HOW MUCH DOES CHEESE COST?

I love cheese. I have slightly higher cholesterol because of my love of cheese. My favorite cheese is Vermont white cheddar cheese. Based off the monthly prices of cheese, I'm studying to see if I can predict how much it will cost me to make grilled cheese sandwiches in the future.

What I did

I used data showing the monthly prices of per pound for a 40 lb. block of cheddar cheese from January 1996 to April 2013. First, I had to test the data to see if it was stationary. I used the below equation to calculate the autocorrelation of the data:

$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

Then, I graphed the autocorrelations. Since my data was not stationary, I used first and second differences to transform the data. Again, I graphed the autocorrelations to determine if it was stationary. Applying second differences resulted in stationary data.

My next step was to build several AR(p) models and fit them to my data. Using the regression function built into Excel, I was able to determine the coefficients of each model. I compared the R^2 and standard error values to determine which model was the best fit and therefore would be the most effective predictor of the price of cheddar cheese. Finally, I compared my predicted values to my actual data values to confirm my hypothesis on which model would be a good fit.

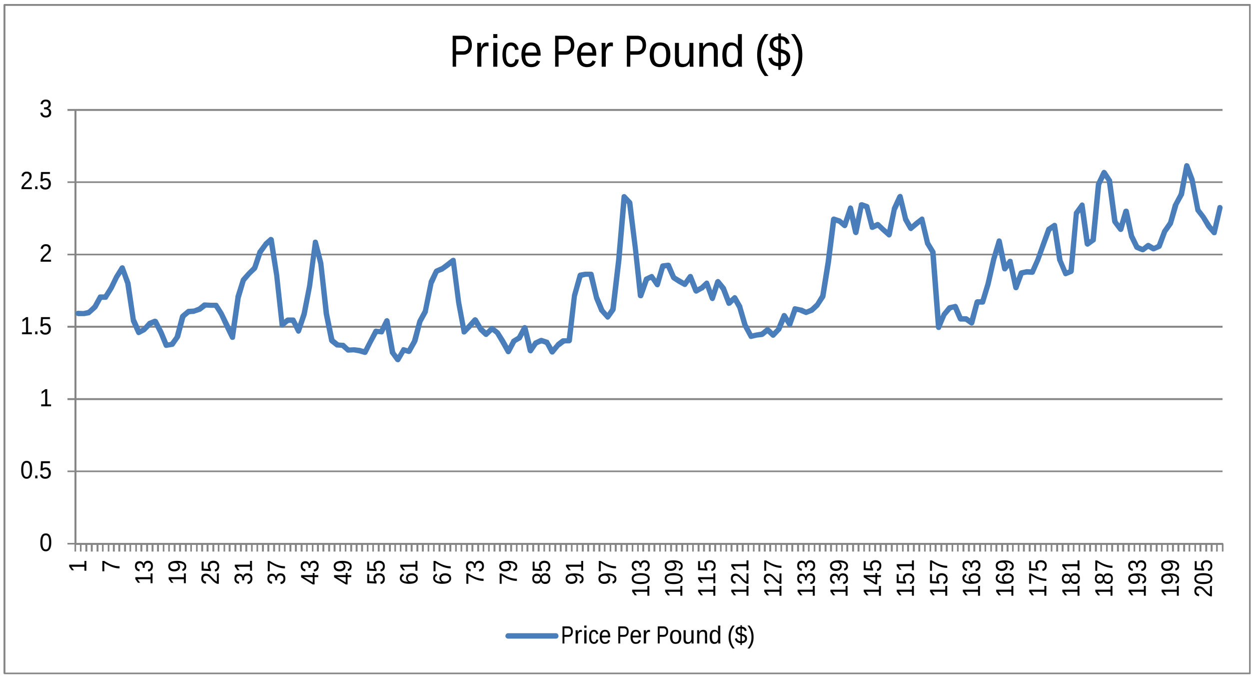
What data did I use?

The monthly retail price of a 40 lb. block of Cheddar cheese from January 1996 to April 2013 was downloaded from http://future.aae.wisc.edu/data/monthly_values/by_area/810?area=NE&grid=true&tab=prices. The data is summarized below:

Date	$ per Pound	Date	$ per Pound	Date	$ per Pound	Date	$ per Pound	Date	$ per Pound	Date	$ per Pound
Jan-1996	1.5919	Jan-1999	1.8568	Jan-2002	1.5467	Jan-2005	1.8387	Jan-2008	2.1874	Jan-2011	1.8808
Feb-1996	1.59	Feb-1999	1.513	Feb-2002	1.4841	Feb-2005	1.8152	Feb-2008	2.2053	Feb-2011	2.2843
March-1996	1.5989	March-1999	1.545	March-2002	1.4476	March-2005	1.7931	March-2008	2.1714	March-2011	2.341
April-1996	1.6368	April-1999	1.545	April-2002	1.4873	April-2005	1.8463	April-2008	2.1356	April-2011	2.0711
May-1996	1.705	May-1999	1.4706	May-2002	1.4581	May-2005	1.7485	May-2008	2.3132	May-2011	2.098
June-1996	1.705	June-1999	1.5885	June-2002	1.3988	June-2005	1.7655	June-2008	2.3999	June-2011	2.4841
July-1996	1.7675	July-1999	1.7826	July-2002	1.328	July-2005	1.8004	July-2008	2.2443	July-2011	2.565
Aug-1996	1.8439	Aug-1999	2.0841	Aug-2002	1.4002	Aug-2005	1.6951	Aug-2008	2.1782	Aug-2011	2.508
Sept-1996	1.9065	Sept-1999	1.9381	Sept-2002	1.4228	Sept-2005	1.8114	Sept-2008	2.2139	Sept-2011	2.2277
Oct-1996	1.8045	Oct-1999	1.5886	Oct-2002	1.4934	Oct-2005	1.7628	Oct-2008	2.2441	Oct-2011	2.1728
Nov-1996	1.5449	Nov-1999	1.4039	Nov-2002	1.3354	Nov-2005	1.663	Nov-2008	2.075	Nov-2011	2.3
Dec-1996	1.4599	Dec-1999	1.3738	Dec-2002	1.3875	Dec-2005	1.7007	Dec-2008	2.0138	Dec-2011	2.1215
Jan-1997	1.4804	Jan-2000	1.3716	Jan-2003	1.4039	Jan-2006	1.6348	Jan-2009	1.4965	Jan-2012	2.0498
Feb-1997	1.5224	Feb-2000	1.3378	Feb-2003	1.3927	Feb-2006	1.5055	Feb-2009	1.5833	Feb-2012	2.0326
March-1997	1.537	March-2000	1.3414	March-2003	1.3266	March-2006	1.4336	March-2009	1.6295	March-2012	2.0615
April-1997	1.4602	April-2000	1.3344	April-2003	1.3738	April-2006	1.4428	April-2009	1.6399	April-2012	2.039
May-1997	1.3724	May-2000	1.325	May-2003	1.4026	May-2006	1.4486	May-2009	1.5533	May-2012	2.0581
June-1997	1.3788	June-2000	1.3966	June-2003	1.4025	June-2006	1.479	June-2009	1.5527	June-2012	2.1581
July-1997	1.4286	July-2000	1.4683	July-2003	1.714	July-2006	1.4418	July-2009	1.5267	July-2012	2.2138
Aug-1997	1.5725	Aug-2000	1.4644	Aug-2003	1.8551	Aug-2006	1.4818	Aug-2009	1.6718	Aug-2012	2.3421
Sept-1997	1.605	Sept-2000	1.5404	Sept-2003	1.8638	Sept-2006	1.577	Sept-2009	1.6719	Sept-2012	2.4159
Oct-1997	1.6073	Oct-2000	1.3201	Oct-2003	1.8638	Oct-2006	1.5156	Oct-2009	1.7957	Oct-2012	2.6126
Nov-1997	1.6207	Nov-2000	1.2726	Nov-2003	1.7031	Nov-2006	1.6243	Nov-2009	1.968	Nov-2012	2.514
Dec-1997	1.6513	Dec-2000	1.3402	Dec-2003	1.6118	Dec-2006	1.6154	Dec-2009	2.0928	Dec-2012	2.3068
Jan-1998	1.6475	Jan-2001	1.3293	Jan-2004	1.5679	Jan-2007	1.5998	Jan-2010	1.9005	Jan-2013	2.2605
Feb-1998	1.6475	Feb-2001	1.4008	Feb-2004	1.6212	Feb-2007	1.6163	Feb-2010	1.9513	Feb-2013	2.1928
March-1998	1.5914	March-2001	1.5395	March-2004	1.9613	March-2007	1.6514	March-2010	1.7699	March-2013	2.1499
April-1998	1.505	April-2001	1.6088	April-2004	2.3978	April-2007	1.7124	April-2010	1.8707	April-2013	2.3245
May-1998	1.4274	May-2001	1.8056	May-2004	2.3597	May-2007	1.9449	May-2010	1.88		
June-1998	1.7035	June-2001	1.883	June-2004	2.0489	June-2007	2.2426	June-2010	1.8768		
July-1998	1.8243	July-2001	1.9008	July-2004	1.7151	July-2007	2.2314	July-2010	1.9622		
Aug-1998	1.87	Aug-2001	1.9279	Aug-2004	1.8273	Aug-2007	2.2001	Aug-2010	2.0706		
Sept-1998	1.9037	Sept-2001	1.9572	Sept-2004	1.846	Sept-2007	2.3192	Sept-2010	2.1731		
Oct-1998	2.0164	Oct-2001	1.6668	Oct-2004	1.79	Oct-2007	2.1518	Oct-2010	2.2003		
Nov-1998	2.0689	Nov-2001	1.4654	Nov-2004	1.9185	Nov-2007	2.3427	Nov-2010	1.9593		
Dec-1998	2.1038	Dec-2001	1.5064	Dec-2004	1.9247	Dec-2007	2.329	Dec-2010	1.8667		

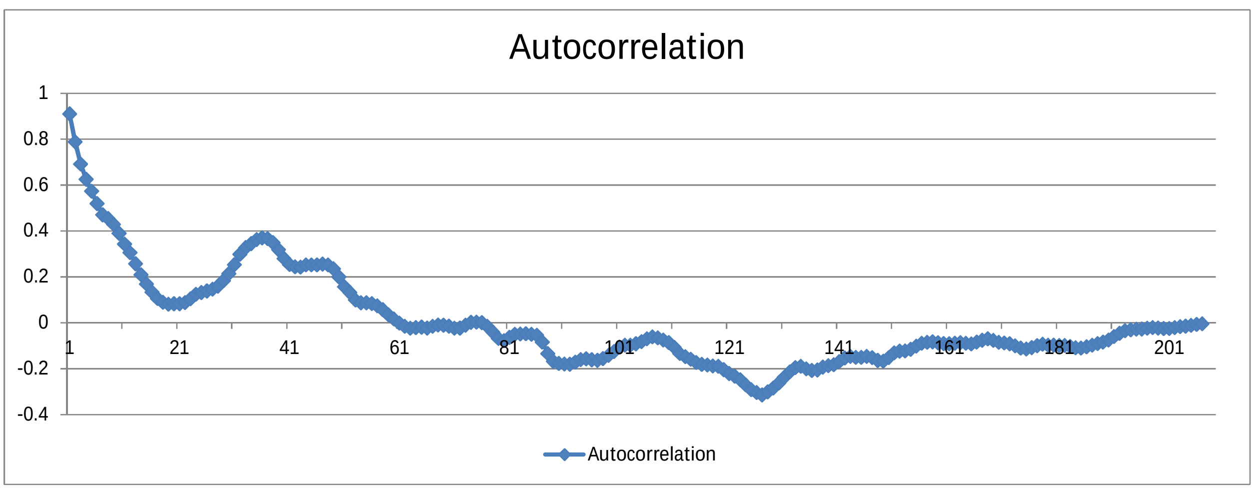
Is my data stationary?

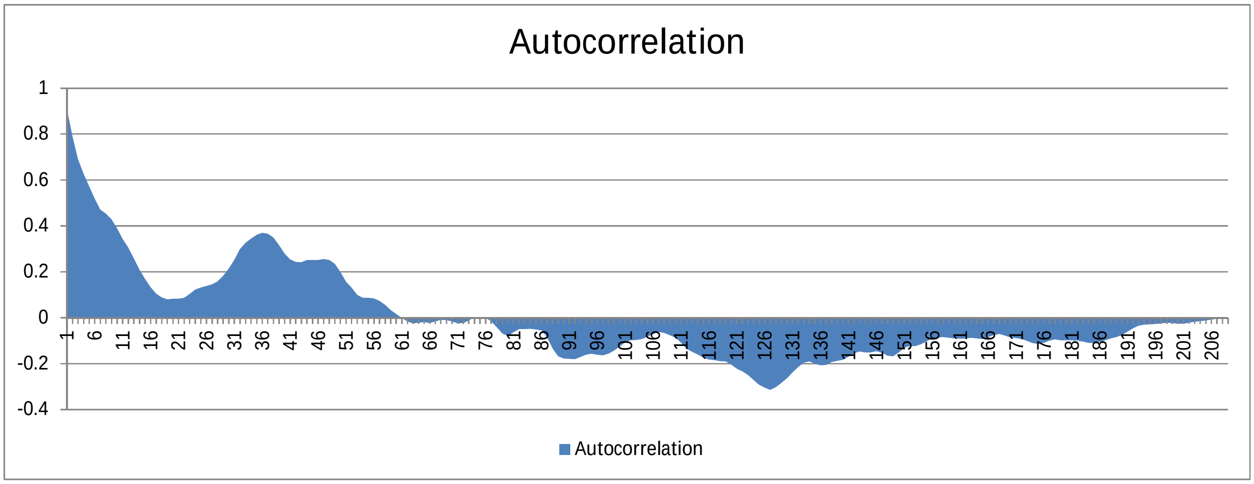
I graphed the raw data from above over the lag periods.



I observe that while the price fluctuates, it appears to be increasing in recent months and non-stationary.

To verify suspicion of non-stationarity, autocorrelation values were calculated and graphed across time lags, as shown below. The graphs represent the same autocorrelation values, but are expressed differently.

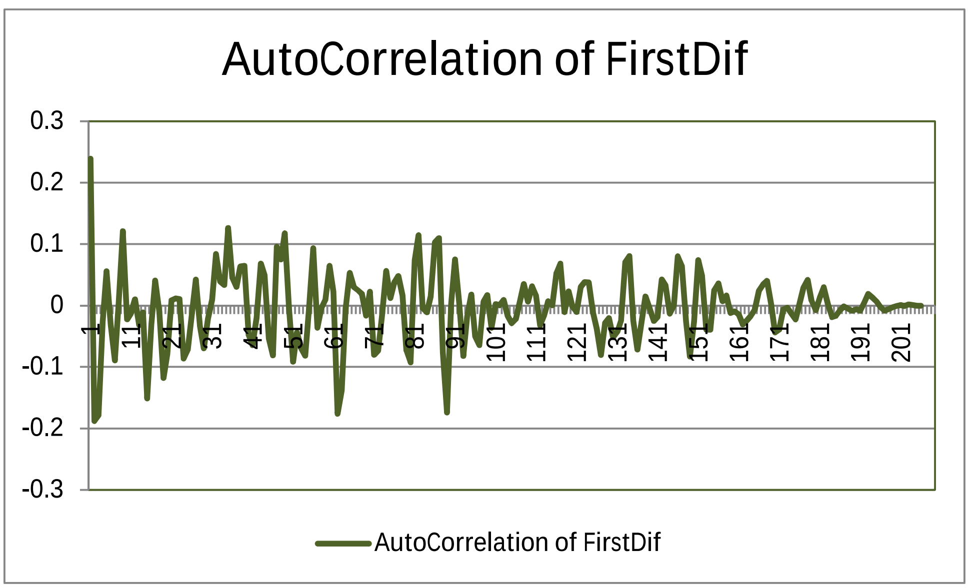




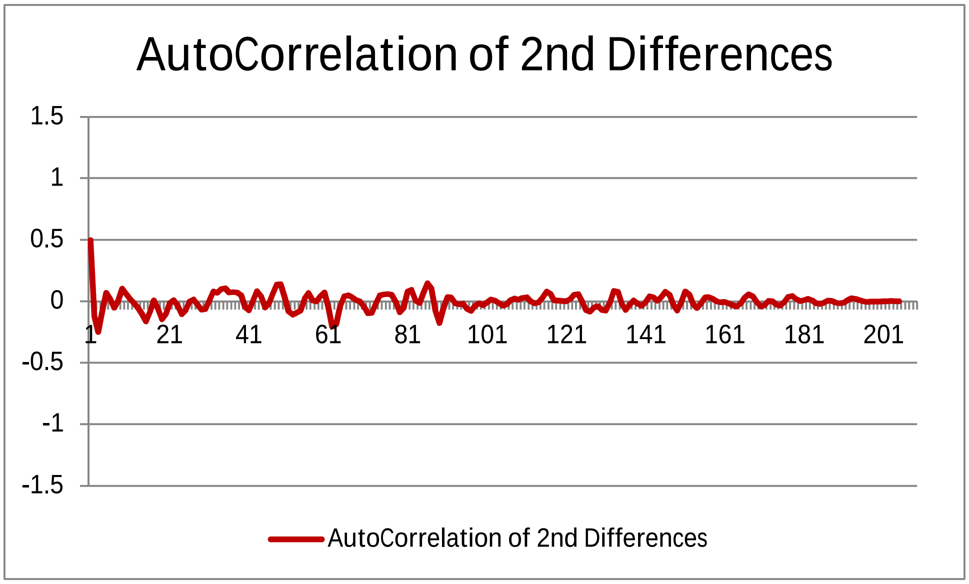
It appears that the autocorrelations slowly converge to 0. Since it converges slowly and has fluctuations, values may still deviate and it's confirmed that the data is not stationary.

Since my data is not stationary, I apply first and second differences to transform the data and graph those resultant autocorrelations as seen below.

First Differences:



Second Differences:



In the second differences graph, I see that the data quickly converges to zero and oscillates around zero. As a result, the second difference autocorrelation data is now stationary.

Fitting models to my data

My first model is an AR(1) with this general formula:

$$Y_t - Y_{t-1} = \varepsilon_t + \phi_1 (Y_{t-1} - Y_{t-2})$$

Using the excel regression package, I fit the model to my data and the results are shown below:

Regression Statistics	
Multiple R	0.553179
R Square	0.306007
Adjusted F	0.300224
Standard E	0.173583
Observatio	122

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressior	1	1.594308	1.594308	52.91237	3.91E-11
Residual	120	3.615732	0.030131		
Total	121	5.21004			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	0.004676	0.015736	0.29717	0.76685	-0.02648	0.035833	-0.02648	0.035833
X Variable	0.553184	0.076049	7.274089	3.91E-11	0.402613	0.703755	0.402613	0.703755

The formula for my AR(1) model is:

$Y_t - Y_{t-1}$ = 0.004676 + 0.553184$(Y_{t-1} - Y_{t-2})$

The R^2 is close to 1 and my standard error is low. To make sure I use the best model to predict the future prices of cheese, I'm going to continue looking at other models.

My next model is an AR(2) with this general formula:

$$Y_t - Y_{t-1} = \varepsilon_t + \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3})$$

Using the excel regression package, I fit the model to my data and the results are shown below:

Regression Statistics	
Multiple R	0.707843
R Square	0.501042
Adjusted R Square	0.492585
Standard Error	0.148408
Observations	121

ANOVA

	df	SS	MS	F	Significance F
Regression	2	2.609784	1.304892	59.24638	1.53E-18
Residual	118	2.598931	0.022025		
Total	120	5.208715			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.006332	0.013512	0.46863	0.640199	-0.02043	0.033091	-0.02043	0.033091
X Variable 1	0.847146	0.078113	10.84513	1.96E-19	0.692461	1.001831	0.692461	1.001831
X Variable 2	-0.53111	0.078224	-6.78964	4.81E-10	-0.68602	-0.37621	-0.68602	-0.37621

The formula for my AR(2) model is:
$Y_t - Y_{t-1} = 0.006332 + 0.847146(Y_{t-1} - Y_{t-2}) - 0.531111(Y_{t-2} - Y_{t-3})$

The R^2 is even larger than in my AR(2) model, although my standard error is lower than the AR(1).

My next model is an AR(3) with this general formula:

$$Y_t - Y_{t-1} = \varepsilon_t + \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \phi_3(Y_{t-3} - Y_{t-4})$$

Using the excel regression package, I fit the model to my data and the results are shown below:

Regression Statistics	
Multiple R	0.719919
R Square	0.518284
Adjusted R Square	0.505826
Standard Error	0.146942
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	3	2.69479	0.898263	41.60189	2.51E-18
Residual	116	2.504659	0.021592		
Total	119	5.199448			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	0.004315	0.013453	0.32073	0.748992	-0.02233	0.030961	-0.02233	0.030961
X Variable 1	0.947145	0.091712	10.32737	3.99E-18	0.765498	1.128793	0.765498	1.128793
X Variable 2	-0.69177	0.110509	-6.2599	6.71E-09	-0.91065	-0.4729	-0.91065	-0.4729
X Variable 3	0.188569	0.09228	2.043451	0.043273	0.005797	0.371341	0.005797	0.371341

The formula for my AR(3) model is:

$$Y_t - Y_{t-1} = 0.004315 + 0.947145(Y_{t-1} - Y_{t-2}) - 0.691777(Y_{t-2} - Y_{t-3}) + 0.188569(Y_{t-3} - Y_{t-4})$$

The R^2 is even larger than in my AR(2) model, although my standard error is lower than the AR(1).

Comparing my results

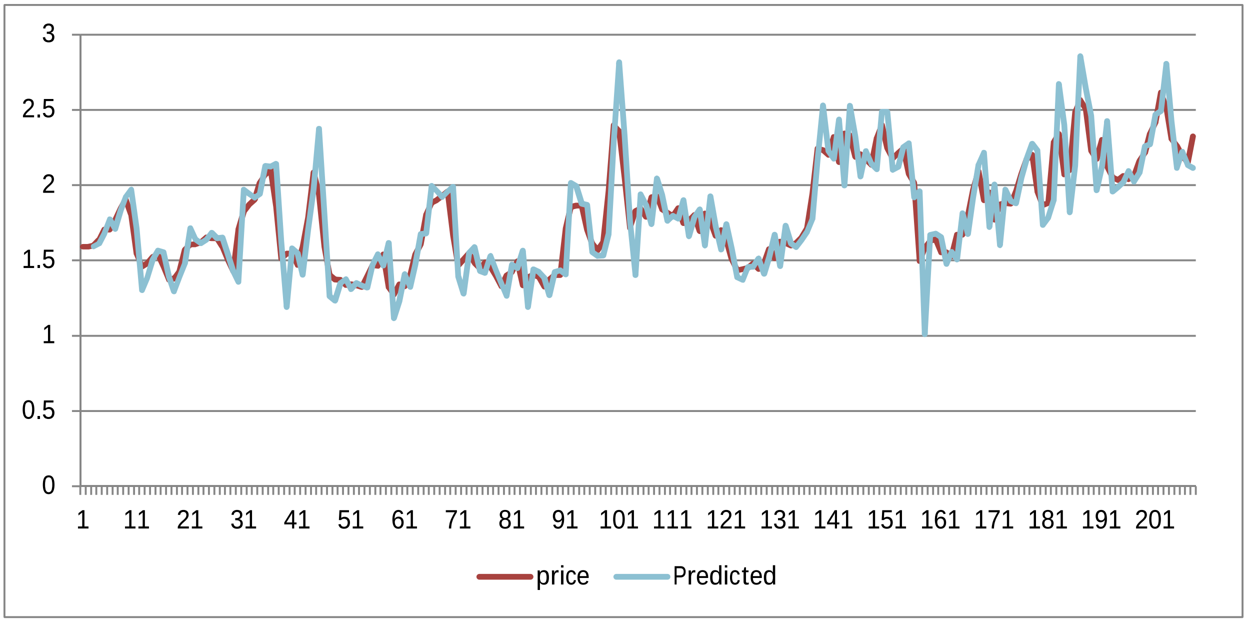
Let me compare my models:

	R^2	Adjusted R^2	Standard Error
AR(1)	0.3060	0.3002	0.1736
AR(2)	0.5010	0.4926	0.1484
AR(3)	0.5183	0.5058	0.1469

Although the standard error is the highest, I am going to use the AR(1) model since it produced the lowest R^2 values and the differences in the standard errors is minor. I think the AR(1) model is the best model to use to predict the cost of cheddar cheese.

Reviewing my results

I then took my chosen model and fit it to the data. I graphed the actual and predicted values, as seen below:



.

$$Y_t = Y_{t-1} + 0.004676 + 0.553184(Y_{t-1} - Y_{t-2})$$

The absolute difference of the actual and predicted prices ranges from 0 to 0.57. Between the small absolute difference and the visual comparison of the actual and predicted prices, my model is a good fit for the data.

In Conclusion

The raw data gathered was not already stationary. In order to make the data stationary, I took the first differences. Then, I fit AR(p) models to the data using the Excel regression function. The regression function helped determine the coefficients for models AR(1), AR(2) and AR(3). At this point, I compared the R^2 and standard error of each model. Based off those, I determined the AR(1) has the best fit for the data. The final formula, $\boldsymbol{Y_t = Y_{t-1} + 0.004676 + 0.553184(Y_{t-1} - Y_{t-2})}$ can now be used as a predictor equation for the price per pound of a 40 lb. block of cheddar cheese . Based on the graph of actual and predicted values, the model is a good fit. The only question remaining is how many grilled cheese sandwiches can I make with a 40 lb. block of cheddar cheese?

Bibliography

University of Wisconsin Dairy Marketing and Risk Management Program. *Country STAT Philippines*. Retrieved May 24, 2013 from http://future.aae.wisc.edu/data/monthly_values/by_area/810?area=NE&grid=true&tab=prices.