Keely M Davison
VEE – Time Series Course
Student Project – Fall 2009
July 6th, 2011

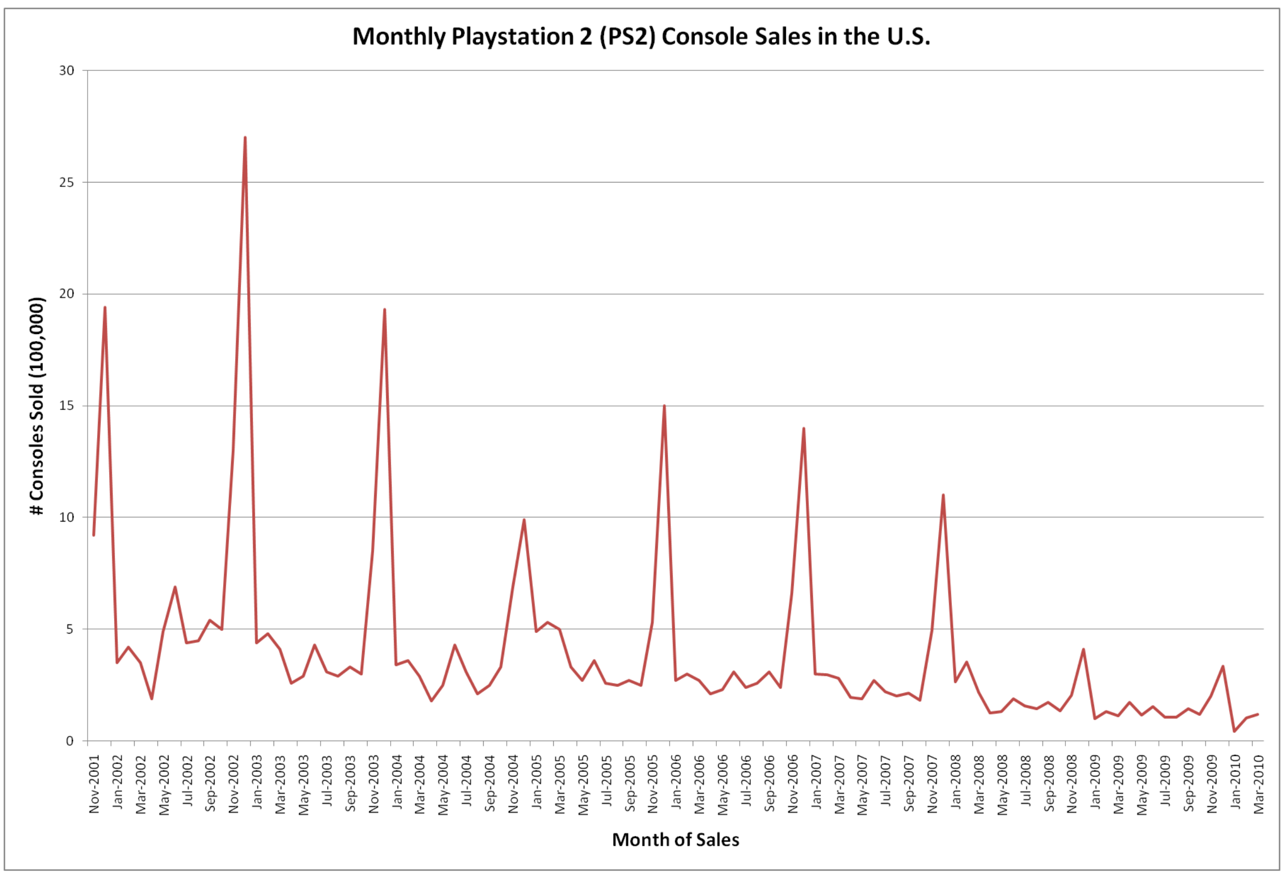
Introduction:

We live in an era where computer technology affects every aspect of our lives. From analyzing spreadsheets all day on our Dell computers to checking Facebook on our Apple iPhones, the American society is constantly using electronic hardware to get things done. This doesn't just affect our productivity or social networking, but our recreation as well. Gaming consoles have become a familiar addition to many living rooms in the United States.

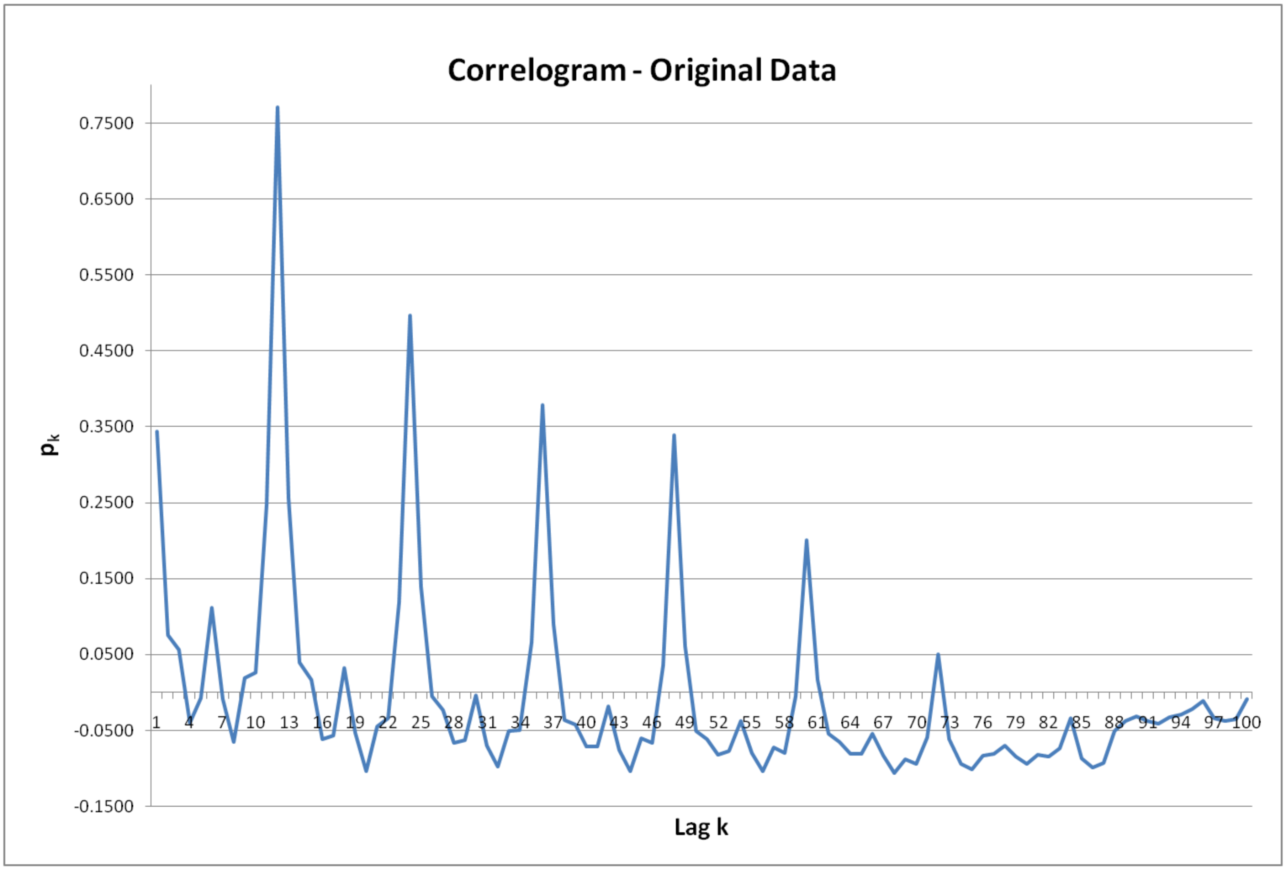
My goal is to model the monthly sales of a popular gaming console, the Playstation 2 (PS2). Not only am I interested in this subject, but the data will allow me to explore some very important time series topics such as seasonality and stationarity.

Data:

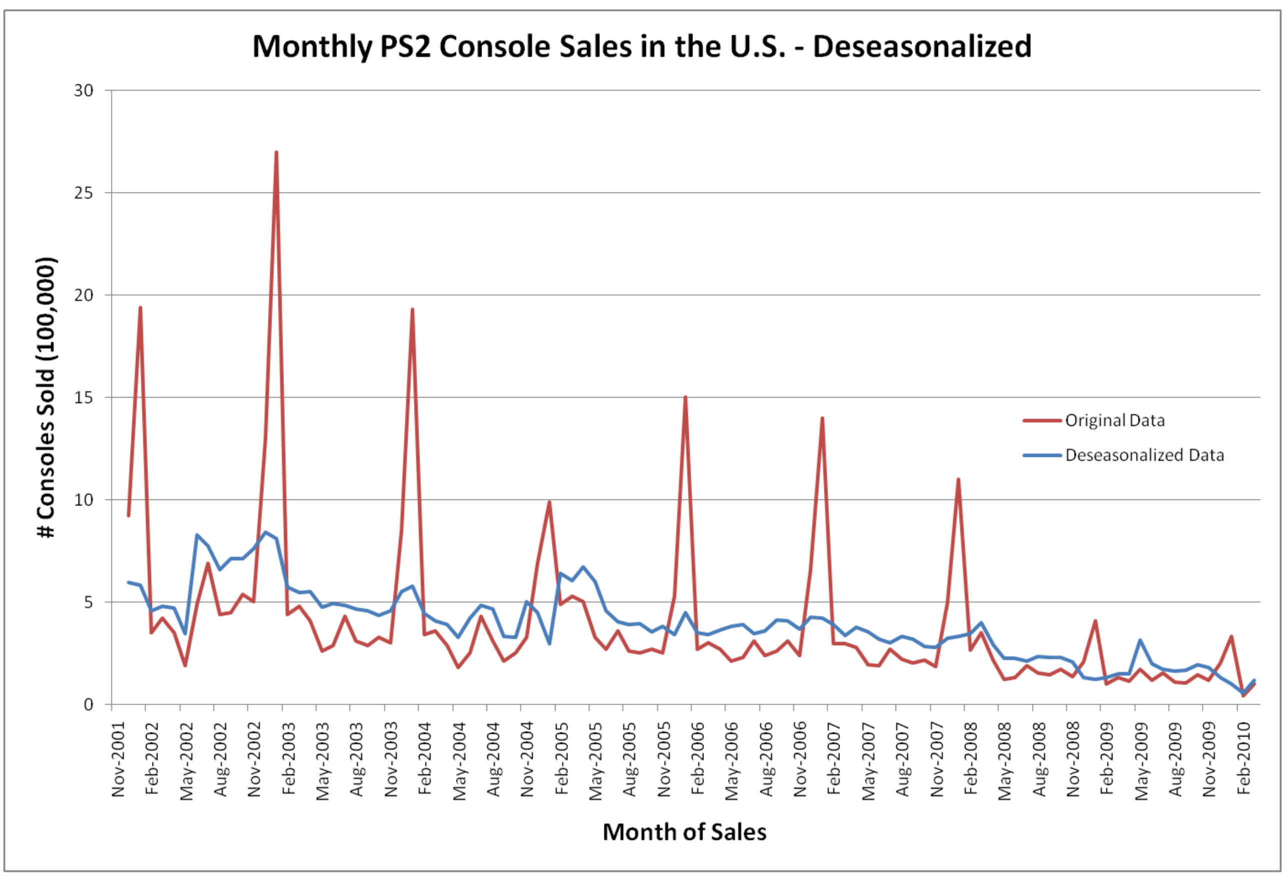
My monthly PS2 sales data was obtained from the PVC Museum website located at www.pvcmuseum.com/games/charts/monthly-console-hardware-sales-in-america.htm.
The timeframe available was from November 2001 to March 2010. Below is a simple graph of this data with respect to month-year of sales:



As one can see by the high annual peaks around December, seasonality is definitely present. This supports the common knowledge that U.S. retail shopping dramatically increases during the holiday season. We need to deseasonalize this data, but let's first look at the corresponding correlogram (plot of the sample autocorrelation function p_k).



This graph also shows high peaks every 12 months, confirming our assumption of seasonality. However, at this point, it is difficult to make any conclusion about stationarity. Let's first deseasonalize the observations by computing monthly seasonal indices and removing the seasonal variations (as shown below). Then we can address the issue of stationarity.



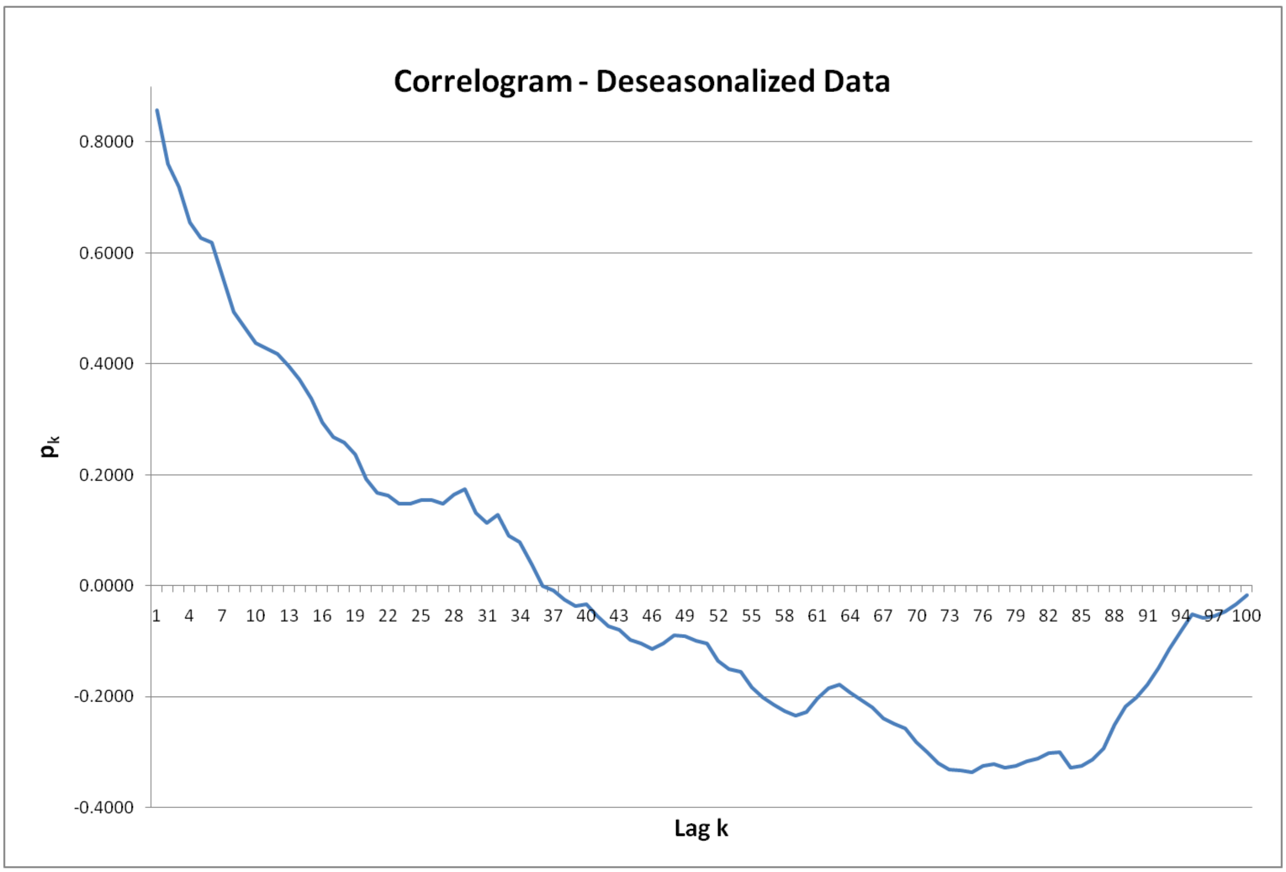
Month	Seasonal Index
January	0.76599
February	0.87743
March	0.74195
April	0.54920
May	0.59056
June	0.89073
July	0.66722
August	0.63164
September	0.75835
October	0.65530
November	1.54244
December	3.32918

To the left, we have the monthly seasonal indices computed using the method on page 483 of *Econometric Models and Economic Forecasts*, 4th Edition; Pindyck and Rubinfeld. Please refer to my Excel workbook for additional details on these computations.

The resulting deseasonalized data is shown as a blue line in the graph above. Notice that most of the seasonal variation has been removed, while the long-term downward trend and short-term irregular fluctuations still remain.

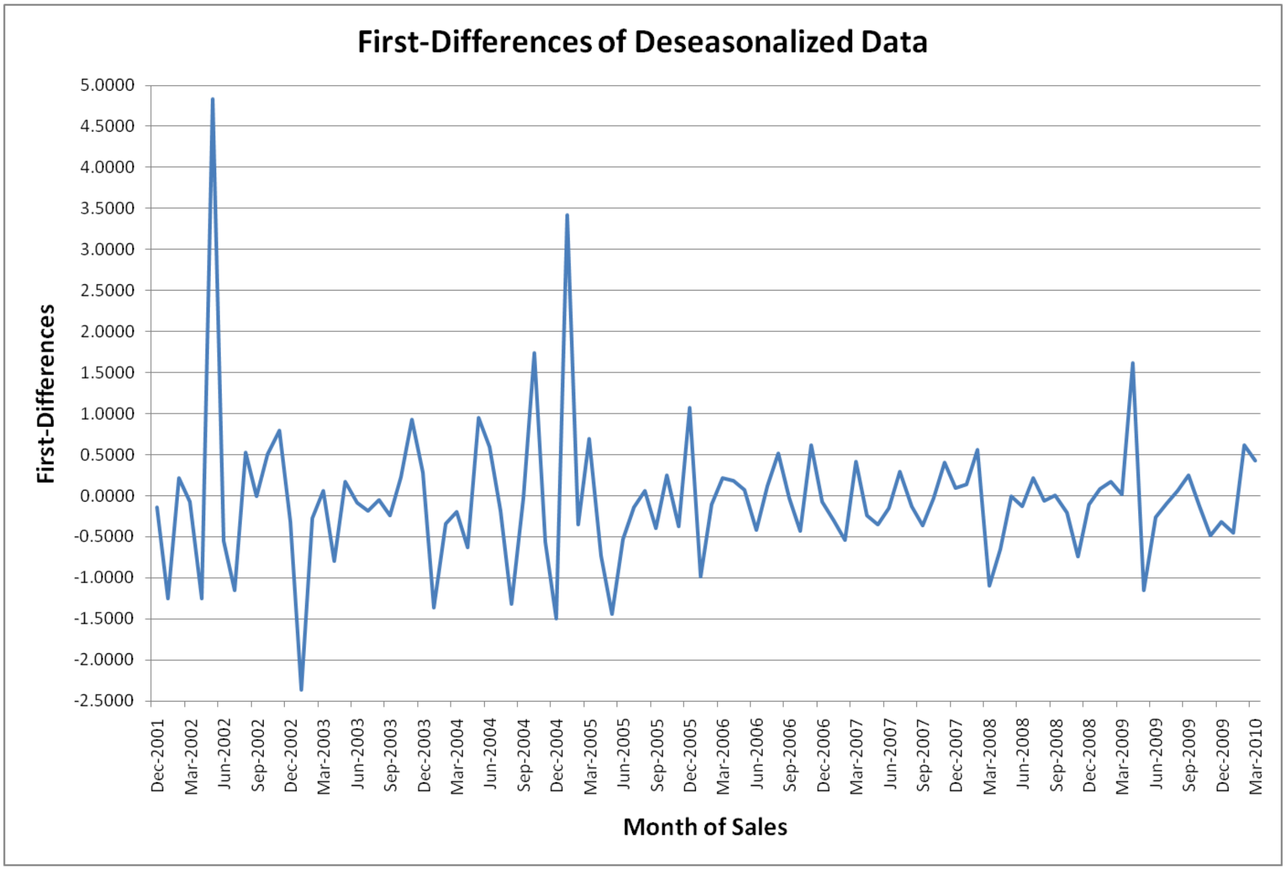
The long-term downward trend can be explained by a gradual decrease in demand. This is reasonable, considering the release of other popular gaming consoles, such as the Playstation 3 and other competitor devices.

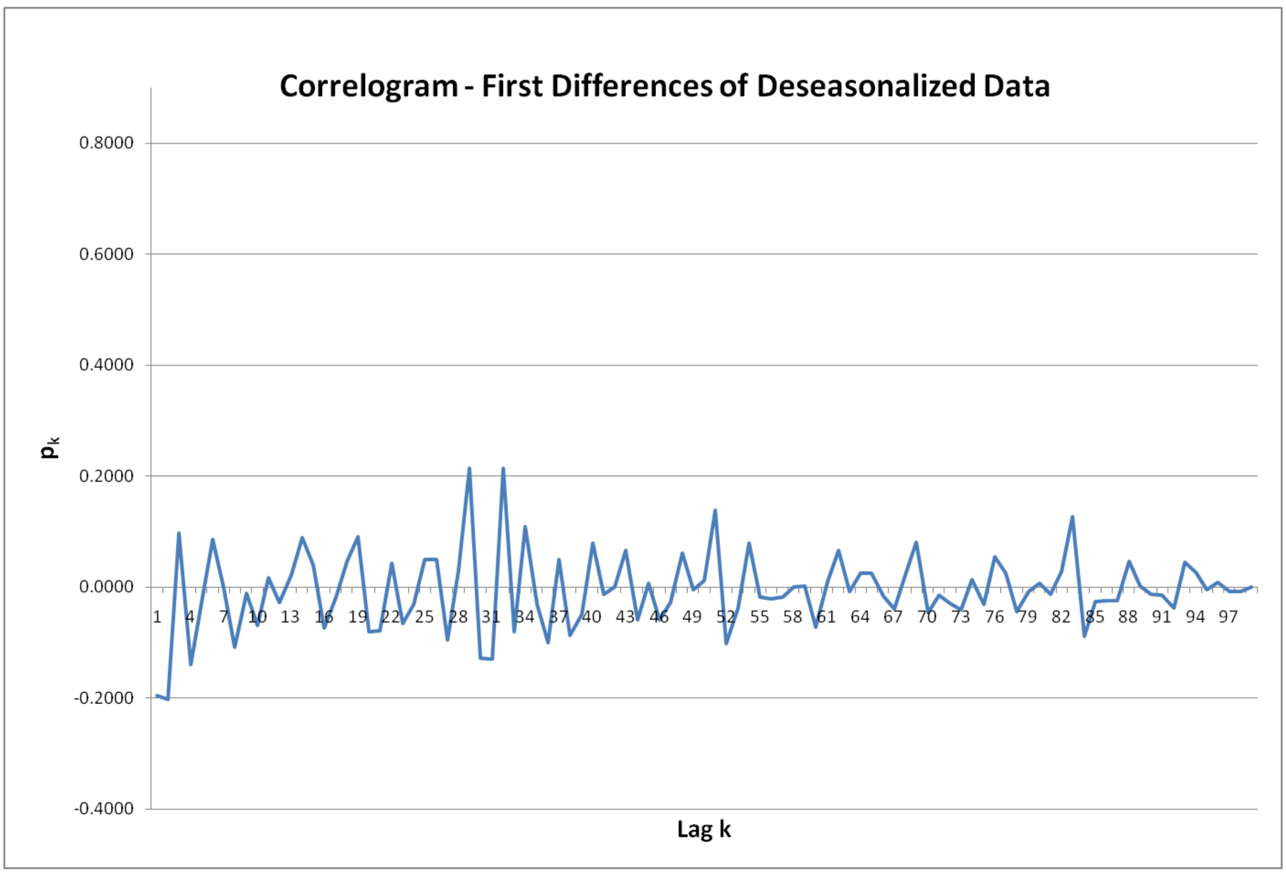
Now let's examine the correlogram corresponding to our deseasonalized data:



This sample autocorrelation function does not exhibit strong seasonality. The high peaks have been removed, confirming that our method of deseasonalization was successful. However, it only declines slowly and doesn't even cross p_k=0 until lag 36. There is definitely doubt as to whether this is a stationary series.

To address this issue, let's first-difference the deseasonalized data and examine its correlogram. If we observe that p_k decreases rapidly and remains small, we can be confident that the resulting series is a stationary, non-seasonal time series.





To better examine changes in the sample autocorrelation function, the correlogram above maintains the same scale as the previous deseasonalized-only correlogram. One can see that p_k now rapidly declines and settles about zero (recall that p_0=1 for any stochastic process). Therefore, the transformed data is a stationary, non-seasonal series. We are now ready to construct a time-series model.

Model Specification:

Define $\acute{y}_t$ to be the first-differences of the deseasonalized data which was shown to be a stationary, non-seasonal series. Since the correlogram of $\acute{y}_t$ seems to peak roughly every 3 periods (see the graph immediately above), we might suspect that $\acute{y}_t$ is autoregressive of order 3. Let's explore this theory by estimating the parameters of an AR(3) model through Microsoft Excel and examining our results.

Autoregressive Model AR(3)

$$\acute{y}_t = \Phi_1\acute{y}_{t-1} + \Phi_2\acute{y}_{t-2} + \Phi_3\acute{y}_{t-3} + \delta + \varepsilon_t$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.319524885
R Square	0.102096152
Adjusted R Square	0.073131512
Standard Error	0.846160851
Observations	97

Excel worksheet "AR(3)" contains the data used to generate this regression, as well as the calculation of related statistics.

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	7.571263233	2.523754411	3.524854828	0.017992696
Residual	93	66.58690122	0.715988185		
Total	96	74.15816445			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.055907519	0.086539744	-0.646032872	0.519849035
X Variable 1	-0.251752028	0.103906748	-2.422865045	0.017335505
X Variable 2	-0.255673893	0.103095095	-2.479981151	0.01493837
X Variable 3	-0.003908148	0.103158216	-0.037884993	0.969860583

As part of the regression output, Excel also provides a set of residuals. I used this information to calculate two statistics: the Durbin-Watson statistic and the Box and Pierce (Q) statistic.

The Durbin-Watson statistic was found to be 1.9969, which is very close to 2.0. This would suggest that we accept the null hypothesis H_0: no serial correlation amongst residuals.

The Box and Pierce (Q) statistic for this model with K=25 was 10.0854. At a 10% significance level, this statistic is far below the X^2 critical value of 30.8133. The X^2 p-value for 10.0854 with K=25 (which was 0.9855) is also well above 10%. Therefore, we cannot reject the null hypothesis H_0: the residuals are a white noise process.

These two statistics favor our model. However, the R^2 value from our regression (0.1021) does not. By definition, the R^2 value is the proportion of the total variation in Y explained by the regression of Y on X. One must keep in mind that R^2 is only a descriptive statistic and there may be several reasons for a low R^2. Even so, I wanted to examine another model to see if I could improve on it.

I chose an autoregressive model of order 4, which would give me an additional independent variable within my model. My hope was that the conclusions based off of the Durbin-Watson statistic and Box and Pierce statistic would stay the same, while my R^2 value would increase.

Below are the regression results for this AR(4) model.

Autoregressive Model AR(4)

$$\acute{y}_t = \Phi_1\acute{y}_{t-1} + \Phi_2\acute{y}_{t-2} + \Phi_3\acute{y}_{t-3} + \Phi_4\acute{y}_{t-4} + \delta + \varepsilon_t$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.367916846
R Square	0.135362805
Adjusted R Square	0.097356775
Standard Error	0.839405518
Observations	96

Excel worksheet "AR(4)" contains the data used to generate this regression, as well as the calculation of related statistics.

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	10.03807565	2.509518913	3.561613869	0.009567444
Residual	91	64.1187477	0.704601623		
Total	95	74.15682336			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	-0.066813814	0.086511294	-0.772313192	0.441930149	-0.238657869
X Variable 1	-0.252525851	0.103081951	-2.449758165	0.016209141	-0.45728548
X Variable 2	-0.308721458	0.106569183	-2.896911168	0.004719962	-0.520408044
X Variable 3	-0.052807573	0.105619292	-0.49998037	0.618295787	-0.262607317
X Variable 4	-0.188399734	0.102365313	-1.840464596	0.068959235	-0.391735848

Again, I used the set of residuals provided by Excel to calculate the Durbin-Watson statistic and the Box and Pierce (Q) statistic.

The Durbin-Watson statistic was 2.0215, which is also very close to 2.0. This would cause us to accept the null hypothesis H_0: no serial correlation amongst residuals.

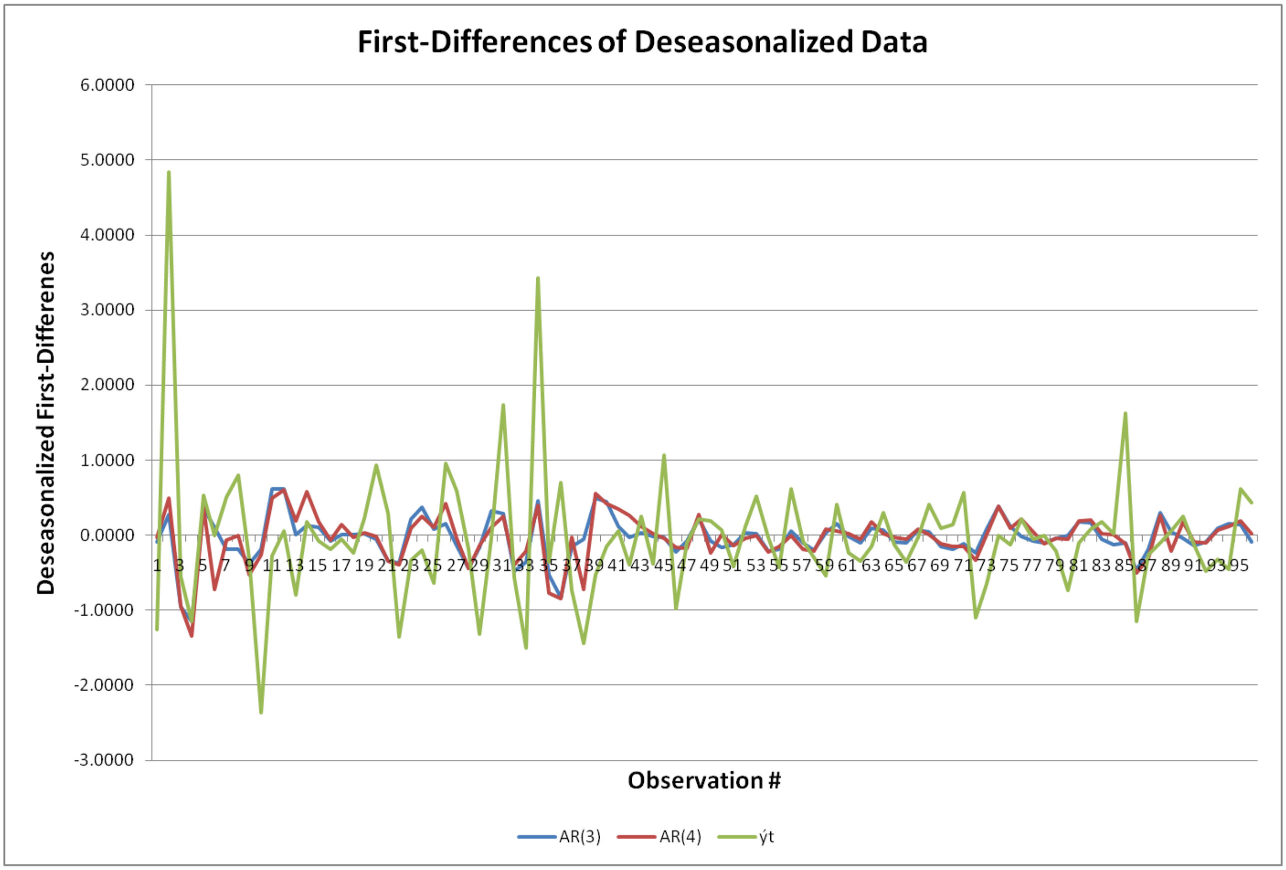
The Box and Pierce (Q) statistic for this model with K=25 was 11.7868. At a 10% significance level, this statistic is far below the X^2 critical value of 29.6151. The X^2 p-value for 11.7868 with K=25 (which was 0.9452) is much greater than 10%. Therefore, the conclusion remains the same - we cannot reject the null hypothesis H_0: the residuals are a white noise process.

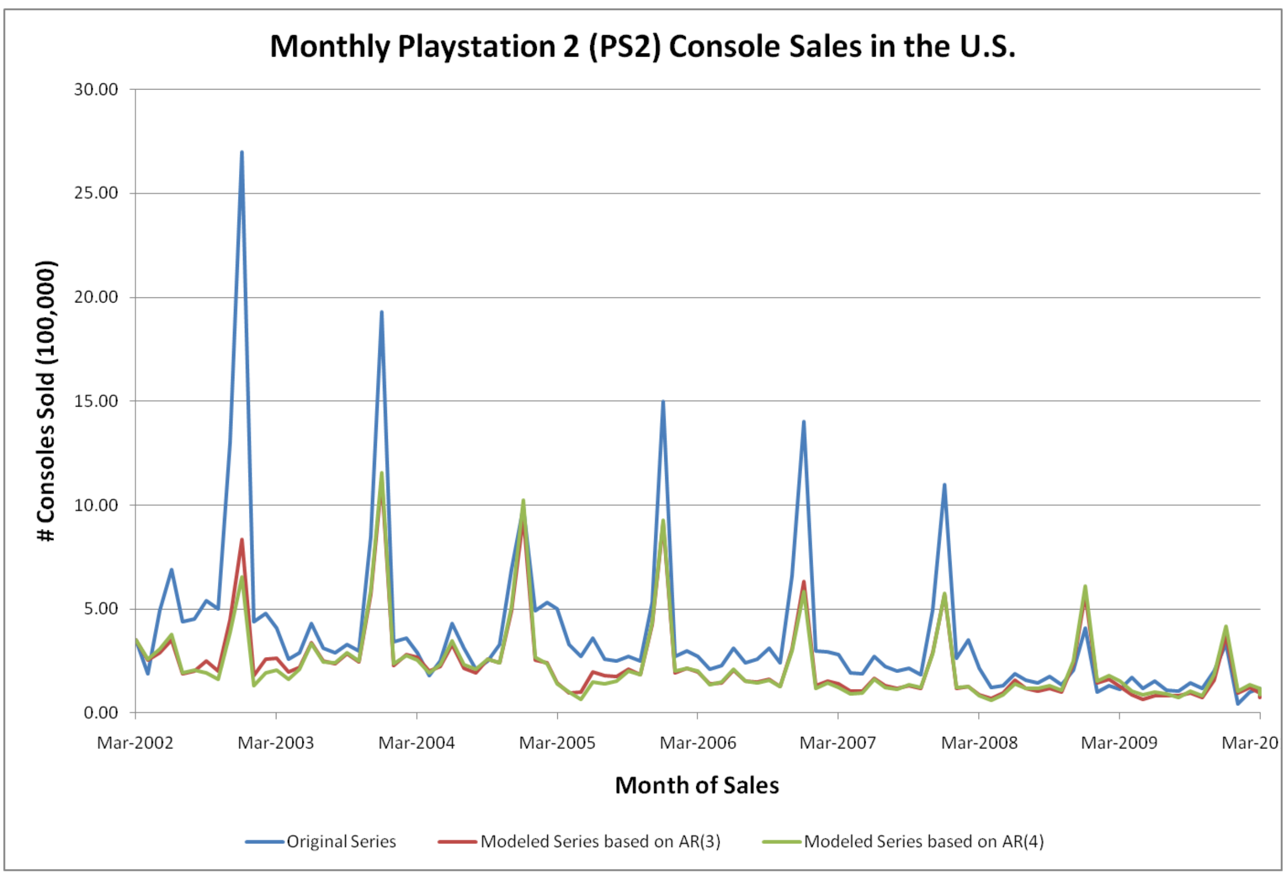
Looking at the regression output above, we see that the R^2 value has become 0.1354. This is slightly better, but not by much. This suggests that we need to significantly increase the order of the autoregressive model or explore the possibility of adding moving average terms. However, finding the optimal model was not the intent of this project. I leave the opportunity for further analysis to upcoming student projects.

Conclusion:

I have created two graphs which compare the original series with the predicted values of our AR(3) and AR(4) models. The first graph shows the first-differences of the deseasonalized data, both actual and modeled. This would be $\acute{y}_t$ as defined above and the predicted values resulting from our Excel regressions.

The second graph shows the monthly Playstation (PS2) console sales in the United States, both actual and modeled. The modeled values were found by performing a reverse transformation on the Excel regression predicted values: adding the modeled first-differences to the actual, initial observation and re-seasonalizing the data. As one can see, the modeled data doesn't give a perfect fit. However, the overall behavior and slight downward trend are apparent.





Supporting Documents:

"KeelyMDavison-TimeSeriesVEEproject.xlsx"