Multiple Linear Regression Analysis of the Demand for Beef in the United States

Michael Streyle

Carly Mester

Matt Foundos

## Abstract

The dataset that we chose to analyze examines the various variables that may or may not affect the demand for beef in the United States. When we use the term demand, we are not trying to estimate the price or the amount of beef sold in the given year, but rather the general demand for beef corresponding to beef consumption in past years - with units in pounds per capita. We have nine predictor variables that we will initially use to examine how they affect demand for beef. Some of these variables are adjusted for inflation, while some intentionally leave out this adjustment. There are 36 observations in our data set, corresponding to data from each year between 1965 and 2000. None of our variables need to be converted to categorical data. Our objective is to determine which of these variables affect the beef market, and if so, to what extent? The first of our initial questions is which of our predictor variables has the largest influence on the demand for beef? Which variable has the least influence on the demand for beef? Finally, which variables are essential to estimating the demand for beef with 80% confidence, and is this a useful confidence level? After much exploration, we found a model that fits our data quite well and provides us with satisfying predictions for the demand for beef. Through the process of creating our final model, we learned a substantial amount about the American market for beef and movement within it.

## Introduction

In our analysis, we examine the demand for beef in the United States, particularly attempting to find the best set of variables to predict the demand for beef for a given year. It is helpful to note some of the basic economic principles before diving into our data, because demand is one of the major pillars in economic study. Khan Academy defines economic demand as “the amount of some good or service consumers are willing and able to purchase at each price. [...] The total number of units purchased at that price is called the quantity demanded. One of the most important factors that generally influences demand is price. Usually, when price of a good goes up, the quantity demanded of that good goes down (Khan Academy Online, Demand). Although there is a difference between demand and quantity demanded, for the purposes of our analysis, we will use the terms loosely and to mean roughly the same thing.

As previously stated, our goal is to create a model that can predict the demand/consumption of beef, given a certain combination of the 9 variables from our data. Our data was found on ww2.amstat.org and was derived from 6 different online sources and contains 36 different observations. When using our variables we need to keep in mind the units which they are in, such as cents per pound or pounds per capita. All of the variables and units that we will be referring to can be found in Appendix A with their explanations. Theoretically, once we establish a final predictive model for the demand of beef, we can apply it to other things such as chicken and pork. However, there is a limit to what we can apply our model to, because beef is a product that not everybody has the leisure of buying. There are cheaper products that can be purchased as a substitute, which means we would need to take into account the economy as a whole if we tried examining other markets with our model. When we decide on our best model, we want to examine which variables are the most significant, which are the least significant, and see if we are able to estimate future beef consumption/demand with 80% confidence. Before any analysis, we believe year and beef price will be among the most significant predictors, as well as possibly chicken price due to the economic theory surrounding the relationship between demand for a product and the price of that products substitutes.

## Our Data

We utilize a data set which includes the price of beef, the price of chicken, consumer price index (CPI), disposable personal income (DPI), year, and inflation adjusted beef and chicken prices in the United States. Throughout our analysis, beef consumption is our response variable and its distribution can be seen in the boxplot in Appendix B. The distribution of beef consumption is slightly skewed and not perfectly normal, but there is no obvious transformation needed. In our presentation we showed the boxplot of log(BeefConsumption), which is nearly the same boxplot as beef consumption without transformation. Beef consumption ranges from 92 to 127.5 pounds per capita. By analyzing the correlation matrix and the pairs plot, we discovered that there a quite a few linear relationships in our data, mainly due to the fact that many of our variables are derived from each other, or represent the same information but with an added inflation adjustment. One variable that was supposed to be the RDPI subtracted by its mean and then squared had a computational error, so we created a new one to replace it called myRealDPIsq. According to its boxplot, this new variable is skewed, however we did not see any benefit in transforming it yet, and never actually did because it turned out to not be a very significant variable. The rest of our variables are relatively normal and do not require a transformation to move further in our analysis.

## Model Selection and Interpretation

After some of our initial exploration, we discovered that selecting a final group of predictor variables would be the toughest part of this analysis. First, we fit a first order model using every variable (with our new added variable and the exclusion of the variable with the computational error), and call this model myfit1. According to our plots, this model seems to have constant variance with normally distributed residuals and no outliers. There seems to be a slight curve to the residuals but a linear model still fit very well. The model shows that Year, BeefPrice, RealBeefPrice, and CPI are all significant predictors while ChickenPrice and DPI were also shown to be approaching significance. We decided to do a quick residual analysis of our myfit1 to examine it a little more closely. All the variables in our myfit1 look good enough to continue, but we still decided to check with a Box-Cox analysis to see if Beef Consumption needed a transformation. The Box-Cox concurs that a transformation of BeefConsump is not needed. At this point, we decided to do a more in-depth search for a better model, and we used automatic backward elimination method. This resulted in a model with Year, BeefPrice, CPI, RealChickenPrice, and RealBeefPrice as significant predictor variables. All of these show slightly better improvements compared to our first order model. We used this method to construct our myfit4 which is by far the best model that we created. At this point, our R-squared value is already over .95, so we began thinking that this model might be as precise as possible. Interpretations of the variable coefficients can be found in an explanation segment in Appendix B, under the Backwards Elimination Method heading.

Although we were quite pleased with myfit4, we still wanted to know if there were any significant interaction effects that we could use to make our model even slightly better. We first created a function to center Year, BeefPrice, RealBeefPrice, and CPI, since they were our most significant continuous variables. We then created interaction variables with the four centered variables, fitting them to a model with all the significant predictors and all the interaction effect variables called fit.int. However, according to the results of the summary of the model, only centered Year was significant, and BeefPrice\*RealBeefPrice was slightly significant. We then explored and experimented by creating a few different models with multiple variations of interactions. However, none of these models fit as well as myfit4. One of our stepwise elimination processes that we ran on a model that started with all variables and the significant interaction variables resulted in a model that was identical to myfit4. This result was very reassuring to us that myfit4 was our best model. We were slightly disappointed that interaction effects were not useful, because we wanted our model to be more exciting, but in reality it is beneficial to have as simple of a model as possible.

## Results/Conclusion

In order for us to determine if our model was useful or not, we made some predictions for data points that were on the lower, middle, and upper ranges of beef consumption. Since so many of our models had such similar residual standard error and R-squared values, we made these predictions using three different models. We found that our initial goal of predictions with 80 percent confidence could easily be surpassed. We settled on 95 percent confidence and found that several of our models produced predictions that we considered precise. As a simple, straightforward way to compare the predictions made by each of our models, we summed the absolute value of the difference between the prediction and the actual values for all six predictions made for each model. Although this may not be a standard method for comparing models, we thought that it would suffice in this case because of how similar the models were in terms of residual standard error and R-squared. This process resulted in myfit4 producing the best predictions. These predictions are precise enough to be useful and with 95 percent confidence, one can feel secure about the accuracy of the predictions made using our model. Since myfit4 was the best model in terms of residual analysis, simplicity, and producing the best predictions, we are very confident in it as a predictive model.

Thinking back to the brief overview of the economic theories at play while modeling the consumption of beef, some new questions arise. Although chicken was included in our best model, it was not as significant as economic theory might suggest. This leads us to wonder if there is another substitute that has a bigger effect on beef that we could add to our data in the future? Another interesting aspect of our model that does not align perfectly with economic theory is the relationship between beef price and beef consumption. As noted in the explanation of coefficients in Appendix B (when myfit4 is created), as beef price increases, so does beef consumption. However the opposite is true when considering the relationship between inflation adjusted beef price and beef consumption. This leads us to believe that a larger emphasis should be put on the effects of inflation when describing the law of demand, since we have found that the inflation adjustment changes the direction of the relationship. Overall, our analysis was very interesting and put to work some economic concepts, something we cannot say confidently about projects required for some Economics classes.

### Appendix A

VARIABLE DESCRIPTIONS:

|  |  |
| --- | --- |
| Year | Calendar year |
| ChickPrice | Chicken Retail Price in cents per pound |
| BeefPrice | Beef Retail Price in cents per pound |
| BeefConsump | Beef Consumption per capita in pounds |
| CPI | Consumer Price Index for food |
| DPI | Disposable Personal Income per capita in dollars |
| RealChickPrice | Inflation-adjusted Chicken Retail Price in cents per pound |
| RealBeefPrice | Inflation-adjusted Beef Retail Price in cents per pound |
| RealDPI | Inflation-adjusted Disposable Personal Income per capita in dollars |
| myRealDPIsq | RealDPI minus the mean of RealDPI, squared |

### Appendix B

KNITED SECTION

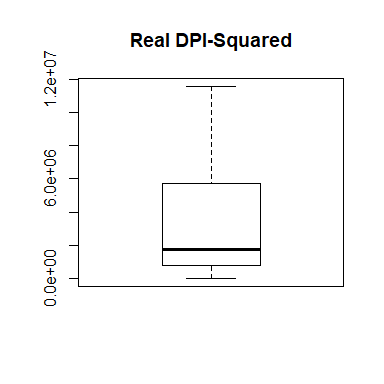
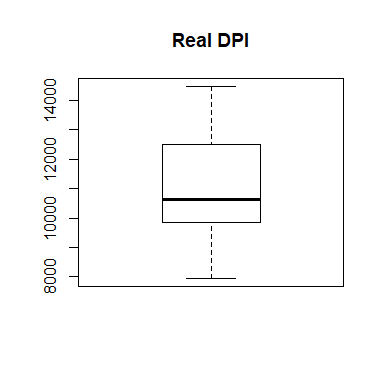
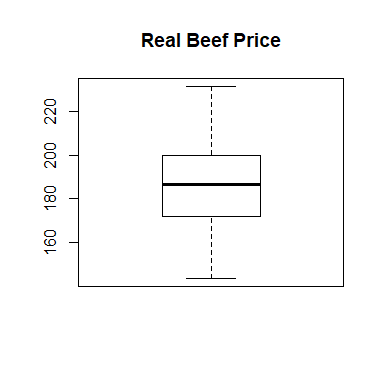
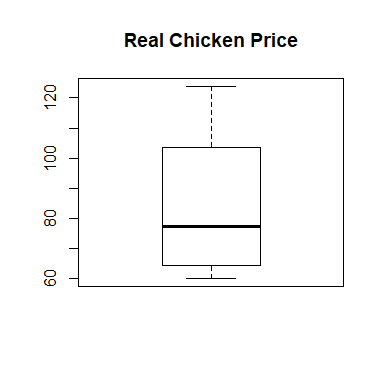
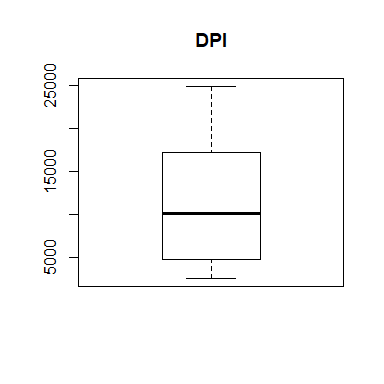
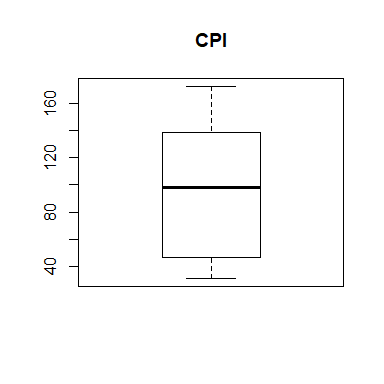
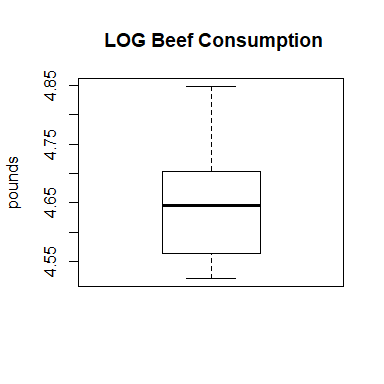
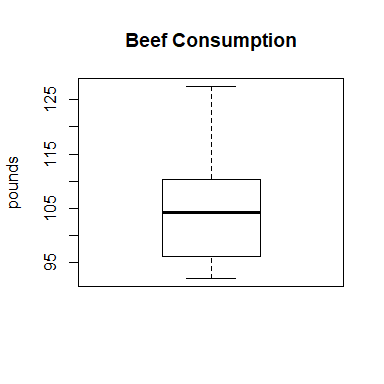
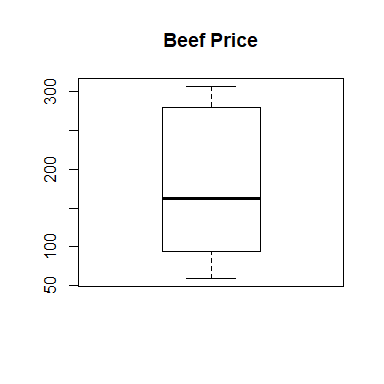
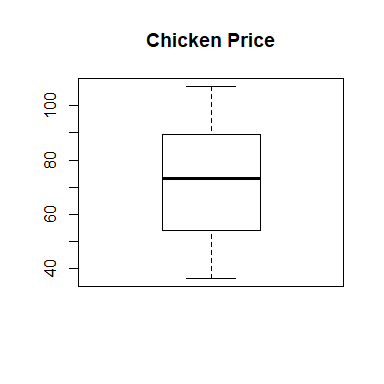
*# Set global figure size*  
knitr::opts\_chunk$**set**(fig.width=6, fig.height=3.5, echo = FALSE, message=FALSE)

## Introduction

First, we import our beef data straight from the web address and attach it. The data has 9 predictor variables (Chicken price,Year, DPI, etc..) that we will use to analyze how they affect the price of beef. Some of the variables are adjusted for inflation and some intentionally leave the inflation adjustment out. There are 36 observations from 1965-2000 in the data set, none of which need to be converted to categorical data.

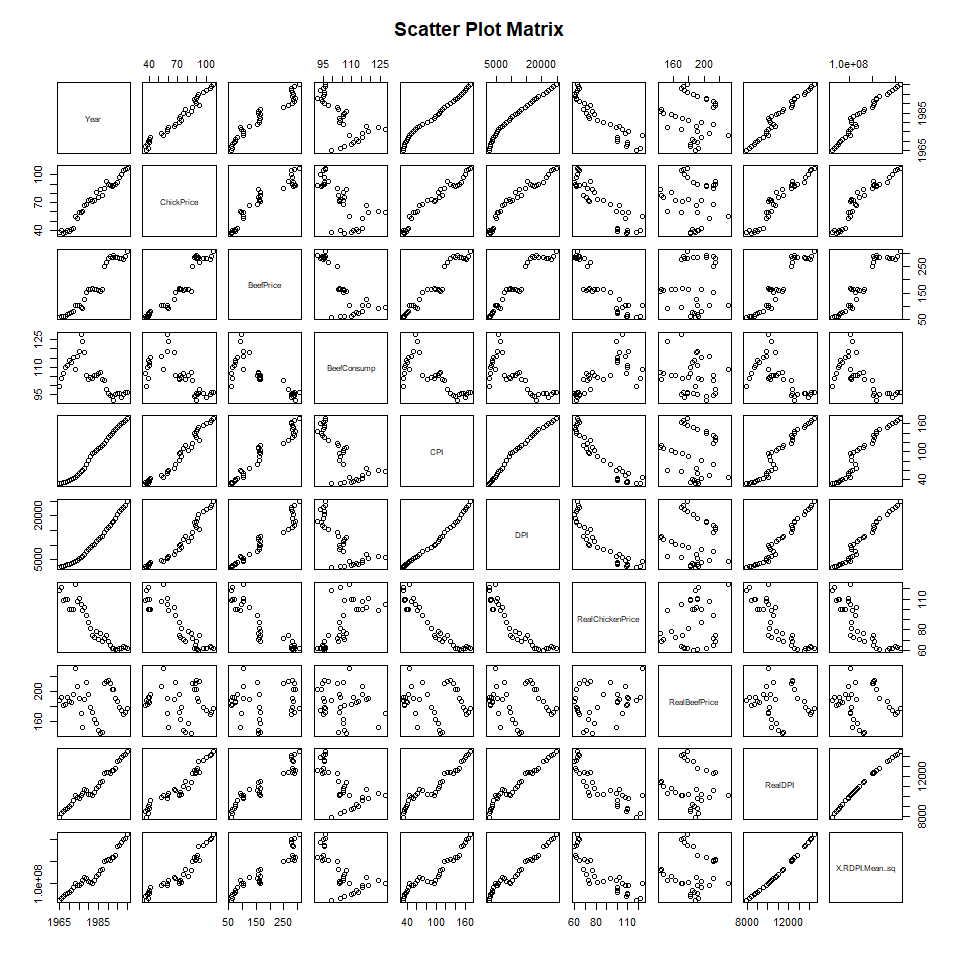
## Exploratory Analysis

Due to a computational error in the data, we created a new variable called myRealDPIsq to replace X.RDPI.Mean..sq. The new variable is the RealDPI minus the mean of RealDPI, squared. The distributions of the variables seem to look pretty good. BeefConsumption is slightly right skewed, but not enough to require a transformation, as we will see later. The new variable, myRealDPIsq, is also skewed right, but we decided a transformation is not needed at this time.

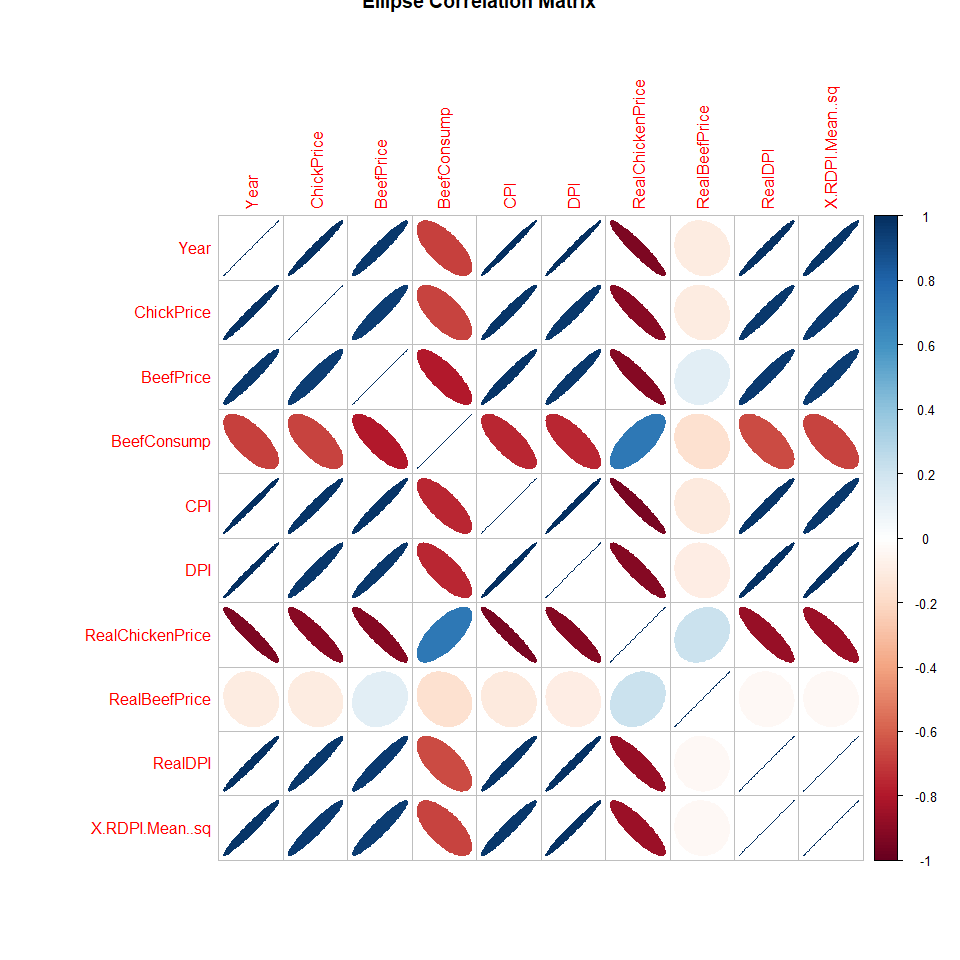


The pairs plot shows quite a few linear relationships. Many of these, such as between DPI and CPI, as well as between myRealDPIsq and DPI, are due to the relation between the variables. Since DPI and CPI describe relatively the same thing, they have a close linear relationship. Many of the variables are derived from each other, such as ChickenPrice and RealChickenPrice, which is why they seem to be related in the pairs plot. Similar such relationships are found within the correlation matrix, and ellipse correlation matrix; DPI and CPI are very highly correlated, as well as other relationships. Another aspect of these plots to note is that none of the variables seem to have any non-linear relationships with BeefConsumption. This suggests that a transformation of the response variable is probably not needed and a transformation of any predictor variable is not obviously needed at this point.

## Year ChickPrice BeefPrice BeefConsump CPI DPI  
## Year 1.00 0.98 0.96 -0.69 0.99 0.99  
## ChickPrice 0.98 1.00 0.94 -0.67 0.97 0.96  
## BeefPrice 0.96 0.94 1.00 -0.79 0.97 0.96  
## BeefConsump -0.69 -0.67 -0.79 1.00 -0.76 -0.76  
## CPI 0.99 0.97 0.97 -0.76 1.00 0.99  
## DPI 0.99 0.96 0.96 -0.76 0.99 1.00  
## RealChickenPrice -0.94 -0.90 -0.91 0.71 -0.95 -0.91  
## RealBeefPrice -0.11 -0.11 0.12 -0.17 -0.12 -0.10  
## RealDPI 0.98 0.96 0.95 -0.66 0.97 0.98  
## X.RDPI.Mean..sq 0.97 0.95 0.94 -0.67 0.96 0.98  
## RealChickenPrice RealBeefPrice RealDPI X.RDPI.Mean..sq  
## Year -0.94 -0.11 0.98 0.97  
## ChickPrice -0.90 -0.11 0.96 0.95  
## BeefPrice -0.91 0.12 0.95 0.94  
## BeefConsump 0.71 -0.17 -0.66 -0.67  
## CPI -0.95 -0.12 0.97 0.96  
## DPI -0.91 -0.10 0.98 0.98  
## RealChickenPrice 1.00 0.21 -0.87 -0.86  
## RealBeefPrice 0.21 1.00 -0.04 -0.04  
## RealDPI -0.87 -0.04 1.00 1.00  
## X.RDPI.Mean..sq -0.86 -0.04 1.00 1.00

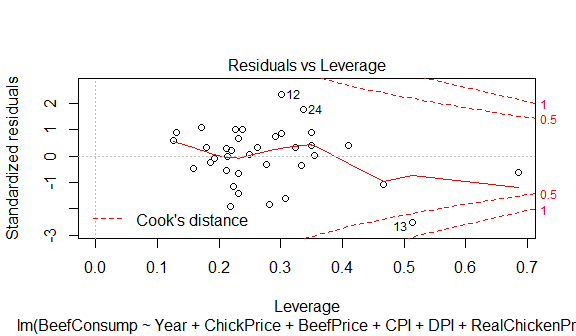
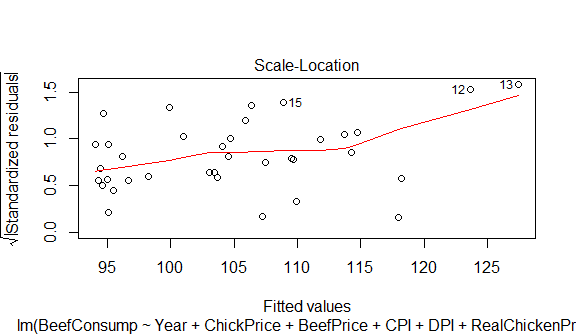
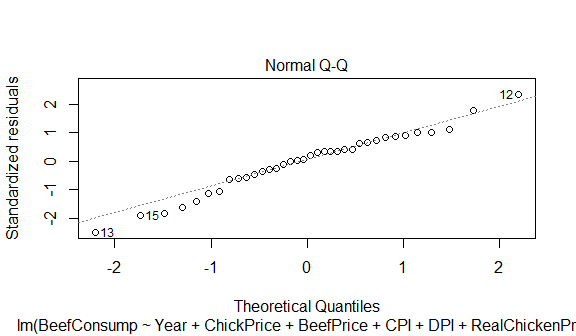
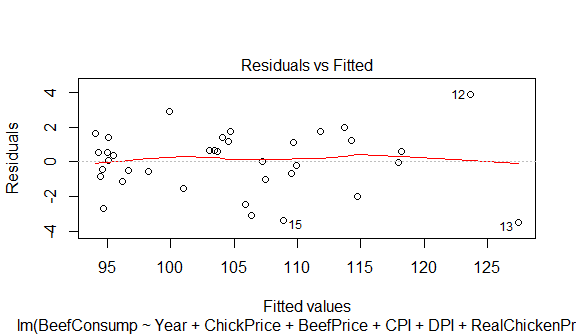


## Warning: package 'corrplot' was built under R version 3.4.2



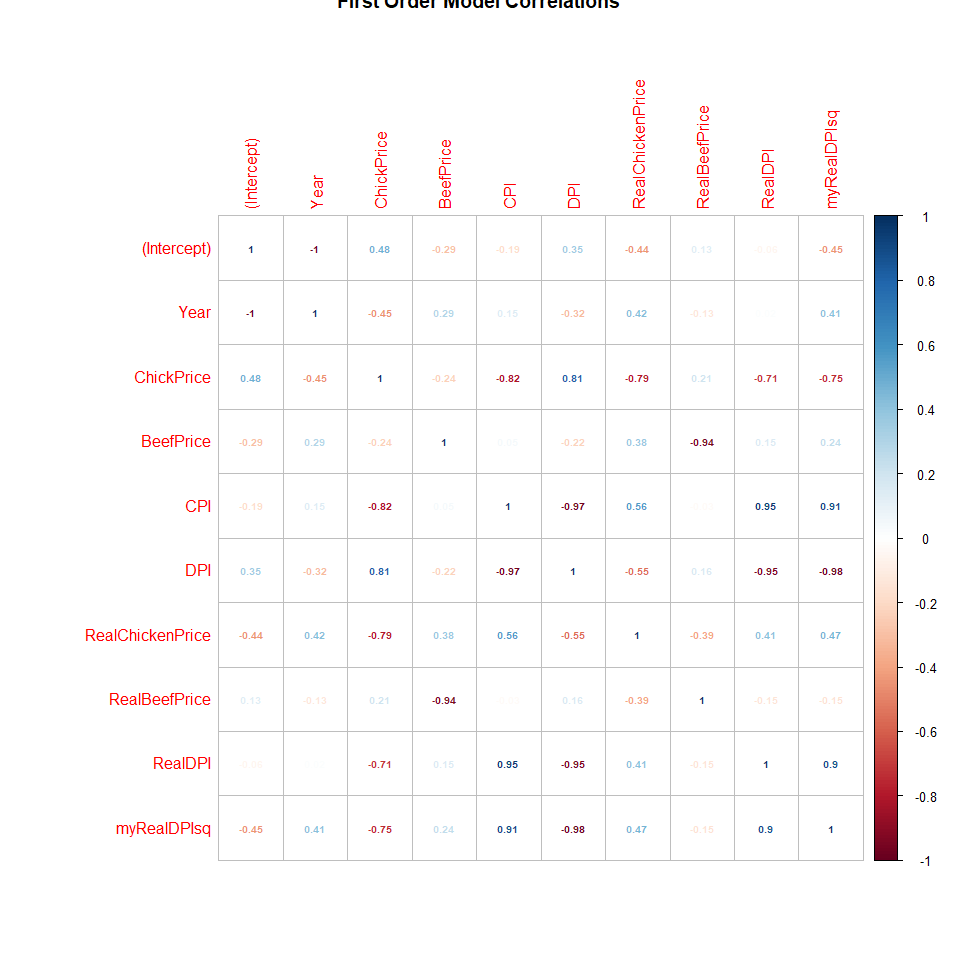
## First Order Model

To create an initial model, we fit a first-order linear model with all nine predictors. Note that we used our variable called myRealDPIsq instead of the flawed X.RDPI.Mean..sq variable.

The analysis of variance table suggests that in our first order model, the significant predictor variables are Year, BeefPrice, CPI, and RealBeefPrice. ChickenPrice and DPI were also marked as slightly significant. According to the summary of the model, only Year, BeefPrice, and RealBeefPrice are significant, with RealChickenPrice being only slightly significant. The Residual standard error for this initial model is 1.989 which is surprisingly small considering the range of the response variable, beefConsumption, is 90 to 130. The R-squared value for this model is 0.9644 and adjusted is 0.952, which tells us that nearly all of the variablility in BeefConsumption is explained by this model. 

##   
## Call:  
## lm(formula = BeefConsump ~ Year + ChickPrice + BeefPrice + CPI +   
## DPI + RealChickenPrice + RealBeefPrice + RealDPI + myRealDPIsq)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4656 -0.8820 0.2151 1.1653 3.8570   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.695e+03 1.863e+03 -5.205 1.96e-05 \*\*\*  
## Year 5.010e+00 9.437e-01 5.308 1.49e-05 \*\*\*  
## ChickPrice -1.783e-01 2.162e-01 -0.825 0.4171   
## BeefPrice 2.854e-01 5.810e-02 4.912 4.24e-05 \*\*\*  
## CPI -1.035e+00 6.892e-01 -1.501 0.1453   
## DPI -4.834e-03 6.016e-03 -0.804 0.4289   
## RealChickenPrice 2.390e-01 1.355e-01 1.764 0.0895 .   
## RealBeefPrice -4.116e-01 6.128e-02 -6.716 3.99e-07 \*\*\*  
## RealDPI 3.299e-03 6.629e-03 0.498 0.6229   
## myRealDPIsq 1.005e-06 1.208e-06 0.832 0.4130   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.989 on 26 degrees of freedom  
## Multiple R-squared: 0.9644, Adjusted R-squared: 0.952   
## F-statistic: 78.16 on 9 and 26 DF, p-value: < 2.2e-16

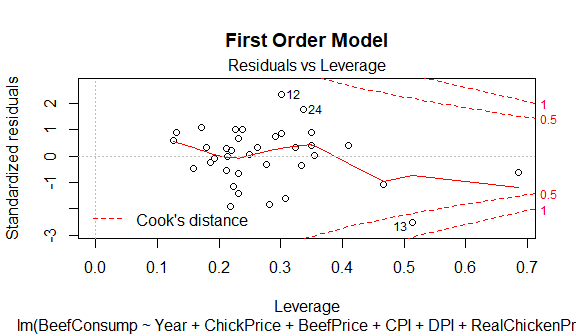
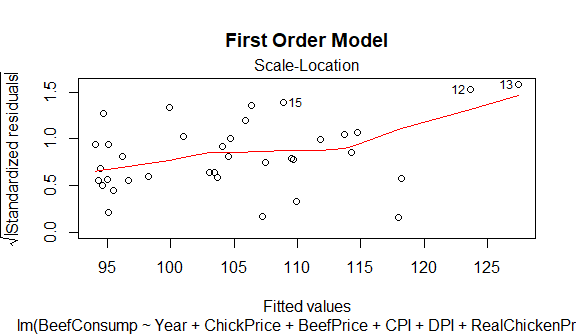
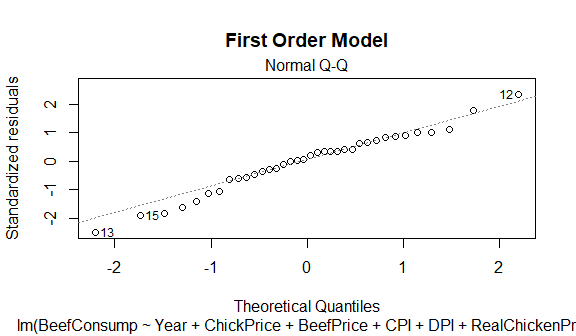
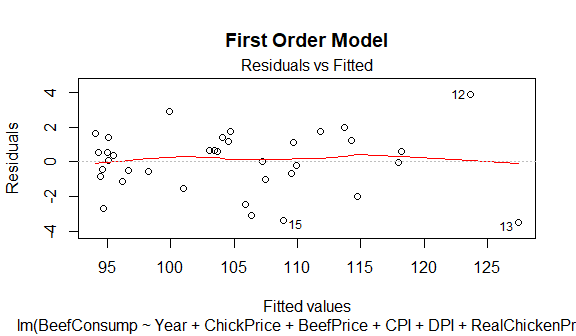
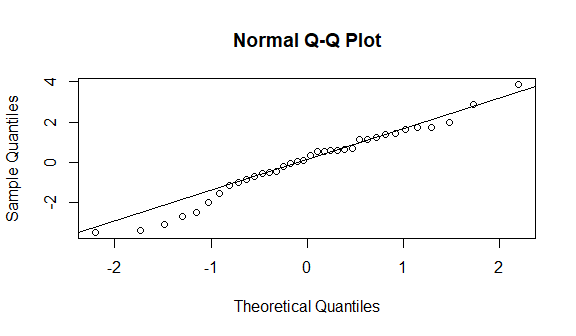
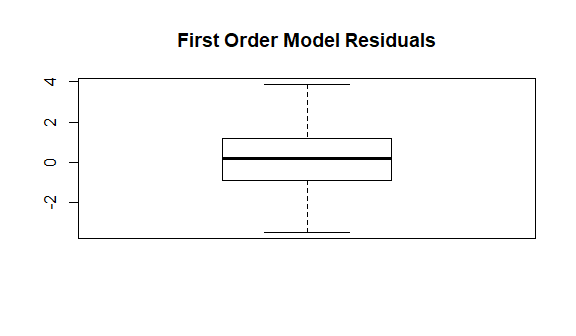
## Analysis of Variance Table  
##   
## Response: BeefConsump  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Year 1 1386.17 1386.17 350.4102 < 2.2e-16 \*\*\*  
## ChickPrice 1 12.25 12.25 3.0957 0.09027 .   
## BeefPrice 1 615.57 615.57 155.6117 1.766e-12 \*\*\*  
## CPI 1 571.95 571.95 144.5842 4.008e-12 \*\*\*  
## DPI 1 12.24 12.24 3.0931 0.09039 .   
## RealChickenPrice 1 4.02 4.02 1.0170 0.32251   
## RealBeefPrice 1 176.66 176.66 44.6580 4.333e-07 \*\*\*  
## RealDPI 1 1.25 1.25 0.3157 0.57904   
## myRealDPIsq 1 2.74 2.74 0.6922 0.41299   
## Residuals 26 102.85 3.96   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



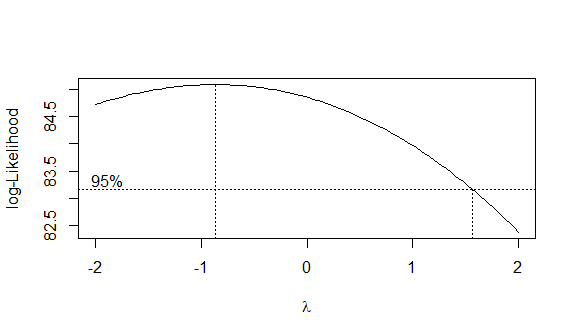
## Residual Analysis of the First Order Model

Residual Analysis of our first order model turned out to be quite good. The Residuals vs Fitted plot looks very good, with no obvious outliers or patterns of non-linearity with a fitted line very tight along y=0. The First Order Model Normal Q-Q plot also looks quite excellent. There appears to be no obvious outliers and the data appears to be normally distributed with only small deviations from normality at the extremes. The First Order Model Scale-Location plot also looks pretty good. The higher fitted values do seem to have higher residuals, which might be something to keep our eyes on, but for now we do not think it is significant enough for a transformation. Finally, the First Order Model Residuals vs Leverage plot shows that there are a few points with high leverage, of which only one is near Cook's distance and is therefore significant. We will keep an eye or two on row 13, and see if this gets remedied.

We also did a quick boxcox plot for our data to make sure a transformation on BeefConsumption wasn't needed. As we suspected, the boxcox plot was inconclusive and did not suggest we needed to transform our response variable.



## Warning: package 'MASS' was built under R version 3.4.2



## Backward elimination method

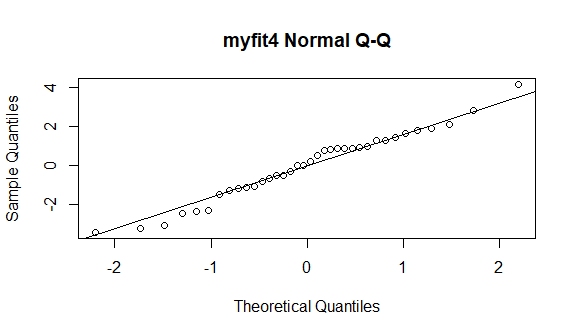
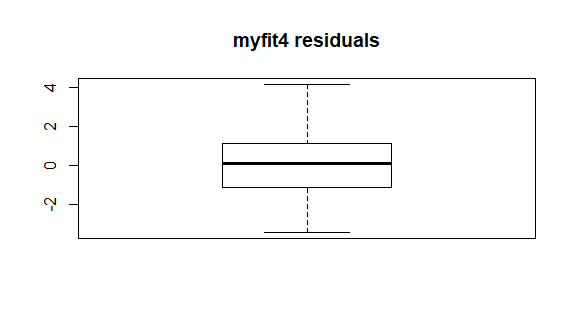
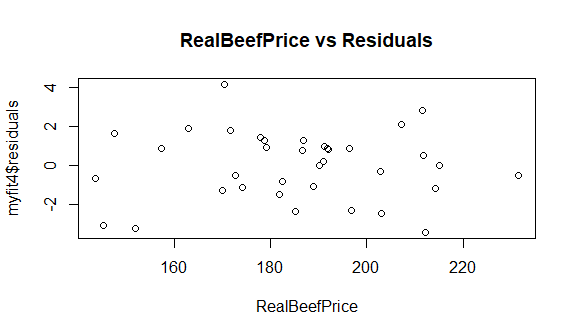
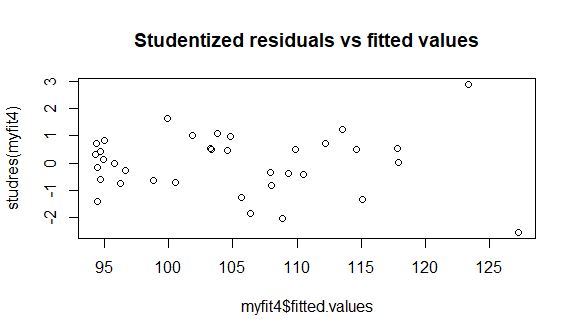
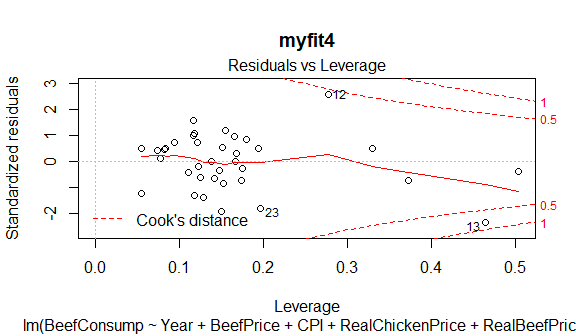
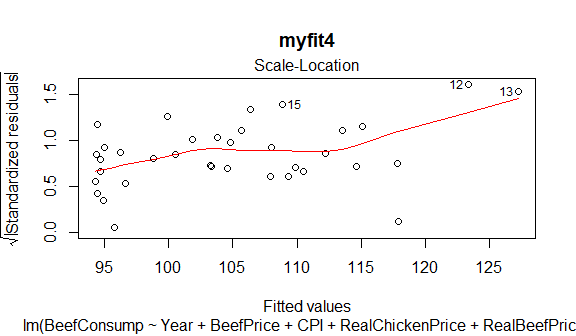
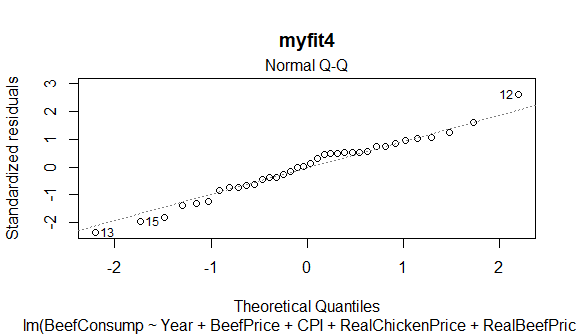
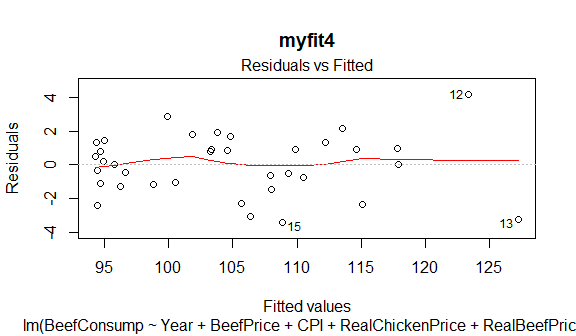
Next, we used the automatic backwards elimination method using the step function with direction set to backward. This resulted in a model with Year, BeefPrice, CPI, RealChickenPrice, and RealBeefPrice as significant predictor variables. The Residual Standard Error of this model (myfit4) is 1.899, the R-squared value is 0.9625, and the adjusted R-squared is 0.9563; all slight improvements from our first order model. All of the plots of this model are quite good, but row 13 still seems to have high leverage. This will have to be looked into more later.

Each of the following statements is made in the context of the other predictors being held at fixed values: As Year increases one year, beefConsumption increases by roughly 4 pounds per capita. As BeefPrice increases one cent per pound, BeefConsumption increases by roughly 0.27 pounds per capita. As CPI increases one unit, BeefConsumption decreases roughly 1.5 pounds per capita. As the RealBeefPrice goes up one cent per pound, BeefConsumption decreases roughly 0.4 pounds per capita.

The only statement above that is not quite what we expected is the fact that as BeefPrice goes up, BeefConsumption increases. This is countered by the relationship between RealBeefPrice and BeefConsumption, which shows that the inflation adjustment makes a difference with BeefPrice vs RealBeefPrice. Also note that in the Studentized residuals vs fitted values plot, none of the studentized residuals are greater than the absolute value of 4 which confirms that there are no outliers.

##   
## Call:  
## lm(formula = BeefConsump ~ Year + BeefPrice + CPI + RealChickenPrice +   
## RealBeefPrice)  
##   
## Coefficients:  
## (Intercept) Year BeefPrice CPI   
## -7955.7024 4.1470 0.2681 -1.5211   
## RealChickenPrice RealBeefPrice   
## 0.1298 -0.4026

##   
## Call:  
## lm(formula = BeefConsump ~ Year + BeefPrice + CPI + RealChickenPrice +   
## RealBeefPrice)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4016 -1.1130 0.1231 1.0582 4.1643   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.956e+03 5.221e+02 -15.236 1.16e-15 \*\*\*  
## Year 4.147e+00 2.671e-01 15.525 7.00e-16 \*\*\*  
## BeefPrice 2.681e-01 5.279e-02 5.077 1.87e-05 \*\*\*  
## CPI -1.521e+00 1.162e-01 -13.095 6.11e-14 \*\*\*  
## RealChickenPrice 1.298e-01 5.980e-02 2.170 0.0381 \*   
## RealBeefPrice -4.026e-01 5.715e-02 -7.045 7.86e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.899 on 30 degrees of freedom  
## Multiple R-squared: 0.9625, Adjusted R-squared: 0.9563   
## F-statistic: 154.1 on 5 and 30 DF, p-value: < 2.2e-16



### Check into row 13's leverage

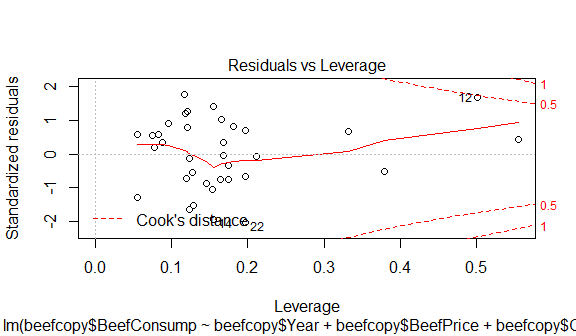
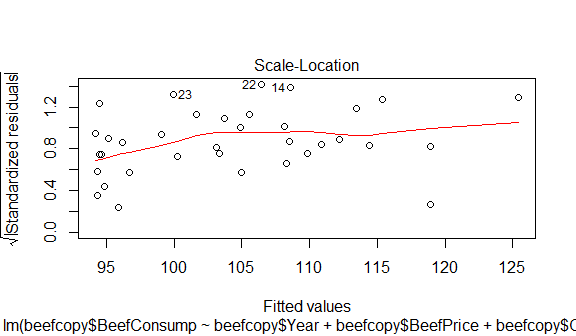
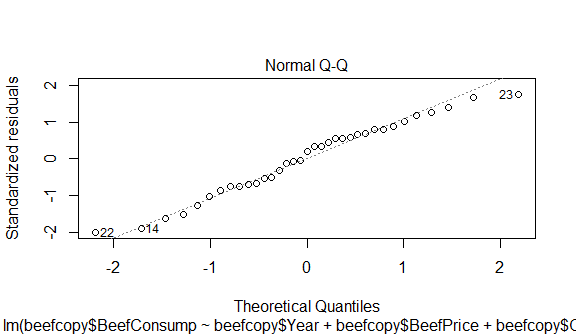
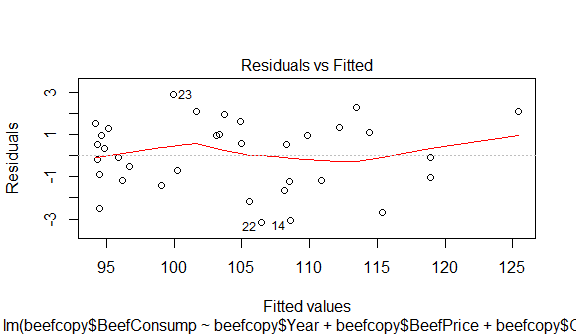
Since row 13 showed some concern in the myfit4 Residuals vs Leverage plot, we made a copy of our data and removed row 13 from it. Then we refit myfit4 as myfit4.1 to see if it significantly changed the model, remembering to put the full column names from beefcopy in myfit4.1. The results of removing this row are slightly better Residual Standard Error (1.765) and a slightly higher R-squared value and adjusted R-squared value. The plots of this new model also look very good. In comparison to the plots of the model with row 13, the new residuals vs fitted plot looks a little worse, the new scale-location plot looks better, and the Residuals vs Leverage plot now shows row 12 inside of Cook's distance. Our conclusion: removing row 13 was not as significant as it could have been and it produces a possible problem with row 12 now, so we are going to stick with including row 13. Plus, the best fit line of the Residuals vs Leverage plot with row 13 included (myfit4) does not cross Cook's distance so it doesn't appear to be of too much concern.

## [1] FALSE

##   
## Call:  
## lm(formula = beefcopy$BeefConsump ~ beefcopy$Year + beefcopy$BeefPrice +   
## beefcopy$CPI + beefcopy$RealChickenPrice + beefcopy$RealBeefPrice)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.138 -1.174 0.329 1.183 2.869   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -8.543e+03 5.336e+02 -16.009 6.20e-16 \*\*\*  
## beefcopy$Year 4.453e+00 2.740e-01 16.254 4.16e-16 \*\*\*  
## beefcopy$BeefPrice 3.521e-01 5.886e-02 5.982 1.67e-06 \*\*\*  
## beefcopy$CPI -1.728e+00 1.345e-01 -12.847 1.70e-13 \*\*\*  
## beefcopy$RealChickenPrice 1.840e-01 5.907e-02 3.115 0.00412 \*\*   
## beefcopy$RealBeefPrice -5.052e-01 6.640e-02 -7.608 2.18e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.748 on 29 degrees of freedom  
## Multiple R-squared: 0.9647, Adjusted R-squared: 0.9586   
## F-statistic: 158.4 on 5 and 29 DF, p-value: < 2.2e-16

## Warning in anova.lmlist(object, ...): models with response '"beefcopy  
## $BeefConsump"' removed because response differs from model 1

## Analysis of Variance Table  
##   
## Response: BeefConsump  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Year 1 1386.17 1386.17 384.4077 < 2.2e-16 \*\*\*  
## BeefPrice 1 620.11 620.11 171.9668 5.888e-14 \*\*\*  
## CPI 1 579.16 579.16 160.6099 1.410e-13 \*\*\*  
## RealChickenPrice 1 13.14 13.14 3.6432 0.06591 .   
## RealBeefPrice 1 178.95 178.95 49.6256 7.863e-08 \*\*\*  
## Residuals 30 108.18 3.61   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



## Interaction Analysis

Now we will examine two-way interactions among the 4 most significant predictor variables. We first created a fuction to center Year, BeefPrice, RealBeefPrice, and CPI. We then created interaction variables with these four centered variables. Fitting a model with all the significant predictors and all the interaction effect variables called fit.int. This model had a Residual Standard Error of 2.025 and an adjusted R-squared value of 0.9503. However, according to the results of the summary of the model, only centered year was significant, and BeefPrice\*RealBeefPrice was slightly significant. Next, we will try a step-wise elimination on the model with interactions.

## tracemem[0x000000001b4cacb8 -> 0x0000000017e62980]: eval eval withVisible withCallingHandlers handle timing\_fn evaluate\_call evaluate in\_dir block\_exec call\_block process\_group.block process\_group withCallingHandlers process\_file <Anonymous> <Anonymous>   
## tracemem[0x0000000017e62980 -> 0x0000000017d7f418]: $<-.data.frame $<- eval eval withVisible withCallingHandlers handle timing\_fn evaluate\_call evaluate in\_dir block\_exec call\_block process\_group.block process\_group withCallingHandlers process\_file <Anonymous> <Anonymous>   
## tracemem[0x0000000017d7f418 -> 0x0000000017d7f568]: $<-.data.frame $<- eval eval withVisible withCallingHandlers handle timing\_fn evaluate\_call evaluate in\_dir block\_exec call\_block process\_group.block process\_group withCallingHandlers process\_file <Anonymous> <Anonymous>

##   
## Call:  
## lm(formula = BeefConsump ~ Year.c + BeefPrice.c + RealBeefPrice.c +   
## CPI.c + Year.BeefPrice + Year.RealBeefPrice + Year.CPI +   
## BeefPrice.RealBeefPrice + BeefPrice.CPI + RealBeefPrice.CPI)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5213 -0.7809 0.1750 1.1316 4.7943   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 105.089928 0.909330 115.569 < 2e-16 \*\*\*  
## Year.c 5.399060 1.032955 5.227 1.85e-05 \*\*\*  
## BeefPrice.c 0.256187 0.738985 0.347 0.732   
## RealBeefPrice.c -0.441598 0.685710 -0.644 0.525   
## CPI.c -1.816713 1.359463 -1.336 0.193   
## Year.BeefPrice 0.033251 0.023263 1.429 0.165   
## Year.RealBeefPrice -0.003906 0.035975 -0.109 0.914   
## Year.CPI -0.039994 0.025888 -1.545 0.134   
## BeefPrice.RealBeefPrice -0.001780 0.001040 -1.711 0.099 .   
## BeefPrice.CPI -0.002482 0.002426 -1.023 0.316   
## RealBeefPrice.CPI NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.025 on 26 degrees of freedom  
## Multiple R-squared: 0.9631, Adjusted R-squared: 0.9503   
## F-statistic: 75.34 on 9 and 26 DF, p-value: 2.488e-16

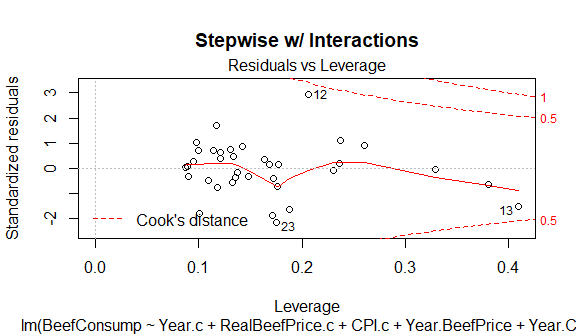
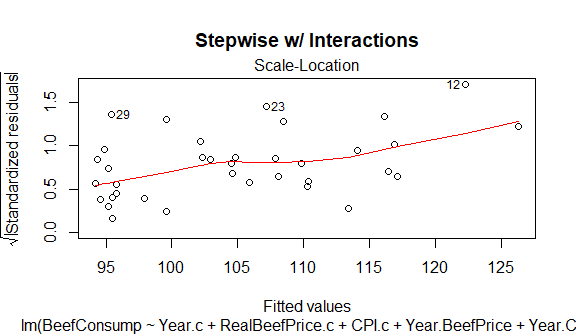
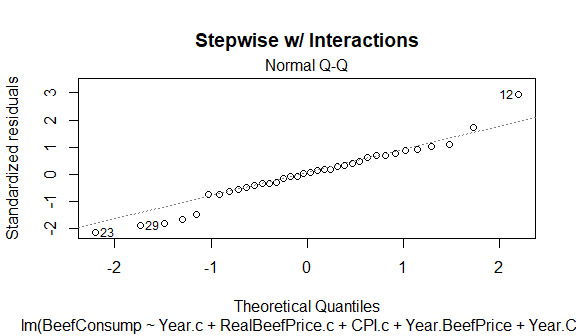
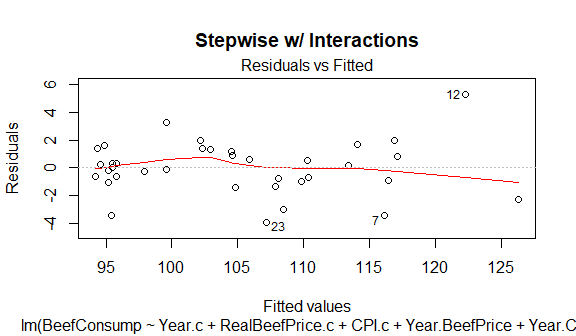
The stepwise elimination procedure ended with the following model:

step.int = lm(formula = BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c + Year.BeefPrice + Year.CPI)

The summary of the model indicates that all of the variables are significant, and has a Residual Standard Error of 2.013 and an adjusted R-squared value of 0.9509. Usually, for every interaction effect in a model, we should have both primary predictors in the model that are involved in that interaction. Here, we have every primary variable that is involved in an interaction except BeefPrice. Because of this, we tested this full model (step.int) with a model where the interaction involving BeefPrice is removed (step.int.red).

## Start: AIC=74.91  
## BeefConsump ~ Year.c + BeefPrice.c + RealBeefPrice.c + CPI.c +   
## Year.BeefPrice + Year.RealBeefPrice + Year.CPI + BeefPrice.RealBeefPrice +   
## BeefPrice.CPI + RealBeefPrice.CPI  
##   
##   
## Step: AIC=74.91  
## BeefConsump ~ Year.c + BeefPrice.c + RealBeefPrice.c + CPI.c +   
## Year.BeefPrice + Year.RealBeefPrice + Year.CPI + BeefPrice.RealBeefPrice +   
## BeefPrice.CPI  
##   
## Df Sum of Sq RSS AIC  
## - Year.RealBeefPrice 1 0.048 106.62 71.339  
## - BeefPrice.c 1 0.493 107.06 71.488  
## - RealBeefPrice.c 1 1.700 108.27 71.892  
## - BeefPrice.CPI 1 4.290 110.86 72.743  
## - CPI.c 1 7.320 113.89 73.714  
## - Year.BeefPrice 1 8.374 114.95 74.046  
## - Year.CPI 1 9.783 116.35 74.484  
## <none> 106.57 74.906  
## - BeefPrice.RealBeefPrice 1 11.995 118.57 75.162  
## - Year.c 1 111.980 218.55 97.178  
##   
## Step: AIC=71.34  
## BeefConsump ~ Year.c + BeefPrice.c + RealBeefPrice.c + CPI.c +   
## Year.BeefPrice + Year.CPI + BeefPrice.RealBeefPrice + BeefPrice.CPI  
##   
## Df Sum of Sq RSS AIC  
## - BeefPrice.c 1 3.130 109.75 68.797  
## - BeefPrice.CPI 1 4.244 110.86 69.161  
## - Year.BeefPrice 1 8.431 115.05 70.495  
## - Year.CPI 1 10.190 116.81 71.041  
## <none> 106.62 71.339  
## - BeefPrice.RealBeefPrice 1 13.059 119.68 71.915  
## - RealBeefPrice.c 1 18.518 125.14 73.520  
## + Year.RealBeefPrice 1 0.048 106.57 74.906  
## - CPI.c 1 93.217 199.84 90.372  
## - Year.c 1 111.979 218.60 93.602  
##   
## Step: AIC=68.8  
## BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c + Year.BeefPrice +   
## Year.CPI + BeefPrice.RealBeefPrice + BeefPrice.CPI  
##   
## Df Sum of Sq RSS AIC  
## - BeefPrice.CPI 1 7.964 117.71 67.735  
## - BeefPrice.RealBeefPrice 1 10.489 120.24 68.499  
## <none> 109.75 68.797  
## + BeefPrice.c 1 3.130 106.62 71.339  
## + RealBeefPrice.CPI 1 3.130 106.62 71.339  
## + Year.RealBeefPrice 1 2.686 107.06 71.488  
## - Year.BeefPrice 1 21.537 131.29 71.664  
## - Year.CPI 1 30.516 140.27 74.045  
## - Year.c 1 154.121 263.87 96.795  
## - RealBeefPrice.c 1 164.127 273.88 98.134  
## - CPI.c 1 203.489 313.24 102.969  
##   
## Step: AIC=67.74  
## BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c + Year.BeefPrice +   
## Year.CPI + BeefPrice.RealBeefPrice  
##   
## Df Sum of Sq RSS AIC  
## - BeefPrice.RealBeefPrice 1 3.79 121.50 65.293  
## <none> 117.71 67.735  
## + BeefPrice.CPI 1 7.96 109.75 68.797  
## + BeefPrice.c 1 6.85 110.86 69.161  
## + RealBeefPrice.CPI 1 6.85 110.86 69.161  
## + Year.RealBeefPrice 1 6.32 111.40 69.334  
## - Year.CPI 1 25.07 142.79 71.104  
## - Year.BeefPrice 1 27.26 144.97 71.649  
## - RealBeefPrice.c 1 178.61 296.32 97.387  
## - Year.c 1 736.25 853.96 135.490  
## - CPI.c 1 962.21 1079.93 143.942  
##   
## Step: AIC=65.29  
## BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c + Year.BeefPrice +   
## Year.CPI  
##   
## Df Sum of Sq RSS AIC  
## <none> 121.50 65.293  
## + BeefPrice.RealBeefPrice 1 3.79 117.71 67.735  
## + Year.RealBeefPrice 1 2.10 119.40 68.248  
## + RealBeefPrice.CPI 1 1.53 119.98 68.421  
## + BeefPrice.c 1 1.53 119.98 68.421  
## + BeefPrice.CPI 1 1.27 120.24 68.499  
## - Year.BeefPrice 1 76.90 198.40 79.362  
## - Year.CPI 1 79.96 201.46 79.913  
## - RealBeefPrice.c 1 309.09 430.59 107.257  
## - Year.c 1 742.31 863.82 132.320  
## - CPI.c 1 963.97 1085.47 140.543

##   
## Call:  
## lm(formula = BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c +   
## Year.BeefPrice + Year.CPI)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.8797 -0.9471 0.0948 1.2222 5.2434   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 104.086415 0.613353 169.701 < 2e-16 \*\*\*  
## Year.c 4.486867 0.331425 13.538 2.59e-14 \*\*\*  
## RealBeefPrice.c -0.173964 0.019914 -8.736 9.65e-10 \*\*\*  
## CPI.c -1.152199 0.074685 -15.427 8.29e-16 \*\*\*  
## Year.BeefPrice 0.012659 0.002905 4.357 0.000142 \*\*\*  
## Year.CPI -0.021400 0.004816 -4.443 0.000112 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.013 on 30 degrees of freedom  
## Multiple R-squared: 0.9579, Adjusted R-squared: 0.9509   
## F-statistic: 136.5 on 5 and 30 DF, p-value: < 2.2e-16

 The variance inflation factors range from 1.0 to 85. They are highest Year and CPI which we know are correlated from our very first pairs plot.

##   
## Call:  
## lm(formula = BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c +   
## Year.CPI)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.2086 -0.9534 0.4590 0.9190 7.5572   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.052e+02 6.947e-01 151.502 < 2e-16 \*\*\*  
## Year.c 3.842e+00 3.728e-01 10.306 1.55e-11 \*\*\*  
## RealBeefPrice.c -1.239e-01 2.045e-02 -6.059 1.04e-06 \*\*\*  
## CPI.c -1.004e+00 8.363e-02 -12.009 3.41e-13 \*\*\*  
## Year.CPI -7.955e-04 1.150e-03 -0.692 0.494   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.53 on 31 degrees of freedom  
## Multiple R-squared: 0.9312, Adjusted R-squared: 0.9224   
## F-statistic: 105 on 4 and 31 DF, p-value: < 2.2e-16

## Analysis of Variance Table  
##   
## Model 1: BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c + Year.CPI  
## Model 2: BeefConsump ~ Year.c + RealBeefPrice.c + CPI.c + Year.BeefPrice +   
## Year.CPI  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 31 198.4   
## 2 30 121.5 1 76.899 18.987 0.0001418 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Warning: package 'car' was built under R version 3.4.2

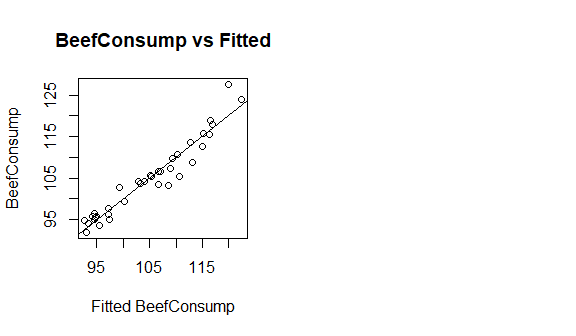
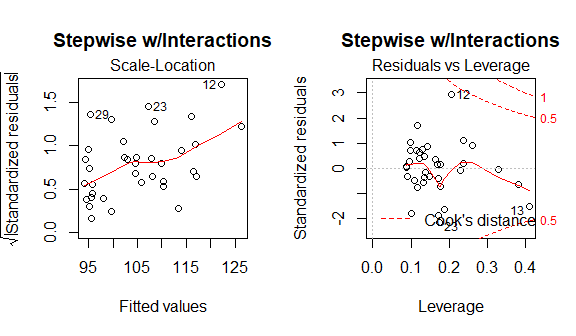
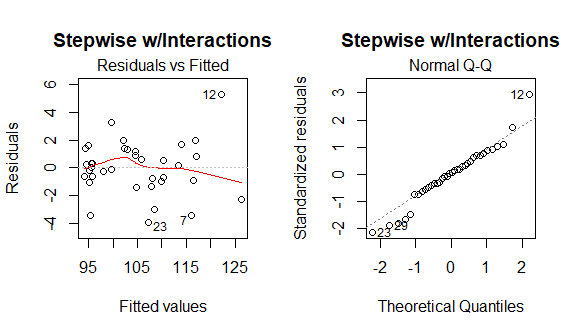
## Year.c RealBeefPrice.c CPI.c Year.CPI   
## 84.377798 1.031044 85.129500 1.165407

## [1] 56.45455

The full step-wise model fits better than the reduced model. The difference between models has a p-value of 0.0001418. While it is a very small p-value, neither of these models fits as well as our myfit4. Since our original stepwise elimination model (myfit4) turned out to be better than any of our other models with interactions, we return to focusing on it.

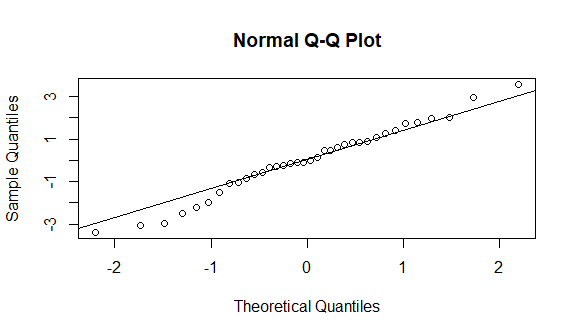
Some notes about the plots of Stepwise w/ Interactions plots: The residuals vs fitted looks pretty good. There is little evidence of outliers, non-constant variance, or significant curvature. The Q-Q plot also looks pretty good. All points are pretty close to the line, indicating that a normal distribution is a reasonable assumption for these residuals.

The one aspect of the model with interaction effects that is better than myfit4 is that the interaction effects seem to fix the issues with row 13. The Stepwise w/ Interactions residuals vs Leverage plot shows none of the data is past Cook's distance. However, since we saw that removal of row 13 only made the model slightly better, we continue to believe that myfit4 is our best model.



Next, we made a model with all variables (significant and insignificant) and the two interaction effects that were found to be significant. After running a step-wise elimination on this model, it resulted in the exact same model that backwards elimination produced from our first order model (myfit4). This confirms to us that myfit4 is our best model.

##   
## Call:  
## lm(formula = BeefConsump ~ Year + ChickPrice + BeefPrice + CPI +   
## DPI + RealChickenPrice + RealBeefPrice + RealDPI + myRealDPIsq +   
## BeefPrice.RealBeefPrice + Year.CPI)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3746 -0.8838 -0.0485 0.9533 3.5685   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.906e+03 2.417e+03 -3.271 0.00323 \*\*  
## Year 4.161e+00 1.219e+00 3.415 0.00227 \*\*  
## ChickPrice -2.121e-01 2.591e-01 -0.819 0.42109   
## BeefPrice 4.577e-01 1.959e-01 2.337 0.02812 \*   
## CPI -2.272e+00 1.362e+00 -1.668 0.10835   
## DPI 5.127e-03 1.343e-02 0.382 0.70599   
## RealChickenPrice 2.911e-01 1.690e-01 1.722 0.09790 .   
## RealBeefPrice -5.868e-01 1.995e-01 -2.941 0.00714 \*\*  
## RealDPI -5.021e-03 1.376e-02 -0.365 0.71838   
## myRealDPIsq -1.636e-07 1.566e-06 -0.104 0.91764   
## BeefPrice.RealBeefPrice -1.125e-03 1.122e-03 -1.003 0.32602   
## Year.CPI -1.203e-02 1.641e-02 -0.733 0.47052   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.991 on 24 degrees of freedom  
## Multiple R-squared: 0.967, Adjusted R-squared: 0.9519   
## F-statistic: 63.97 on 11 and 24 DF, p-value: 4.923e-15

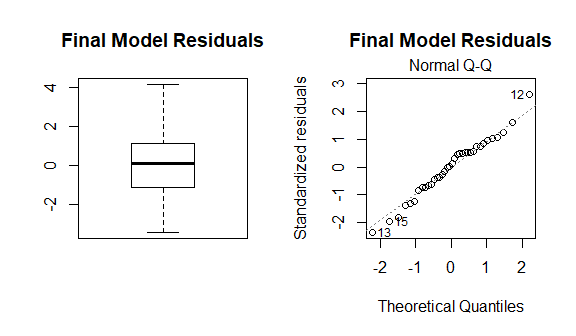
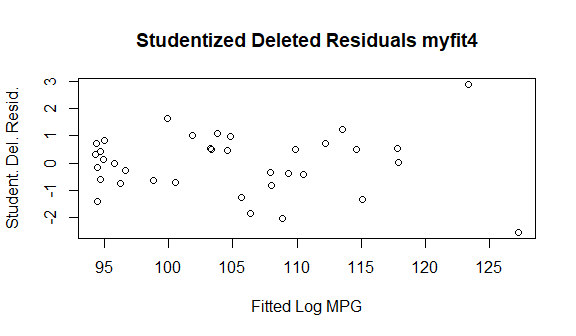


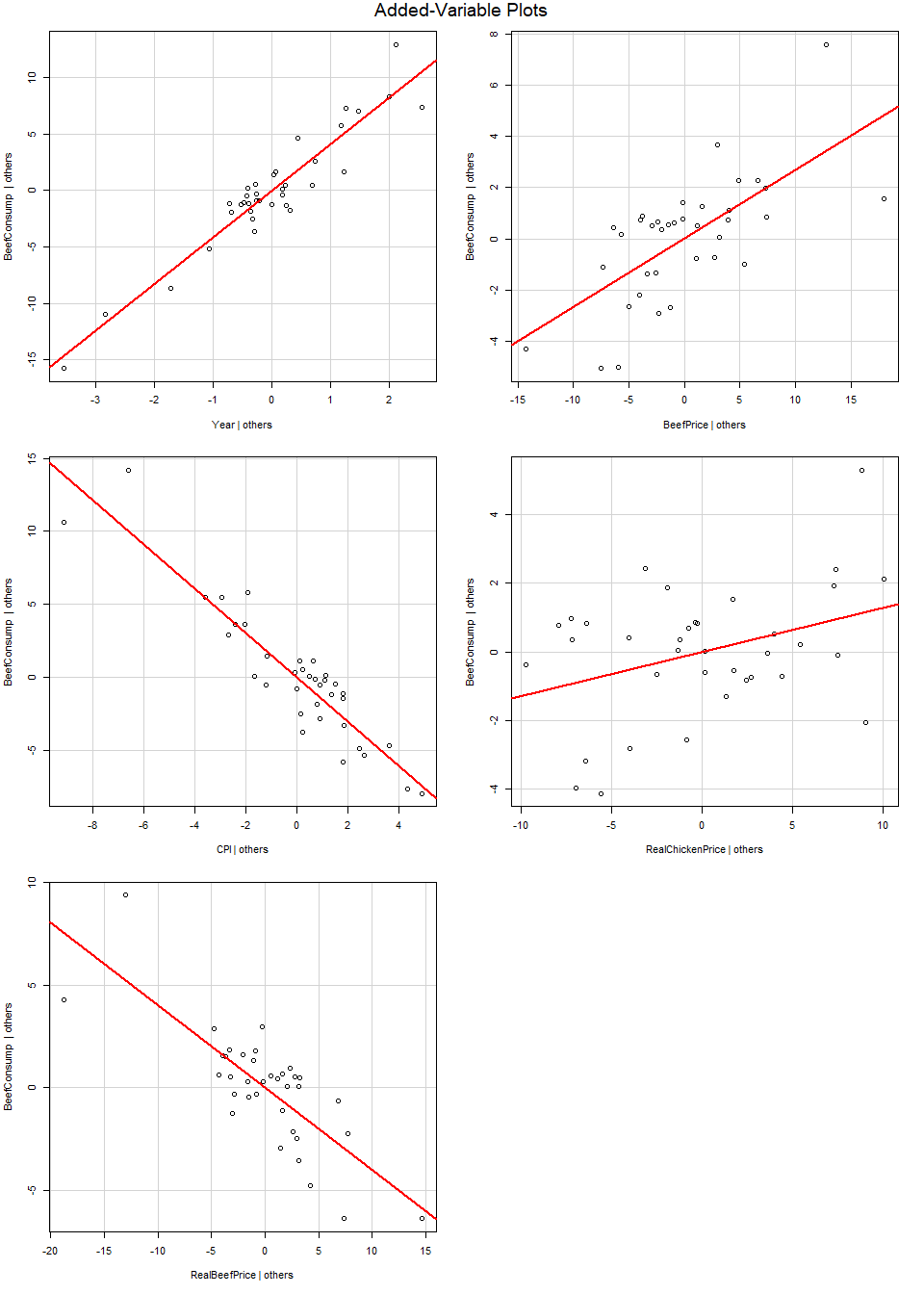
## Start: AIC=59  
## BeefConsump ~ Year + ChickPrice + BeefPrice + CPI + DPI + RealChickenPrice +   
## RealBeefPrice + RealDPI + myRealDPIsq + BeefPrice.RealBeefPrice +   
## Year.CPI  
##   
## Df Sum of Sq RSS AIC  
## - myRealDPIsq 1 0.043 95.216 57.015  
## - RealDPI 1 0.528 95.701 57.198  
## - DPI 1 0.578 95.751 57.216  
## - Year.CPI 1 2.132 97.305 57.796  
## - ChickPrice 1 2.657 97.830 57.990  
## - BeefPrice.RealBeefPrice 1 3.987 99.160 58.476  
## <none> 95.173 58.998  
## - CPI 1 11.031 106.203 60.946  
## - RealChickenPrice 1 11.761 106.934 61.193  
## - BeefPrice 1 21.657 116.830 64.379  
## - RealBeefPrice 1 34.300 129.473 68.078  
## - Year 1 46.246 141.419 71.255  
##   
## Step: AIC=57.01  
## BeefConsump ~ Year + ChickPrice + BeefPrice + CPI + DPI + RealChickenPrice +   
## RealBeefPrice + RealDPI + BeefPrice.RealBeefPrice + Year.CPI  
##   
## Df Sum of Sq RSS AIC  
## - RealDPI 1 0.968 96.184 55.379  
## - DPI 1 1.377 96.593 55.532  
## - Year.CPI 1 2.672 97.888 56.011  
## - ChickPrice 1 3.017 98.233 56.138  
## - BeefPrice.RealBeefPrice 1 3.952 99.168 56.479  
## <none> 95.216 57.015  
## + myRealDPIsq 1 0.043 95.173 58.998  
## - RealChickenPrice 1 11.742 106.958 59.201  
## - BeefPrice 1 21.615 116.831 62.380  
## - RealBeefPrice 1 34.279 129.495 66.084  
## - CPI 1 50.540 145.756 70.343  
## - Year 1 55.592 150.808 71.570  
##   
## Step: AIC=55.38  
## BeefConsump ~ Year + ChickPrice + BeefPrice + CPI + DPI + RealChickenPrice +   
## RealBeefPrice + BeefPrice.RealBeefPrice + Year.CPI  
##   
## Df Sum of Sq RSS AIC  
## - DPI 1 0.569 96.753 53.591  
## - ChickPrice 1 2.121 98.305 54.164  
## - Year.CPI 1 2.514 98.698 54.308  
## <none> 96.184 55.379  
## + RealDPI 1 0.968 95.216 57.015  
## - BeefPrice.RealBeefPrice 1 10.604 106.788 57.144  
## + myRealDPIsq 1 0.484 95.701 57.198  
## - RealChickenPrice 1 12.736 108.921 57.856  
## - BeefPrice 1 40.716 136.900 66.086  
## - RealBeefPrice 1 65.268 161.452 72.025  
## - Year 1 100.834 197.019 79.192  
## - CPI 1 120.795 216.980 82.666  
##   
## Step: AIC=53.59  
## BeefConsump ~ Year + ChickPrice + BeefPrice + CPI + RealChickenPrice +   
## RealBeefPrice + BeefPrice.RealBeefPrice + Year.CPI  
##   
## Df Sum of Sq RSS AIC  
## - ChickPrice 1 1.60 98.35 52.181  
## <none> 96.75 53.591  
## - Year.CPI 1 5.69 102.45 53.649  
## - BeefPrice.RealBeefPrice 1 10.08 106.83 55.159  
## + myRealDPIsq 1 0.99 95.76 55.220  
## + DPI 1 0.57 96.18 55.379  
## + RealDPI 1 0.16 96.59 55.532  
## - RealChickenPrice 1 12.23 108.98 55.875  
## - BeefPrice 1 40.60 137.35 64.204  
## - RealBeefPrice 1 66.45 163.21 70.414  
## - CPI 1 159.69 256.44 86.681  
## - Year 1 535.22 631.97 119.152  
##   
## Step: AIC=52.18  
## BeefConsump ~ Year + BeefPrice + CPI + RealChickenPrice + RealBeefPrice +   
## BeefPrice.RealBeefPrice + Year.CPI  
##   
## Df Sum of Sq RSS AIC  
## - Year.CPI 1 5.55 103.90 52.158  
## <none> 98.35 52.181  
## - BeefPrice.RealBeefPrice 1 9.24 107.59 53.412  
## + ChickPrice 1 1.60 96.75 53.591  
## + RealDPI 1 0.10 98.26 54.146  
## + myRealDPIsq 1 0.07 98.28 54.155  
## + DPI 1 0.05 98.31 54.164  
## - RealChickenPrice 1 18.63 116.98 56.426  
## - BeefPrice 1 39.17 137.52 62.249  
## - RealBeefPrice 1 64.86 163.21 68.415  
## - CPI 1 175.18 273.54 87.005  
## - Year 1 569.73 668.08 119.152  
##   
## Step: AIC=52.16  
## BeefConsump ~ Year + BeefPrice + CPI + RealChickenPrice + RealBeefPrice +   
## BeefPrice.RealBeefPrice  
##   
## Df Sum of Sq RSS AIC  
## - BeefPrice.RealBeefPrice 1 4.28 108.18 51.610  
## <none> 103.90 52.158  
## + Year.CPI 1 5.55 98.35 52.181  
## + myRealDPIsq 1 5.15 98.75 52.328  
## + DPI 1 4.87 99.04 52.431  
## + RealDPI 1 2.93 100.97 53.128  
## + ChickPrice 1 1.46 102.45 53.649  
## - RealChickenPrice 1 13.09 116.99 54.429  
## - BeefPrice 1 34.38 138.28 60.448  
## - RealBeefPrice 1 62.89 166.79 67.197  
## - CPI 1 172.57 276.47 85.389  
## - Year 1 868.44 972.34 130.663  
##   
## Step: AIC=51.61  
## BeefConsump ~ Year + BeefPrice + CPI + RealChickenPrice + RealBeefPrice  
##   
## Df Sum of Sq RSS AIC  
## <none> 108.18 51.610  
## + BeefPrice.RealBeefPrice 1 4.28 103.90 52.158  
## + RealDPI 1 1.84 106.34 52.991  
## + DPI 1 0.75 107.43 53.358  
## + Year.CPI 1 0.59 107.59 53.412  
## + myRealDPIsq 1 0.18 108.00 53.551  
## + ChickPrice 1 0.04 108.14 53.596  
## - RealChickenPrice 1 16.98 125.16 54.858  
## - BeefPrice 1 92.96 201.14 71.937  
## - RealBeefPrice 1 178.95 287.13 84.751  
## - CPI 1 618.34 726.52 118.171  
## - Year 1 869.13 977.31 128.846

## Analysis of Variance Table  
##   
## Model 1: BeefConsump ~ Year + BeefPrice + CPI + RealChickenPrice + RealBeefPrice  
## Model 2: BeefConsump ~ Year + BeefPrice + CPI + RealChickenPrice + RealBeefPrice  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 30 108.18   
## 2 30 108.18 0 0

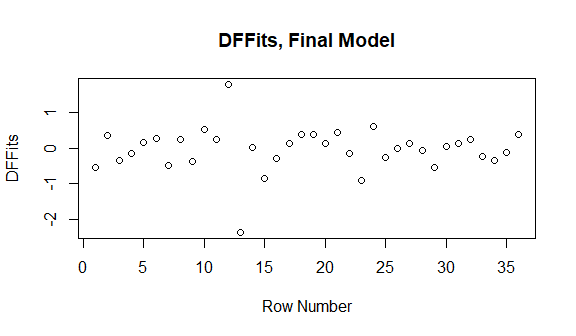
## More Analysis of Myfit4 (Final Model)

The studentized deleted residuals plot does not show any concerns. We also double checked the final model's residuals and Normal Q-Q plot - all of which look good. The added variable plots show a trend for each predictor in the model, given all of the other predictors. The directions of these trends correspond to the parameter estimates.





The DFFits plot shows some years with high DFFits values. Rows 12 and 13 had high and low predictions respectively. These years were also the furthest from normal on the Q-Q plot.



## Example Predictions

We decided to do some example predictions with 95% confidence using myfit4, step.int, and step.int.red because all of them seemed to be pretty good models. Initially, we set out asking ourselves if we could predict BeefConsumption with 80 percent significance. After realizing how small of a residual standard error we had, we decided to bump that up to 95% confidence for our prediction intervals. We ended up settling on myfit4, our model that resulted from a backwards elimination process on our original first order model, as our final model. We wanted to see if this model actually did a better job predicting BeefConsumption, compared to the model with interactions and the reduced model with interactions. We predicted response values for rows 3, 4, 23, 24, 27, 28 using all three models. Summing the absolute differences between the predictive model and the actual values serves as a way to compare relatively how useful each model is in predicting the response variable (BeefConsumption). The sum of the differences using step.int is 10.374, the sum of the differences using step.int.red is 12.234, and the sum of the differences using myfit4 is 8.989. This indicates that for predictive purposes, myfit4 is probably the most useful (although all models show strong standard error and R squared values). At the end, we used the same three models to see what the predicted values were for the outlier rows 12 and 13. All of the models seem to do a decent job predicting them, but as you can see in the DFFits, Final model plot, row 12 is underpredicted and row 13 is overpredicted using our best model myfit4. Looking back, we think the models did a good enough job predicting outliers on each end that remedial measures were not needed. The rest of our predictions are precise enough that they make 12 and 13 look like poor predictions, but considering the spread of our response variable, the model (myfit4) predicts quite well.

## Warning in predict.lm(myfit4, interval = "prediction", level = 0.95): predictions on current data refer to \_future\_ responses

## [1] 107.972 110.481 106.348 99.947 94.333 94.457

## [1] 106.500 109.700 103.300 102.799 94.856 94.144

## Warning in predict.lm(step.int, interval = "prediction", level = 0.95): predictions on current data refer to \_future\_ responses

## [1] 107.816 110.355 107.180 99.562 94.586 95.160

## [1] 106.500 109.700 103.300 102.799 94.856 94.144

## Warning in predict.lm(step.int.red, interval = "prediction", level = 0.95): predictions on current data refer to \_future\_ responses

## [1] 106.984 109.395 108.509 99.360 92.724 93.479

## [1] 106.500 109.700 103.300 102.799 94.856 94.144

## [1] 10.374

## [1] 8.989

## [1] 12.234

## [1] 123.336 127.237

## [1] 127.5 124.0

## [1] 119.943 122.406

## [1] 127.5 124.0

## [1] 122.257 126.301

## [1] 127.5 124.0