Name: Elizabeth Dailey
Course: Time Series
Session: Summer 2015
E-mail: emdailey15@gmail.com

Time Series Study Project: Personal Income

Introduction

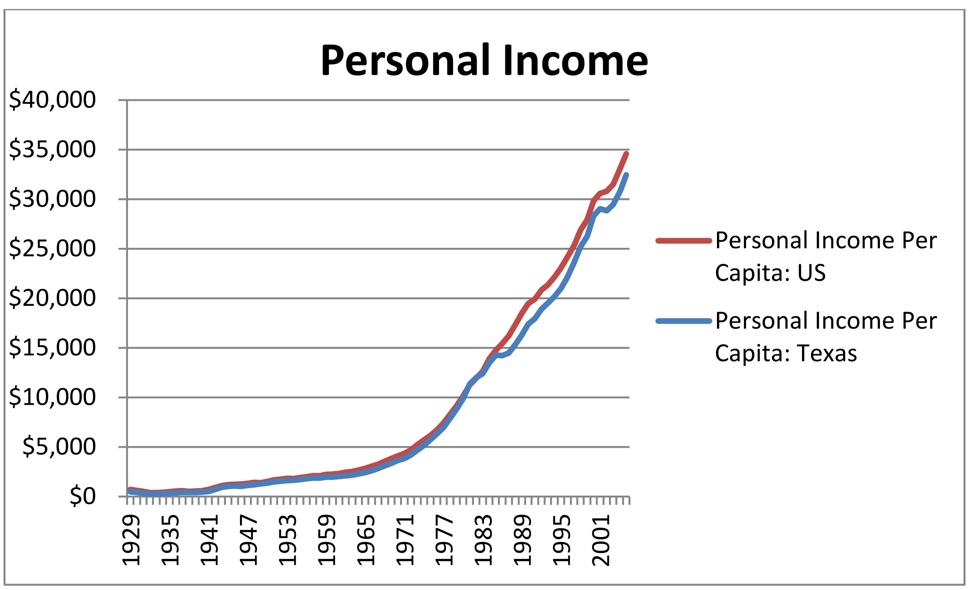
For my time series project, I chose study how personal income has changed over time for the entire United States and an individual state. I decided to analyze the data for the state of Texas since that is where my largest consulting client is located. Note that detailed calculations for the results shown below have been provided in the attached Excel workbook.

Data

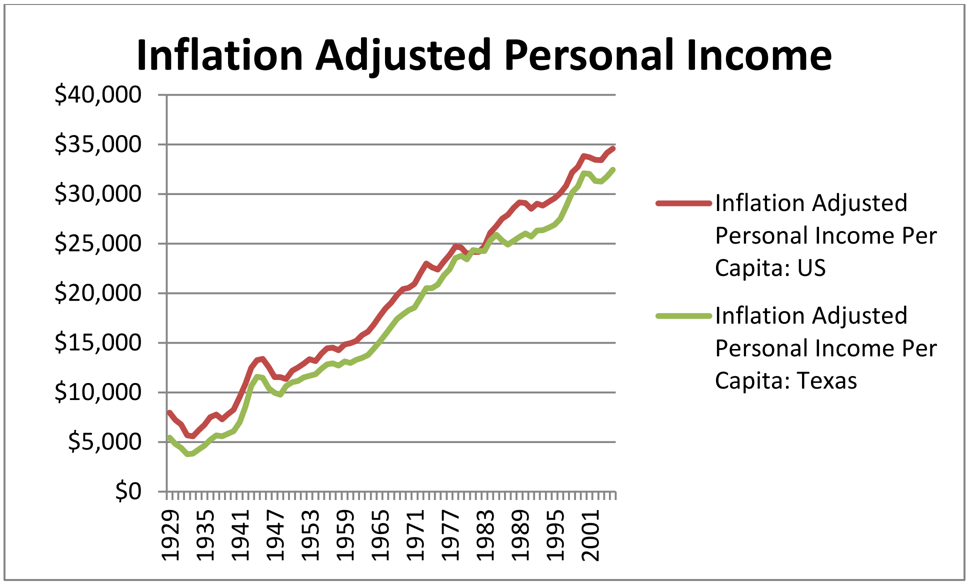
The data I collected on personal income for the US and Texas was from the "Project Template on ARIMA Modeling of Personal Income" post on the NEAS VEE Time Series Student Project message board (http://tempforum.neas-seminars.com/Topic7751.aspx). I was also interested in modeling personal income after inflation, so I collected data on the Consumer Price Index (CPI) for all urban consumers. This data was available from the Bureau of Labor Statistics on the following webpage: http://data.bls.gov/cgi-bin/surveymost?cu.

Analysis

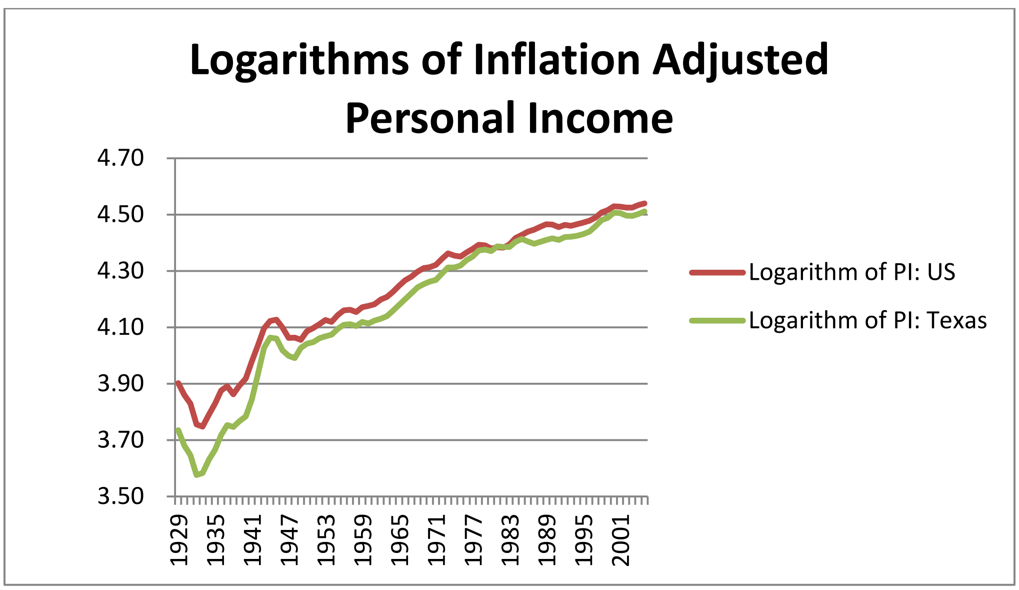
The graph below shows personal income for the US and Texas from 1929 to 2005. For both the US total and Texas, the curve increases slowly until the 1970s, at which point the growth in personal income rates appears to be exponential.



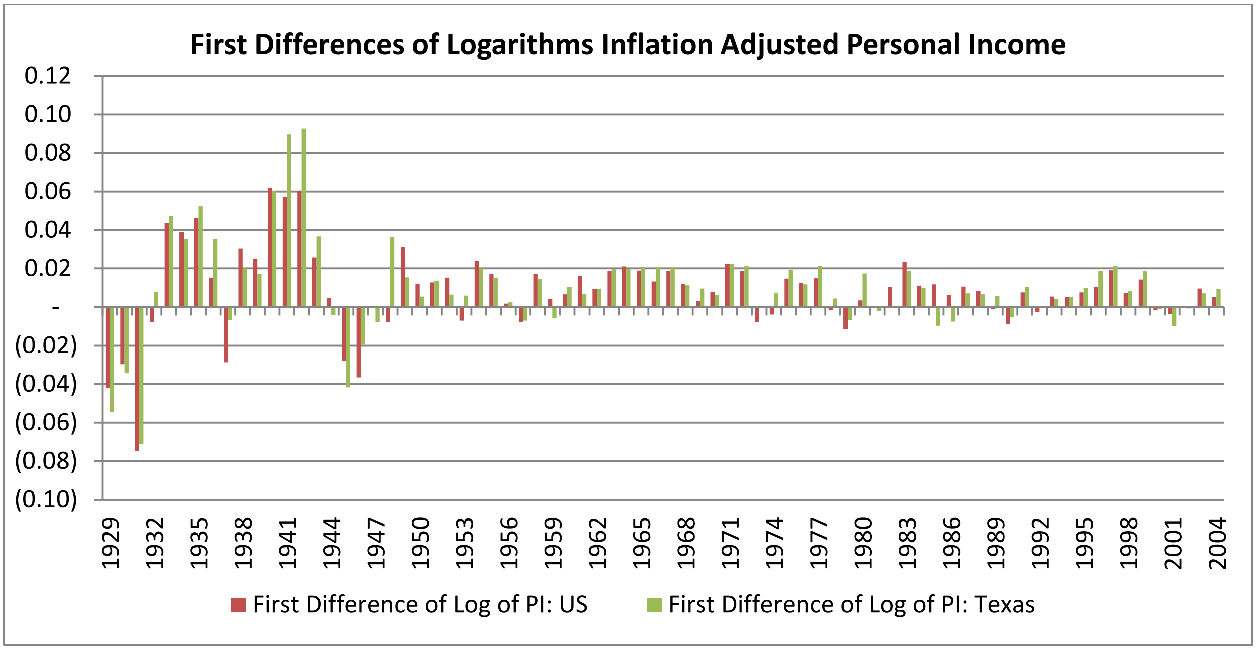
I felt that some of the steep growth could be caused by inflation, so I wanted to view personal income after taking this into account. Using the CPI data, I adjusted personal income in each year to represent the values after inflation. The resulting graph shows a more linear relationship of personal income over the period.



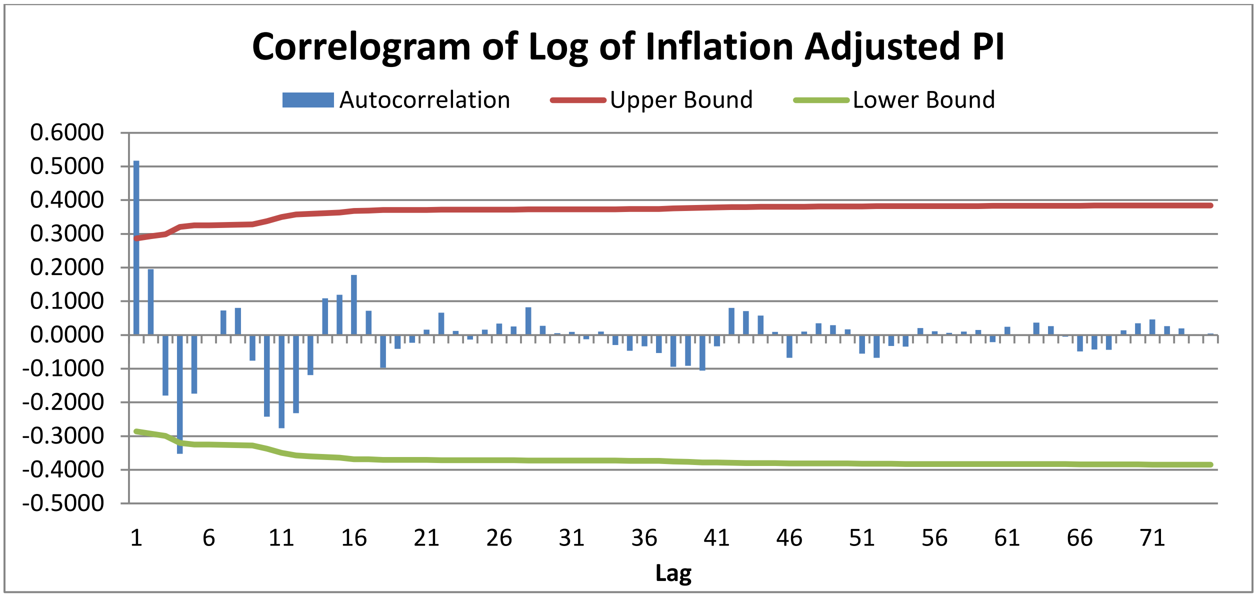
Stationarity
In order to fit an autoregressive model to my time series, I first needed to ensure stationarity. The upward trend of the inflation adjusted personal income led me to believe this series was not stationary, so I began by taking the logarithm of each data point. The results are displayed in the graph below.



Even after taking logarithms of the inflation adjusted personal income data points, there was still an upward sloping trend. Therefore, I decided to take the first difference of the logarithms which is graphed below.



Though there are larger deviations in early years, the time series appears to be stationary since there does not appear to be a trend and the values are all around zero. In order to confirm this hypothesis, I calculated the sample autocorrelation function for each lag and examined the correlogram of the first differences of logarithms of inflation adjusted personal income. Since the time series appear to move in a similar pattern, I have focuses on the US information to develop all following information.



The bounds are determined by the complex standard error formula found in “Time Series Analysis” by Cryer and Chan as Equation 6.1.11 on page 112. With the exception of lag 1 and 4, the sample autocorrelations fall within the bounds. Since the process has autocorrelations close to zero after a few lags, the process appears to be stationary.

The sample autocorrelations appear to have a reasonably geometric decline over time, so I felt an AR(1) would be the best model. Since the first observation is above the confidence interval, a rapid decline could be concluded which suggests a moving average process. However, since lag 4 also falls outside the confidence interval, I have decided a moving average process may not be the best fit so I have focused on fitting autoregressive models for this time series.

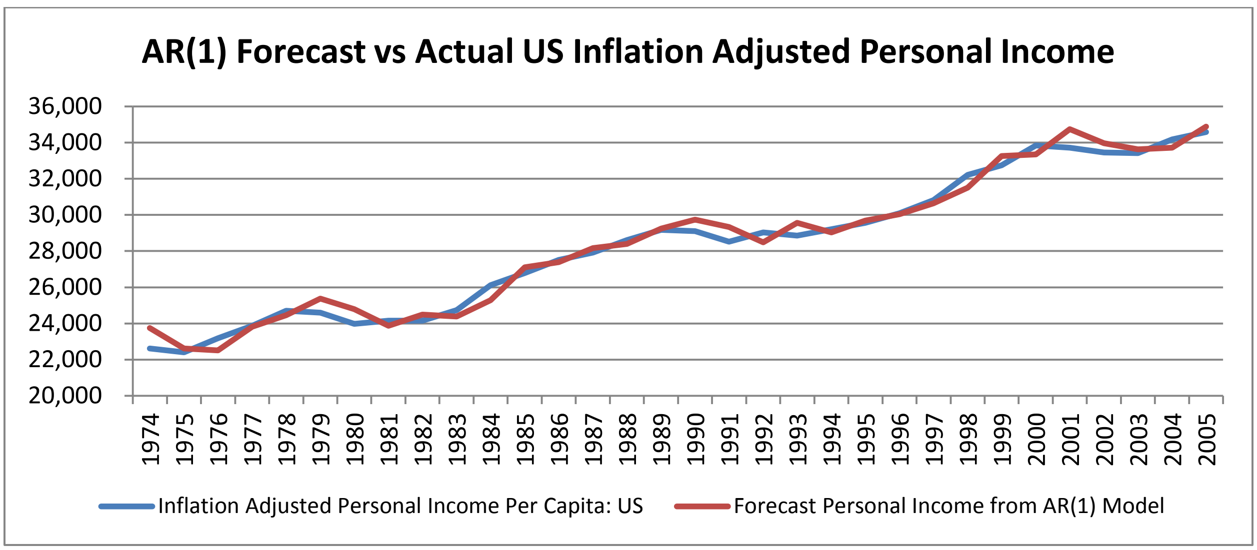
First-Order Autoregressive Model – AR(1)

The equation for an AR(1) model is $Y_t = \Phi Y_{t-1} + e_t$.

Using parameter estimation, we can estimate Φ to be the sample autocorrelation of lag 1 since $\widehat{\Phi} = \rho_1 \approx r_1$. Based on the calculations in the attached workbook, Φ is approximately 0.5169 and the model can be rewritten as $Y_t = 0.5169 Y_{t-1} + e_t$.

The AR(1) forecast formula is $\widehat{Y_t(\ell)} = \mu + \Phi^{\ell}(Y_t - \mu)$ which estimates the value of $Y_{t+\ell}$ using time t as the forecast origin and ℓ as the lead time. The mean of the series, μ, is equal to 0.00839. Using the previously determined parameters, the forecast for time t+1 can be written as $\widehat{Y_t(1)} = 0.00839 + 0.5169(Y_t - 0.00839)$.

The graph below shows the actual time series compared to these forecast estimates over the past 30 years.



Second-Order Autoregressive Model – AR(2)

The equation for an AR(2) model is $Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + e_t$.

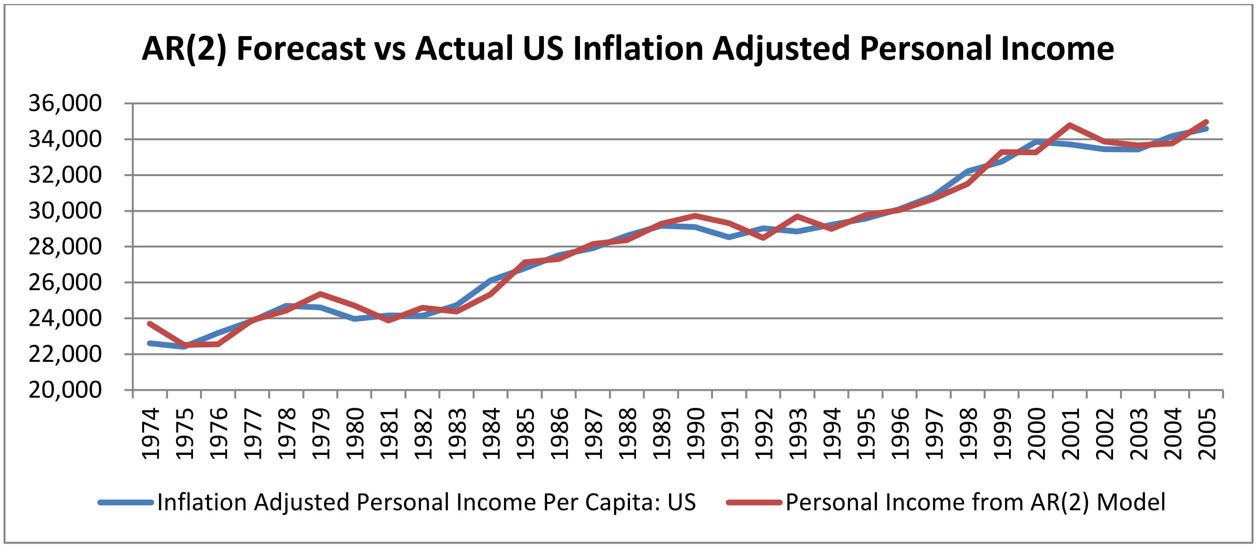
Using parameter estimation, we can estimate Φ_1 and Φ_2 based on the Yule-Walker equations as follows:

$$\widehat{\Phi_1} = \frac{r_1 * (1 - r_2)}{1 - {r_1}^2} = \frac{0.5169 * (1 - 0.1948)}{1 - (0.5169)^2} = 0.5680$$

$$\widehat{\Phi_2} = \frac{r_2 - {r_1}^2}{1 - {r_1}^2} = \frac{0.1948 - (0.5169)^2}{1 - (0.5169)^2} = -0.0988$$

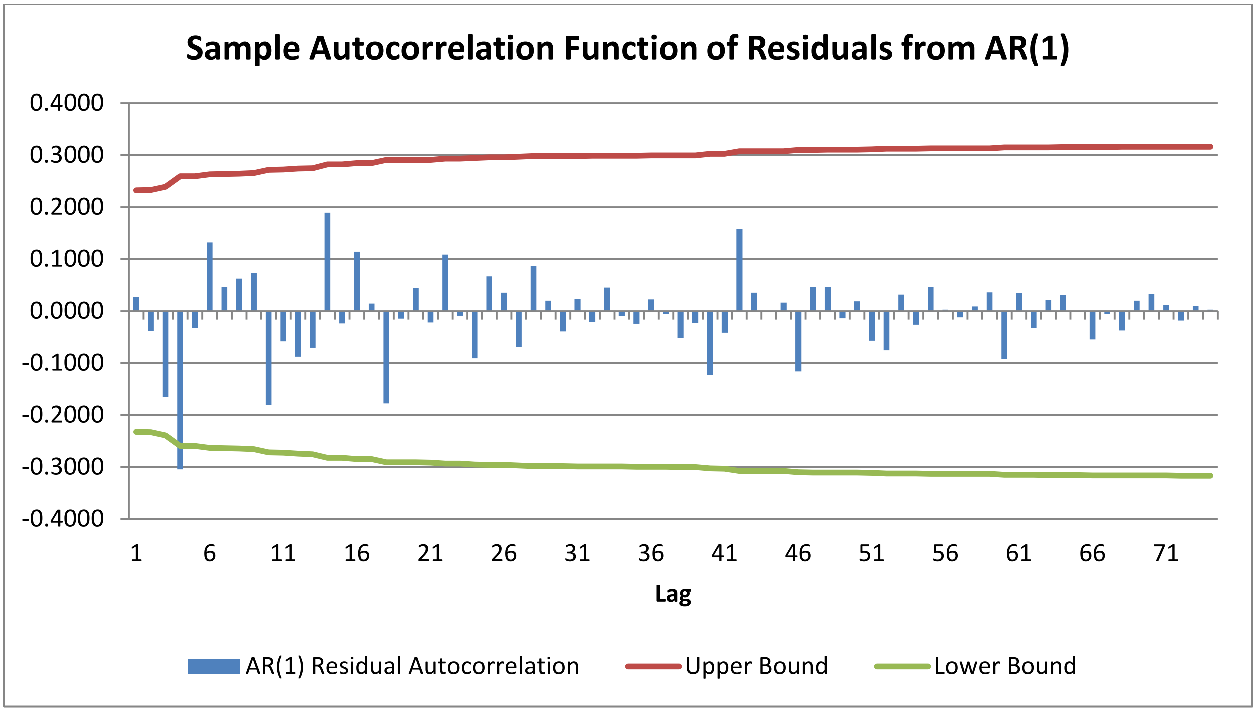
Using the estimated parameters, the model can be rewritten as $Y_t = 0.5680\, Y_{t-1} - 0.0988\, Y_{t-2} + e_t$.

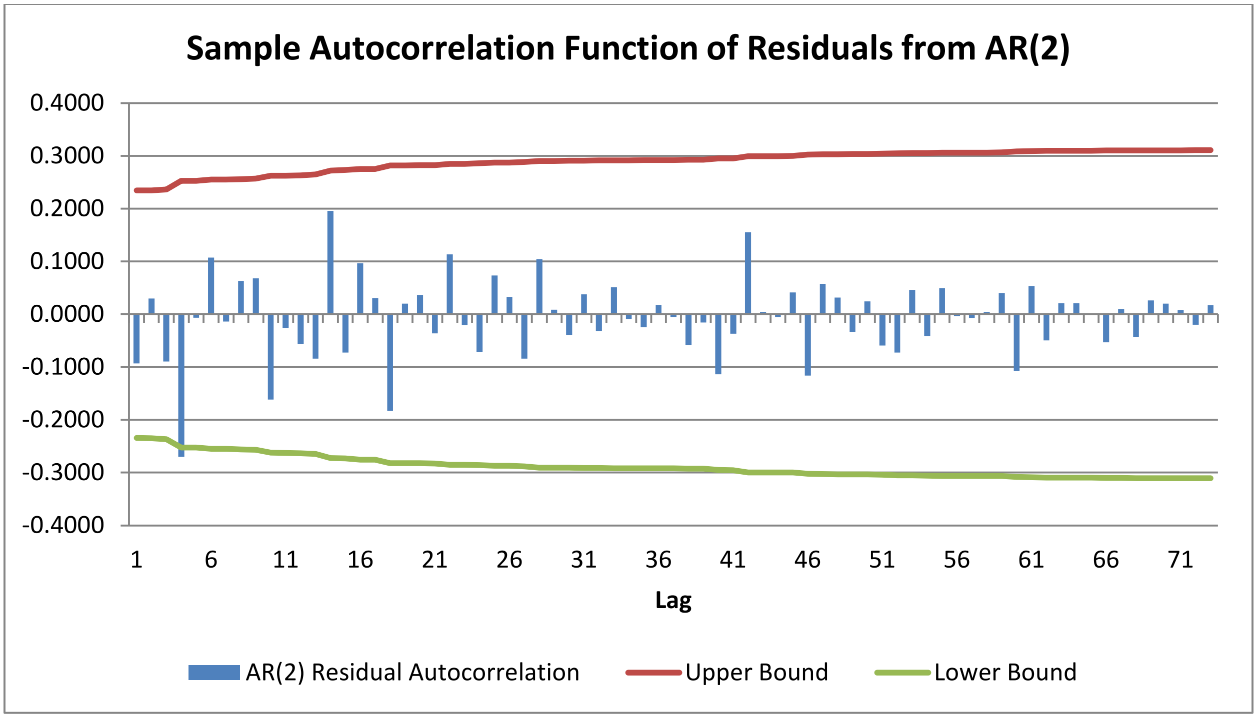
The AR(2) forecast formula with lead time of 1 is $\widehat{Y_t(1)} = \mu + \Phi_1(Y_{t-1} - \mu) + \Phi_2(Y_{t-2} - \mu)$. Using the mean of the time series, the following graph shows the forecast values using the fit parameters for the past 30 years.



Residual Analysis

Based on the forecast graphs alone, it is hard to determine whether the AR(1) or AR(2) model is a better fit for this time series. Therefore, I will use residual analysis to compare these two models.





Using the sample autocorrelation functions of the residuals for the two different models, we see that both of them fall within the bounds except at lag 4.

<u>Conclusion</u>

I have created inflation adjusted personal income models based on first- and second-order autoregressive time series. Since the residuals for both models indicate a relatively good fit, I would recommend using the AR(1) based on the principle of parsimony.