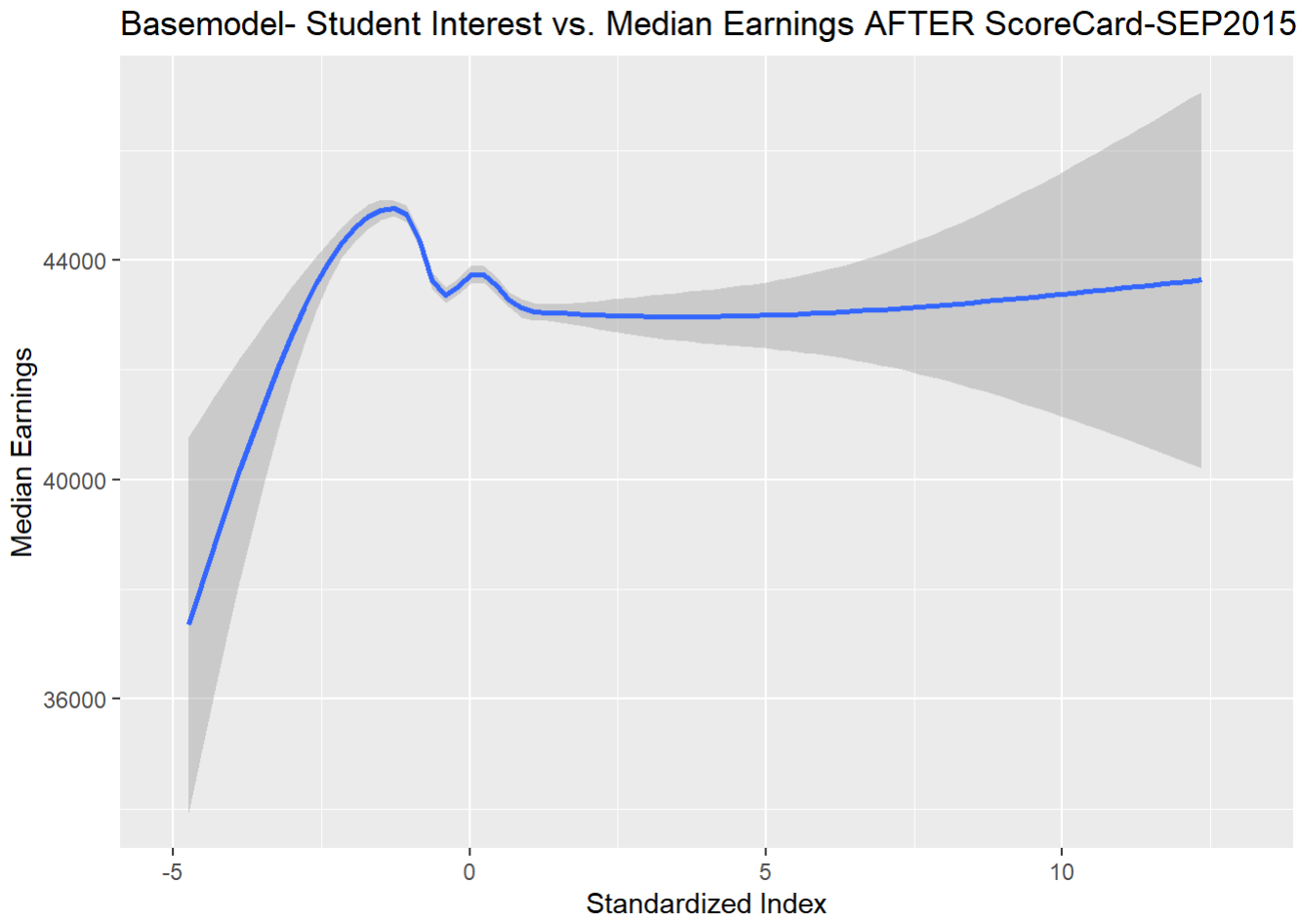


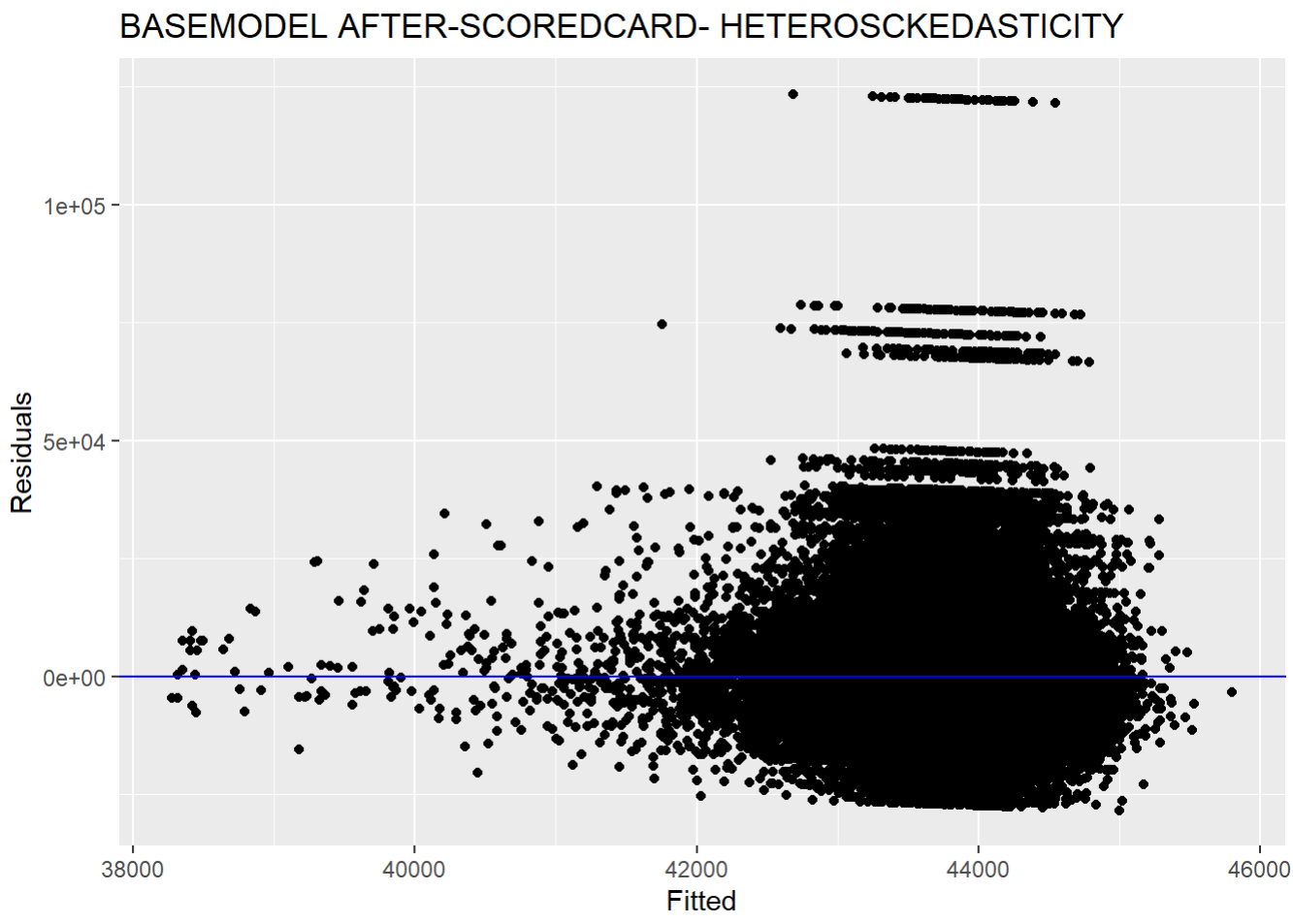


BASEMODEL BEFORE SCOREDCARD- Histogram of standardized residuals

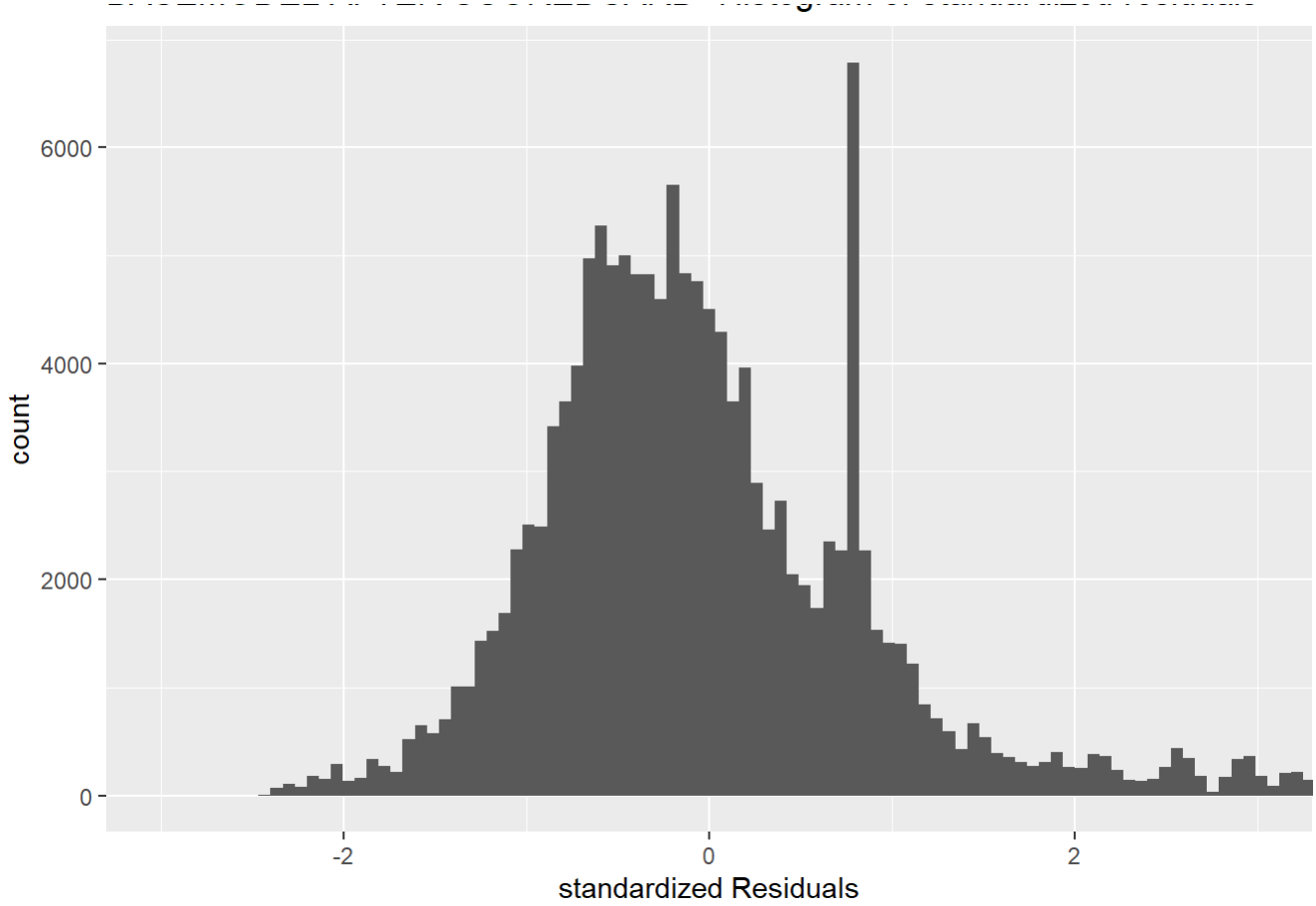



```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```





BASEMODEL AFTER SCOREDCARD- Histogram of standardized residuals

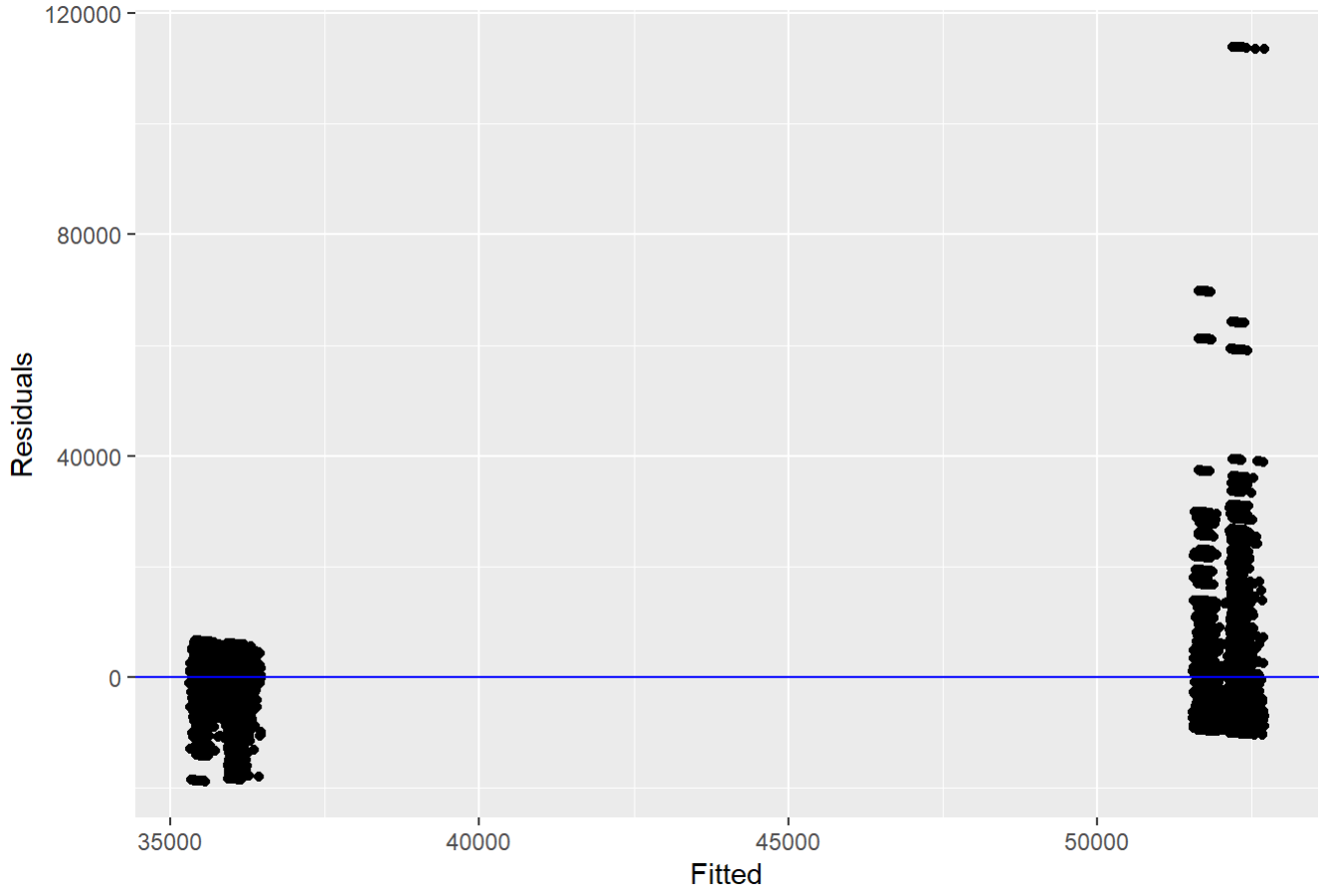


The table below includes all the models that were produced. My independent variable here is ‘median_earnings_10’ and independent variable is ‘standardized_index’, both pulling from pre scorecard(‘PRE_SC_Model’) and post scorecard(‘POOST_SC_MODEL’ to compare between the two time frames. Standardized Index was chosen as the independent variable because we want to see the effect of student interest on how much money alumni make after graduating from certain colleges. I’ve included the binary variable of ‘high_earning_school’ to show the effect of the Index on high vs. low earning schools. The categorical variable ‘public_private’ notes three different values: Public School, Private School (Non-Profit), and Private School (For-Profit). The categorical variable ‘school_geography’ was transformed into 4 values: City, Suburb, town, rural.

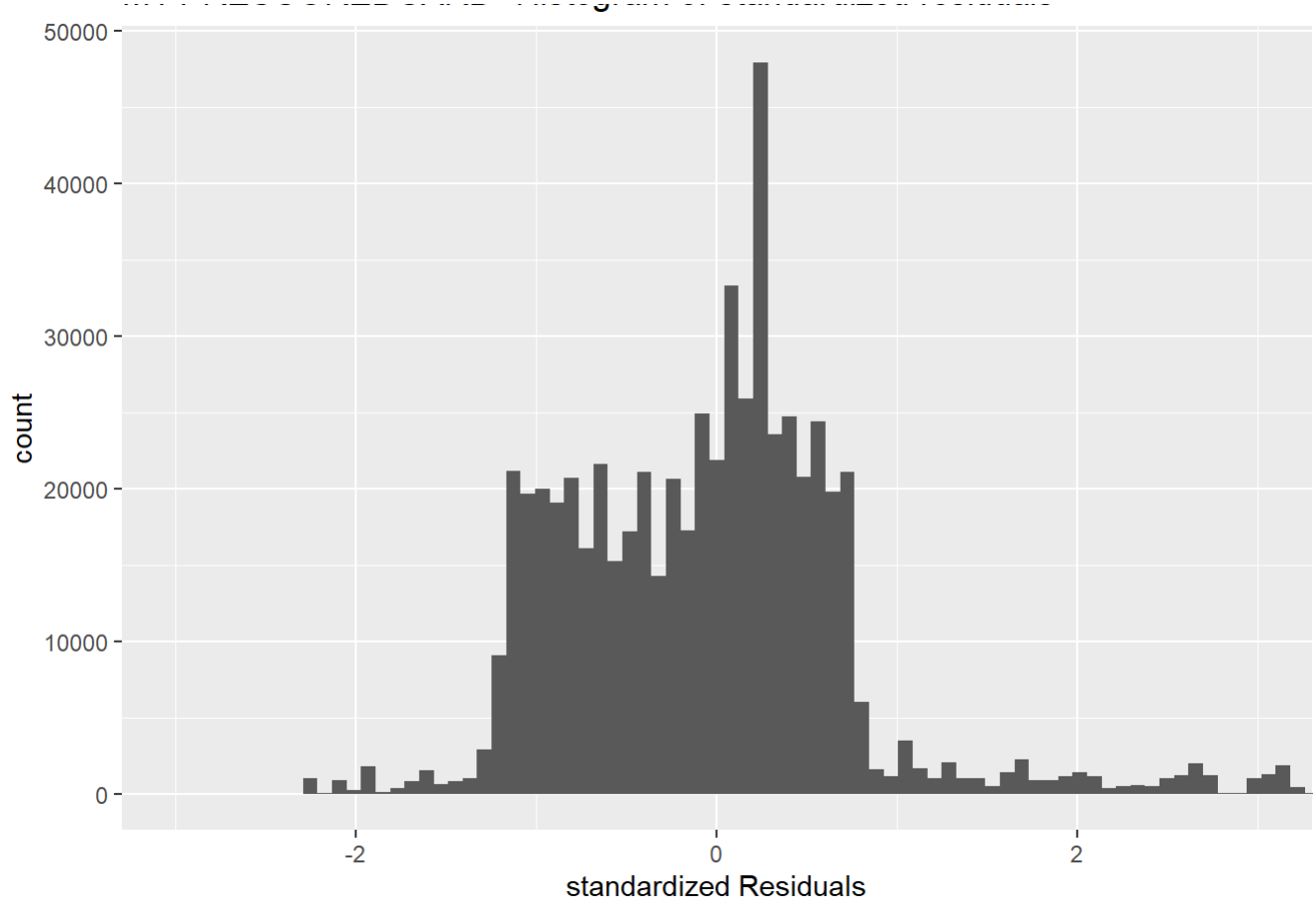
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

M1- Student Interest vs. Median Earnings Before ScoreCard-SEP2015:

M1 PRESCOREDCARD- HETEROSCKEDASTICITY

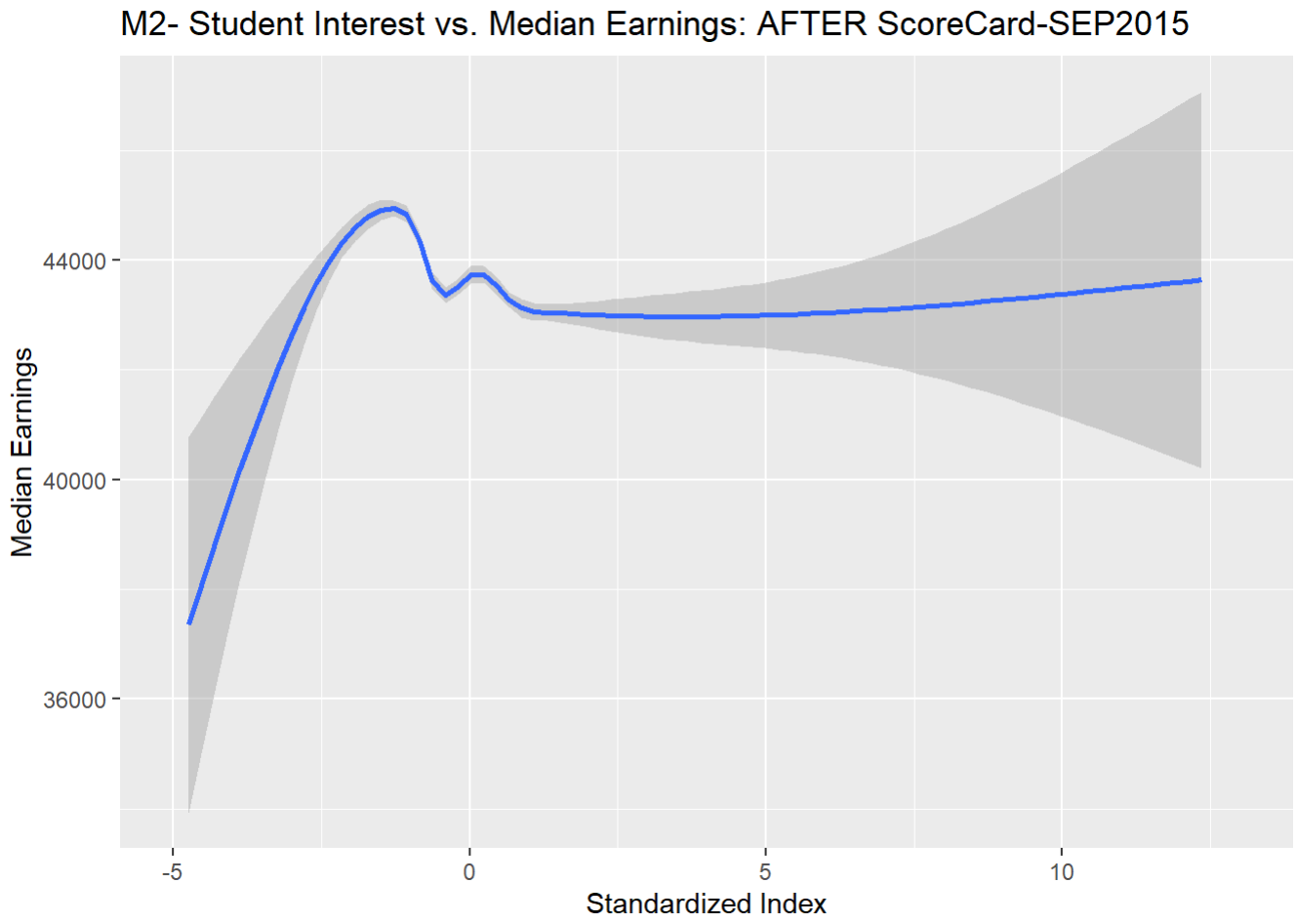


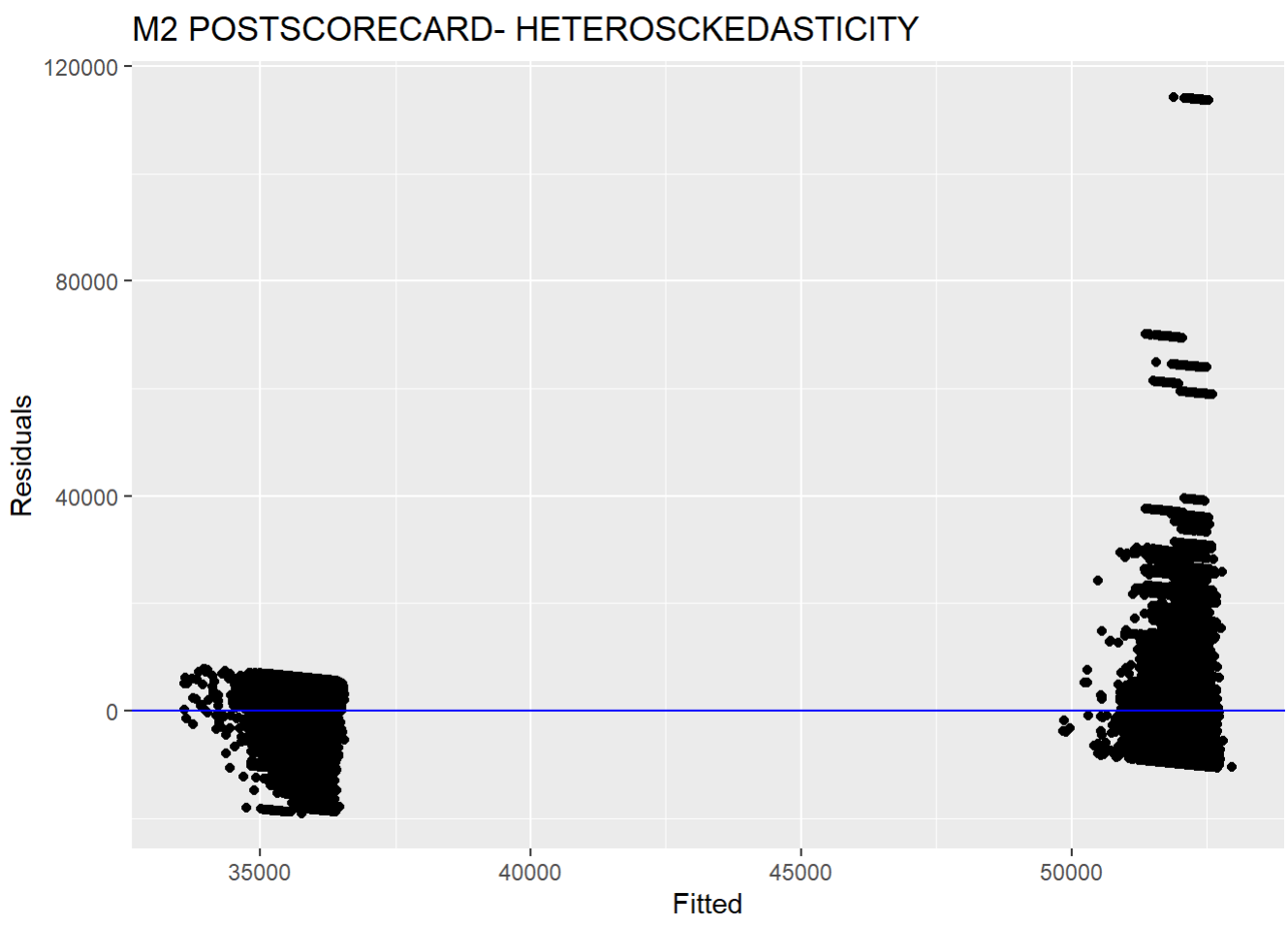
M1 PRESCOREDCARD- Histogram of standardized residuals



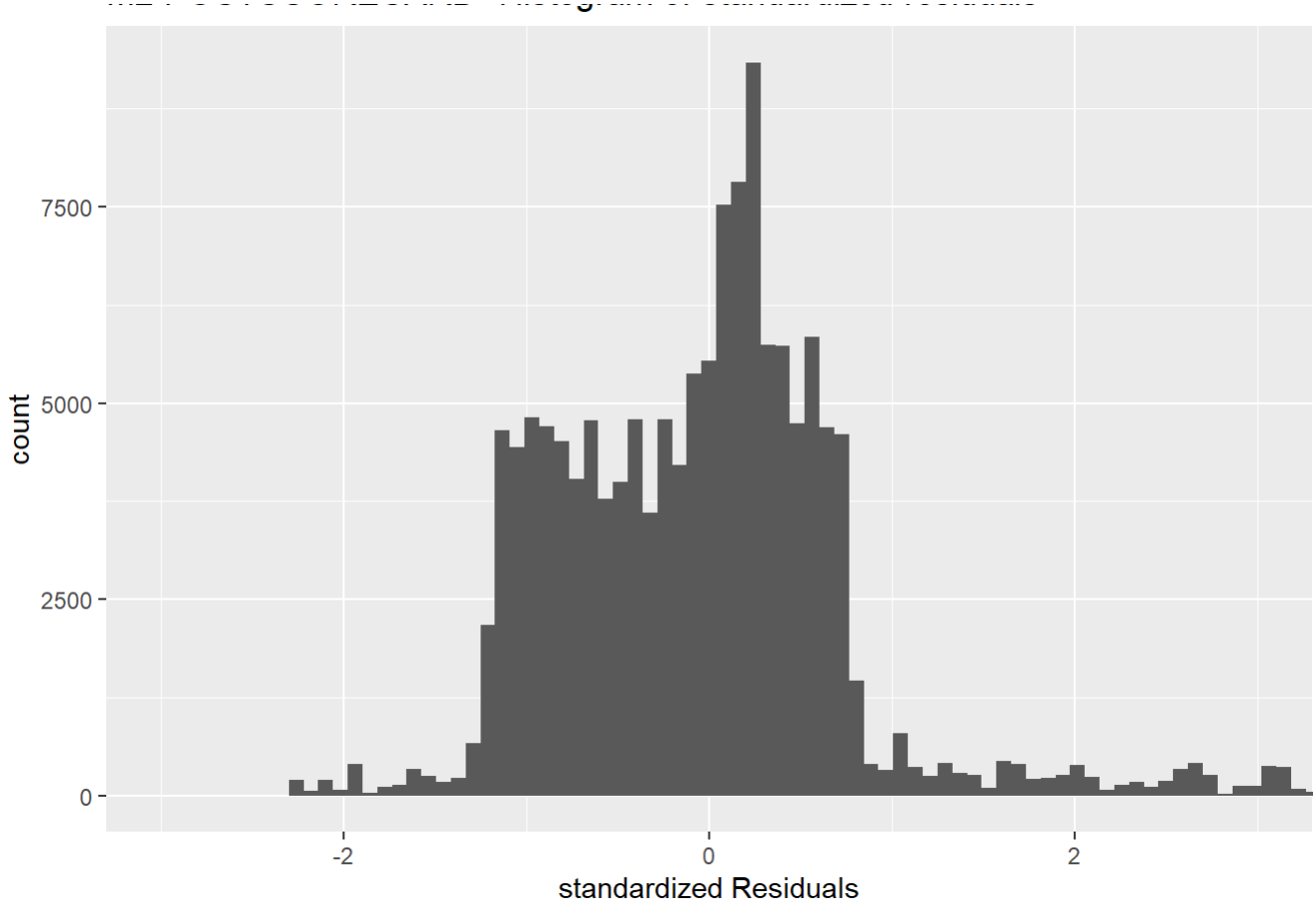

```
## Wald test, H0: joint nullity of public_privateprivate_non_profit and public_privatepublic
##  stat = 87.4, p-value < 2.2e-16, on 2 and 140,260 DoF, VCOV: Heteroskedasticity-robust.
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



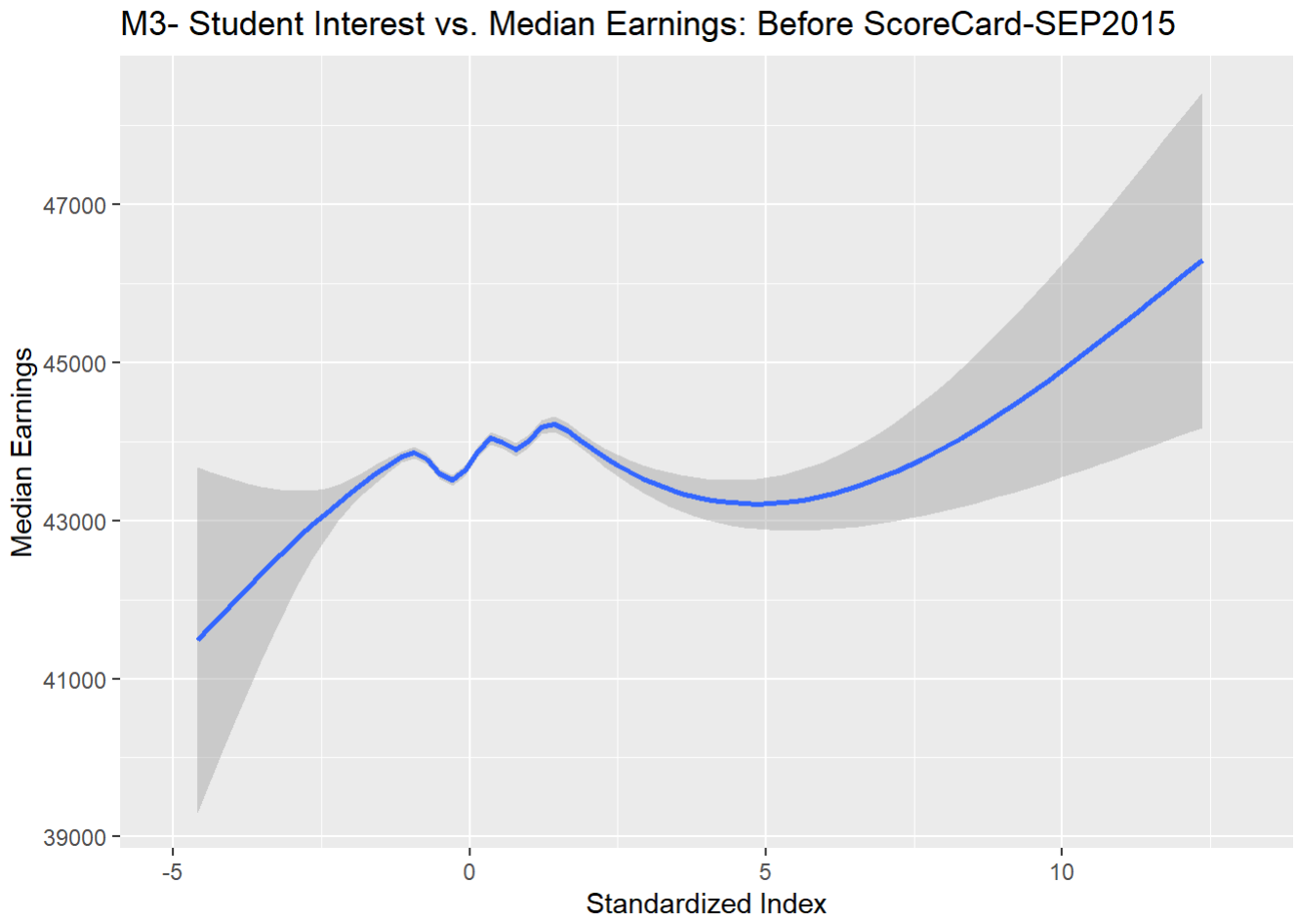


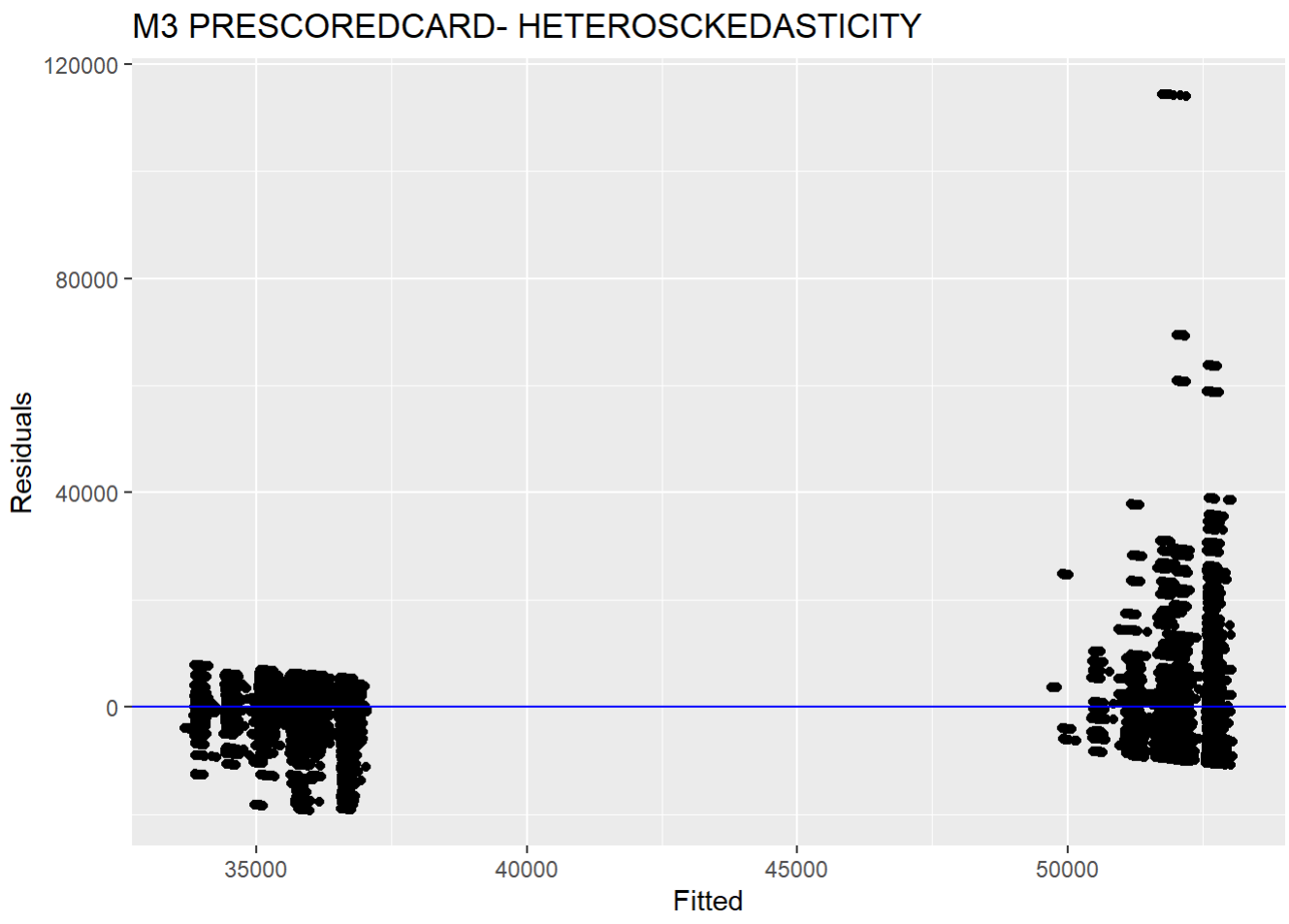
M2 POSTSCORECARD- Histogram of standardized residuals



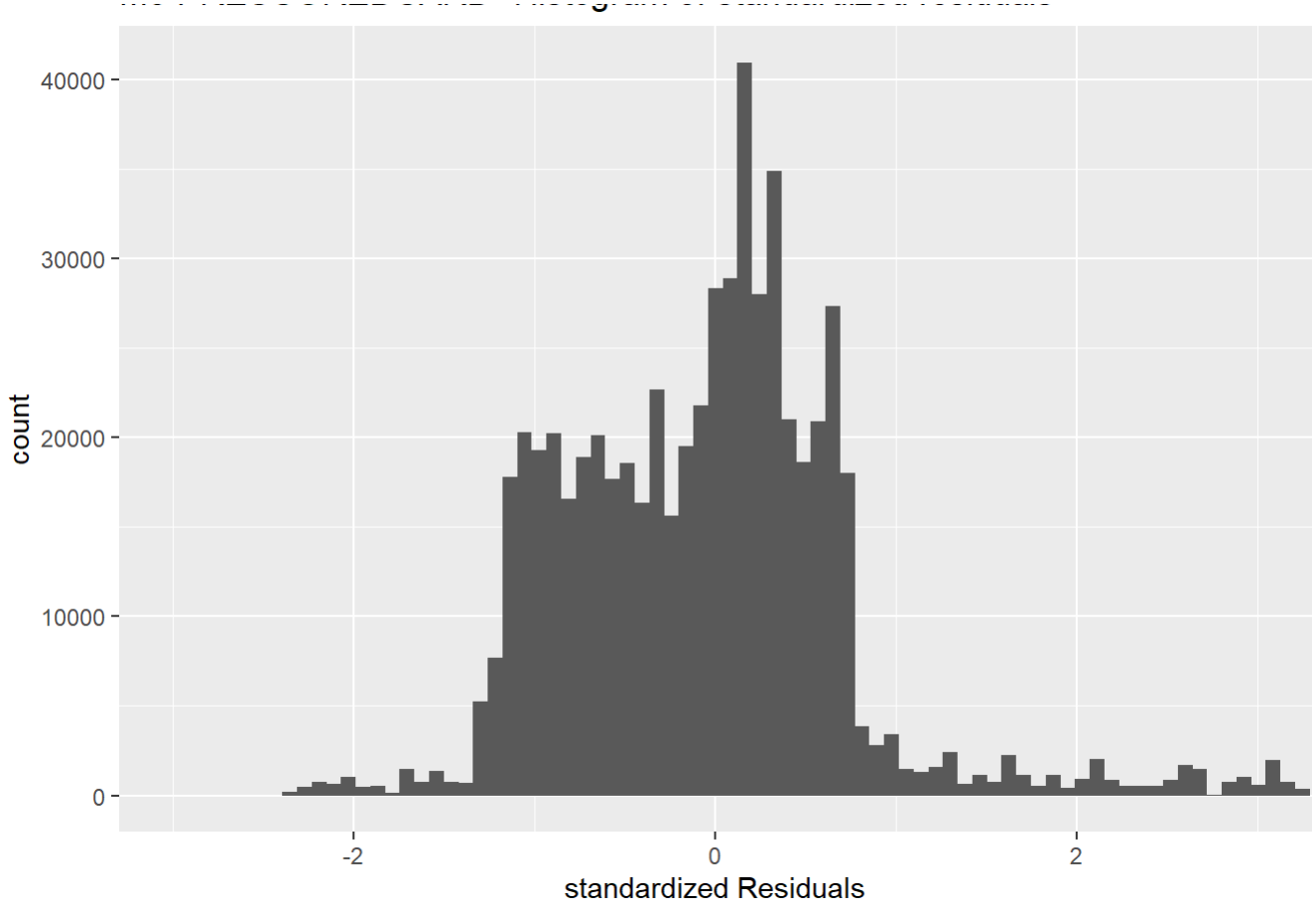

```
## Wald test, H0: joint nullity of school_geographyrural, school_geographySuburb and school_geog
raphytown
##  stat = 1,520.7, p-value < 2.2e-16, on 3 and 601,325 DoF, VCOV: Heteroskedasticity-robust.
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



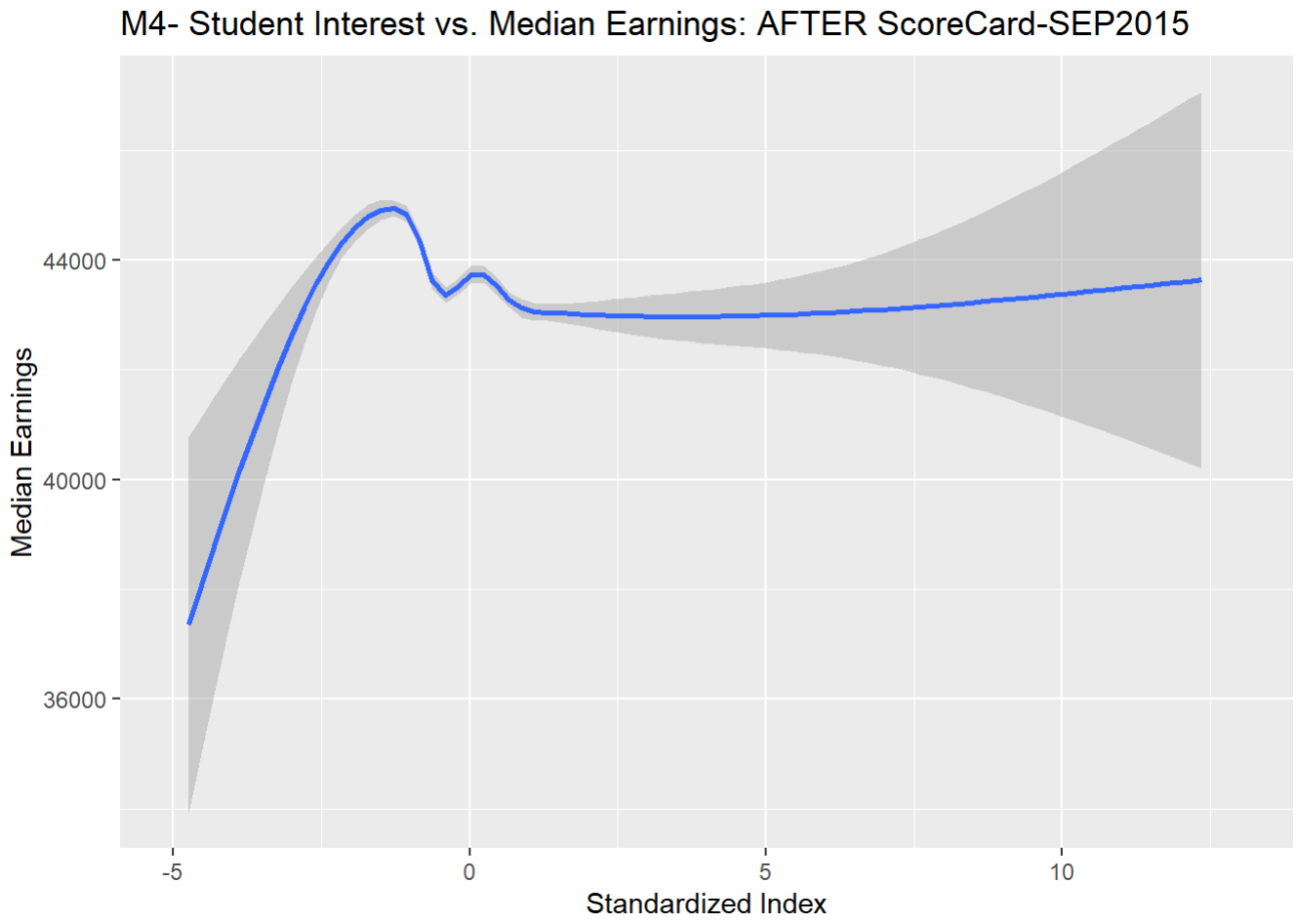


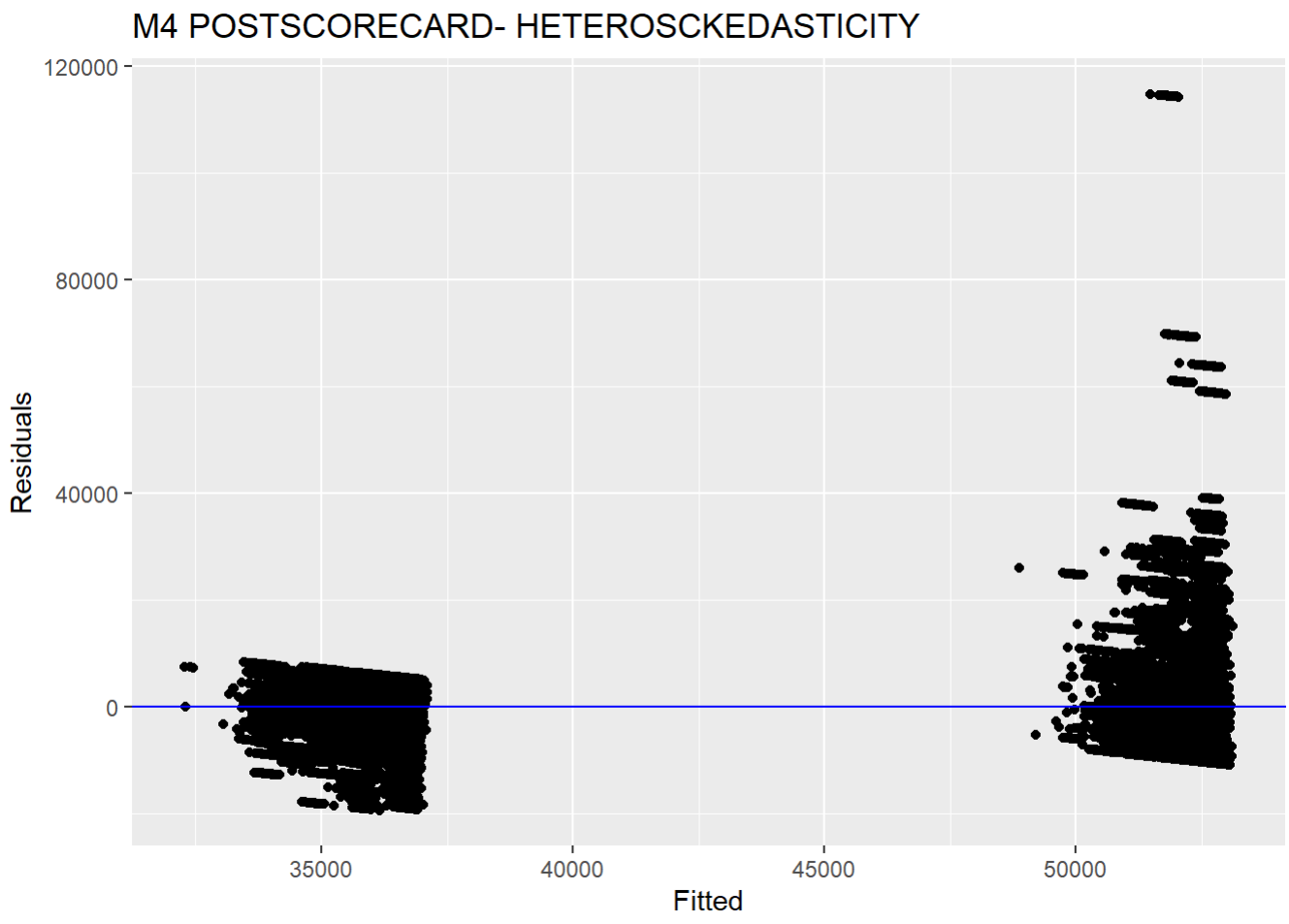
M3 PRESCOREDCARD- Histogram of standardized residuals



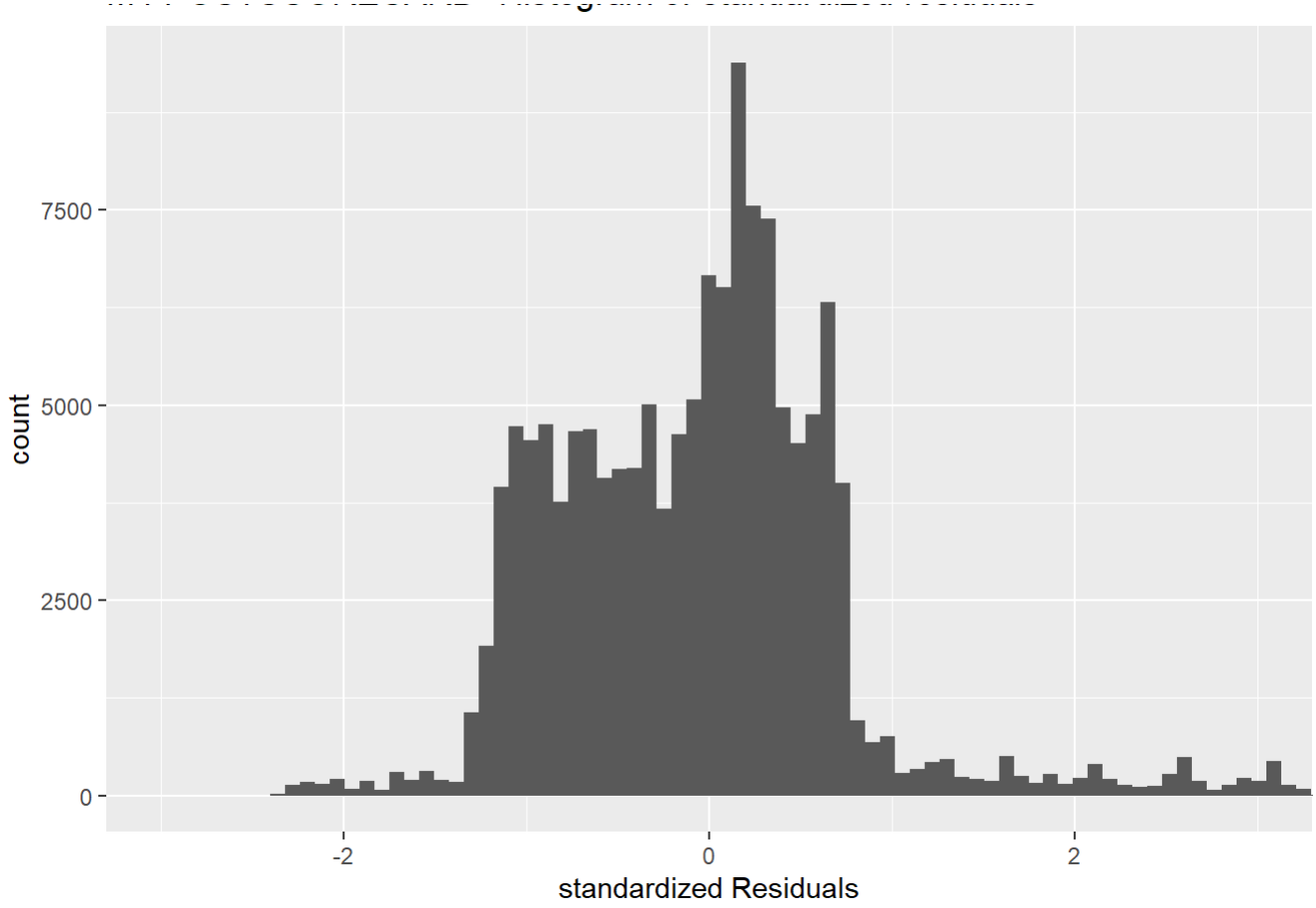

```
## Wald test, H0: joint nullity of school_geographyrural, school_geographySuburb and school_geog
raphytown
##  stat = 348.9, p-value < 2.2e-16, on 3 and 140,257 DoF, VCOV: Heteroskedasticity-robust.
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```





M4 POSTSCORECARD- Histogram of standardized residuals


7500
5000
2500
0
count
-2
0
2
standardized Residuals

```
##                                  basemodel_preScor.. basemodel_postSco..
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      43,800.2*** (14.97) 43,716.3*** (31.31)
## standardized_index                  107.3*** (15.12)   -440.3*** (30.54)
## high_earning_schoolTRUE
## public_privateprivate_non_profit
## public_privatepublic
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ____________________ ____________________
## S.E. type                                          IID                  IID
## Observations                                   601,333              140,265
## R2                                            8.36e-5              0.00148
## Adj. R2                                        8.2e-5              0.00147
##
##                                               m1_feols             m2_feols
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      35,438.3*** (27.55) 35,324.2*** (59.50)
## standardized_index                  36.22*** (10.97)   -151.9*** (21.08)
## high_earning_schoolTRUE          16,221.9*** (22.54) 16,230.4*** (46.72)
## public_privateprivate_non_profit    592.8*** (29.82)    687.2*** (63.50)
## public_privatepublic                  61.14* (27.71)      144.7* (59.58)
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ____________________ ____________________
## S.E. type                        Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                   601,333              140,265
## R2                                            0.48909              0.48883
## Adj. R2                                       0.48908              0.48881
##
##                                               m3_feols             m4_feols
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      35,866.9*** (28.81) 35,761.8*** (62.14)
## standardized_index                   32.15** (10.94)   -135.4*** (21.08)
## high_earning_schoolTRUE          16,010.9*** (22.48) 16,025.7*** (46.61)
## public_privateprivate_non_profit    783.6*** (30.33)    867.5*** (64.44)
## public_privatepublic                211.6*** (28.45)    284.5*** (60.92)
## school_geographyrural            -2,139.0*** (33.99) -2,125.3*** (70.45)
## school_geographySuburb             -853.4*** (28.63)   -867.0*** (59.48)
## school_geographytown               -937.2*** (22.72)   -922.9*** (47.03)
## _________________________________ ____________________ ____________________
## S.E. type                        Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                   601,333              140,265
## R2                                            0.49137              0.49109
## Adj. R2                                       0.49136              0.49107
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis

The table above/below includes all the models that were produced. My independent variable here is ‘median_earnings_10’ and independent variable is ‘standardized_index’, both pulling from pre scorecard(‘PRE_SC_Model’) and post scorecard(‘POOST_SC_MODEL’ to compare between the two time frames. Standardized Index was chosen as the independent variable because we want to see the effect of student interest on how much money alumni make after graduating from certain colleges. I’ve included the binary variable of ‘high_earning_school’ to show the effect of the Index on high vs. low earning schools. The categorical variable ‘public_private’ notes three different values: Public School, Private School (Non-Profit), and Private School (For-Profit). The categorical variable ‘school_geography’ was transformed into 4 values: City, Suburb, town, rural.

basemodel_preScorecard_feols:

Before 9/1/2015, the mean median_earnings was $43,800.20 with a standard error of $14.97 and statistically significant at the .1% level. For a one-unit increase in ‘standardized_index’ is associated with a $107.3 increase in alumni earnings (’median_earnings_10), the standard errors were 15.12 and statistically significant at the .1% level. There was some heteroskedasticity (“../figures/”) as evidenced graphically, thus heteroskedastic robust adjustments were made in all feols models. Histogram of standardized residuals looked relatively normal.

basemodel_postScorecard_feols:

After scorecard implementation on 9/1/2015, the mean median_earnings was $43,716.30 with a standard error of $31.31 and statistically significant at the .1% level. For a one-unit increase in ‘standardized_index’ is associated with a $440.30 decrease in alumni earnings (’median_earnings_10), the standard errors were 30.54 and statistically significant at the .1% level. There was some heteroskedasticity (“../figures/”) as evidenced graphically, thus heteroskedastic robust adjustments were made in all feols models. Histogram of standardized residuals looked relatively normal.

m1_feols (BEFORE SCORECARD):

```
##                                  basemodel_preScor.. basemodel_postSco..
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      43,800.2*** (14.97) 43,716.3*** (31.31)
## standardized_index                  107.3*** (15.12)   -440.3*** (30.54)
## high_earning_schoolTRUE
## public_privateprivate_non_profit
## public_privatepublic
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ____________________ ____________________
## S.E. type                                          IID                  IID
## Observations                                   601,333              140,265
## R2                                            8.36e-5              0.00148
## Adj. R2                                        8.2e-5              0.00147
##
##                                               m1_feols             m2_feols
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      35,438.3*** (27.55) 35,324.2*** (59.50)
## standardized_index                  36.22*** (10.97)   -151.9*** (21.08)
## high_earning_schoolTRUE          16,221.9*** (22.54) 16,230.4*** (46.72)
## public_privateprivate_non_profit    592.8*** (29.82)    687.2*** (63.50)
## public_privatepublic                  61.14* (27.71)      144.7* (59.58)
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ____________________ ____________________
## S.E. type                        Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                   601,333              140,265
## R2                                            0.48909              0.48883
## Adj. R2                                       0.48908              0.48881
##
##                                               m3_feols             m4_feols
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      35,866.9*** (28.81) 35,761.8*** (62.14)
## standardized_index                   32.15** (10.94)   -135.4*** (21.08)
## high_earning_schoolTRUE          16,010.9*** (22.48) 16,025.7*** (46.61)
## public_privateprivate_non_profit    783.6*** (30.33)    867.5*** (64.44)
## public_privatepublic                211.6*** (28.45)    284.5*** (60.92)
## school_geographyrural            -2,139.0*** (33.99) -2,125.3*** (70.45)
## school_geographySuburb             -853.4*** (28.63)   -867.0*** (59.48)
## school_geographytown               -937.2*** (22.72)   -922.9*** (47.03)
## _________________________________ ____________________ ____________________
## S.E. type                        Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                   601,333              140,265
## R2                                            0.49137              0.49109
## Adj. R2                                       0.49136              0.49107
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

categorical variables note: public_private = Both categories are relative to Private For-Profit University school_geography = all 3 categories are relative to Universities located in cities (aggregated 3 categories with case_when into city)

Before Scorecard, the mean ‘median_earnings_10’ was $35,438.30 standard error of $27.55 and statistically significant at the .1% level. Previous to September 2015,a one-unit increase in ‘standardized_index’ is associated with a $36.22 increase in alumni earnings (‘median_earnings_10), holding all else constant. The standard errors were 10.97 and statistically significant at the .1% level. For every one-unit increase in ’standardized_index’, median earnings for alumni from a ‘high_earning_school’ increased $16,221.90 prior to September 2015. The standard errors were 22.54 and this was statically significant at the .1% level. For every one-unit increase in ‘standardized_index’ graduates from a private, non-profit school earned $592.80 more than graduates from a private, for-profit university. The standard error was 29.82 and this was statically significant at the .1% level. For every one-unit increase in ‘standardized_index’ graduates from a public university earned $61.14 more than graduates from a private for-profit university. The standard errors were 27.71 and this was statically significant at the 5% level. Heteroskedasticity noticed in plot and histogram of standardized residuals looked relatively normal(“../figures/”).

m2_feols (AFTER SCORECARD):

```
##                                  basemodel_preScor.. basemodel_postSco..
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      43,800.2*** (14.97) 43,716.3*** (31.31)
## standardized_index                  107.3*** (15.12)   -440.3*** (30.54)
## high_earning_schoolTRUE
## public_privateprivate_non_profit
## public_privatepublic
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ____________________ ____________________
## S.E. type                                          IID                  IID
## Observations                                   601,333              140,265
## R2                                            8.36e-5              0.00148
## Adj. R2                                        8.2e-5              0.00147
##
##                                               m1_feols             m2_feols
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      35,438.3*** (27.55) 35,324.2*** (59.50)
## standardized_index                  36.22*** (10.97)   -151.9*** (21.08)
## high_earning_schoolTRUE          16,221.9*** (22.54) 16,230.4*** (46.72)
## public_privateprivate_non_profit    592.8*** (29.82)    687.2*** (63.50)
## public_privatepublic                  61.14* (27.71)      144.7* (59.58)
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ____________________ ____________________
## S.E. type                        Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                   601,333              140,265
## R2                                            0.48909              0.48883
## Adj. R2                                       0.48908              0.48881
##
##                                               m3_feols             m4_feols
## Dependent Var.:                   median_earnings_10  median_earnings_10
##
## (Intercept)                      35,866.9*** (28.81) 35,761.8*** (62.14)
## standardized_index                   32.15** (10.94)   -135.4*** (21.08)
## high_earning_schoolTRUE          16,010.9*** (22.48) 16,025.7*** (46.61)
## public_privateprivate_non_profit    783.6*** (30.33)    867.5*** (64.44)
## public_privatepublic                211.6*** (28.45)    284.5*** (60.92)
## school_geographyrural            -2,139.0*** (33.99) -2,125.3*** (70.45)
## school_geographySuburb             -853.4*** (28.63)   -867.0*** (59.48)
## school_geographytown               -937.2*** (22.72)   -922.9*** (47.03)
## _________________________________ ____________________ ____________________
## S.E. type                        Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                   601,333              140,265
## R2                                            0.49137              0.49109
## Adj. R2                                       0.49136              0.49107
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

categorical variables note: public_private = Both categories are relative to Private For-Profit University school_geography = all 3 categories are relative to Universities located in cities (aggregated 3 categories with case_when into city)

After Scorecard, the mean ‘median_earnings_10’ was $35,324.20 with a standard error of $59.50 and statistically significant at the .1% level. After September 2015,a one-unit increase in ‘standardized_index’ is associated with a -$151.90 decrease in alumni earnings (‘median_earnings_10), holding all else constant. The standard errors were 21.08 and statistically significant at the .1% level. For every one-unit increase in ’standardized_index’, median earnings for alumni from a ‘high_earning_school’ increased $16,230.40 after September 2015. This was statically significant at the .1% level. For every one-unit increase in ‘standardized_index’ graduates from a private, non-profit school earned $687.20 more than graduates from a private for-profit university. This was statically significant at the .1% level. For every one-unit increase in ‘standardized_index’ graduates from a public university earned $144.70 more than graduates from a private for-profit university, standard errors were 59.58 and this was statically significant at the 5% level. Heteroskedasticity noticed in plot and histogram of standardized residuals looked relatively normal(“../figures/”).

m3_feols (BEFORE SCORECARD):

```
##                                 basemodel_preScor.. basemodel_postSco..
## Dependent Var.:                  median_earnings_10  median_earnings_10
##
## (Intercept)                     43,800.2*** (14.97) 43,716.3*** (31.31)
## standardized_index                 107.3*** (15.12)   -440.3*** (30.54)
## high_earning_schoolTRUE
## public_privateprivate_non_profit
## public_privatepublic
## school_geographyrural
## school_geographySuburb
## school_geographytown
## ________________________________ ____________________ ____________________
## S.E. type                                         IID                  IID
## Observations                                  601,333              140,265
## R2                                           8.36e-5              0.00148
## Adj. R2                                       8.2e-5              0.00147
##
##                                              m1_feols             m2_feols
## Dependent Var.:                  median_earnings_10  median_earnings_10
##
## (Intercept)                     35,438.3*** (27.55) 35,324.2*** (59.50)
## standardized_index                 36.22*** (10.97)   -151.9*** (21.08)
## high_earning_schoolTRUE         16,221.9*** (22.54) 16,230.4*** (46.72)
## public_privateprivate_non_profit   592.8*** (29.82)    687.2*** (63.50)
## public_privatepublic                 61.14* (27.71)      144.7* (59.58)
## school_geographyrural
## school_geographySuburb
## school_geographytown
## ________________________________ ____________________ ____________________
## S.E. type                       Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                  601,333              140,265
## R2                                           0.48909              0.48883
## Adj. R2                                      0.48908              0.48881
##
##                                              m3_feols             m4_feols
## Dependent Var.:                  median_earnings_10  median_earnings_10
##
## (Intercept)                     35,866.9*** (28.81) 35,761.8*** (62.14)
## standardized_index                  32.15** (10.94)   -135.4*** (21.08)
## high_earning_schoolTRUE         16,010.9*** (22.48) 16,025.7*** (46.61)
## public_privateprivate_non_profit   783.6*** (30.33)    867.5*** (64.44)
## public_privatepublic               211.6*** (28.45)    284.5*** (60.92)
## school_geographyrural           -2,139.0*** (33.99) -2,125.3*** (70.45)
## school_geographySuburb            -853.4*** (28.63)   -867.0*** (59.48)
## school_geographytown              -937.2*** (22.72)   -922.9*** (47.03)
## ________________________________ ____________________ ____________________
## S.E. type                       Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                  601,333              140,265
## R2                                           0.49137              0.49109
## Adj. R2                                      0.49136              0.49107
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

categorical variables note: public_private = Both categories are relative to Private For-Profit University school_geography = all 3 categories are relative to Universities located in cities (aggregated 3 categories with case_when into city)

Before Scorecard, the mean ‘median_earnings_10’ was $35,866.90 standard error of $28.81 and statistically significant at the .1% level. Previous to September 2015,a one-unit increase in ‘standardized_index’ is associated with a $32.15 increase in alumni earnings (‘median_earnings_10), holding all else constant. The standard errors were 10.94 and statistically significant at the .1% level. For every one-unit increase in ’standardized_index’, median earnings for alumni from a ‘high_earning_school’ increased $16,010.90 prior to September 2015. The standard errors were 22.48 and this was statically significant at the .1% level. For every one-unit increase in ‘standardized_index’ graduates from a private, non-profit school earned $783.60 more than graduates from a private, for-profit university. The standard error was 30.33 and this was statically significant at the .1% level. For every one-unit increase in ‘standardized_index’ graduates from a public university earned $211.60 more than graduates from a private for-profit university. The standard errors were 28.45 and this was statically significant at the 5% level. For everyone one -unit increase in ‘standardized_index’ universities located in rural areas earned -$2,139 less than graduates from universities in the city. The standard errors were 33.00 and statistically significant at the .1% level. For everyone one-unit increase in ‘standardized_index’ universities located in suburban areas earned -$853.40 less than graduates from universities in the city. The standard errors were 28.63 and statistically significant at the .1% level. For everyone-unit increase in ‘standardized_index’ universities located in towns earned -$937.20 less than graduates from universities in the city. The standard errors were 22.72 and statistically significant at the .1% level. F-stat was 1,520 additional explanatory power scaled by 601,435 degrees of freedom and p-value of 2.2e-16. Heteroskedasticity noticed in plot and histogram of standardized residuals looked relatively normal(“../figures/”). Heteroskedasticity noticed in plot and histogram of standardized residuals looked relatively normal(“../figures/”).

m4_feols:

```
##                                 basemodel_preScor.. basemodel_postSco..
## Dependent Var.:                  median_earnings_10  median_earnings_10
##
## (Intercept)                     43,800.2*** (14.97) 43,716.3*** (31.31)
## standardized_index                 107.3*** (15.12)   -440.3*** (30.54)
## high_earning_schoolTRUE
## public_privateprivate_non_profit
## public_privatepublic
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ___________________ ___________________
## S.E. type                                         IID                 IID
## Observations                                  601,333             140,265
## R2                                           8.36e-5             0.00148
## Adj. R2                                       8.2e-5             0.00147
##
##                                              m1_feols            m2_feols
## Dependent Var.:                  median_earnings_10  median_earnings_10
##
## (Intercept)                     35,438.3*** (27.55) 35,324.2*** (59.50)
## standardized_index                 36.22*** (10.97)   -151.9*** (21.08)
## high_earning_schoolTRUE         16,221.9*** (22.54) 16,230.4*** (46.72)
## public_privateprivate_non_profit   592.8*** (29.82)    687.2*** (63.50)
## public_privatepublic                 61.14* (27.71)      144.7* (59.58)
## school_geographyrural
## school_geographySuburb
## school_geographytown
## _________________________________ ___________________ ___________________
## S.E. type                       Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                  601,333             140,265
## R2                                            0.48909             0.48883
## Adj. R2                                       0.48908             0.48881
##
##                                              m3_feols            m4_feols
## Dependent Var.:                  median_earnings_10  median_earnings_10
##
## (Intercept)                     35,866.9*** (28.81) 35,761.8*** (62.14)
## standardized_index                  32.15** (10.94)   -135.4*** (21.08)
## high_earning_schoolTRUE         16,010.9*** (22.48) 16,025.7*** (46.61)
## public_privateprivate_non_profit   783.6*** (30.33)    867.5*** (64.44)
## public_privatepublic               211.6*** (28.45)    284.5*** (60.92)
## school_geographyrural           -2,139.0*** (33.99) -2,125.3*** (70.45)
## school_geographySuburb            -853.4*** (28.63)   -867.0*** (59.48)
## school_geographytown              -937.2*** (22.72)   -922.9*** (47.03)
## _________________________________ ___________________ ___________________
## S.E. type                       Heteroskedast.-rob. Heteroskedast.-rob.
## Observations                                  601,333             140,265
## R2                                            0.49137             0.49109
## Adj. R2                                       0.49136             0.49107
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

categorical variables note: public_private = Both categories are relative to Private For-Profit University school_geography = all 3 categories are relative to Universities located in cities (aggregated 3 categories with case_when into city)

After Scorecard, the mean 'median_earnings_10' was $35,761.80 standard error of $62.14 and statistically significant at the .1% level. After September 2015,a one-unit increase in 'standardized_index' is associated with a -$135.40 decrease in alumni earnings ('median_earnings_10), holding all else constant. The standard errors were 21.08 and statistically significant at the .1% level. For every one-unit increase in 'standardized_index', median earnings for alumni from a 'high_earning_school' increased $16,025.70 after September 2015. The standard errors were 46.61 and this was statically significant at the .1% level. For every one-unit increase in 'standardized_index' graduates from a private, non-profit school earned $867.50 more than graduates from a private, for-profit university. The standard error was 64.44 and this was statically significant at the .1% level. For every one-unit increase in 'standardized_index' graduates from a public university earned $284.50 more than graduates from a private for-profit university. The standard errors were 60.92 and this was statically significant at the 5% level. For everyone one -unit increase in 'standardized_index' universities located in rural areas earned -$2,125.30 less than graduates from universities in the city. The standard errors were 70.45 and statistically significant at the .1% level. For everyone one-unit increase in 'standardized_index' universities located in suburban areas earned -$867.00 less than graduates from universities in the city. The standard errors were 59.48 and statistically significant at the .1% level. For everyone-unit increase in 'standardized_index' universities located in towns earned -$922.90 less than graduates from universities in the city. The standard errors were 47.03 and statistically significant at the .1% level. Our F-test stat is 348.9 additional explanatory power scaled by 140,257 degrees of freedom and p-value of 2.2e-16. Heteroskedasticity noticed in plot and histogram of standardized residuals looked relatively normal("../figures/").

Conclustion & Recommendation(s):

There is little empirical evidence supporting the notion that the release of scorecard shifted student interest, as measured by Google trends keyword search, into colleges where median earnings are greater than $42,000 (a.k.a. 'high_earning_school'). It is evident in base model comparison with simple bivariate regression before and after Sept 2015; as well as in m3 and m4 models with controls. Future models would have incorporated median_debt after 10 years from alumni (mean = $24k w/sd 5,418) and SAT_average (mean = 1070 w/sd 132) as those are salient variables and interesting to add to WorkingData, but I think it would not have furthered in answering the current research question. Using College Scorecard can help answer student questions about descriptive statistics, but providing information on alumni salaries after 10 years didn't facilitate Google trends search history.