Cross Validation and Subset Selection

Maureen Renaud

6/9/2020

INTRODUCTION

The purpose of this study was to compare various alternative model selection methods with simulated, random data. According to "An Introduction to Statistical Learning"[1], using alternatives to least squares fitting can result in better prediction accuracy and model interpretability. In this study, I focused on the Best Subset Selection Method, Forward Stepwise Selection Method, and Backwards Stepwise Selection Method. Methods were compared based on C_p, Bayesian information criterion (BIC), and Adjusted R^2.

In this particular study, the three-variable model had the lowest BIC for all three selection methods. The four-variable model had the lowest C_p and highest Adjusted R^2 for all all three selection methods.

The best three variable model for all three selection methods consisted of x^1, x^2, and x^3. The best four variable model for the Best Subset Selection Method and the Forward Stepwise Selection Method consisted of x^1, x^2, x^3, and x^5. The best four variable model for the Backward Stepwise Selection Method consisted of x^1, x^2, x^3, and x^9.

I selected the three-variable model as the best model overall due to its relative simplicity compared to the four variable model and its adherence to the hierarchy principle.

This study highlighted the fact that the various methods often lead us to the same results. However, ultimately we have to use our best judgement, expertise, and subject matter knowledge to choose the most appropriate model.

DATA

The data for this study was generated first by using the rnorm() function to generate a predictor X of length n=100 and a noise vector of length n=100.

Next, the response variable y was created according to the model below, where coefficients were assigned by my choice.

```
y <- 4 + 2*x + -3*x^2 + 5*x^3 + Noise
```

x and y were then combined into a data frame, summarized below:

```
##        x                 y
##  Min.   :-2.2147   Min.   :-70.109
##  1st Qu.:-0.4942   1st Qu.:  1.384
##  Median : 0.1139   Median :  4.186
##  Mean   : 0.1089   Mean   :  2.801
##  3rd Qu.: 0.6915   3rd Qu.:  6.177
##  Max.   : 2.4016   Max.   : 61.185
```

This is random, simulated data. So there are no obvious outliers or problematic values.

ANALYSES

This particular study focuses on best subset selection, forward stepwise selection, and backwards stepwise selection.

BEST SUBSET SELECTION

According to "An Introduction to Statistical Learning"[1], the best subset selection fits a least squares regression for all combinations of predictors. It then picks the best model at each level of predictors (the best one predictor model, the best two predictor model, etc.) based on the Residual Sum of Squares (RSS). It is then up to the analyst to choose the final model.

This is an excellent method to use, because it compares all possible models. However, it is more complex and does require the analyst to do some work comparing the models provided by the method.

The full printout for the best subset selection is provided in Appendix A.

A table of the values for BIC, C_p, and Adjusted R^2 is given in Table 1. Looking at this table and a plot of the values given in Figure 1, the lowest BIC is in the three-variable model, the lowest C_p is in the four-variable model, and the highest adjusted R^2 is in the four-variable model.

The three variable model contains x^1, x^2, and x^3 The four variable model contains x^1, x^2, x^3, and x^5

Based on this, I select the three-variable model as the best model. It has the lowest BIC and while it doesn't have the lowest C_p or highest adjusted R^2, it is adequately close to the four-variable method. Because of that, it makes sense to go with the simpler three-variable model.

Additionally, the four-variable model selected would actually be an unwise choice. It contains x^1, x^2, x^3, and x^5 instead of x^4. In linear regression, models must adhere to the hierarchy principle. This states that if you include x^n in your model, you must include x^1 through $x\hat{}(n - 1)\hat{}$.

The coefficients for this model are:

```
## (Intercept)           x      I(x^2)      I(x^3)
##    4.061507    1.975280   -3.123791    5.017639
```

FORWARD STEPWISE SELECTION

Now we turn to forward stepwise selection. According to "An Introduction to Statistical Learning"[1], This method begins with an empty model and adds predictors one by one. At each state, the predictor resulting in the model with the lowest RSS is chosen.

A benefit of this method is that it is much less complex than best subset selection since not all of the models are being compared. However, because variables are added one at a time, once a variable has been added it cannot be taken away. So, a variable added in the second step must remain in the seven variable model, even if it is no longer a worthwhile addition.

A full print out of the forward stepwise selection is included in Appendix A.

A table of the values for BIC, C_p, and Adjusted R^2 is given in Table 2. Looking at this table and a plot of the values given in Figure 2, we see that the lowest BIC is in the three-variable model, the lowest C_p is in the four-variable model, and the highest adjusted R^2 is in the four-variable model.

The three-variable model contains x^1, x^2, and x^3. The four-variable model contains x^1, x^2, x^3, and x^5.

Based on this, I would once again select the three-variable model as the best model. It has the lowest BIC and while it doesn't have the lowest C_p or highest adjusted R^2, it is close enough to the four-variable method that there isn't a significant difference. There is not much additional benefit to including the fourth variable. By sticking with the three-variable model, we end up with a simpler model and one that adheres to the hierarchy principle.

The coefficients for this model are:

```
## (Intercept)           x       I(x^2)      I(x^3)
##    4.061507    1.975280    -3.123791    5.017639
```

BACKWARD STEPWISE SELECTION

Next, we take a look at backward stepwise selection. According to "An Introduction to Statistical Learning"[1], this method starts with a full model and then removes one predictor at a time until the model has no predictors. Like forward stepwise selection, it is much less complex than best subset selection. However, once a variable has been taken away, it cannot be added back into the model. So, a variable that was removed in the first step cannot be added back in the fourth step, even if it would now be a logical addition.

The full print out of the backward stepwise selection is included in Appendix A.

A table of the values for BIC, C_p, and Adjusted R^2 is given in Table 3. Looking at this table and a plot of the values given in Figure 3, we see that the lowest BIC is in the three-variable model, the lowest C_p is in the four-variable model, and the highest adjusted R^2 is in the four-variable model.

The three-variable model contains x^1, x^2, and x^3. The four-variable model contains x^1, x^2, x^3, and x^9.

Again I would select the three-variable model as the best model using backward stepwise selection. It has the lowest BIC and while it doesn't have the lowest C_p or highest adjusted R^2, it is close enough to the four-variable method that there isn't a significant difference.

There is not much additional benefit to including the fourth variable. By sticking with the three-variable model, we end up with a simpler model and one that adheres to the hierarchy principle. In this four-variable model, it jumps all the way to up to include x^9. So, to keep the model simple and accurate, I will select the three-variable model.

The coefficients for this model are:

```
## (Intercept)           x       I(x^2)      I(x^3)
##    4.061507    1.975280    -3.123791    5.017639
```

COMPARISONS

With all three models, using BIC as the primary criterion led to us choosing the three-variable model, while using C_p or Adjusted R^2 as the primary criterion led to us choosing the four variable model.

For all three selection methods, the three variable model consisted of x^1, x^2, and x^3.

For the best subset method and the forward stepwise selection method, the best four-variable model consisted of x^1, x^2, x^3, and x^5. Using the backward selection method, the best four-variable model consisted of x^1, x^2, x^3, and x^9.

With all three methods, I chose the three-variable model as the best model to maintain simplicity and remain in accordance with the hierarchy principle.

PLOTS AND TABLES

Table 1: Evaluation Metrics Using Best Subset Selection

#	BIC	Cp	Adj. R^2^
1	-256.5225	1459.0302677	0.9291490
2	-455.3615	109.3256041	0.9906405
3	-527.7101	2.1859433	0.9956193
4	-526.9792	0.6067483	0.9957414
5	-522.8481	2.1782005	0.9957164
6	-518.4456	3.9955812	0.9956791
7	-514.0726	5.7869063	0.9956423
8	-510.1575	7.1694092	0.9956247
9	-505.5701	9.1535580	0.9955769
10	-501.1374	11.0000000	0.9955349

Figure 1: Best Subset Selection
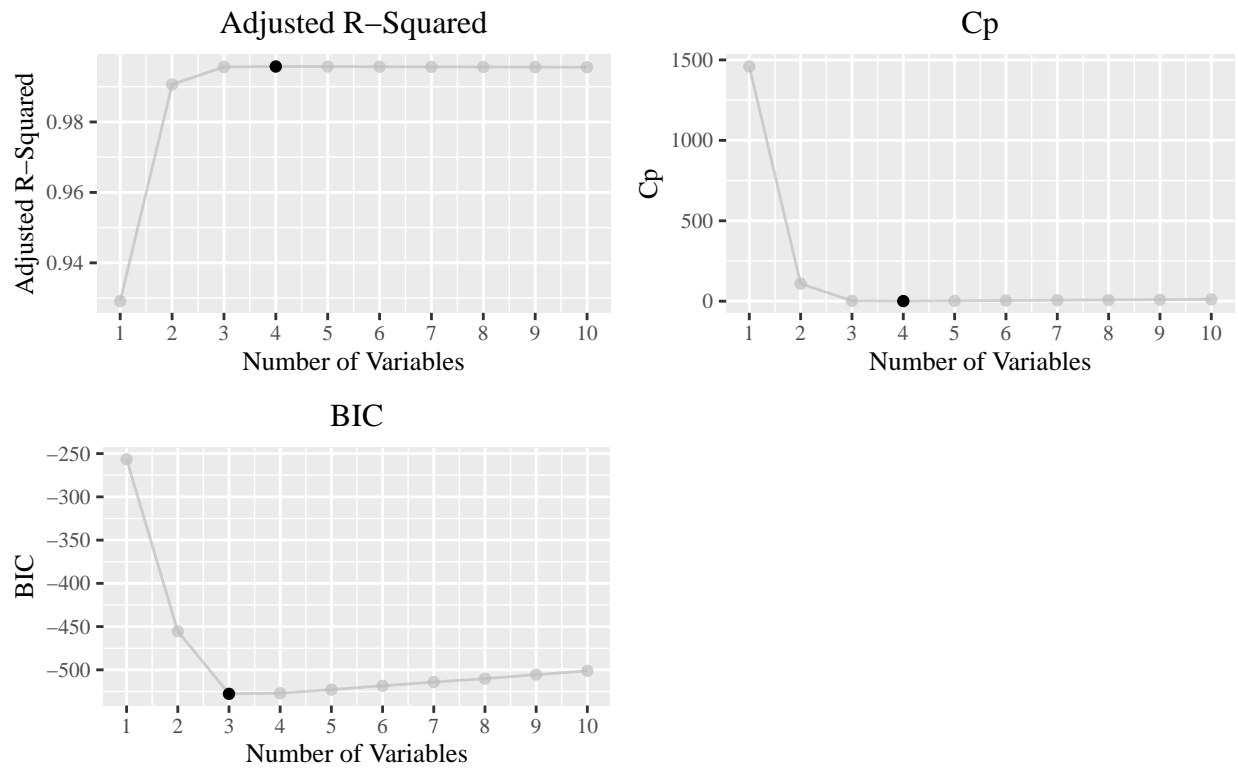
Table 2: Evaluation Metrics Using Forward Stepwise Selection

#	BIC	Cp	Adjusted R^2
1	-256.5225	1459.0302677	0.9291490
2	-455.3615	109.3256041	0.9906405
3	-527.7101	2.1859433	0.9956193
4	-526.9792	0.6067483	0.9957414
5	-522.8481	2.1782005	0.9957164
6	-518.3717	4.0621068	0.9956759
7	-513.8172	6.0165107	0.9956311
8	-509.2768	7.9582570	0.9955860
9	-504.8652	9.7842795	0.9955456
10	-501.1374	11.0000000	0.9955349

Figure 2: Forward Stepwise Selection
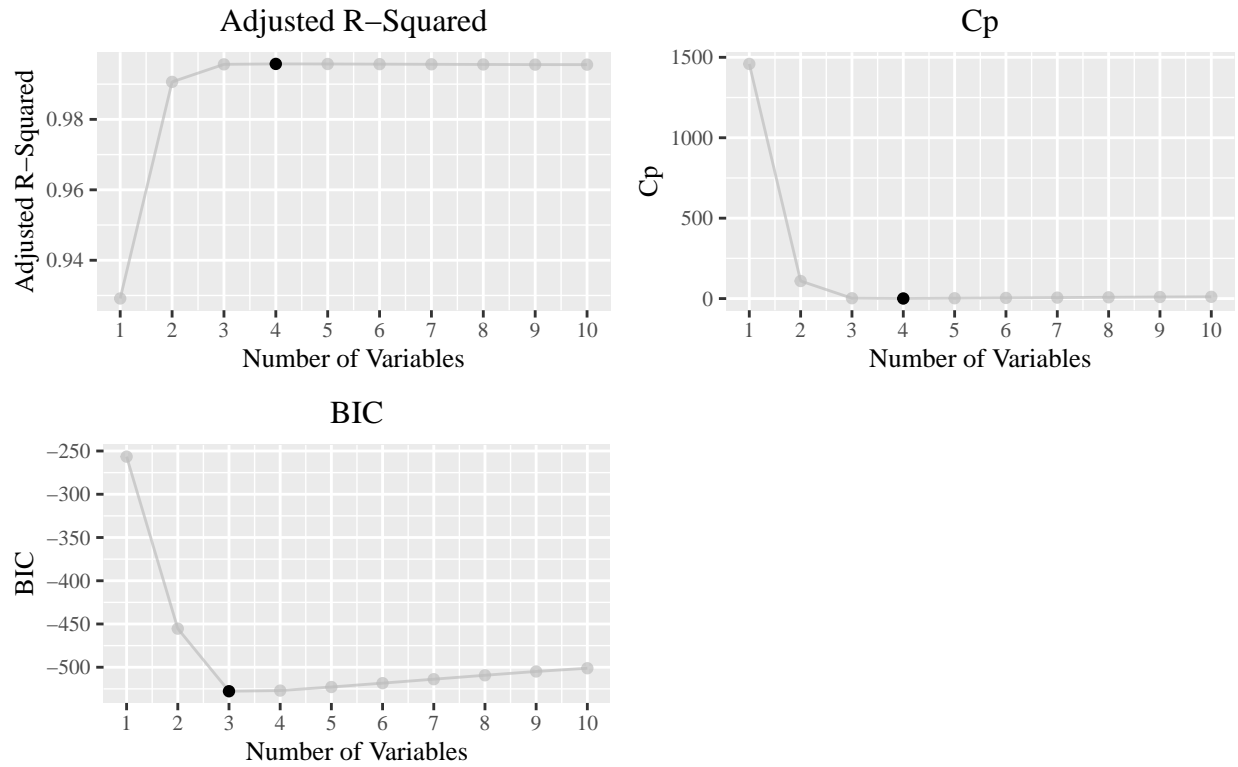
Table 3: Evaluation Metrics Using Backward Stepwise Selection

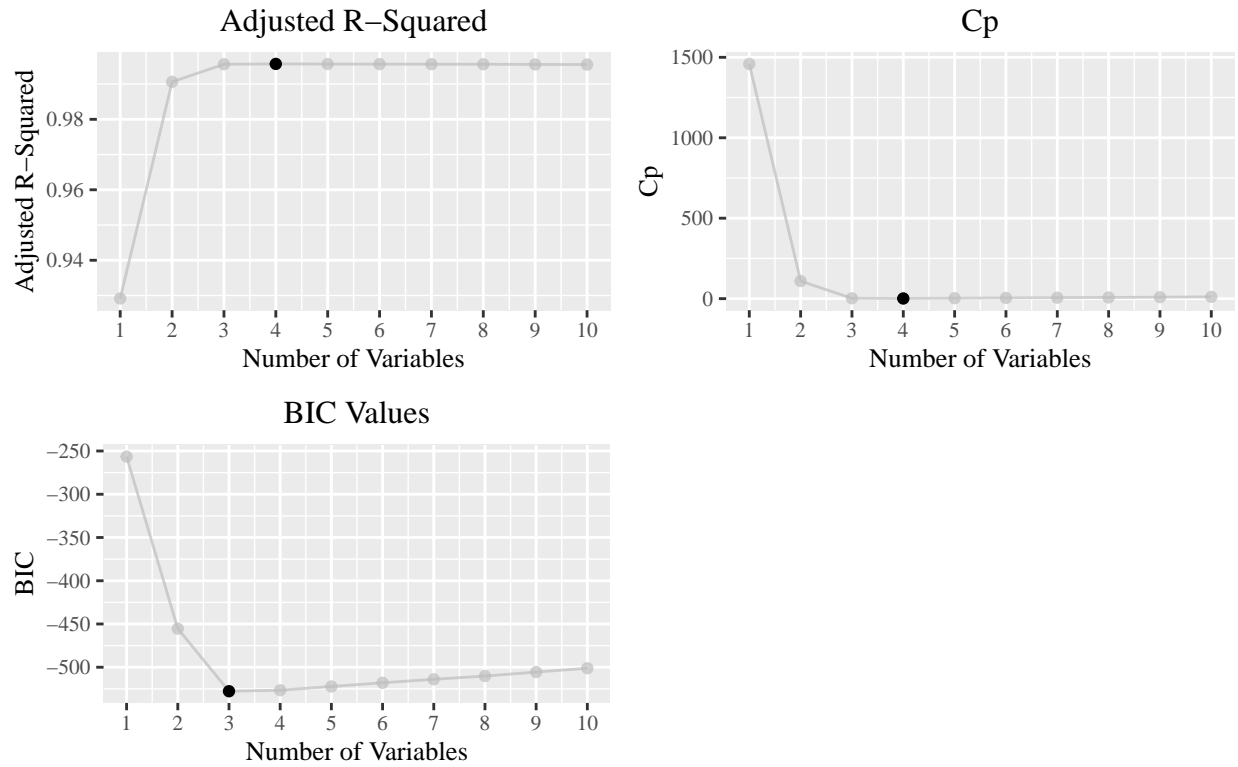| #	BIC	Cp	Adjusted R^2
1	-256.5225	1459.0302677	0.9291490
2	-455.3615	109.3256041	0.9906405
3	-527.7101	2.1859433	0.9956193
4	-526.5671	0.9808795	0.9957238
5	-522.2369	2.7310150	0.9956902
6	-517.9772	4.4181003	0.9956588
7	-513.9189	5.9250326	0.9956356
8	-510.1575	7.1694092	0.9956247
9	-505.5701	9.1535580	0.9955769
10	-501.1374	11.0000000	0.9955349

Figure 3: Backward Stepwise Selection

Figure 3

CONCLUSION

This study compared alternative model selection methods. Here, we saw that the-three variable model had the lowest BIC for all three selection methods. The four-variable model had the lowest C_p and highest Adjusted R^2 for all all three selection methods.

Despite the four-variable model being better based on two evaluations metrics, I selected the three-variable model as the best model overall. When you consider the fact that the three-variable model is simpler and adheres to the hierarchy principle, I feel that it is the clear choice for best model

This study highlights the fact that the various methods often lead us to the similar results. However, numbers alone cannot guide all of our decisions. We must use that data combined with our best judgement, expertise, and subject matter knowledge to choose the most appropriate model.

SOURCES

1) James, G et al. (2017). *An Introduction To Statistical Learning.* New York, NY.

APPENDIX A

Best subset selection

```
## Subset selection object
## Call: regsubsets.formula(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) +
##     I(x^6) + I(x^7) + I(x^8) + I(x^9) + I(x^10), data = my.data,
##     nvmax = 10)
## 10 Variables  (and intercept)
##          Forced in Forced out
## x            FALSE      FALSE
## I(x^2)       FALSE      FALSE
## I(x^3)       FALSE      FALSE
## I(x^4)       FALSE      FALSE
## I(x^5)       FALSE      FALSE
## I(x^6)       FALSE      FALSE
## I(x^7)       FALSE      FALSE
## I(x^8)       FALSE      FALSE
## I(x^9)       FALSE      FALSE
## I(x^10)      FALSE      FALSE
## 1 subsets of each size up to 10
## Selection Algorithm: exhaustive
##          x   I(x^2) I(x^3) I(x^4) I(x^5) I(x^6) I(x^7) I(x^8) I(x^9) I(x^10)
## 1  ( 1 ) " " " "    "*"    " "    " "    " "    " "    " "    " "    " "
## 2  ( 1 ) " " "*"    "*"    " "    " "    " "    " "    " "    " "    " "
## 3  ( 1 ) "*" "*"    "*"    " "    " "    " "    " "    " "    " "    " "
## 4  ( 1 ) "*" "*"    "*"    " "    "*"    " "    " "    " "    " "    " "
## 5  ( 1 ) "*" "*"    "*"    " "    "*"    "*"    " "    " "    " "    " "
## 6  ( 1 ) "*" "*"    "*"    " "    " "    " "    "*"    "*"    "*"    " "
## 7  ( 1 ) "*" "*"    "*"    " "    "*"    "*"    " "    "*"    " "    "*"
## 8  ( 1 ) "*" "*"    "*"    "*"    " "    "*"    " "    "*"    "*"    "*"
## 9  ( 1 ) "*" "*"    "*"    "*"    "*"    "*"    " "    "*"    "*"    "*"
## 10  ( 1 ) "*" "*"    "*"    "*"    "*"    "*"    "*"    "*"    "*"    "*"
```

Forward stepwise selection

```
## Subset selection object
## Call: regsubsets.formula(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) +
##     I(x^6) + I(x^7) + I(x^8) + I(x^9) + I(x^10), data = my.data,
##     nvmax = 10, method = "forward")
## 10 Variables  (and intercept)
##          Forced in Forced out
## x            FALSE      FALSE
## I(x^2)       FALSE      FALSE
```

```
## I(x^3)        FALSE       FALSE
## I(x^4)        FALSE       FALSE
## I(x^5)        FALSE       FALSE
## I(x^6)        FALSE       FALSE
## I(x^7)        FALSE       FALSE
## I(x^8)        FALSE       FALSE
## I(x^9)        FALSE       FALSE
## I(x^10)       FALSE       FALSE
## 1 subsets of each size up to 10
## Selection Algorithm: forward
##           x   I(x^2) I(x^3) I(x^4) I(x^5) I(x^6) I(x^7) I(x^8) I(x^9) I(x^10)
## 1  ( 1 )  " " " "   "*"    " "    " "    " "    " "    " "    " "    " "
## 2  ( 1 )  " " "*"   "*"    " "    " "    " "    " "    " "    " "    " "
## 3  ( 1 )  "*" "*"   "*"    " "    " "    " "    " "    " "    " "    " "
## 4  ( 1 )  "*" "*"   "*"    " "    "*"    " "    " "    " "    " "    " "
## 5  ( 1 )  "*" "*"   "*"    " "    "*"    "*"    " "    " "    " "    " "
## 6  ( 1 )  "*" "*"   "*"    " "    "*"    "*"    " "    " "    "*"    " "
## 7  ( 1 )  "*" "*"   "*"    " "    "*"    "*"    "*"    " "    "*"    " "
## 8  ( 1 )  "*" "*"   "*"    " "    "*"    "*"    "*"    "*"    "*"    " "
## 9  ( 1 )  "*" "*"   "*"    " "    "*"    "*"    "*"    "*"    "*"    "*"
## 10 ( 1 )  "*" "*"   "*"    "*"    "*"    "*"    "*"    "*"    "*"    "*"
```

Backward stepwise selection

```
## Subset selection object
## Call: regsubsets.formula(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) +
##     I(x^6) + I(x^7) + I(x^8) + I(x^9) + I(x^10), data = my.data,
##     nvmax = 10, method = "backward")
## 10 Variables  (and intercept)
##            Forced in Forced out
## x              FALSE       FALSE
## I(x^2)         FALSE       FALSE
## I(x^3)         FALSE       FALSE
## I(x^4)         FALSE       FALSE
## I(x^5)         FALSE       FALSE
## I(x^6)         FALSE       FALSE
## I(x^7)         FALSE       FALSE
## I(x^8)         FALSE       FALSE
## I(x^9)         FALSE       FALSE
## I(x^10)        FALSE       FALSE
## 1 subsets of each size up to 10
## Selection Algorithm: backward
##           x   I(x^2) I(x^3) I(x^4) I(x^5) I(x^6) I(x^7) I(x^8) I(x^9) I(x^10)
## 1  ( 1 )  " " " "   "*"    " "    " "    " "    " "    " "    " "    " "
## 2  ( 1 )  " " "*"   "*"    " "    " "    " "    " "    " "    " "    " "
## 3  ( 1 )  "*" "*"   "*"    " "    " "    " "    " "    " "    " "    " "
## 4  ( 1 )  "*" "*"   "*"    " "    " "    " "    " "    " "    "*"    " "
## 5  ( 1 )  "*" "*"   "*"    " "    " "    " "    " "    "*"    "*"    " "
## 6  ( 1 )  "*" "*"   "*"    " "    " "    " "    " "    "*"    "*"    "*"
## 7  ( 1 )  "*" "*"   "*"    " "    " "    "*"    " "    "*"    "*"    "*"
## 8  ( 1 )  "*" "*"   "*"    "*"    " "    "*"    " "    "*"    "*"    "*"
## 9  ( 1 )  "*" "*"   "*"    "*"    "*"    "*"    " "    "*"    "*"    "*"
## 10 ( 1 )  "*" "*"   "*"    "*"    "*"    "*"    "*"    "*"    "*"    "*"
```

APPENDIX B

All R code for the above project is included below.

Setting x and Noise:

```r
set.seed(1)
x <- rnorm(100)
Noise <- rnorm(100)
```

Creating Y

```r
y <- 4 + 2*x + -3*x^2 + 5*x^3 + Noise
```

Creating and summarizing the data frame:

```r
my.data <- data.frame(x, y)
summary(my.data)
```

Using the Best Subset Method:

```r
library (leaps)
regfit.full <- regsubsets(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6) + I(x^7)
                          + I(x^8) + I(x^9) + I(x^10), data = my.data, nvmax = 10)
reg.summary <- summary(regfit.full)
```

Showing the coefficients for the three variable model found using the Best Subset Selection Method:

```r
coef(regfit.full, which.min(reg.summary$bic))
```

Using the Forward Stepwise Selection Method:

```r
regfit.fwd= regsubsets(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6) + I(x^7)
                       + I(x^8) + I(x^9) + I(x^10), data = my.data, nvmax = 10, method = "forward")
forward.summary <- summary(regfit.fwd )
```

Showing the coefficients for the three variable model found using the Forward Stepwise Selection Method:

```r
coef(regfit.fwd, which.min(forward.summary$bic))
```

Using the Backward Stepwise Selection Method:

```r
regfit.bwd= regsubsets(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6) + I(x^7)
                       + I(x^8) + I(x^9) + I(x^10), data = my.data, nvmax = 10, method = "backward")
backward.summary <- summary(regfit.bwd)
```

Showing the coefficients for the three variable model found using the Backward Stepwise Selection Method:

```r
coef(regfit.bwd, which.min(backward.summary$bic))
```

Creating the table of BIC, C_p, and Adjusted R^2 Values for the Best Subset Method:

```r
regfitvalues <- matrix(cbind(1:10, reg.summary$bic, reg.summary$cp, reg.summary$adjr2),
                        ncol=4,byrow=FALSE)
colnames(regfitvalues) <- c("#", "BIC","C~p~","Adj. R^2^")
rownames(regfitvalues) <- c(1:10)
regfitvalues <- as.table(regfitvalues)
```

Creating the table of BIC, C_p, and Adjusted R^2 Values for the Best Subset Method:

```r
library(knitr)
kable(regfitvalues, caption = "Evaluation Metrics Using Best Subset Selection", align = 'c')
```

Creating the plots of BIC, C_p, and Adjusted R^2 values for the Best Subset Method:

```r
subset.data <- data.frame(Var = 1:10, BIC = reg.summary$bic, Cp = reg.summary$cp,
                          AdjR = reg.summary$adjr2)
library(ggplot2)
library(gghighlight)
subset.adjr.plot <- ggplot(subset.data, aes(x = Var, y = AdjR))
one <- subset.adjr.plot + geom_point() + geom_line()
                        + labs(title="Adjusted R-Squared", x ="Number of Variables",
                               y = "Adjusted R-Squared")
                        +  theme(plot.title = element_text(hjust = 0.5),
                                 text=element_text(size=10,  family="serif"))
                        + gghighlight(AdjR > 0.99574) + scale_x_continuous(breaks = 1:10)


subset.cp.plot <- ggplot(subset.data, aes(x = Var, y = Cp))
two <- subset.cp.plot + geom_point() + geom_line()  + labs(title="Cp",
       x ="Number of Variables", y = "Cp") + theme(plot.title = element_text(hjust = 0.5),
                                                   text=element_text(size=10,  family="serif"))
                                          + gghighlight(Cp < 1)
                                          + scale_x_continuous(breaks = 1:10)

subset.bic.plot <- ggplot(subset.data, aes(x = Var, y = BIC))
three <- subset.bic.plot + geom_point() + geom_line()  + labs(title="BIC",
       x ="Number of Variables", y = "BIC") +  theme(plot.title = element_text(hjust = 0.5),
                                                     text=element_text(size=10,  family="serif"))
       + gghighlight(BIC < -527) + scale_x_continuous(breaks = 1:10)

library(cowplot)
subset.plot.row <- plot_grid(one, two, three, align = 'h')
title <- ggdraw() + draw_label("Figure 1: Best Subset Selection",
    x = 0,
    hjust = 0, , fontfamily = "serif", size = 10)
    + theme(plot.margin = margin(0, 0, 0, 7))
plot_grid(title, subset.plot.row, ncol = 1, rel_heights = c(0.1, 1))
```

Creating the table for BIC, C_p, and Adjusted R^2 for the Forward Stepwise Method:

```r
forward.values <- matrix(cbind(1:10, forward.summary$bic, forward.summary$cp,
                          forward.summary$adjr2), ncol=4,byrow=FALSE)
colnames(forward.values) <- c("#","BIC","Cp","Adjusted R^2")
```

```
rownames(forward.values) <- c(1:10)
forward.values <- as.table(forward.values)
```

Creating the table for BIC, C_p, and Adjusted R^2 for the Forward Stepwise Method:

```
kable(forward.values, caption = "Evaluation Metrics Using Forward Stepwise Selection", align = 'c')
```

Creating the plots for the Forward Stepwise Method:

```
forward.data <- data.frame(Var = 1:10, BIC = forward.summary$bic,
                           Cp = forward.summary$cp, AdjR = forward.summary$adjr2)
forward.adjr.plot <- ggplot(forward.data, aes(x = Var, y = AdjR))
fone <- forward.adjr.plot + geom_point() + geom_line()
        + labs(title="Adjusted R-Squared",
               x ="Number of Variables", y = "Adjusted R-Squared")
        +  theme(plot.title = element_text(hjust = 0.5)) + gghighlight(AdjR > 0.99574)
        + scale_x_continuous(breaks = 1:10)

forward.cp.plot <- ggplot(forward.data, aes(x = Var, y = Cp))
ftwo <- forward.cp.plot + geom_point() + geom_line()
        + labs(title="Cp Values",
        x ="Number of Variables", y = "Cp") + theme(plot.title = element_text(hjust = 0.5))
        + gghighlight(Cp < 1) + scale_x_continuous(breaks = 1:10)

forward.bic.plot <- ggplot(forward.data, aes(x = Var, y = BIC))
fthree <- forward.bic.plot + geom_point() + geom_line()
          + labs(title="BIC Value",
        x ="Number of Variables", y = "BIC") +  theme(plot.title = element_text(hjust = 0.5))
        + gghighlight(BIC < -527) + scale_x_continuous(breaks = 1:10)

library(cowplot)
forward.plot.row <- plot_grid(fone, ftwo, fthree, align = 'h')

title <- ggdraw() + draw_label("Figure 2: Forward Stepwise Selection", x = 0, hjust = 0,
                               fontfamily = "serif", size = 10) +
  theme(plot.margin = margin(0, 0, 0, 7)
plot_grid(title, forward.plot.row, ncol = 1, rel_heights = c(0.1, 1))
```

Creating the table for the Backward Stepwise Selection Method:

```
backward.values <- matrix(cbind(1:10, backward.summary$bic, backward.summary$cp,
                                backward.summary$adjr2), ncol=4,byrow=FALSE)
colnames(backward.values) <- c("#", "BIC","Cp","Adjusted R^2")
rownames(backward.values) <- c(1:10)
backward.values <- as.table(backward.values)
```

Creating the table for the Backward Stepwise Selection Method:

```
kable(backward.values, caption = "Evaluation Metrics Using Backward Stepwise Selection", align = 'c')
```

Creating the plots for the Backward Stepwise Selection Method:

```r
backward.data <- data.frame(Var = 1:10, BIC = backward.summary$bic, Cp = backward.summary$cp,
                            AdjR = backward.summary$adjr2)
backward.adjr.plot <- ggplot(backward.data, aes(x = Var, y = AdjR))
bone <-backward.adjr.plot + geom_point() + geom_line()
  + labs(title="Adjusted R-Squared",
         x ="Number of Variables", y = "Adjusted R-Squared")
  +  theme(plot.title = element_text(hjust = 0.5))
  + gghighlight(AdjR > 0.99572) + scale_x_continuous(breaks = 1:10)

backward.cp.plot <- ggplot(backward.data, aes(x = Var, y = Cp))
btwo <- backward.cp.plot + geom_point() + geom_line()
  + labs(title="Cp Values",
         x ="Number of Variables", y = "Cp")
  + theme(plot.title = element_text(hjust = 0.5))
  + gghighlight(Cp < 1) + scale_x_continuous(breaks = 1:10)

backward.bic.plot <- ggplot(backward.data, aes(x = Var, y = BIC))
bthree <- backward.bic.plot + geom_point() + geom_line()
  + labs(title="BIC Values",
         x ="Number of Variables", y = "BIC") +  theme(plot.title = element_text(hjust = 0.5))
  + gghighlight(BIC < -527) + scale_x_continuous(breaks = 1:10)

library(cowplot)
backward.plot.row <- plot_grid(bone, btwo, bthree, align = 'h')

title <- ggdraw() +
  draw_label("Figure 3: Backward Stepwise Selection",
    x = 0, hjust = 0, , fontfamily = "serif", size = 10)
  + theme(plot.margin = margin(0, 0, 0, 7))
plot_grid(title, backward.plot.row, ncol = 1, rel_heights = c(0.1, 1))
```