OM337: Supply Chain Analytics

Individual Assignment #3

**Model Selection:**

**A Dual Elasticity Demand Model with Competition**

(40 points)

(Due by Sep. 25, 11:59pm)

This homework assignment is based on the “Soft Drinks.csv” data file we used in class last week. This time we are interested in obtaining a demand model that accounts for the actions of a competitor. We assume we are the firm who owns “Product X” and Product Y competes for the same demand as our product. Hence pricing and promotion decisions regarding Product Y will affect the demand for Product X.

We are interested in fitting a model of the form:

To fit the parameters of this model we take the ln (log(…) in R) transformation on both sides of this equation to obtain:

Before proceeding any further, let us examine the first five rows and the variables included in the “Soft Drinks.csv” data set.

**STORE WEEK OUNCES deal\_X deal\_Y oz\_X oz\_Y pX pY**

1 1 1 72 1 0 13032 2088 0.025 0.028

2 1 2 72 0 0 3384 3960 0.028 0.030

3 1 3 72 0 0 2088 19296 0.030 0.028

4 1 4 72 1 0 35064 576 0.021 0.030

5 1 5 72 1 0 26640 3312 0.022 0.028

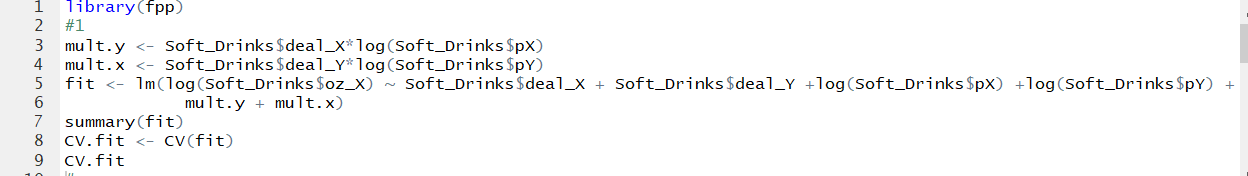
The first three variables (STORE, WEEK and OUNCES) are not used in the model above. Then we can identify the following correspondences:

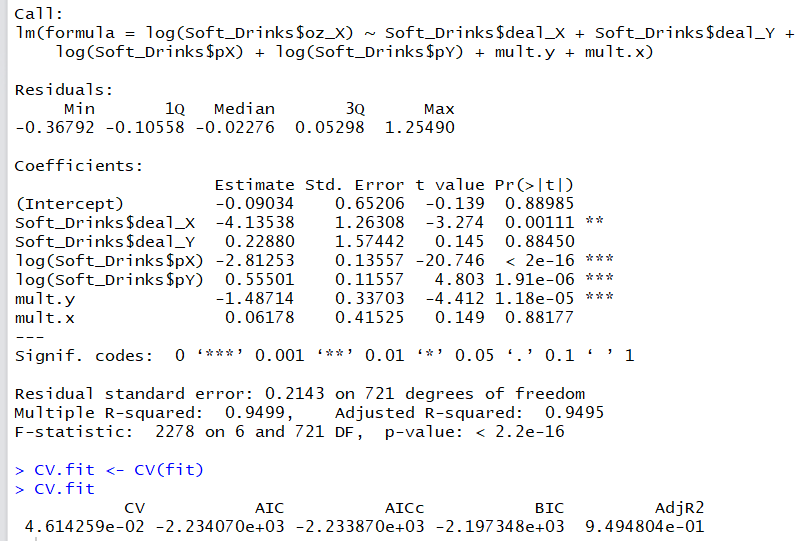
* The variables “deal\_X” and “deal\_Y” correspond to the dummy variables and
* “oz\_X” are the sales of product X; this variable is in the model.
* The variable “oz\_Y” are the sales of product Y and it will not be used in this assignment.
* The variables “pX” and “pY” correspond to and respectively.

It is clear that from these correspondences we can create the column vectors for each of the coefficients in the regression model above.

We are concerned that the above model may be unnecessarily complex and it may be overfitting the data. To explore this idea we will select the variables of the model using the “Backward Selection” procedure as described below:

1. Fit the full (most complex) model as described above.
2. Examine the p-values of all the coefficients and drop from the model the variable corresponding to the coefficient with the largest p-value.
3. Use the CV(…) function to obtain the value of the model in (II). If the value increased as you dropped the last variable, go back to the previous model and stop. If the value decreased repeat Step (II).
4. (5 pts) Fit the full model (i.e., a model with all 7 coefficients) and report the table with p-values and the of this model.

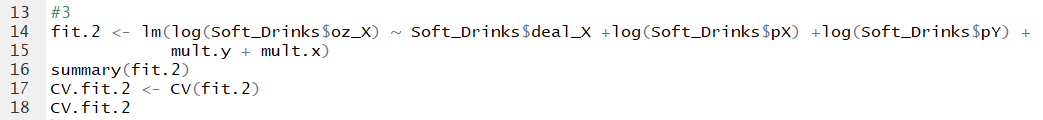


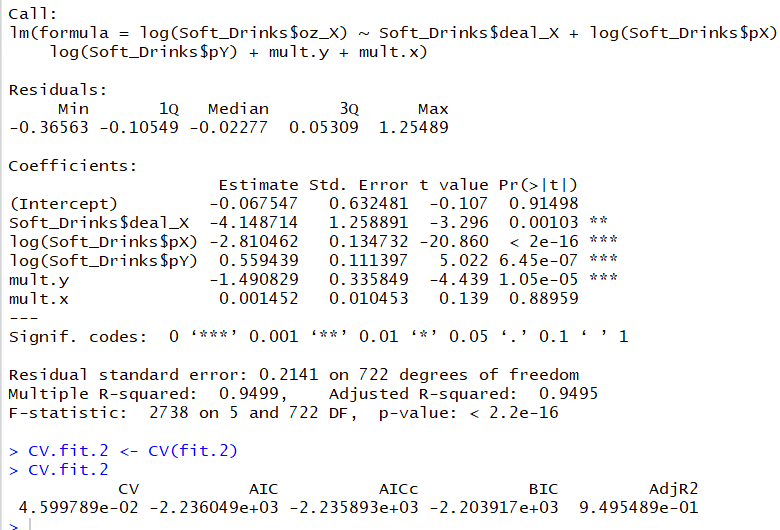


1. (5 pts) According to the “Backward Selection” process described above, what variable should we drop from the model?

According to the backward selection model, we should drop the variable with the highest p-value. This variable is deal\_Y, with a p-value of .8845.

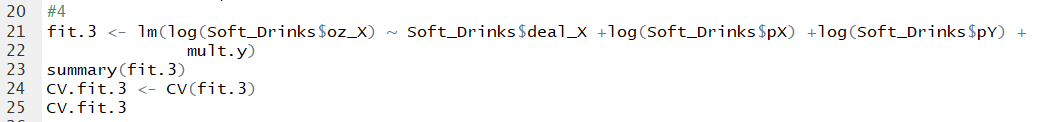
1. (5 pts) Calculate the reduced model according to your answer in (2) and report the p-values and value.

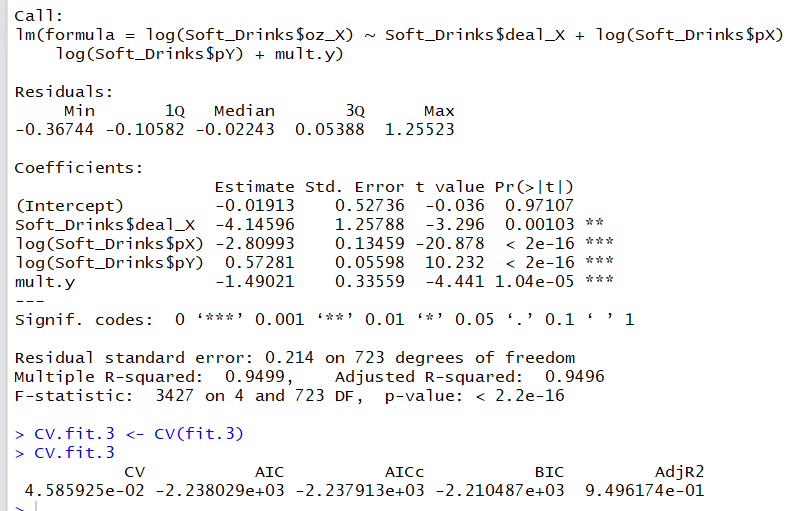




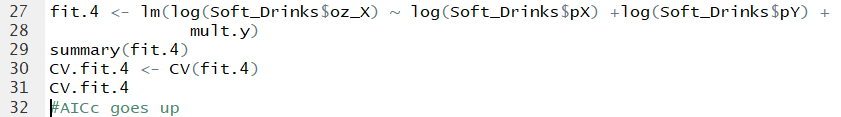
1. (5 pts) Continue executing the Backward process until you meet the stopping criteria in Step (III). What is the best model identified by this process (write it explicitly). What is its and its cross-validation standard error?

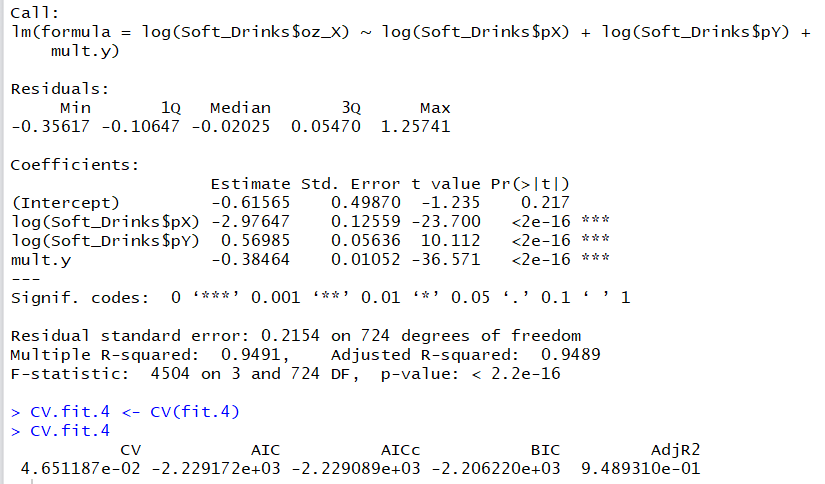
We now dropped mult.x.





We will now drop deal\_x.





Best model is fit.3.

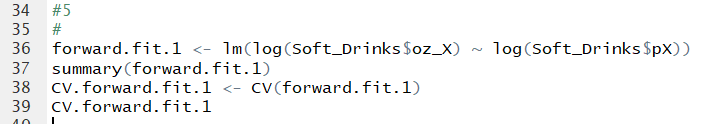
Log(oz\_x) = deal\_x +log(pX) + log(pY) + deal\_x\*log(pX)

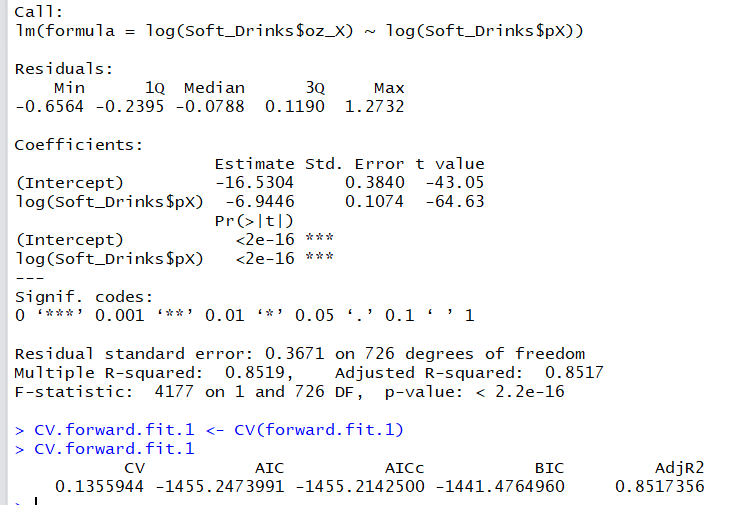
AICc for fit.3 is -2.237913e+03.

CV for fit.3 is 4.585925e-02.

A different alternative for model selection is the “Forward Selection” process. One way to implement this process is as follows:

1. Start with the simplest model you are willing to consider. For this exercise let us start with the isoelastic demand model:
2. Based on the last model considered, we add the predictive variable that increases the value of the most.
3. Fit the model with the additional variable and examine its . If the decreased, repeat Step (II). If the increased, stop and select the previous model (i.e., the one with the lowest ).
4. (5 pts) Fit the base model as described in Step (I) of the “Forward Selection” process above and report the table with p-values and the of this model.

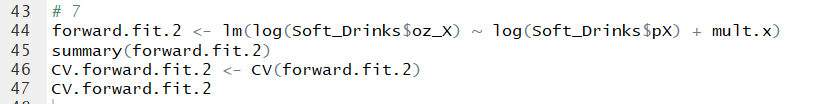


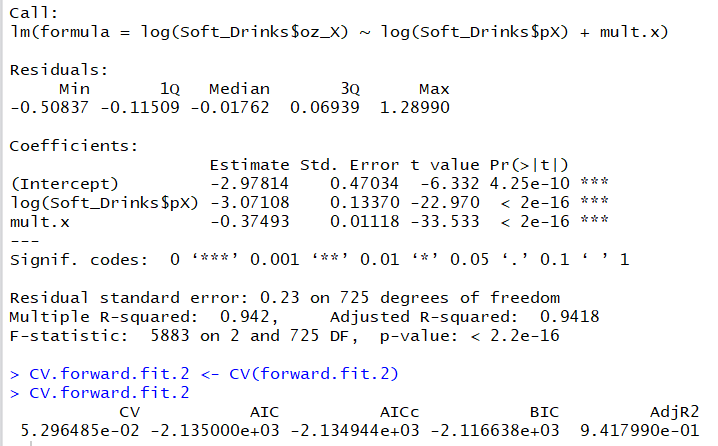


1. (5 pts) What is the variable that, when added to the model, increases the most (as described in Step (II) above?

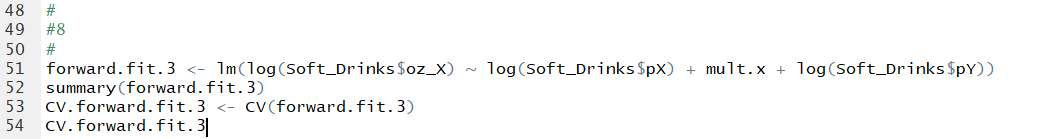
The variable that increases the value of R^2 the most is mult.x (=deal\_X \* pX). It increases the R^2 of the model from 0.8519 to 0.942.

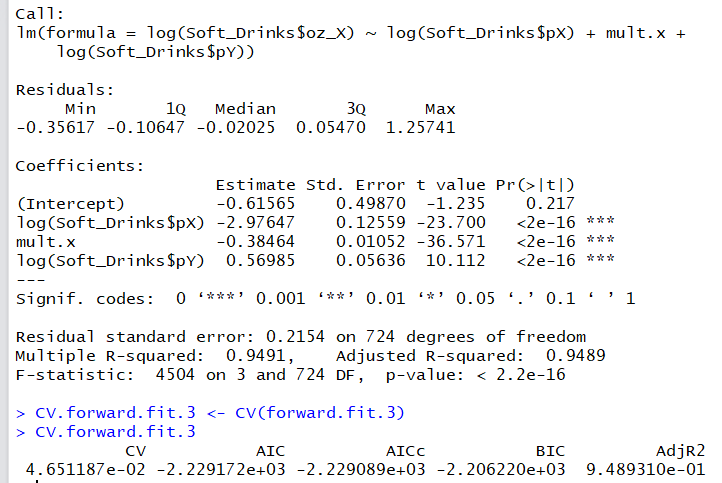
1. (5 pts) Calculate the new enhanced model according to your answer in (6) and report the p-values and value.

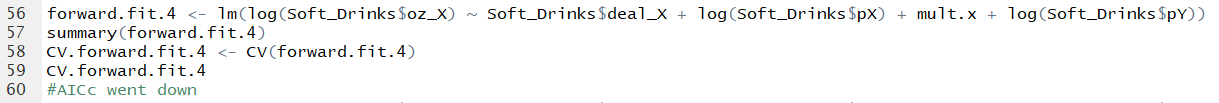


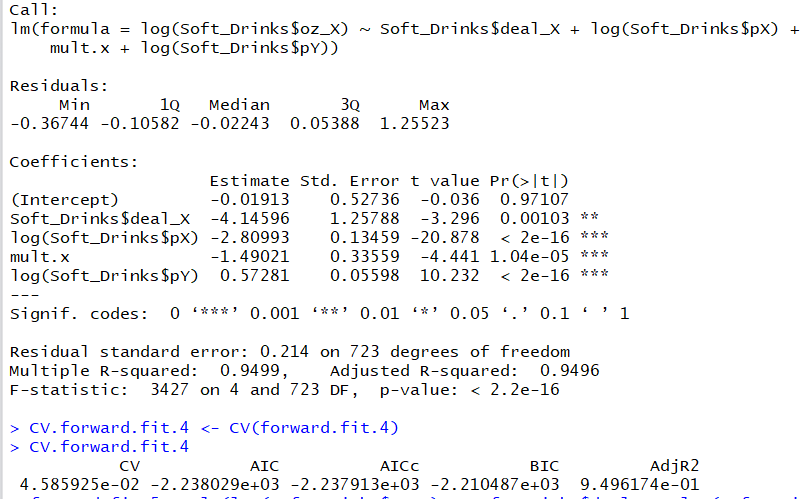


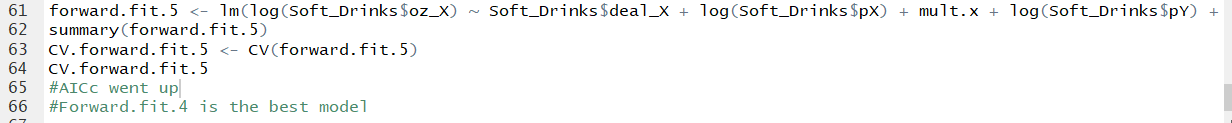
1. (5 pts) Continue executing the Forward Selection process until you meet the stopping criteria in Step (III). Write explicitly the best model identified by this selection process. What is its and its cross-validation standard error?

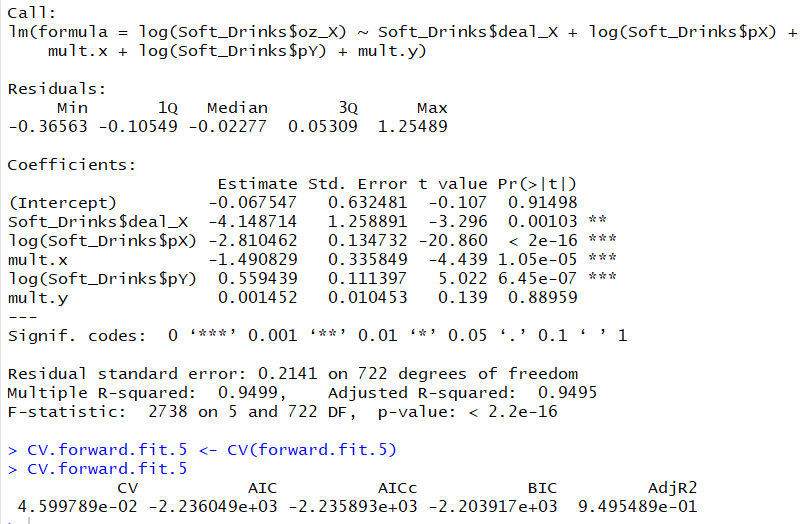












BEST MODEL:

Log(oz\_x) = deal\_x +log(pX) + log(pY) + deal\_x\*log(pX)

AICc for fit.3 is -2.237913e+03.

CV for fit.3 is 4.585925e-02.