Milk and Money Case Study

Chaman Preet Kaur and Chris Atterbury

October 5, 2015

Contents

1 Executive Summary

Dairy farmers have to rely on fluctuating dairy prices for their business. While they cannot control the price, they can buy futures options to mitigate the possibilities of sustaining damaging losses when prices drop. Gerard, a dairy farmer, wants a to choose a put option, which allows him to lock in a minimum price for the products on which he purchases it, for the commodity most closely related to his mailbox price with 95% confidence that it will be in the money (i.e. gain a profit) when his mailbox price drops below \$12.50. By using simple linear regression (SLR) metrics and a correlation matrix we determine that Class III dairy most closely relates. We choose a target value for the mailbox at the top of the 95% confidence interval. Our chosen SLR model allows us to determine the strike price Class III dairy that should generate the desired mailbox (\$13.18). Finally, we include the transaction and premium fees in our desired strike price so that, when those are included, Gerard will still get his desired mailbox (\$13.18 + premium + transaction fees). Thus, he has protection against the volatility of dairy prices and can continue as a dairy farmer.

2 Background

The financial success of dairy farms depends on the price of their milk. The case study helps in understanding how milk prices are decided by the federal and state agencies, how payments are made and focuses on different options that farm owners can explore to hedge the price risk against market fluctuations. The simplest approach is to purchase an option or a futures contract from the Chicago Mercantile Exchange on commodity futures that can ensure minimum floor value for their produce. An option on a futures contract is the right, but not the obligation, to buy or sell a futures contract at a specific price on or before an expiration date.

The price a farm receives for its milk depends on various factors. What the dairy producers actually receive per hundredweight is called the mailbox price. The farmer (Gerard in this case study) needs to determine which of the options available for trade offer the best hedge for his own milk price or his own mailbox price. The federal order groups milk into four classes: Class I (defined as California Class 1), Class II (California Class 2 and Class 3), Class III (California Class 4b), and Class IV (California Class 4a). Options contracts allow hedgers to establish a price floor without giving up the opportunity to benefit from favorable price changes in the future. Buyers pay a price for the option called the premium up front. There are two different types of options:

- Put Options. The right to sell a futures contract at a certain price.

- Call Options. The right to buy a futures contract at a certain price.

The price at which the buyer has the right to buy or sell a futures contract is known as the strike price or exercise price of the option. Payoff options are described in the following ways:

- At the money. When the price of the underlying security (roughly) equals the strike price.

- In the money. When holders do gain a profit if the option is exercised. For a put option, the strike price is more than the market price of the underlying security. For a call option, it is when the strike price is less than the market price of the underlying security.

- Out of the money. When holders do not gain a profit if the option is exercised, so they allow the option to expire unused. The only loss is that of the initial premium and the trading costs of the initial transaction.

Here we use about three and a half years of data from a dairy farmer named Gerard to determine how he can purchase futures options that will be in the money if the mailbox price falls below $12.50.

3 Methods

3.1 Initial Considerations

To help Gerard with his put option deliberations, we must analyze the data that he has collected. We have feature columns *Month*, *Mailbox*, *Class.IV*, *Class.III*, *Butter*, *NFDM*. Please refer to **Codeblock 1** in the *Appendix* to see the first six rows of this data set to get a feel for it. We will not use the *Month* column since we are not doing time series analysis, but simply viewing the data as a collection of measurements of the other feature columns. Also, we must remove the dollar signs from each of the other columns so that we can treat them as numeric columns. Please see **Codeblock 2** in the *Appendix* for details on this approach, including the code itself. With the preprocessing done, we can evaluate the data.

3.2 Determine the Predictor

Gerard plans to purchase put options in just one commodity; the one that is most closely related to his mailbox price. To determine this, we will look at plots of each variable against each other (see **Figure 1** in the *Appendix*), the correlation matrix (see **Codeblock 3** in the *Appendix*), and simple linear regression (SLR) models of each of the potential predictors. We will evaluate the R^2 value, mean squared error (MSE), and distribution of the error terms (see **Codeblock 4** in the *Appendix*). The results obtained through SLR models are viewed in conjunction with the p-values of individual variables in the full multiple linear regression model (see **Codeblock 5** in the *Appendix*). Please see the *Results* section for an explanation of what we found.

3.3 Choose put options

We want to choose a put option that leaves Gerard 95% certain that he will be in the money if the mailbox price falls below $12.50. To obtain this certainty, we construct a 95% Confidence Interval and choose the upper bound as our target mailbox price (see **Codeblock 6** in the *Appendix*). Using our SLR model for the commodity most closely related to the mailbox price, we determine the strike price needed for the commodity (see **Codeblock 7** in the *Appendix*). See the *Results* section for an explanation of what we found.

3.4 Final Adjustments for Fees

Premium and trading fees are a part of purchasing dairy futures options. Thus, we finish our work with Gerard by accounting for those costs when determining the desired strike price. Since those fees are subtracted out right away, we simply increase our strike price so that, after the fees are subtracted, we still have the price of our commodity guaranteed at a high enough level to be in the money if the mailbox price falls below $12.50.

With our methods explained, we look at the results.

4 Results

4.1 Determine the Predictor

While looking at individual SLR models *Class.III* dairy displays the highest multiple R^2 value of 0.9268, followed by *Class.IV* dairy with 0.7121, *Butter* with 0.6178 and *NFDM* with 0.2961 (see **Codeblock 4** in the *Appendix*). The results are in agreement with the correlation matrix and the pairs plot (see **Figure 1** in the *Appendix*) which show highest correlation between *Mailbox* and *Class.III*, which is 0.962720 (see **Codeblock 3** in the *Appendix*). These results are again confirmed by the full multiple linear regression model (MLR) output (see **Codeblock 5** in the *Appendix*). The p-value for *Class.III* in the MLR model was the lowest ($2.73e^{-11}$) and the most significant.

In order to improve the model for *Class.III* further, we also tried transforming the data by taking a log of *Class.III*, which did not make a significant difference to the Multiple R^2 and returned a value of 0.9462. Taking the log of *Mailbox* instead returned a multiple R^2 of 0.8949 (see **Codeblock 7**, **Figure 2**, **Figure 3**, and **Figure 4** in the *Appendix*). Thus, we considered the simple linear model without any transformation for calculation purposes. All these approaches confirmed hierarchical order that *Class.III* was the most related to the mailbox price followed by *Class.IV*, *Butter*, and *NFDM.*

4.2 Choose put options

In an endeavour to choose the correct strike price, the construction of the 95% confidence interval using $\hat{Y}_h = 12.50$ yielded (12.48889, 12.51111) as the lower bound and upper bound values. We use the upper bound as our target mailbox price making Gerard 95% sure that he is in the money if mailbox price falls below $12.50. Using a mailbox price of 12.51111, we find the Class III dairy price from our previous SLR model developed while calculating the most closely related commodity to the mailbox price. This yields a *Class.III* price of $13.18.

4.3 Final Adjustments for Fees

The premium and transactional fees are constants and can be added to the derived strike price in order to accomodate for extra cost while keeping the option in the money. Hence, we arrive at final strike price by adding the extra premium and transactional costs.

5 Conclusions

5.1 General Conclusion

The historic data provided in the case study helps Gerard make an informed decision about the put option he should buy. Linear regression modeling on historic data helps in predicting strike price for put options that can be bought from the Chicago Mercentile Exchange despite of the fact that California milk agencies calculate their prices differently. This contributes significantly to the financial success of small scale dairy farmers like Gerard.

5.2 Initial Considerations

We started by looking at how SLR models work in predicting commodity values when the Mailbox is fixed at \$12.50, since this is the value that garners Gerard's interest (see **Codeblock 9** in the *Appendix*)(**Answer 1**). The values obtained for different products are as folllows:

Class.IV = \$11.99

Class.III = \$13.16

Butter = \$1.38

NFDM = \$0.54

Comparing these predicted values with similar actual values in the data set, we see that *Class.III* and *Class.IV* appear to correspond the best.

5.3 Determine the Predictor

With this in mind, we see which predictor is actually most closely related to the mailbox price (**Answer 2**). As detailed above in the *Methods* and *Results* sections, the Class III dairy commodities appear to be the most closely related to the mailbox price (See **Codeblock 3** for the correlations matrix, **Codeblock 4** for models, **Figure 1** for the pairs plots, and **Figure 2** for the scatter plot of *Class.III* vs *Mailbox*; all in the *Appendix*). With strong correlation, the ability to explain much of the variance in the mailbox price, and the smallest error terms, Class III is the clear choice here.

5.4 Choose Put Options

Gerard wants to be in the money when the mailbox price falls below \$12.50. Using the methods detailed above, we found that a strike price of \$13.18 was the best for our put option so that the predicted mailbox price for the month is at the upper bound of our 95% Confidence Interval. Thus, we can be 95% confident that the option will be in the money if the mailbox price falls below \$12.50 (**Answer 3**).

Specifically, if the mailbox price falls to \$11.50, we see that the Class III option should be in the money by about \$1.29 and the mailbox price is predicted to be in the money by about \$1.01 (this does not include fees; see **Codeblock 10** in the *Appendix* for code details). Based on the way that we came to our put option purchase price for Class III, we can be 95% confident that the net price will exceed \$12.50 (**Answer 4**).

5.5 Final Adjustments for Fees

All put option purchases include premium and transactional fees. Since these will come right out of the strike price, we just account for this when purchasing the option. Thus, we conclude that we want to purchase our Class III put option at a value of \$13.18 + premium + transaction fees (**Answer 5**).

6 Appendix

6.1 Codeblocks

Codeblock 1

```
milk <- read.csv("41330727.csv", header = TRUE, stringsAsFactors = FALSE)
head(milk)
```

```
##         Month Mailbox Class.IV Class.III  Butter    NFDM
## 1  January-04  $11.64   $10.97    $11.61 $1.4320 $0.8366
## 2 February-04  $12.25   $12.21    $11.89 $1.7132 $0.8413
## 3    March-04  $14.31   $14.10    $14.49 $2.1350 $0.8518
## 4    April-04  $17.04   $14.57    $19.66 $2.2204 $0.8808
## 5      May-04  $17.29   $14.50    $20.58 $2.0363 $0.9050
## 6     June-04  $16.14   $13.72    $17.68 $1.9300 $0.9188
```

Codeblock 2

In **Codeblock 1** we anticipated the need to convert the numeric feature columns. When being read into the data.frame `milk`, those columns are normally interpretated as factors because of the dollar signs. By reading them in as strings, we can use a regular expression to remove all of the dollar signs, then convert the columns from character to numeric types. See below for the code doing this.

```
for (i in 2:length(milk)) {
  milk[[i]] <- gsub("\\$", "", milk[[i]])
  milk[[i]] <- as.numeric(milk[[i]])
}
```

Codeblock 3

```
cor(milk[2:length(milk)])
```

```
##             Mailbox  Class.IV Class.III     Butter       NFDM
## Mailbox   1.0000000 0.8438798 0.9627200  0.7859903  0.1720781
## Class.IV  0.8438798 1.0000000 0.8155221  0.5512816  0.5608375
## Class.III 0.9627200 0.8155221 1.0000000  0.7231414  0.2232973
## Butter    0.7859903 0.5512816 0.7231414  1.0000000 -0.2901637
## NFDM      0.1720781 0.5608375 0.2232973 -0.2901637  1.0000000
```

Codeblock 4

```
lm_classiv <- lm(Mailbox ~ Class.IV, data = milk)
print (summary(lm_classiv))
```

```
##
## Call:
## lm(formula = Mailbox ~ Class.IV, data = milk)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max 
## -2.2378 -0.6067 -0.1703  0.5040  2.4422 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  1.25861    1.21771   1.034    0.308    
## Class.IV     0.93719    0.09541   9.822 4.24e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.9673 on 39 degrees of freedom
## Multiple R-squared:  0.7121, Adjusted R-squared:  0.7048 
## F-statistic: 96.48 on 1 and 39 DF,  p-value: 4.24e-12
```

```
lm_classiii <- lm(Mailbox ~ Class.III, data = milk)
print (summary(lm_classiii))
```

```
## 
## Call:
## lm(formula = Mailbox ~ Class.III, data = milk)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.02598 -0.30017 -0.01168  0.26231  1.16295 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  2.18135    0.49832   4.377 8.72e-05 ***
## Class.III    0.78400    0.03527  22.226  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.4877 on 39 degrees of freedom
## Multiple R-squared:  0.9268, Adjusted R-squared:  0.925 
## F-statistic:   494 on 1 and 39 DF,  p-value: < 2.2e-16
```

```
lm_butter <- lm(Mailbox ~ Butter, data = milk)
print (summary(lm_butter))
```

```
## 
## Call:
## lm(formula = Mailbox ~ Butter, data = milk)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -1.9798 -0.6459 -0.1431  0.6679  3.3427 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   5.5167     0.9742   5.663 1.52e-06 ***
## Butter        5.0459     0.6355   7.940 1.14e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.115 on 39 degrees of freedom
## Multiple R-squared:  0.6178, Adjusted R-squared:  0.608
## F-statistic: 63.04 on 1 and 39 DF,  p-value: 1.14e-09
```

```
lm_nfdm <- lm(Mailbox ~ NFDM, data = milk)
print (summary(lm_nfdm))
```

```
##
## Call:
## lm(formula = Mailbox ~ NFDM, data = milk)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9163 -1.3683  0.0148  0.8255  4.3243
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.822      1.228   9.625 7.48e-12 ***
## NFDM           1.264      1.159   1.091    0.282
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.776 on 39 degrees of freedom
## Multiple R-squared:  0.02961,    Adjusted R-squared:  0.004729
## F-statistic:  1.19 on 1 and 39 DF,  p-value: 0.282
```

Codeblock 5

```
lm_full <- lm(Mailbox ~ ., data = milk[-1])
print (summary(lm_full))
```

```
##
## Call:
## lm(formula = Mailbox ~ ., data = milk[-1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.61028 -0.22891  0.04805  0.17654  1.10697
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.13895    0.48970   2.326  0.02578 *
## Class.IV     0.39648    0.11167   3.550  0.00109 **
## Class.III    0.53240    0.05632   9.453 2.73e-11 ***
## Butter       0.46185    0.54053   0.854  0.39851
## NFDM        -1.12574    0.58859  -1.913  0.06378 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3713 on 36 degrees of freedom
## Multiple R-squared:  0.9608, Adjusted R-squared:  0.9565
## F-statistic: 220.8 on 4 and 36 DF,  p-value: < 2.2e-16
```

Codeblock 6

The code below uses the following equation to determine the upper bound of the 95% Confidence Interval: $\hat{Y}_h + t_{(1-\alpha/2;n-2)} s\{\hat{Y}_h\}$

```
s_2 <- anova(lm_classiii)[["Mean Sq"]][[2]]
x_bar <- mean(milk[["Class.III"]])
x_h <- (12.50 - coef(lm_classiii)[[1]]) / coef(lm_classiii)[[2]]
num <- (x_h - x_bar) ^ 2
denom <- sum((milk[["Class.III"]] - x_bar) ^ 2)
val <- 1 / 41 + num / denom
s <- s_2 * val
mbox_upper <- 12.50 + s * qt(0.95, 39)
print (mbox_upper)
```

```
## [1] 12.51111
```

Codeblock 7

```
b_0 <- coef(lm_classiii)[[1]]
b_1 <- coef(lm_classiii)[[2]]
classiii_price <- (mbox_upper - b_0) / b_1
print (classiii_price)
```

```
## [1] 13.1758
```

Codeblock 7

```
lm_lclassiii <- lm(Mailbox ~ log(Class.III), data = milk)
print (summary(lm_lclassiii))
```

```
##
## Call:
## lm(formula = Mailbox ~ log(Class.III), data = milk)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7134 -0.2889 -0.0434  0.1699  1.0264
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -17.1652     1.1587  -14.81   <2e-16 ***
## log(Class.III)   11.5397     0.4407   26.18   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4183 on 39 degrees of freedom
## Multiple R-squared:  0.9462, Adjusted R-squared:  0.9448
## F-statistic: 685.6 on 1 and 39 DF,  p-value: < 2.2e-16
```

```
lm_classiiil <- lm(log(Mailbox) ~ Class.III, data = milk)
print (summary(lm_classiiil))
```

```
##
## Call:
## lm(formula = log(Mailbox) ~ Class.III, data = milk)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10438 -0.03004  0.00229  0.02740  0.09311
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.745586   0.045550   38.32   <2e-16 ***
## Class.III   0.058742   0.003224   18.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04458 on 39 degrees of freedom
## Multiple R-squared:  0.8949, Adjusted R-squared:  0.8922
## F-statistic: 331.9 on 1 and 39 DF,  p-value: < 2.2e-16
```

Codeblock 9

```
x_classiv <- (12.50 - coef(lm_classiv)[[1]]) / coef(lm_classiv)[[2]]
print (x_classiv)
```

```
## [1] 11.9948
```

```
x_classiii <- (12.50 - coef(lm_classiii)[[1]]) / coef(lm_classiii)[[2]]
print (x_classiii)
```

```
## [1] 13.16162
```

```
x_butter <- (12.50 - coef(lm_butter)[[1]]) / coef(lm_butter)[[2]]
print (x_butter)
```

```
## [1] 1.383944
```

```
x_nfdm <- (12.50 - coef(lm_nfdm)[[1]]) / coef(lm_nfdm)[[2]]
print (x_nfdm)
```

```
## [1] 0.5365911
```

Codeblock 10

```
b_0 <- coef(lm_classiii)[[1]]
b_1 <- coef(lm_classiii)[[2]]
classiii_newprice <- (11.50 - b_0) / b_1
print (classiii_newprice)
```

```
## [1] 11.8861
```

```r
diff_classiii <- classiii_price - classiii_newprice
print (diff_classiii)
```

```
## [1] 1.289694
```

```r
diff_mbox <- mbox_upper - 11.50
print (diff_mbox)
```

```
## [1] 1.011115
```

6.2 Figures

Figure 1

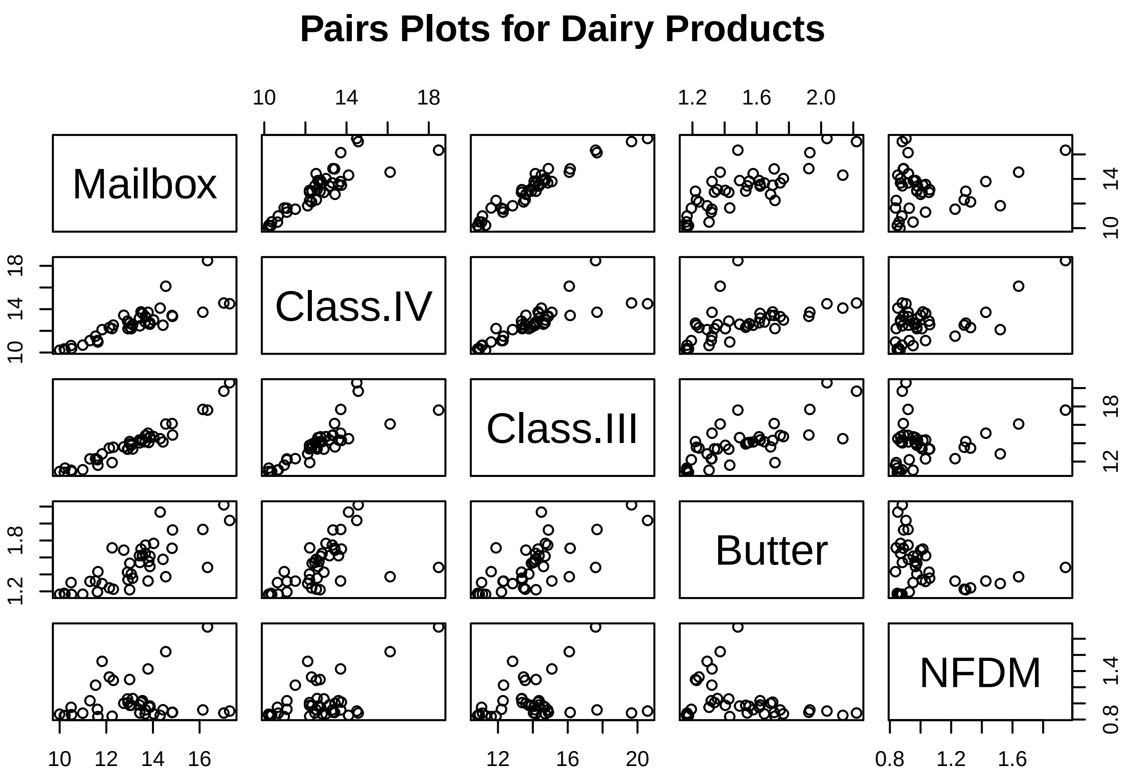


Figure 2

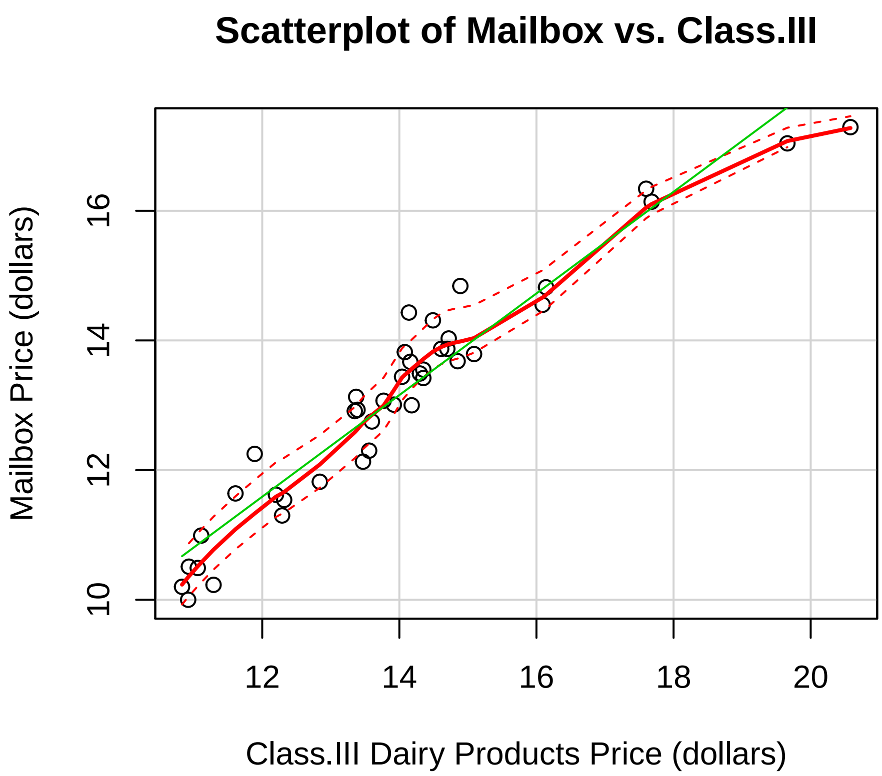


Figure 3

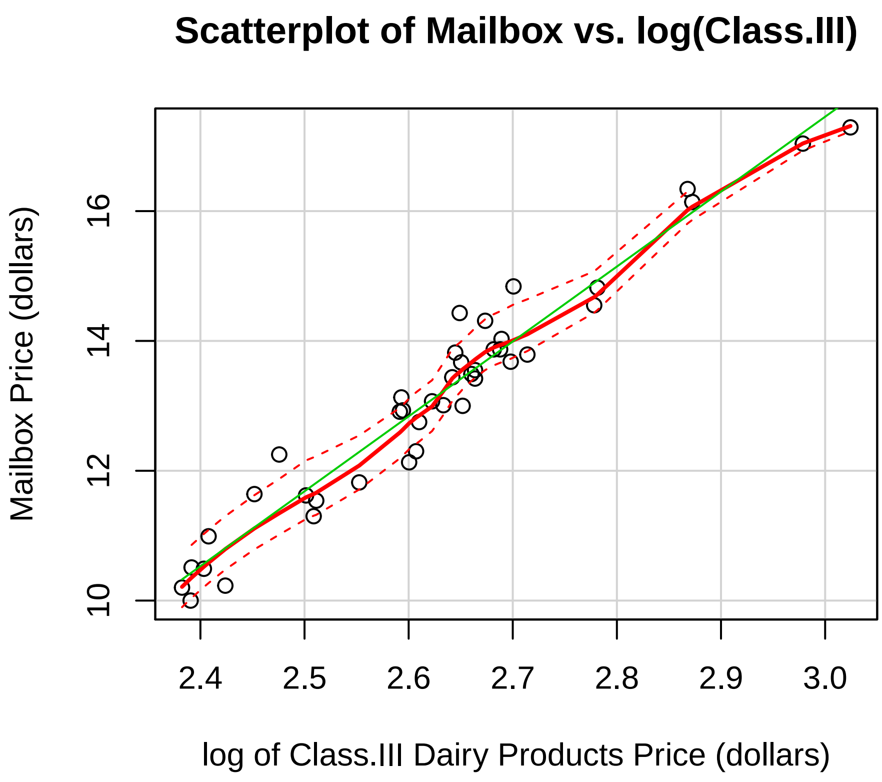
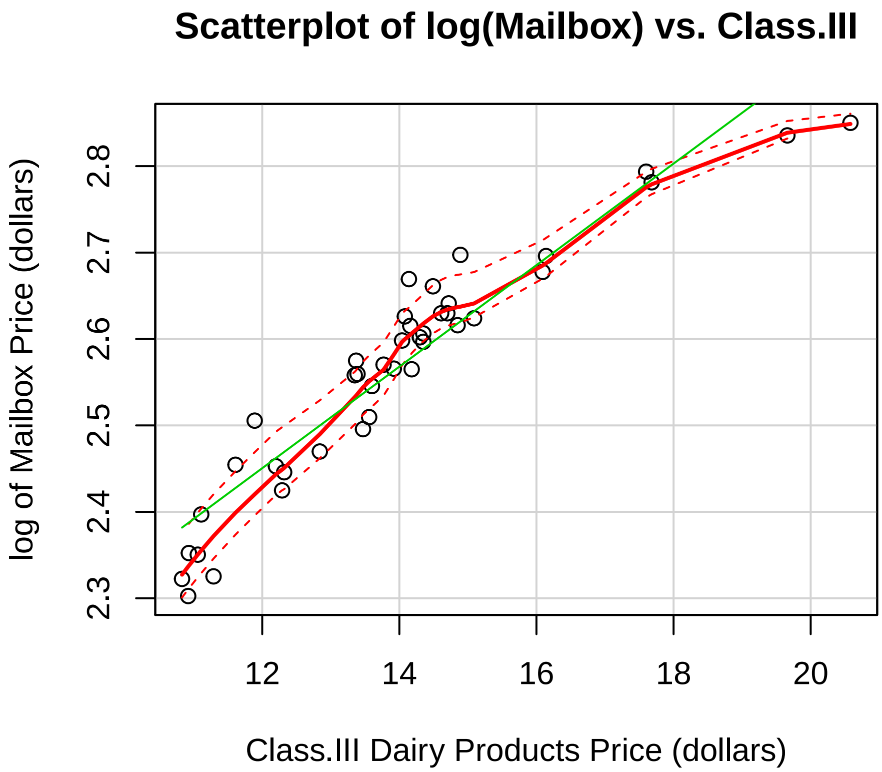


Figure 4


Scatterplot of log(Mailbox) vs. Class.III
log of Mailbox Price (dollars)
2.3
2.4
2.5
2.6
2.7
2.8
12
14
16
18
20
Class.III Dairy Products Price (dollars)