Igor Lavitas
Time Series Project
Summer 2011

Oranges Monthly Prices

Introduction

Orange juice is a breakfast must have. It is made mostly by extraction from fresh fruit. Being a loyal customer of Tropicana, which owns 65% of the US orange juice market, I have long noticed the fluctuation in orange juice prices in the supermarkets.

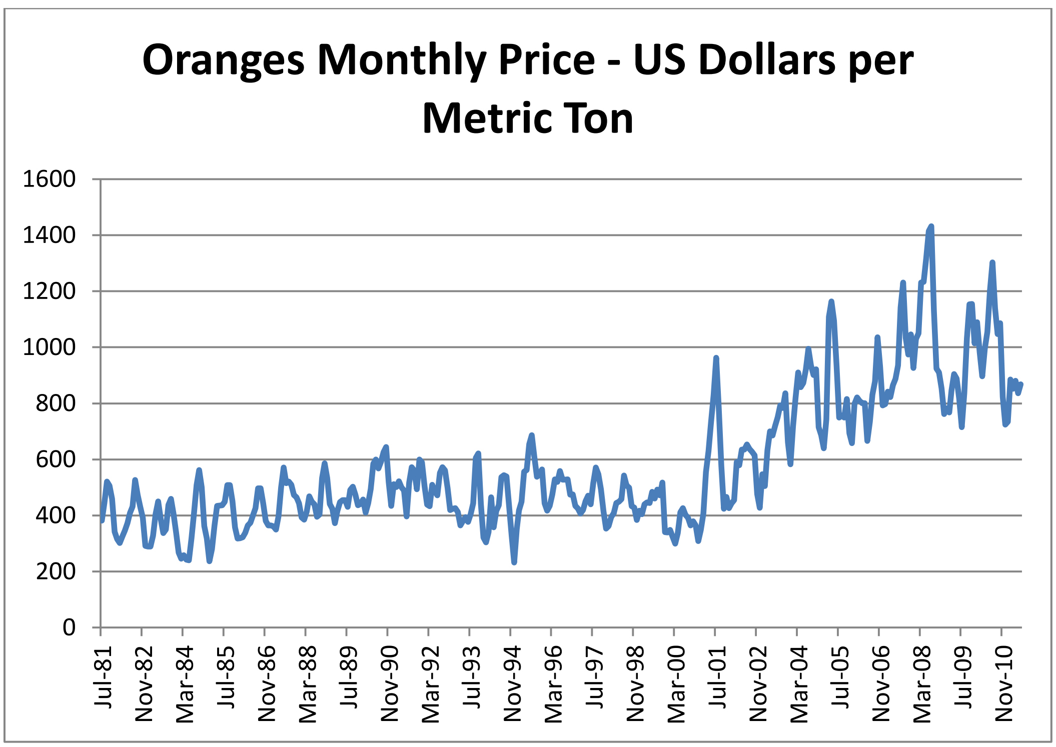
In this study, I will look at 30 years of monthly orange prices data. The data will most likely show seasonal trends as people tend to consume more orange products in the summer. I will review few models and select the best fit.

Data

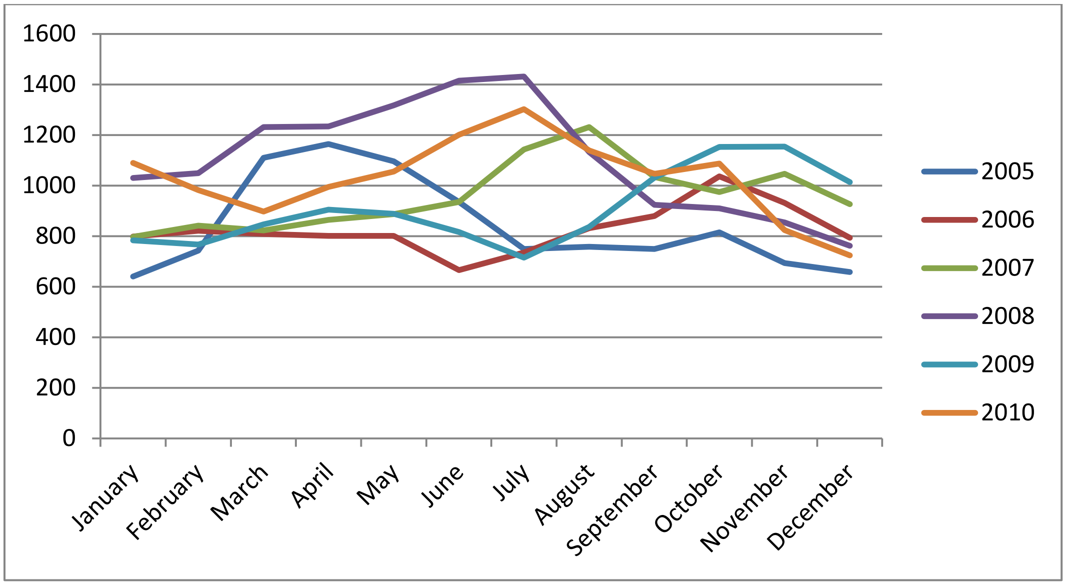
I am using data from
http://www.indexmundi.com/commodities/?commodity=oranges&months=360
The data exhibits monthly price of oranges in US dollars per metric ton. I am using 30 years of price data, from July 1981 through June 2011.

Model specifications

Graphing the data, there appears to be a general upward price trend in the past 10 years. Additionally, prices have been more volatile in the past 10 years.



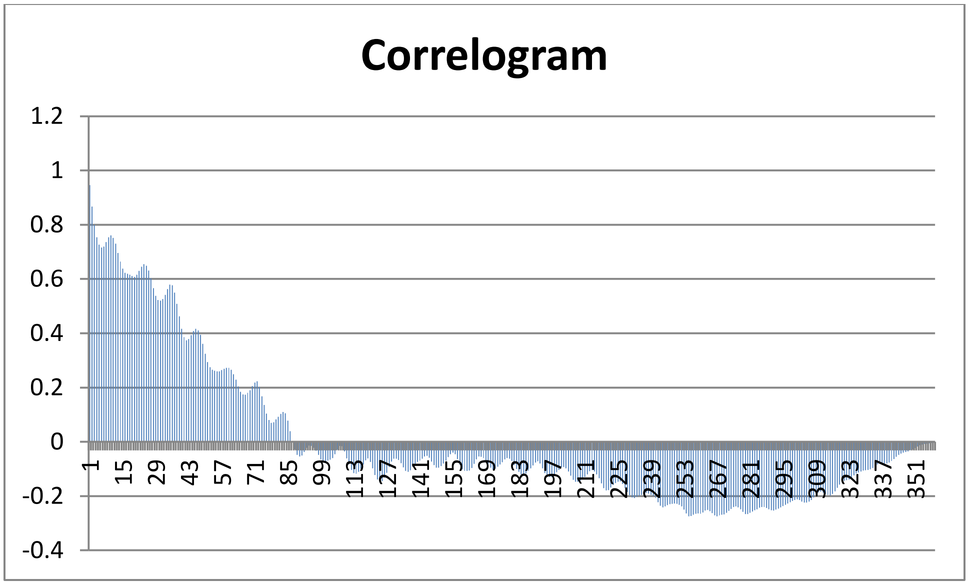
Examining the graphs below, it is hard to detect any seasonality in the data. Seasonal trend is apparent in years 2007, 2008, 2010 and not so apparent in years 2005, 2006, 2009.



Autocorrelations

Next, we will graph a correlogram, i.e. sample autocorrelation function, which will show the correlation for each time lag. The sample autocorrelation function at lag *k* is defined as

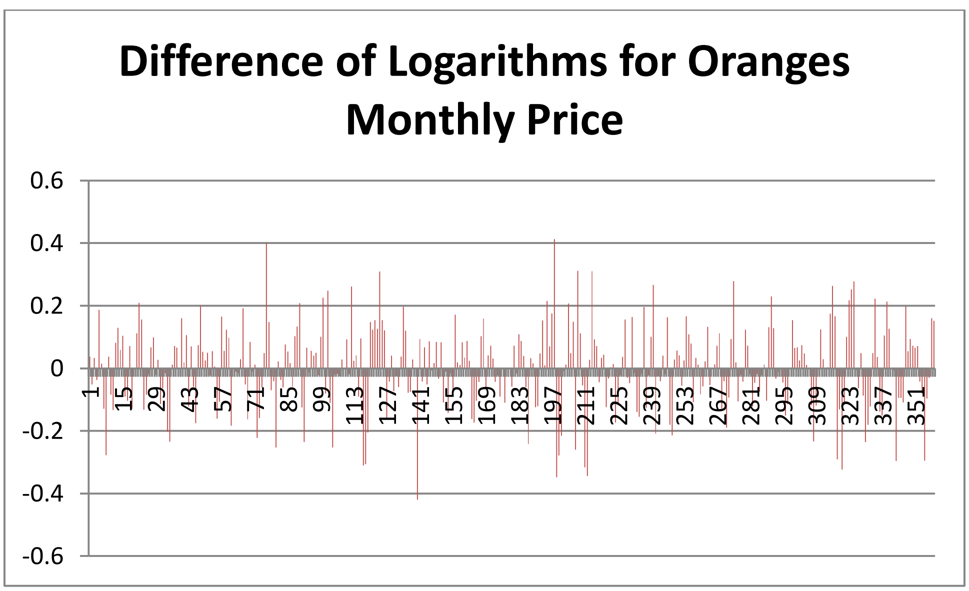
$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} \qquad \text{for } k = 1, 2, \ldots$$

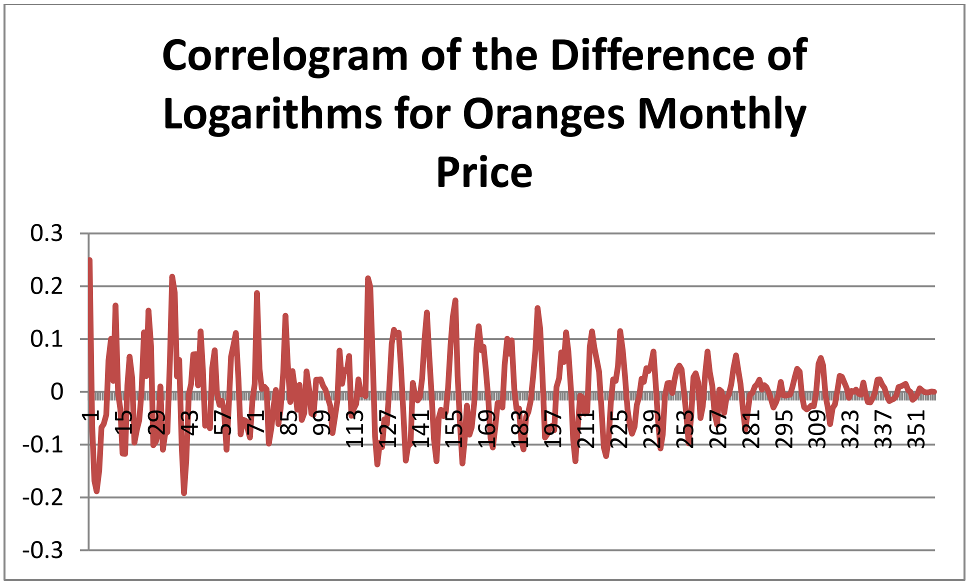


The autocorrelations start very high and gradually grade down to zero around lag 87. Autocorrelations then remain between 0 and -.275. Since autocorrelations do not fluctuate around some mean or quickly drop to zero, this is not a stationary process.

To create a stationary process we assume that prices follow logarithmic pattern. We then create a new series by taking logarithms and then computing first differences.

Below are the plots of the First Differences of Logarithms and Correlogram of the First Differences of Logarithms. The autocorrelation function of the first differences of logarithms decreasingly fluctuates towards zero. On the basis of these plots, we may consider this series a stationary process.





Assuming stationary, we will fit the following models to the data: ARIMA(1,1,0), ARIMA(2,1,0), and ARIMA(3,1,1). The parameters for the series were estimated using method of moments and Yule-Walker equation below:

$$
\begin{aligned}
\phi_1 + r_1\phi_2 + r_2\phi_3 + \cdots + r_{p-1}\phi_p &= r_1 \\
r_1\phi_1 + \phi_2 + r_1\phi_3 + \cdots + r_{p-2}\phi_p &= r_2 \\
&\vdots \\
r_{p-1}\phi_1 + r_{p-2}\phi_2 + r_{p-3}\phi_3 + \cdots + \phi_p &= r_p
\end{aligned}
$$

The parameters were verified in Excel regression analysis tool.

ARIMA(1,1,0): Y_t=2.7950+0.5571Y_{t-1}

Regression Statistics	
Multiple R	0.7387
R Square	0.5456
Adjusted R Square	0.5443
Standard Error	0.2561

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.7950	0.1690	16.5432	0.0000
Y(t-1)	0.5571	0.0269	20.7332	0.0000

ARIMA(2,1,0): Y_t=2.7992-0.6243Y_{t-1}+0.0681Y_{t-2}

Regression Statistics	
Multiple R	0.7412
R Square	0.5493
Adjusted R Square	0.5468
Standard Error	0.2554

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.799236147	0.168514124	16.61128509	0.00000
Y(t-2)	-0.068119261	0.039748959	-1.713736977	0.08745
Y(t-1)	0.624316455	0.047513252	13.13983834	0.00000

ARIMA(3,1,0): $Y_t=2.8097+0.6200Y_{t-1}-0.0198Y_{t-2}-0.0458Y_{t-3}$

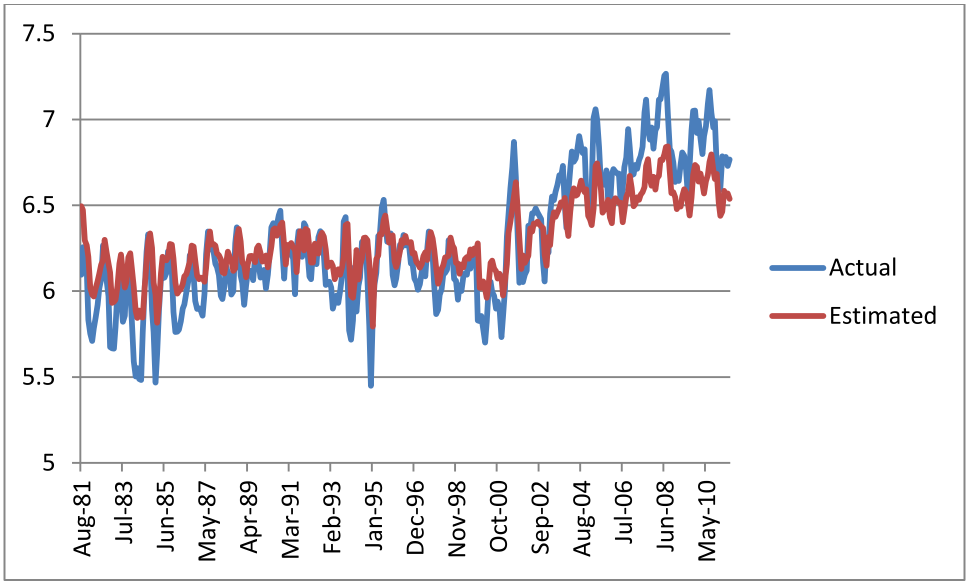
Regression Statistics	
Multiple R	0.7423
R Square	0.5510
Adjusted R Square	0.5472
Standard Error	0.2553

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.8097	0.1687	16.6566	0.00
Y(t-3)	-0.0458	0.0399	-1.1500	0.25
Y(t-2)	-0.0198	0.0578	-0.3422	0.73
Y(t-1)	0.6200	0.0476	13.0154	0.00

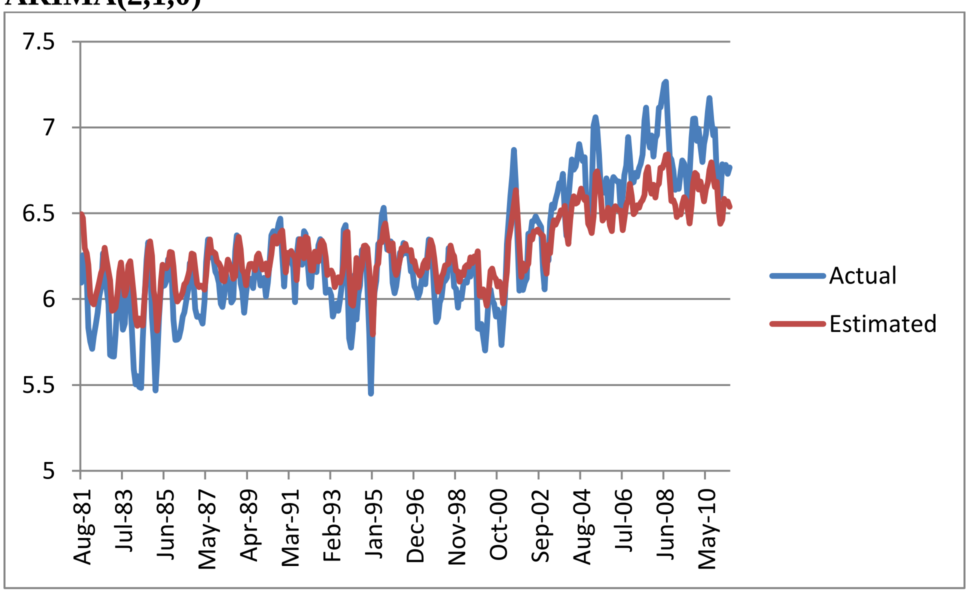
The parameters above were confirmed using Excel regression analysis tool. Additionally, the sum of coefficients is less than 1, confirming stationary nature of series.

Graphs

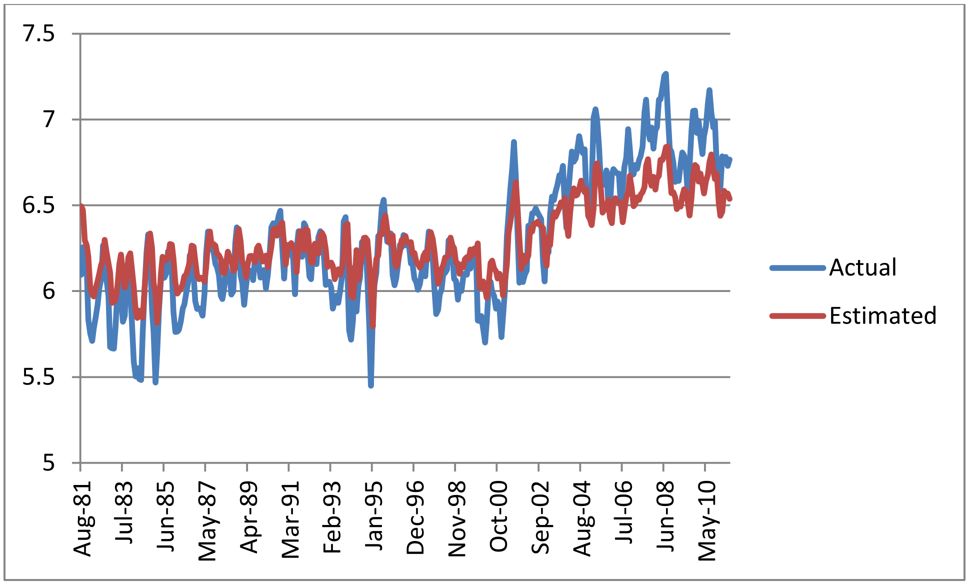
ARIMA(1,1,0)



ARIMA(2,1,0)



ARIMA(3,1,0)



Summary

It is difficult to determine from the graphs which model is the best fit. Moving from ARIMA(1,1,0) to ARIMA(2,1,0) increases R-square value. Moving from ARIMA(2,1,0) to ARIMA(3,1,0) also increases R-square value. However, ARIMA(3,1,0) exhibits large P-values. It appears that the best fit models are either ARIMA(1,1,0) or ARIMA(2,1,0). Durbin-Watson statistic is best for autoregressive models. It indicates the likelihood that the residuals have a first-order autoregression component.

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

	ARIMA(1,1,0)	AR(2,1,0)	ARIMA(3,1,0)
Durbin-Watson	2.17	2.03	2.09

Based on the values in the table above, it appears that residual autocorrelation is unlikely for ARIMA(2,1,0). ARIMA(1,1,0) has somewhat higher than 2 DW-statistics which may indicate possible correlation between residuals. Overall, ARIMA(2,1,0) model appears to be preferable estimator for predicting oranges monthly price.