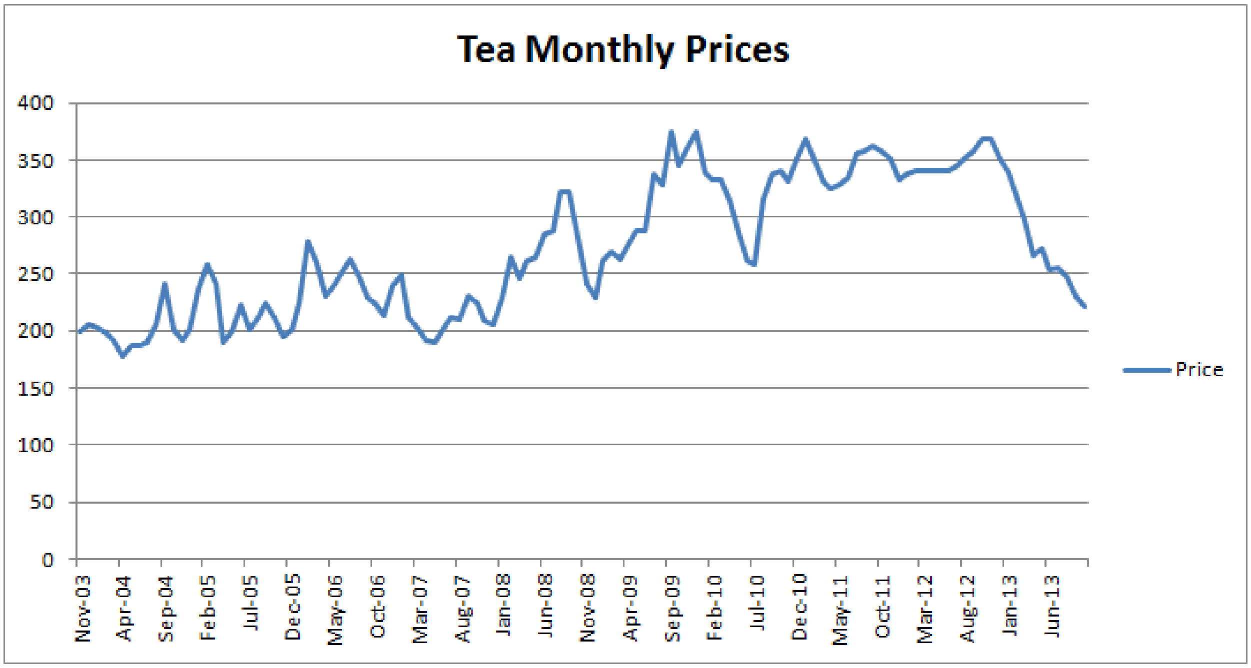
Sae Hee Kim
April 2, 1981
VEE Time Series Student Project
Session: Fall 2012

**Introduction**

I am expecting my first child and one of the changes that I had to make was to cut my consumption of caffeine. This meant I had to cut out coffee and I had to find a good substitute. I found it in a cup of herbal tea. I was never a big fan of tea, but now I came to enjoy the morning cup of tea before starting my day. Therefore, I decided to examine the price of tea over time for my Time Series project.

**Data**

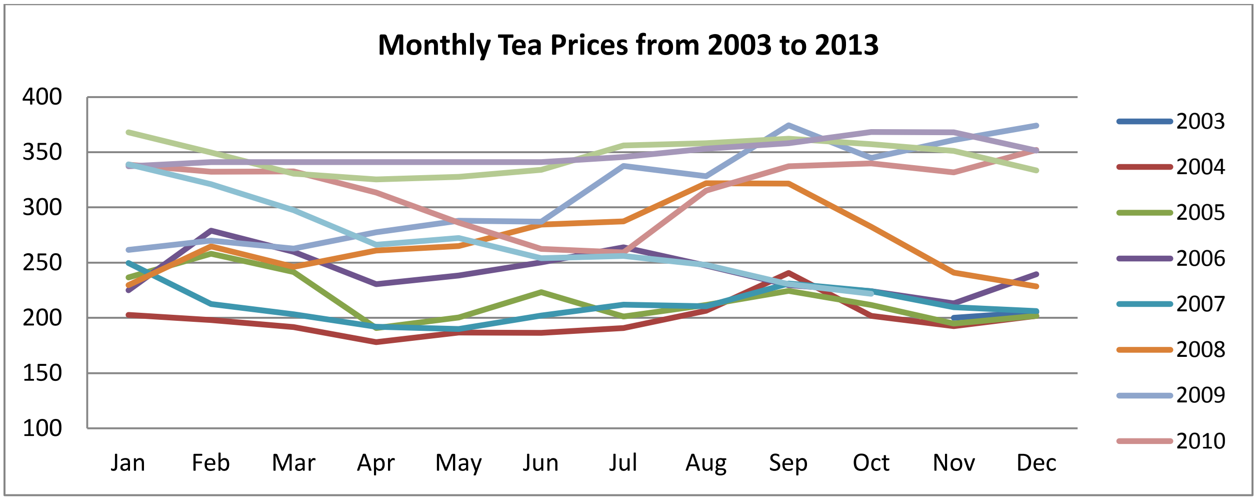
I obtained my data from the Index Mundi website at http://www.indexmundi.com/commodities/?commodity=tea&months=120 and the data that I collected was from November 2003 to November 2013. This data contains the average monthly price in US cents per kilogram.




The prices range from a low of 177.95 cents and high of 374.41 cents. From the graph, you see the price has been rising overall until the end of 2012 and decreasing in 2013.

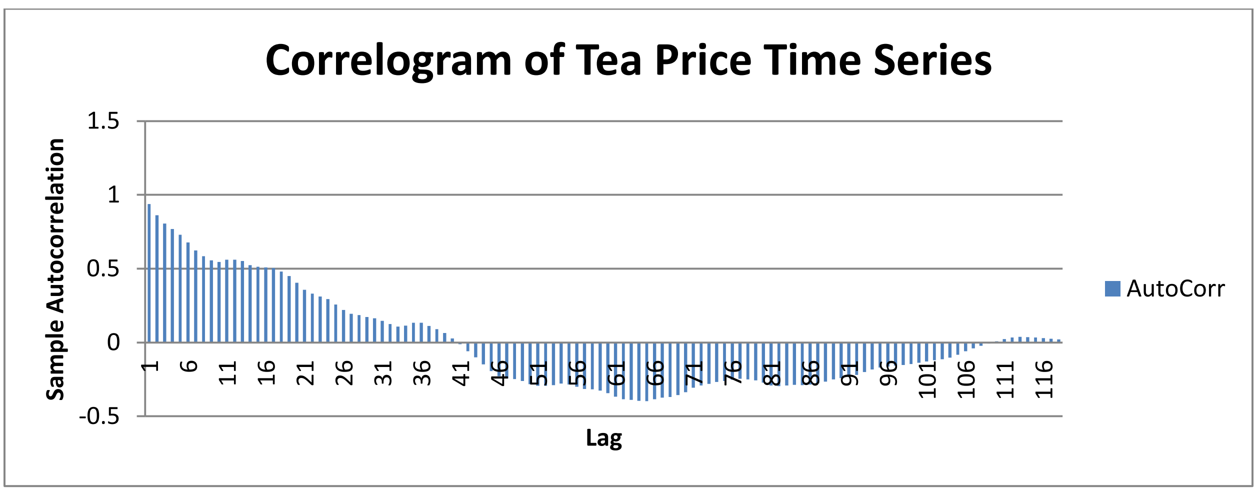
**Seasonality**

To check further for seasonality, I graphed each of the years by month to observe any monthly trends in the graph below. There do not appear to be any seasonal trends that will require adjustments to the data.




**Analysis**

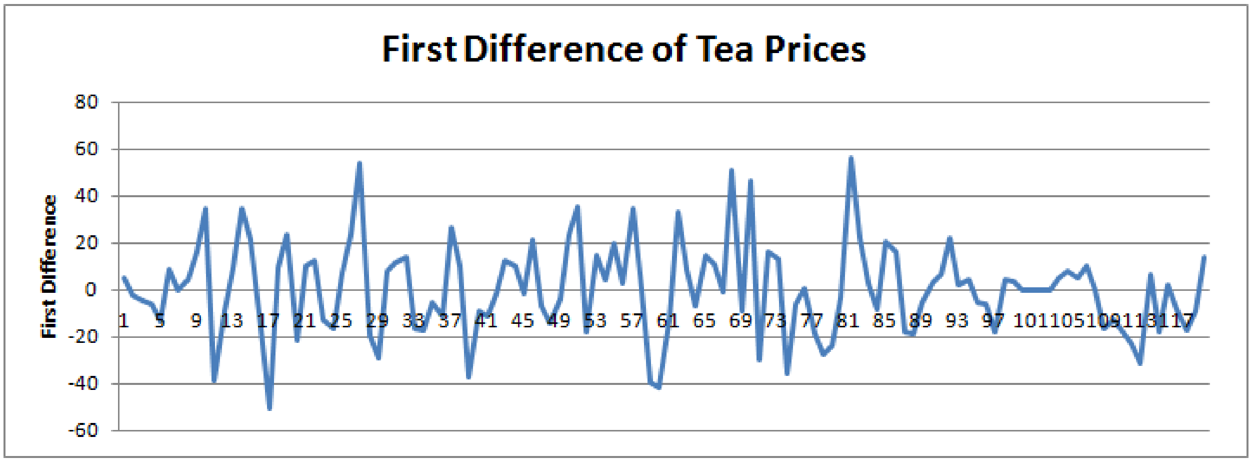
Firstly, it must be determined if the data is stationary. In order to do so, we look at the sample autocorrelations by lag period.




We look for a correlogram that declines to zero after several lags to demonstrate stationarity. The correlogram above does not fall to zero quickly and when it does drop to zero at lag 40, it does not stay zero at subsequent lags with minimum fluctuations. It decreases until negative, then rises back up. Thus the series could be represented by an AR(1) or AR(2) process.

**First Difference**

Below is the graph of the first difference with lags. It does not show any particular trend.




The regression results are as follows:

**SUMMARY OUTPUT**

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.944742 |
| R Square | 0.892537 |
| Adjusted R Square | 0.891627 |
| Standard Error | 19.53097 |
| Observations | 120 |

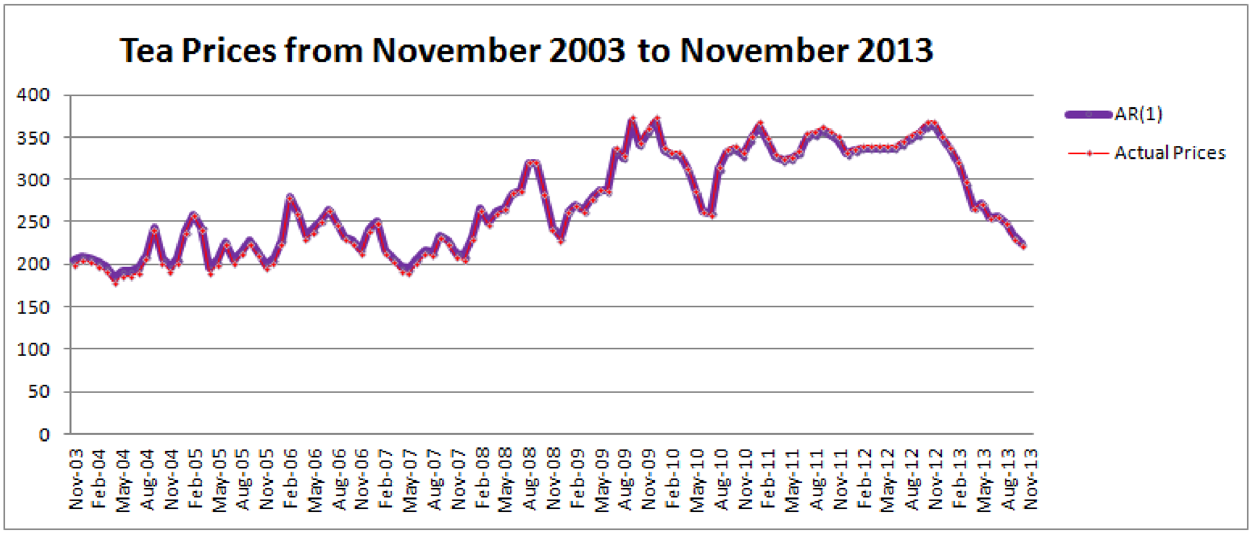
ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 373850.6 | 373850.6 | 980.0548 | 5.41E-59 |
| Residual | 118 | 45012.15 | 381.4589 | | |
| Total | 119 | 418862.8 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 16.51733 | 8.37102 | 1.97315 | 0.05081 | -0.05959 | 33.0942 | -0.05959 | 33.0942 |
| X Var 1 | 0.940414 | 0.03004 | 31.3058 | 5.41E-59 | 0.88092 | 0.999901 | 0.880928 | 0.999901 |

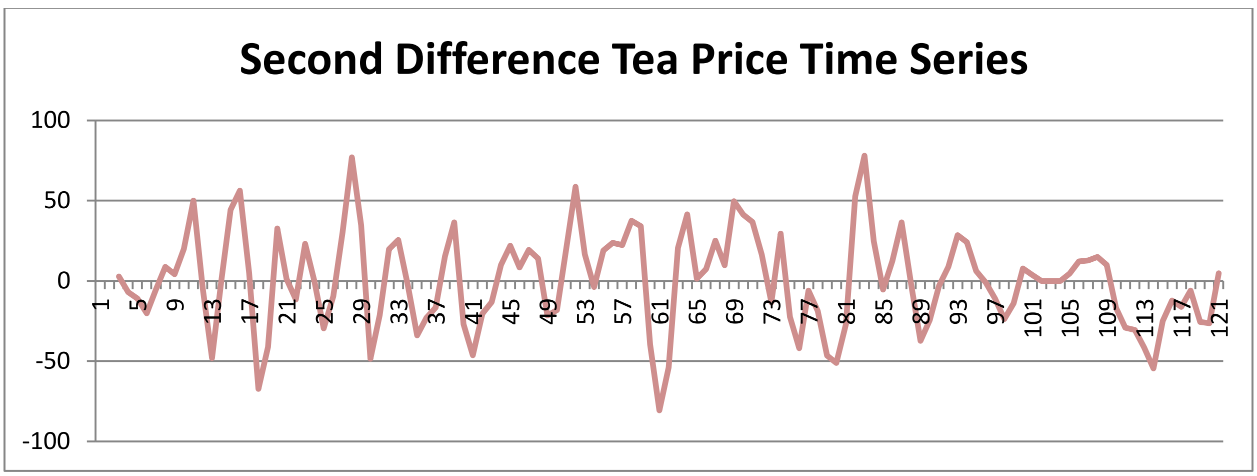
The result of the equation is $\mathbf{Y_t=16.51733+0.940414Y_{t-1}}$. The $R^2$ and adjusted $R^2$ values are fairly high, so this is a good model. Significance value is close to zero and it indicates that regression is significant.

Below is the predicted AR(1) prices and the original prices.




The AR(1) is almost a perfect fit. Second Difference will be examined to see if it provides better fit.

**Second Difference**




Again, the graph above, the second difference with lags, does not show any particular trend.

The AR(2) regression is as follows:

**SUMMARY OUTPUT**

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.945314 |
| R Square | 0.893618 |
| Adjusted R Square | 0.891784 |
| Standard Error | 19.49198 |
| Observations | 119 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 2 | 370216.4114 | 185108.2 | 487.2072 | 3.62E-57 |
| Residual | 116 | 44072.72747 | 379.9373 | | |
| Total | 118 | 414289.1388 | | | |

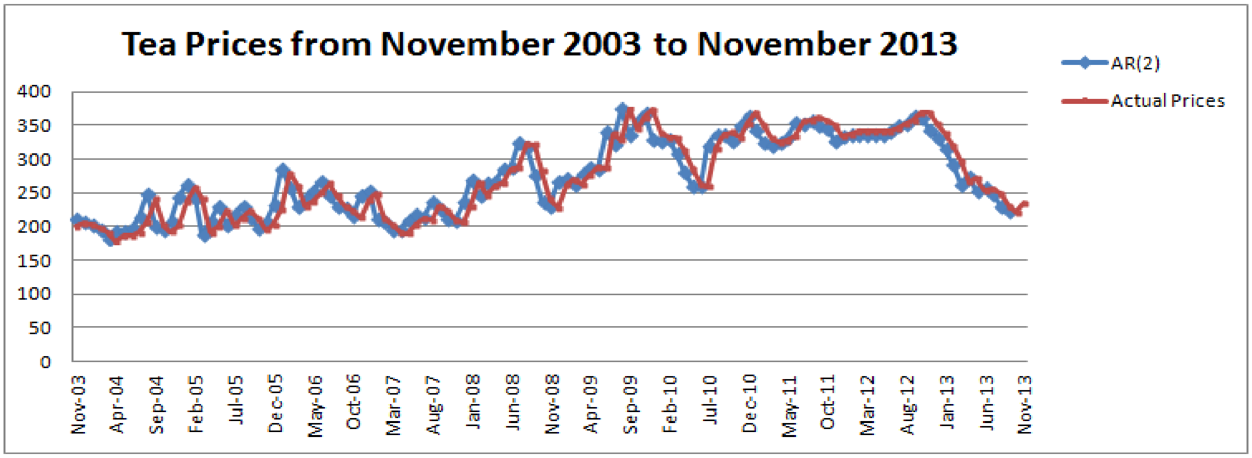
| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 18.51248 | 8.52234442 | 2.172228 | 0.031873 | 1.632899 | 35.39205 | 1.632899 | 35.39205 |
| X Variable 1 | 1.07713 | 0.091982314 | 11.71019 | 2.21E-21 | 0.894948 | 1.259313 | 0.894948 | 1.259313 |
| X Variable 2 | -0.14414 | 0.091687364 | -1.57211 | 0.118648 | -0.32574 | 0.037456 | -0.32574 | 0.037456 |

The equation is **$Y_t$=18.51248+1.07713$Y_{t-1}$-0.14414$Y_{t-2}$**.The R2 and adjusted R2 calues are a little bit higher than AR(1) model. The coefficients follow the rules of AR(2) stationarity:

X1+X2<1 →1.07713+(-0.14414)=0.93299<1

X2<1 → -0.14414<1

X2-X1<1 → -0.14414-1.07713=-1.22127<1.




The graph above shows the predicted prices vs. the actual prices for the AR(2) model. It is also a very good fit, but it is not as good as AR(1).

**<u>Conclusion:</u>**

Based on the analysis, AR(1) model, $\mathbf{Y_t=16.51733+0.940414Y_{t-1}}$, provides a great prediction of the tea price. The model has proven to provide a very close result when compared to the actual prices.