MSDS 6372: Unit 2 HW 2

One of the major advantages of LASSO regression is that the estimate of the regression coefficients (betas) are allowed to be biased whereas the OLS estimates are forced to be unbiased. This highlights the commonly referred to “variance / bias trade-off”. Since MSE = Variance + Bias2, it is easy to see that for the OLS unbiased estimates that the MSE(betas) = Variance(betas) since the Bias is zero. However, with the biased LASSO estimates, one is often able to reduce the variance of the estimate of the betasat the cost of introducing a little bias. Often the reduction in the variance is greater than the increase in the squared bias and we see a reduction in the MSE of the betas.

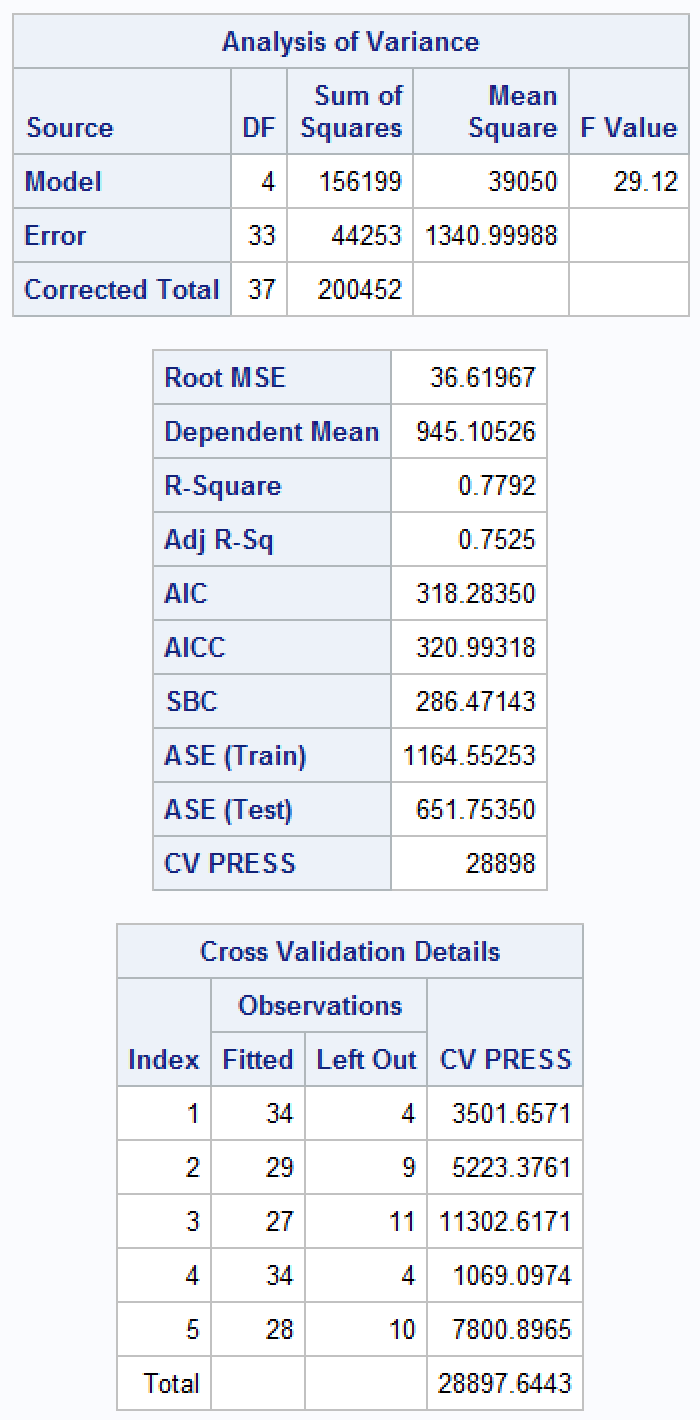
In short, this means that our estimate of the regression equation with the smaller MSE has greater probability of being close to the equation with the real betas: the real trend. This means that if we cross validate our model on a test set (that maintains the same trend as the training set but with different noise), the model with the smaller MSE has greater probability of capturing more of the true trend. Statistically, this will be reflected in statistics such as the ASE (Test) (Average Squared Error for the model trained on the training set and used to fit the test set) and R squared of the test set (basically any goodness of fit statistic that is **with respect to the model and the test set**.)

We will use the ASE (Test) statistic to provide evidence of preferred models for this data set: LASSO or OLS.

Attached HW 2 SAS Code has the data set from the paper you read as well as some code to divide the data set into a training and test set.

The assignment for this week is simple:

1. Run the code and make an argument / discussion as to which model / estimates (LASSO or OLS) will provide better predictability.   
   Make sure and copy and paste all relevant output to support your decision. Don’t overthink this. The answer can reference a single statistic.

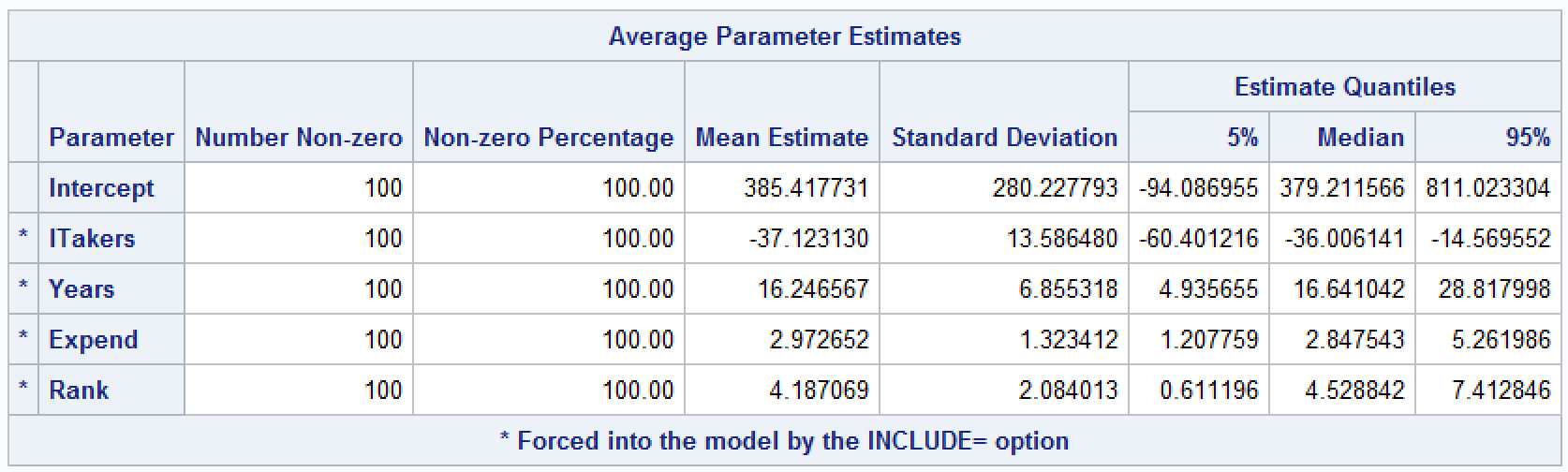


In this instance, the statistical output of the LASSO verifies that it is the preferred model since it clearly illustrates a lower RMSE, AIC, and AICC. The adjusted R² is also higher than the OLS model, which confirms the better fit.

← LASSO OLS →

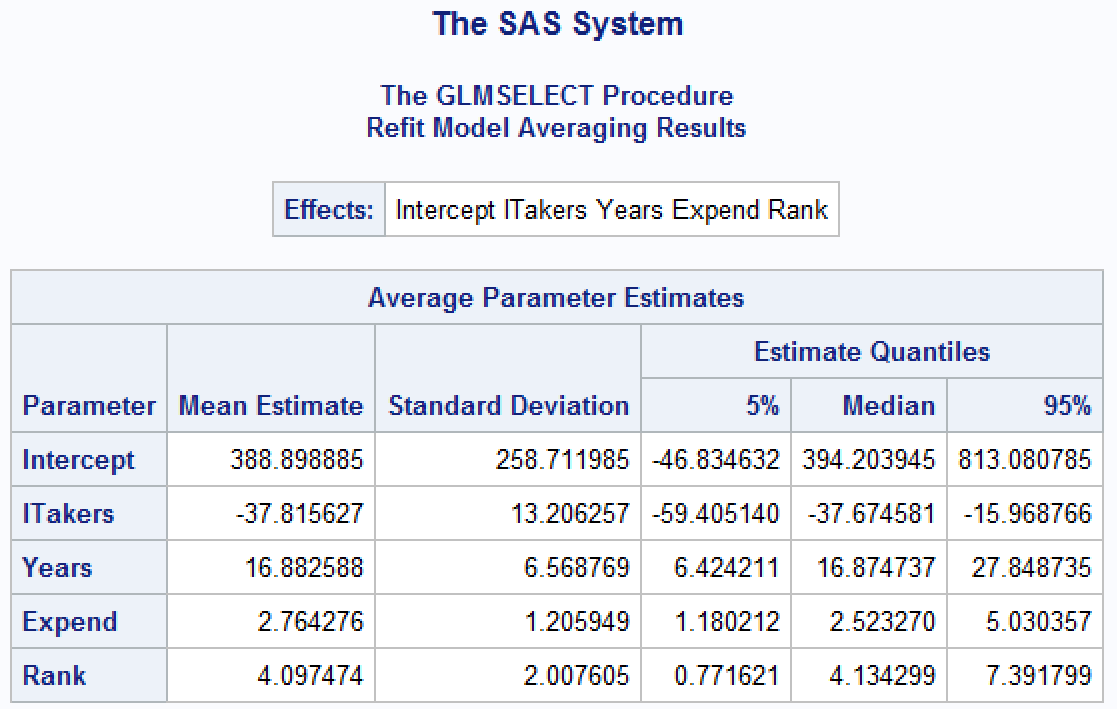
1. Provide confidence intervals for all estimates from both the LASSO and OLS models. Also report the margin of error for each interval and comment on which margins of error seem to be smaller. This should be fun! It uses a cutting age function and method that is related to machine learning topics (bootstrapping) called model averaging. Read the SAS documentation about it … it is a very strait forward yet powerful idea.

LASSO results & code





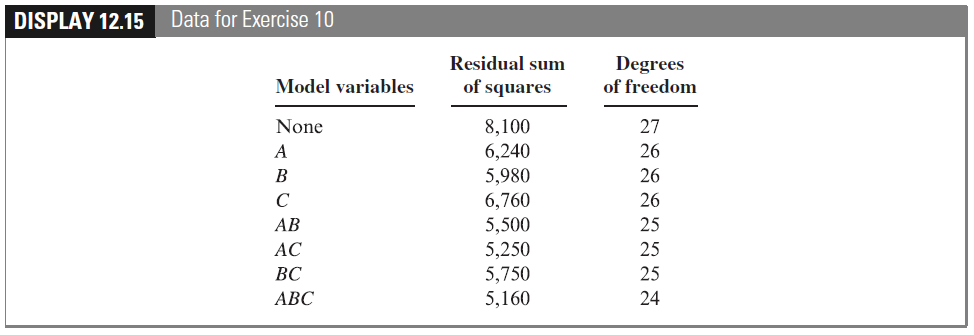
OLS results & code



1. Chapter 12.10 and 12.11

**10.** A, B, and C are three explanatory variables in a multiple linear regression with n = 28 cases.

Display 12.15 shows the residual sums of squares and degrees of freedom for all models.



(a) Calculate the estimate of σ² for each model. (b) Calculate the adjusted R² for each model.

(c) Calculate the Cp statistic for each model. (d) Calculate the BIC for each model. (e) Which model

has (i) the smallest estimate of σ²? (ii) the largest adjusted R²? (iii) the smallest Cp statistic? (iv) the

smallest BIC?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Models | *p* | RSS | df | σ² | R² | Adj R² | Cp | BIC | F-stat |
| None | 1 | 8100 | 27 | 300 | 0 | 0 | 11.674 | 165.352 |  |
| A | 2 | 6240 | 26 | 240 | 0.200 | 0.136 | 5.023 | 161.379 | 7.75 |
| B | 2 | 5980 | 26 | 230 | 0.233 | 0.172 | 3.814 | 160.188 | 9.22 |
| C | 2 | 6760 | 26 | 260 | 0.133 | 0.064 | 7.442 | 163.621 | 5.15 |
| AB | 3 | 5500 | 25 | 220 | 0.267 | 0.175 | 3.581 | 161.177 | 3.36 |
| **AC** | **3** | **5250** | **25** | **210** | **0.300** | **0.213** | **2.419** | **159.875** |  |
| BC | 3 | 5750 | 25 | 230 | 0.233 | 0.138 | 4.744 | 162.422 | 2.13 |
| ABC | 4 | 5160 | 24 | 215 | 0.283 | 0.159 | 4.000 | 162.723 |  |

In this case, the AC model would be ideal since it produced the highest adjusted R2 and the lowest BIC & Cp.

**11.** Using the residual sums of squares from Exercise 10, find the model indicated by forward selection. (Start with the model “None,” and identify the single-variable model that has the smallest residual sum of squares. Then perform an extra-sum-of-squares F -test to see whether that variable is significant. If it is, find the two-variable model that includes the first term and has the smallest residual sum of squares. Then perform an extra-sum-of-squares F -test to see whether the additional variable is significant. Continue until no F -statistics greater than 4 remain for inclusion of another variable.)

The B model has the lowest residual sum of squares (5980) and largest F-statistic (9.22). The F statistic is higher than 4, which would indicate it is significant. So we proceed to a two-variable model using the same extra sum of squares F-test. Since we know B is significant, we focus on AB since it contains B in it. When we look at the models with B in them, AB has the highest F statistic. Yet, it is under 4 which would indicate it is not significant. As a result, we would side with model B.