Case study - Milk and Money

Felipe Chamma Felipe Ferreira

October 7, 2015

1 Executive summary

In this case we'll study how the dairy farmer Gerard van der Lann can overcome milk price fluctuation using put options in order to hedge his business. This will allow him to lock in a floor price without giving up upside price gains.

When using this strategy, the farmer should buy preferably Class III milk put options in the futures exchange - as this category is the one that most closely relates to the California mailbox price. It is possible to fit a linear regression model in which the Class III price can be predicted from the mailbox price with a 92.68% accuracy - the highest among the 4 classes available. By doing that, Gerard will be able to lock in a floor price for his earnings with a relatively low risk.

The main concern of the farmer is to ensure a payoff when his mailbox falls below $12.50 per cwt. For that price level, we can predict the Class III price to be between $12.96 and $13.48 with 90% confidence, using the regression model mentioned above. If the farmer buys the option at a $13.50 strike, there is roughly a 95% chance that it will be in the money if the mailbox falls below $12.50 - which guarantees his target floor price.

However, the premium paid for the option, along with trading fees, must also be taken into account for the net payoff. In that case, the actual strike price would have to be $13.50 plus premium and fees. Besides, he also must have in mind that put options with a $ 13.50 strike simply won't be affordable if the current commodity price is considerably below that level.

2 Predicting prices of different milk classes

To illustrate the problem we'll try to predict all milk classes from Chicago Mercantile Exchange using Californian Mailbox Price. From the data available for 41 months we'll run a single linear regression for each one of them and use the linear fitted line as a predictor.

Some of the classes, however, do not have a decent correlation with the mailbox price, even when using their logs, squared roots or quadratic terms. In this sense, the predicted price will be tied to a high uncertainty (and a wider confidence interval).

In the table below, prices were predicted for each class in a scenario where the mailbox was \$12.50. The columns show the price prediction and associated R^2 statistic (how well the model fits the data, on a 0-1 scale):

Table 1: Price predictions

Category	Predicted price	R^2
Class III	\$13.22	0.9268
Class IV	\$12.19	0.7121
Butter	\$1.43	0.6170
NFDM	\$4.24	0.0325

3 Choosing commodity that best relates to mailbox

The data in the table above is a first indication of the commodity that best relates to the mailbox price, since a model with a higher coefficient of adjustment (R^2) is desired to predict future commodity prices.

However, a few more tests need to be run to assess the most appropriate commodity to pick. Analyzing plots is an essential step in order to determine if a regression model is accurately capturing the overall trend in the data, and if there are any issues with the residuals.

Plots for all milk classes against the mailbox price can be seen in the appendix, as well as their residual plots. The clear winner is the linear model for Class III prices, whose scatter plot can be seen in figure 1.

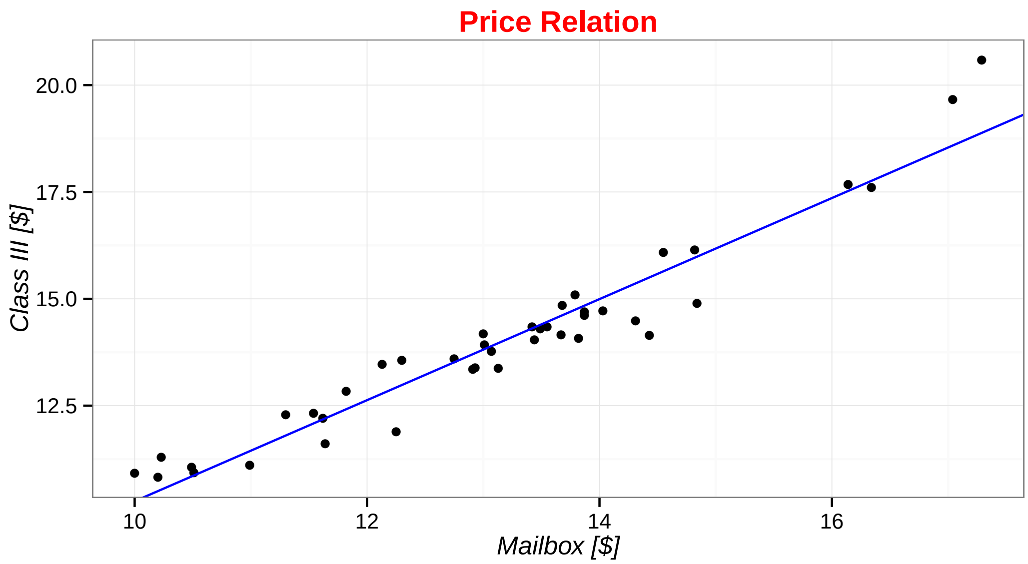

Figure 1: Scatter plot for Mailbox prices and Class III milk prices

Attempting models using transformations in the variables

So far we discussed the fit of the linear model without any transformations on the variables. It is possible, for example, that taking the logarithm (or the square) of the response variable would improve the overall fit of the model.

Indeed, when running a `box-cox` transformation on the response variable, we can see that the optimal fit happens when Y is raised to $\lambda = -0.42$. The plot in figure A.6 shows how the sum of residual errors is distributed along the different powers we raise Y and how it reaches a minimum at -0.42.
The coefficient of determination (R^2) in that situation increases from 0.9268 to 0.9472, and other statistics such as the F-test also indicate a better adjustment.

However, this model has no practical meaning. It doesn't feel right to raise the price of a commodity to a fractional power to predict the price of an essentially similar commodity.

Since the number of data points in the model is not very large (41), it may be possible that a couple of outliers caused that transformation effect, instead of having a linear transformation as the ideal model.

Also, if we wanted to extrapolate the values for the prediction and estimate the Class III price for a \$1000 mailbox price, it is more likely that the linear transformation model will have a better estimate than the one where \$1000 is raised to -0.42.

For the reasons mentioned above, we decide to keep the linear transformation model as the most appropriate for predicting Class III milk prices.

Residual tests for the selected model

The residual plot for the chosen model (figure A.2 in the appendix) shows that there aren't any apparent heteroskedasticity issues. Hence we can continue the analysis of whether Class III milk is an appropriate hedge for mailbox prices.

Because our data was collected in a monthly basis it is also a good practice to inspect the presence of seasonality and autocorrelation. From figure 2, it is possible to notice a seasonal pattern in the residuals.

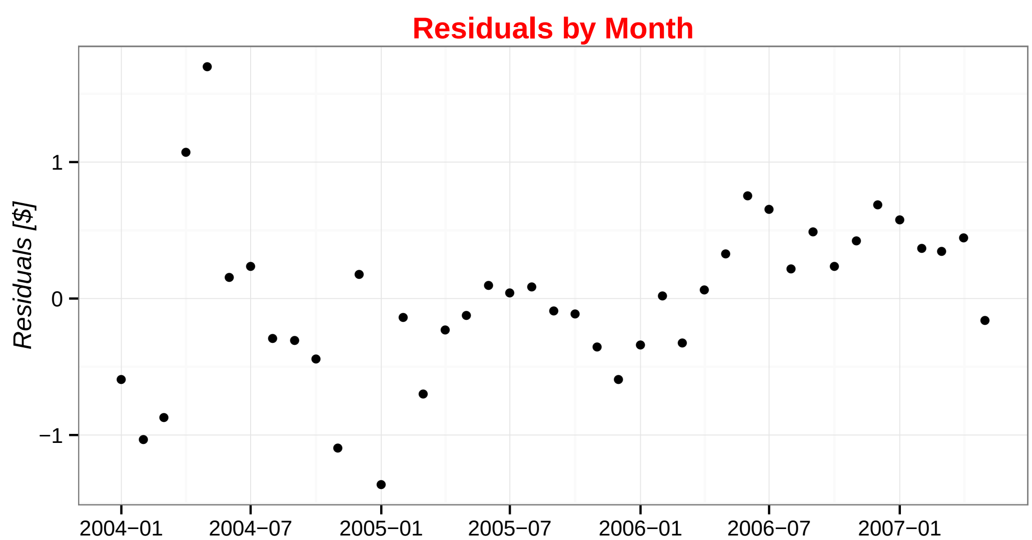

Figure 2: Monthly residuals from Mailbox prices regressed on Class III milk.

The seasonality is also very clear when we look at Figure A.7 in the appendix section.

Besides the seasonality issue, the plot also appears to have autocorrelation. It seems that previous observations have an impact on future ones. To be sure of that a Durbin Watson test was performed.

The test confirmed our intuition and yielded a D = 1.095, which falls below dL = 1.44. That indicates that the error terms have a positive autocorrelation with the terms right before them, with value $\rho = 0.4398$. To adjust for this issue, ρ has to be included in the previous simple regression model associated to the previous error term ϵ_{t-1}:[1]

$$\hat{Y} = -1.557 + 1.182 \cdot X_i + \epsilon_t$$
$$\epsilon_t = 0.4398 \cdot \epsilon_{t-1} + u_t$$
$$u_t \sim N(0, \sigma^2)$$

4 Confidence interval for option strike price

On the previous section we arrived at a single linear regression model with an autocorrelation term as the best fit. The goal of this section is to use this model to estimate the most precise price for Gerard to make his hedge.

Because he plans on hedging six months down the road, and since we do not have the error term for a future period, we will drop the correlation term for this analysis. This leaves us with the simplified model for Class III prices ($\hat{Y}$) in terms of mailbox prices (X_i):

$$\hat{Y} = -1.557 + 1.182 \cdot X_i$$

When $X_i = 12.50$, then $\hat{Y} = 13.22$ - which is the point estimator for Class III prices.

The 90% confidence interval for this prediction indicates that there is a 5% chance of the actual value falling above the upper bound (and 5% of it falling below the lower bound). Thus, the option would be in the money with 95% confidence for a strike price equal to the upper bound value.

The upper bound of the prediction interval, based on $p = 0.1$ and 39 degrees of freedom, yields to the ideal strike price of \$13.48. Since the options are traded in intervals of \$0.25, Gerard should aim for a strike of \$13.50 on Class III milk. This would give him a 95% certainty of \$12.50 minimum return (not accounting for the premium or trading costs).

5 Point estimate for the option value

The value of the put option is given by the difference between the strike price and the commodity price. By using the regression model to predict the Class III price when the mailbox is at \$11.50, we have that:

$\hat{Y} = -1.557 + 1.182 \cdot \$11.50 = \$12.04$

Assuming that the farmer got the option for the suggested strike price of \$13.50, the value

[1] The R code for such tests and ρ calculations can be found on the appendix.

of the put option yields to $\$13.50 - \$12.04 = \$1.46$.

The 90% prediction for the Class III milk price when the mailbox is at $11.50 is [$11.78, $12.30]. The worst case scenario within this interval is the Class III price at the $12.30 upper bound, which would limit the option payoff to $\$13.50 - \$12.30 = \$1.20$. There is a 5% chance that the option payoff would be less than $1.20.

However, even with a $1.20 payoff, the net price would be $\$11.50 + \$1.20 = \$12.70$, which exceeds $12.50 (before premium and fees). Therefore, Gerard can be 95% assured that his net price will exceed the target of $12.50 in this scenario.

6 Including option premium and trading fees

Including the option premium and the trading fees is vital to the analysis of this hedge strategy for a few reasons.

First, the real option payoff is only given after discounting these costs. The net price of $12.50 would be achieved with 95% confidence only if the option strike price was $13.48 plus the premium and trading fees.

Second, buying an option for a certain strike price is probably not feasible if the current spot price is below that price. For example, if a commodity is selling for $10 in the futures market, it would cost way too much to acquire the right to sell it for $13.50, and the target net price becomes unachievable. Therefore, Gerard can only apply his hedge strategy on months where the commodity price is close or above his target floor.

Appendix

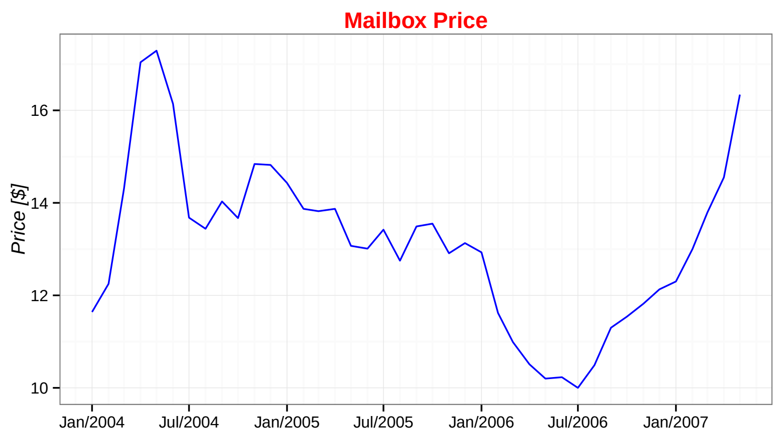

Figure A.1: California Mailbox price from January 2004 to May 2007.

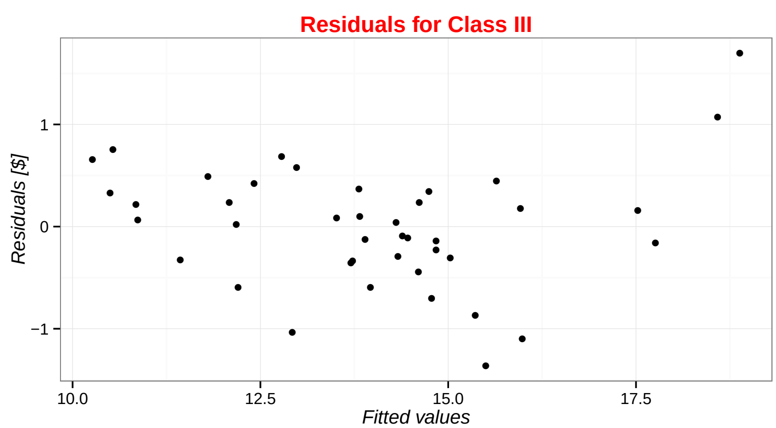

Figure A.2: Residuals against Fitted values for Class III as a regressor.

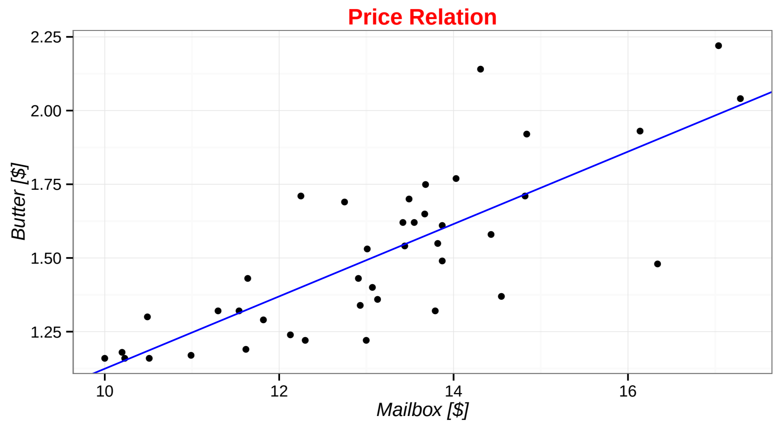

Figure A.3: Scatter plot for Mailbox prices and Butter prices.

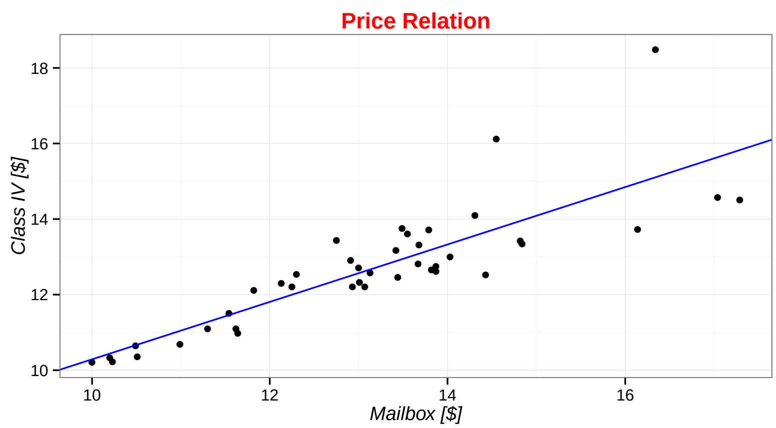

Figure A.4: Scatter plot for Mailbox prices and Class IV milk prices.

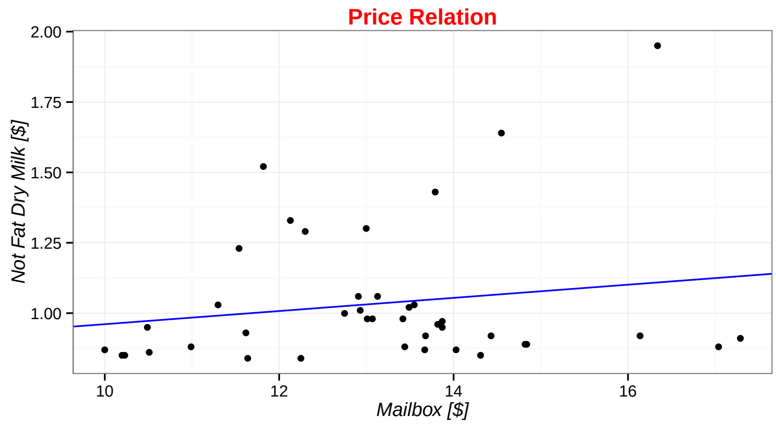

Figure A.5: Scatter plot for Mailbox prices and Non Fat Dry Milk prices

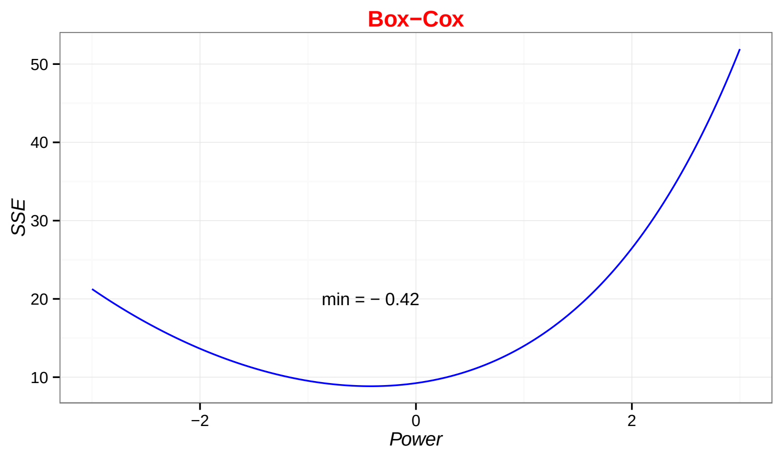

Figure A.6: Box cox plot for various powers

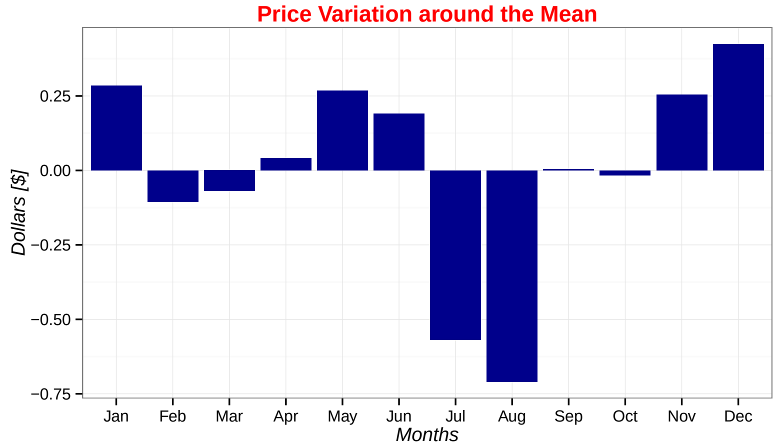

Figure A.7: Monthly variations around the average price.[2]

R code for autocorrelation

```
df <- master_df[,c("Mailbox","Class.III")]
lm_model <- lm(Class.III ~ Mailbox, data = df)
dwtest(lm_model)
n <- nrow(master_df)
ei <- master_df$lm_residuals[-n]
ei_lag <- master_df$lm_residuals[-1]
rho <- sum(ei*ei_lag)/sum(ei ^ 2)
```

[2]Only values from June 2004 to May 2007 were taking into account. This way we have exactly 3 full years