Douglas Kanner
Time Series Course
Summer 2015 Session – Student Project

Examining the Reasonability of the 2% Mean Reverting Model on Personal Income

Introduction

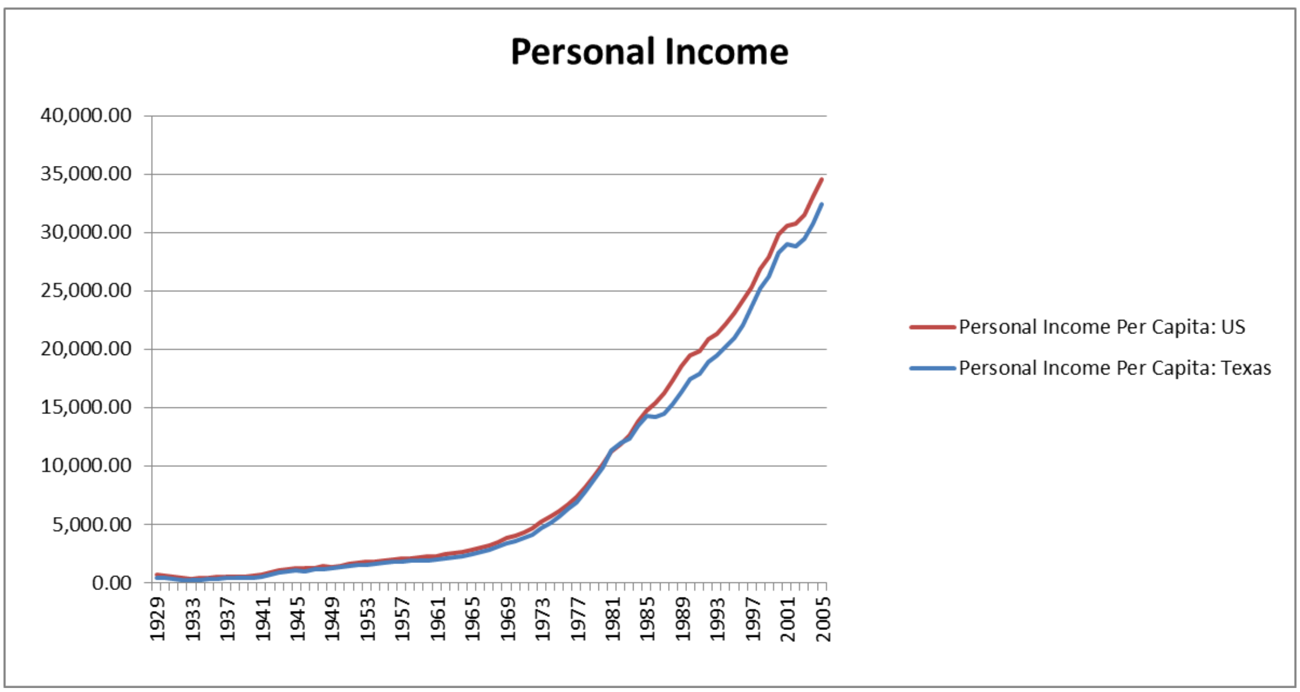
As stated in the "Project Template on ARIMA Modeling of Personal Income," economists believe that the relative income of state per-capita income to national per-capita income (PCI) is mean reverting at approximately 2% per year. I decided to model this for the state of Texas, then examine a couple other ARIMA models for the relative per-capita income of Texas to the national per-capita income in order to see how appropriate the 2% mean reverting model was.

Data Collection

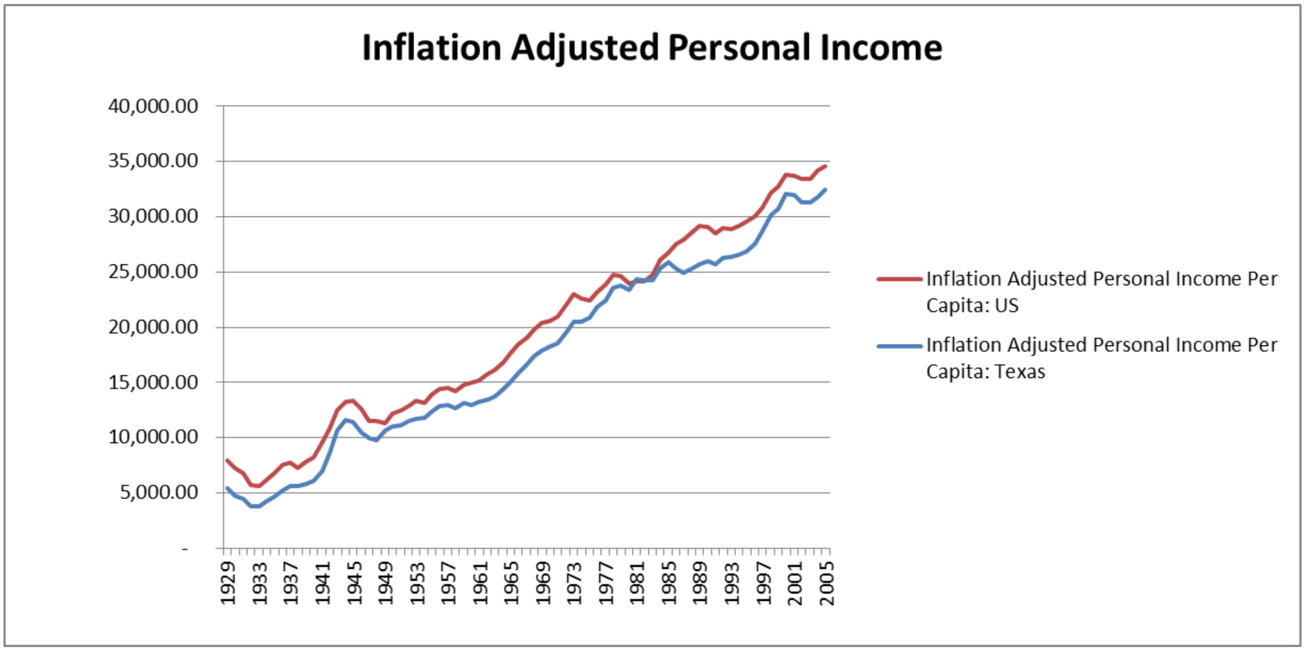
The data used was collected from a couple of sources. For the average per-capita income by year from 1929 to 2005, I collected data from the NEAS Student Project website. In order to inflation-adjust this income, I collected CPI data from the Bureau of Labor Statistics website (http://data.bls.gov/pdq/SurveyOutputServlet).

Creating a Stationary Time Series

Initially, I looked at the individual per-capita incomes without making any adjustment:

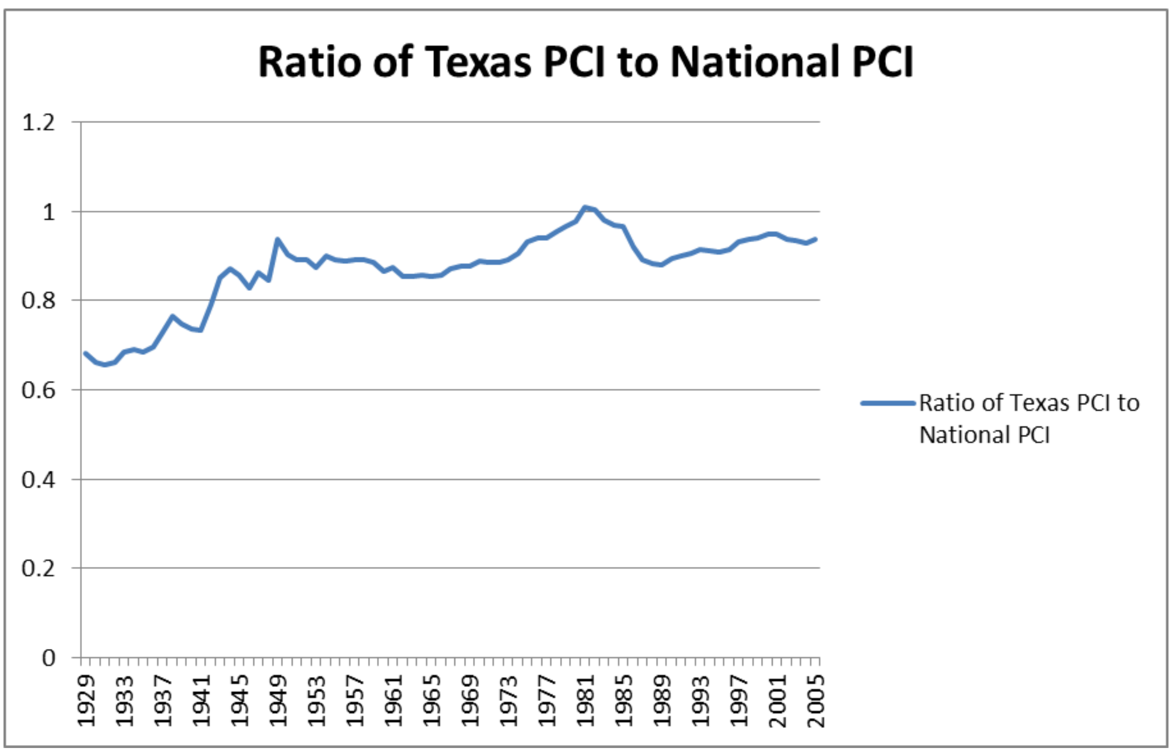


It is quite clear that neither dataset is a stationary time series. One way to begin to adjust this would be to adjust for inflation, giving the following result:

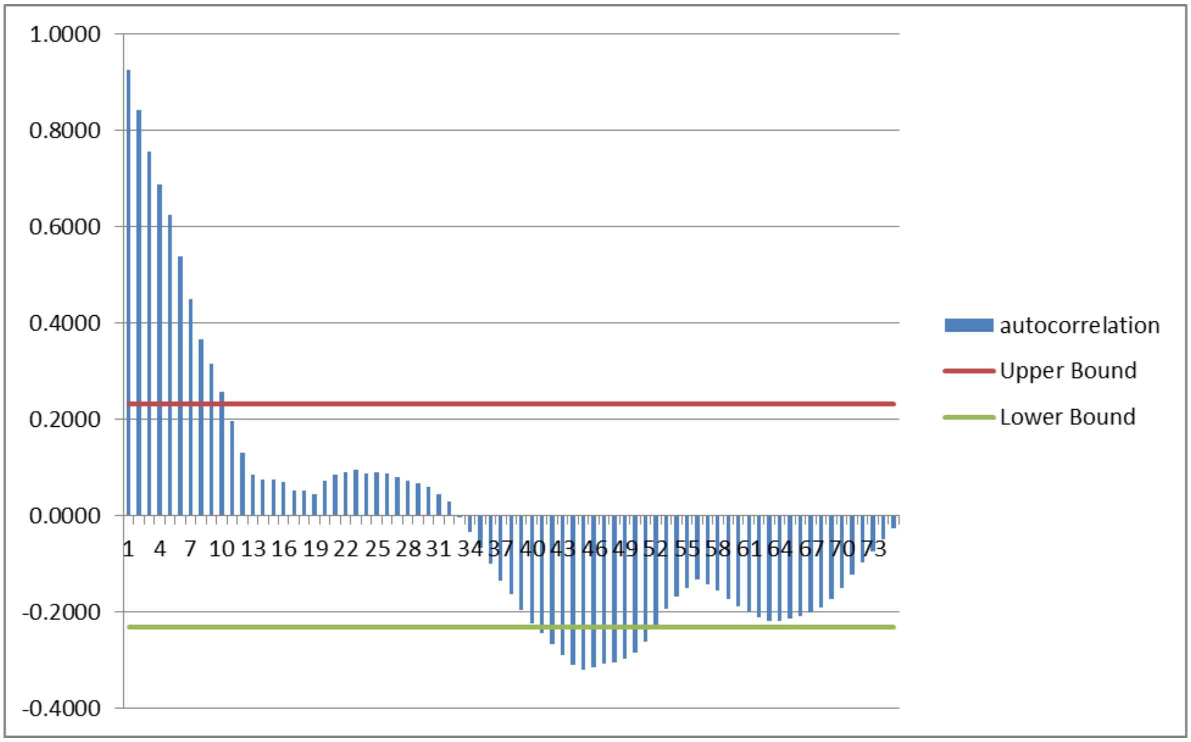


This result could then be analyzed for stationarity, possibly taking log differences if necessary.

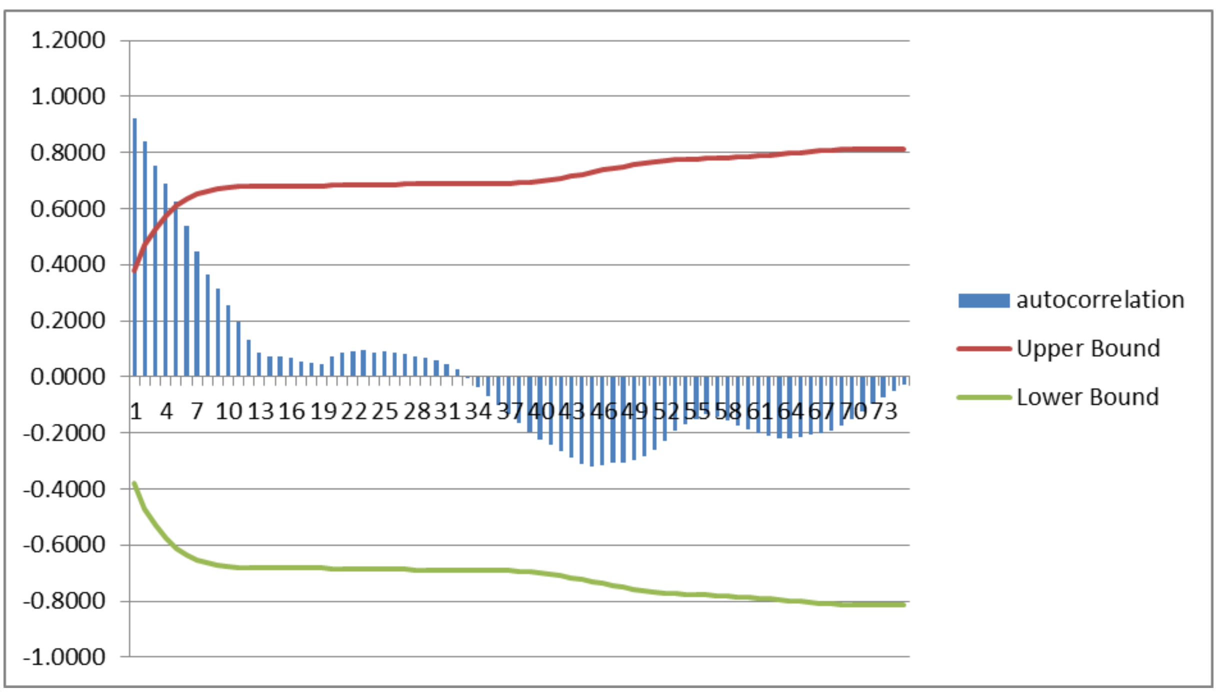
However, for the current analysis, I was more interested in the ratio of Texas PCI to national PCI. Graphing this ratio shows:



Prior to fitting a model, it was useful to the autocorrelation function of the data to see which models should be considered as alternatives. Initially the following graph was produced:



This shows the decaying autorrelation we would expect from an autoregressive process, however, it seems that lags of approximately 1-8 could be significant. Applying the more refined boundaries, however, produced:



This suggests that only lags of 1-4 would possibly be significant, and in fact it is likely only necessary to examine lags of 1-2, since lags beyond that level decay quite quickly.

Fitting the initial model

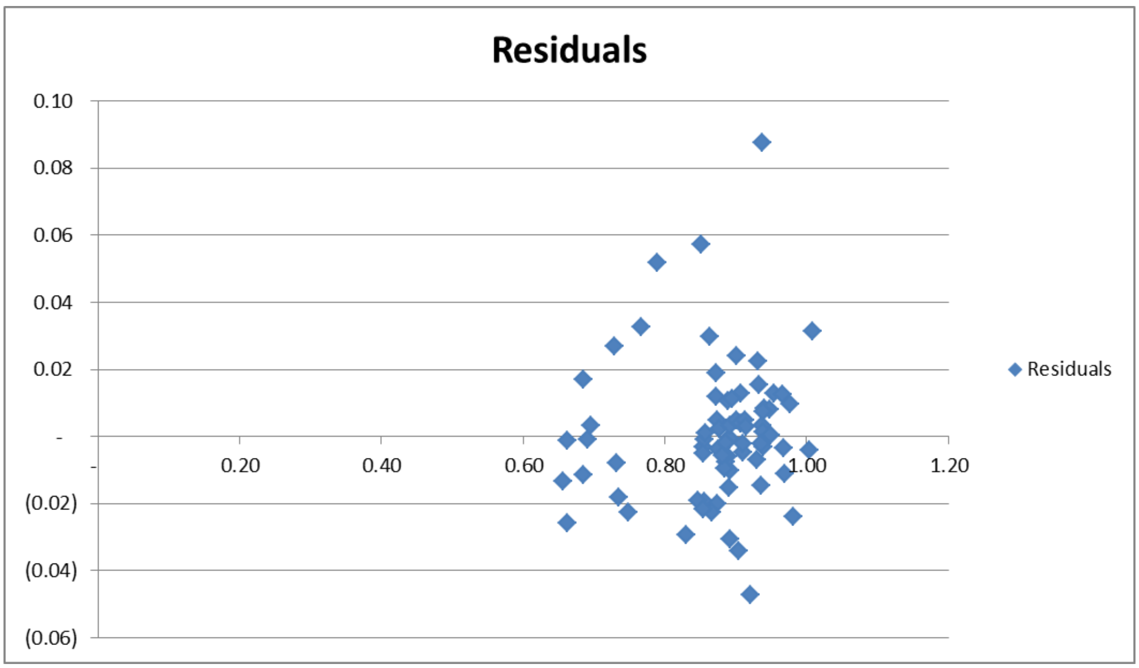
Fitting the initial model was a matter of constructing the mean-reverting autoregressive process with 2% reversion per year. This was done as follows:

$Y_t = 1 - 0.98*(\mu - Y_{t-1})$

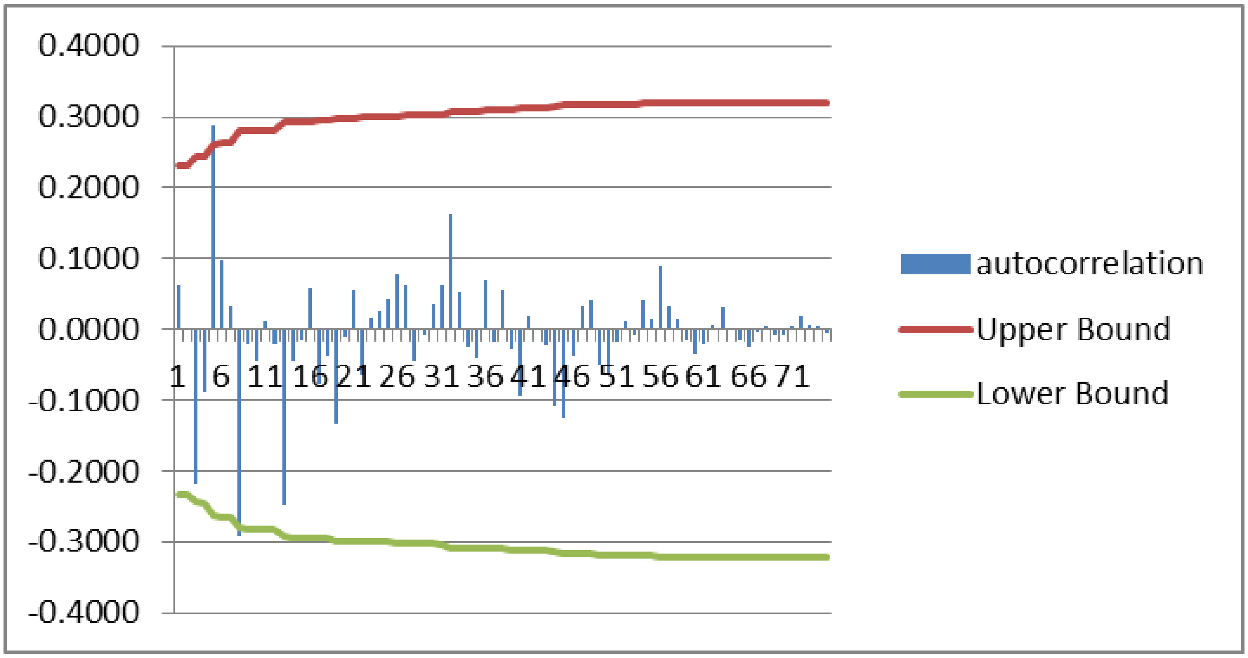
$Y_t = 1 - 0.98*(1 - Y_{t-1})$

$\mathbf{Y_t = 0.02 + 0.98* Y_{t-1}}$

Using the final equation on the dataset, produced a set of residuals that at an initial glance look relatively stationary. However, it is necessary to test this. First, an eye test can be applied to create a residual plot:



This shows no gross trend or cause for concern, thus we proceed to the next test. Looking at a correlogram of the residuals produced the following:



This shows a small potential autocorrelation at lag 5, but this does not seem damaging to the hypothesis that the residuals are white noise given the results of the next tests.

Using excel, the Durbin-Watson Statistic to test for serial correlation of the residuals was calculated to be 1.85. This is near 2, which is the value that suggests no serial correlation.

Next, the Box-Pierce Q statistic was examined. There are 76 residuals, so when calculating the Box-Pierce Q statistic, we expect 75 degrees of freedom. At the 76^{th} residual, the sum of the square of the product of the residuals was 0.00046, the sum of the square of the residuals was 0.03137, and the subsequent chi-square test statistic calculated (using 76 observations and 75 dof) was 34.75. Comparing this to a critical value of 89.95 at 10% significance, we fail to reject the null hypothesis that the residuals are white noise.

Armed with the above information, it is reasonable to conclude that the ARIMA(1,0,0) model with ϕ=0.98 is an appropriate model for the ratio of Texas PCI to national PCI, and it is suggested that the 2% annual mean reversion is an appropriate conclusion.

Additional models

It was also appropriate to test a couple of additional models, to see whether another ϕ is suggested by the data. Accordingly, an ARIMA(1,0,0) and an ARIMA(2,0,0) model inferred from the data:

ARIMA(1,0,0)

In order to derive the appropriate ARIMA(1,0,0) model, the lag-1 ratio of Texas PCI to national PCI was regressed on the unlagged ratio. This model was calculated in excel with the following result:

Regression Statistics	
Multiple R	0.970952319
R Square	0.942748406
Adjusted R Square	0.941974736
Standard Error	0.020331165
Observations	76

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.503691323	0.503691323	1218.540432	1.03861E-47
Residual	74	0.030588364	0.000413356		
Total	75	0.534279687			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.052218631	0.023619113	2.210863338	0.030135289
Y_{t-1}	0.943814752	0.027037516	34.90759849	1.03861E-47

	Lower 95%	*Upper 95%*
Intercept	0.005156521	0.099280741
Y_{t-1}	0.889941325	0.997688178

The ARIMA(1,0,0) model calculated from the data suggests a slightly different ϕ from our initial model (0.94 instead of 0.98) but contains our initially hypothesized model within the 95% confidence intervals of both the intercept and the ϕ coefficient; (0.05, 0.1) and (0.89, 1.0) respectively. This suggests that we would fail to reject the null hypothesis that the correct ARIMA(1,0,0) model for the data is the aforementioned $\mathbf{Y_t = 0.02 + 0.98^* Y_{t-1}}$

ARIMA(2,0,0)

What the above model did not test was whether the appropriate number of coefficients are being used. Regressing on lag-1 and lag-2 ratios produces the following ARIMA(2,0,0) model:

Regression Statistics	
Multiple R	0.969736113
R Square	0.940388129
Adjusted R Square	0.938732243
Standard Error	0.020130621
Observations	75

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.46027905	0.230139525	567.9065578	8.16543E-45
Residual	72	0.029177416	0.000405242		
Total	74	0.489456465			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.062900049	0.024223294	2.596676155	0.011404608
Y_{t-1}	1.010000884	0.115248203	8.763701782	5.75501E-13
Y_{t-2}	-0.078222662	0.111933098	-0.698834062	0.486907036

This model shows an extremely insignificant (P-value of 0.49) lag 2 coefficient. Comparing the Adjusted R-square to the ARIMA(1,0,0) model shows this model to be a worse fit than the ARIMA(1,0,0) model. Using the information from these two statistics shows that the lag-2 term is adding no value to the model, and is actually hurting the fit.

Conclusion

Based on the above analysis, I think it is reasonable to conclude that the 2% mean reverting model is an appropriate model for the ratio of Texas PCI to national PCI. The analysis showed the residuals as white-noise, and the model had a strong fit to the data.