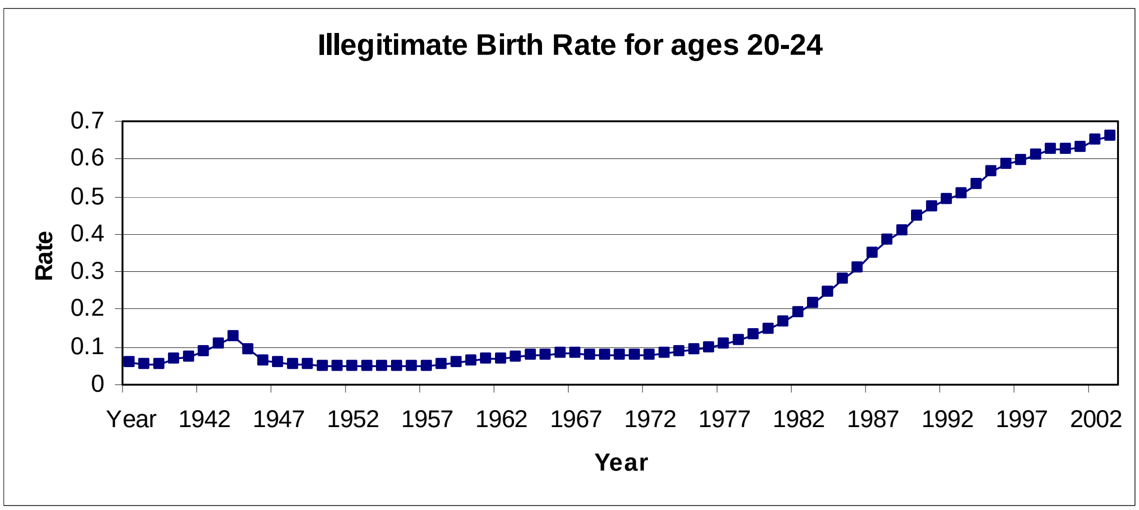
Time Series Project – Illegitimate Birth Rates
Susan Woodward
November 2009

Introduction

Over the past several decades, illegitimate births have become more culturally accepted in the United States and Western Europe. This time series project will deal with illegitimate births from 1938 to 2004 in England and Wales, and will focus specifically on women ages 20 to 24. This project will attempt to fit a time series to this data and will analyze the appropriateness of various models.

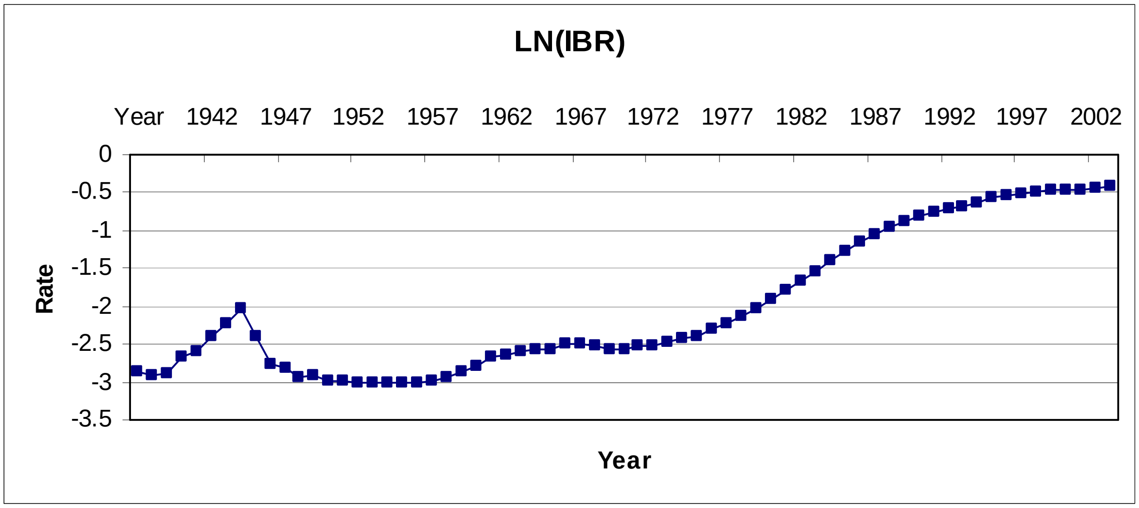
Data Set Analysis

Numbers of births will obviously grow over time as the population of the earth increases, so I created ratios for each year in the period of illegitimate births to total births. The raw data is graphed below.



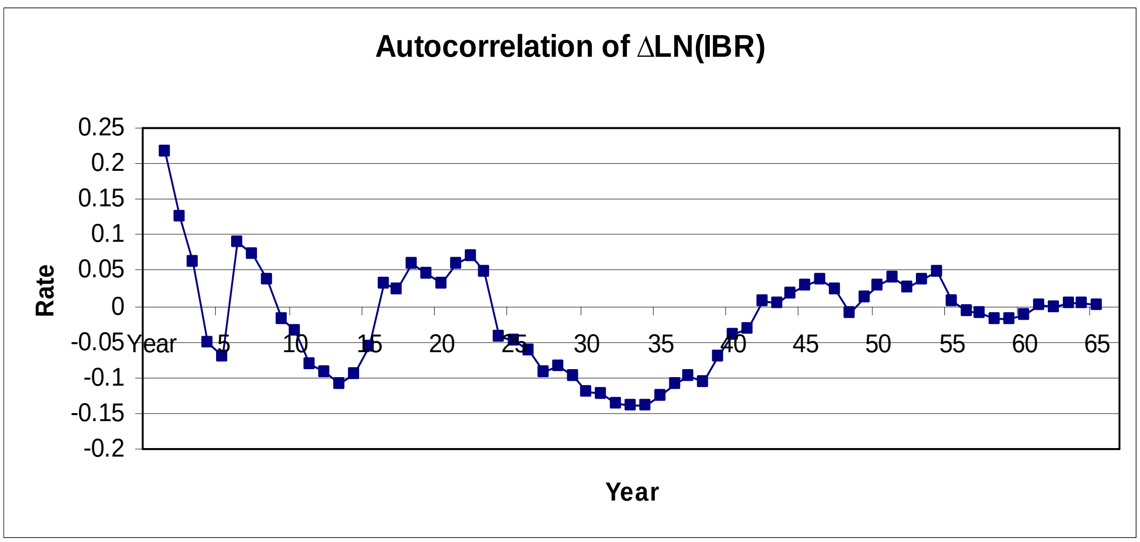
First, we must determine if the series is stationary, as ARIMA models require that the mean and variance of a series be stationary. Our data is obviously increasing, most closely resembling an exponential curve, so the series is not stationary. Therefore, I took the natural log of the data.

Time Series Project – Illegitimate Birth Rates
Susan Woodward
November 2009

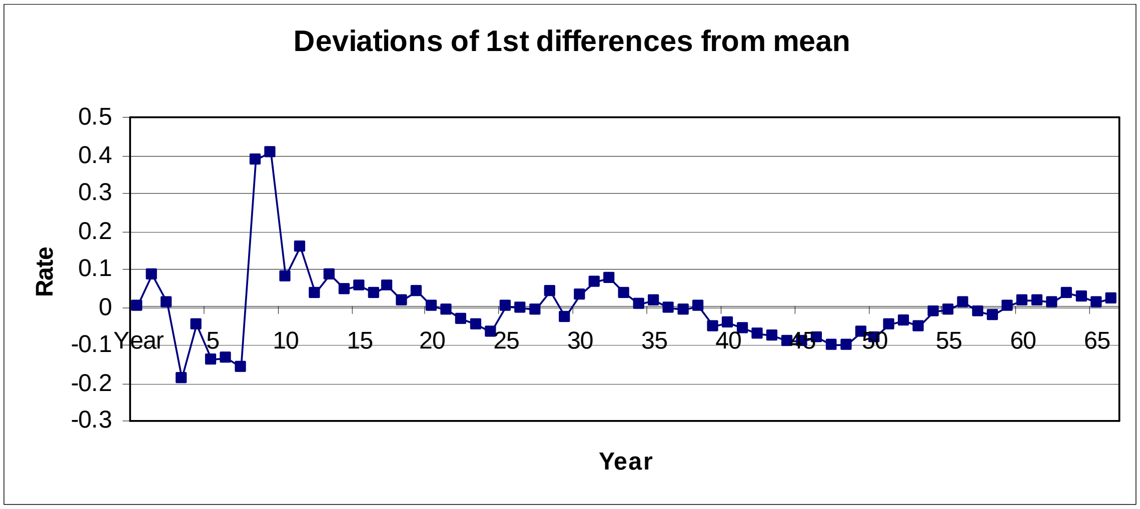


The natural log, while negative, continues to follow the curve of the original data.

To determine if the entire series is stationary, I calculated and graphed the autocorrelation function calculated the deviation of the first differences from the mean.



Time Series Project – Illegitimate Birth Rates
Susan Woodward
November 2009



The correlogram does not go to zero until close to the end of the series. Around point 43 (year 1981), the chart begins to more closely resemble a stationary series. The graph still oscillates but not as much. The deviations of the first differences stay close to the axis from 43 on also.

I decided to only examine the data from 1938 to 1980. I think it will produce a more dynamic model than the later data. From 1980 on, it's seems to be steadily increasing – maybe marriage will be done away with altogether in the UK!

1938 to 1980

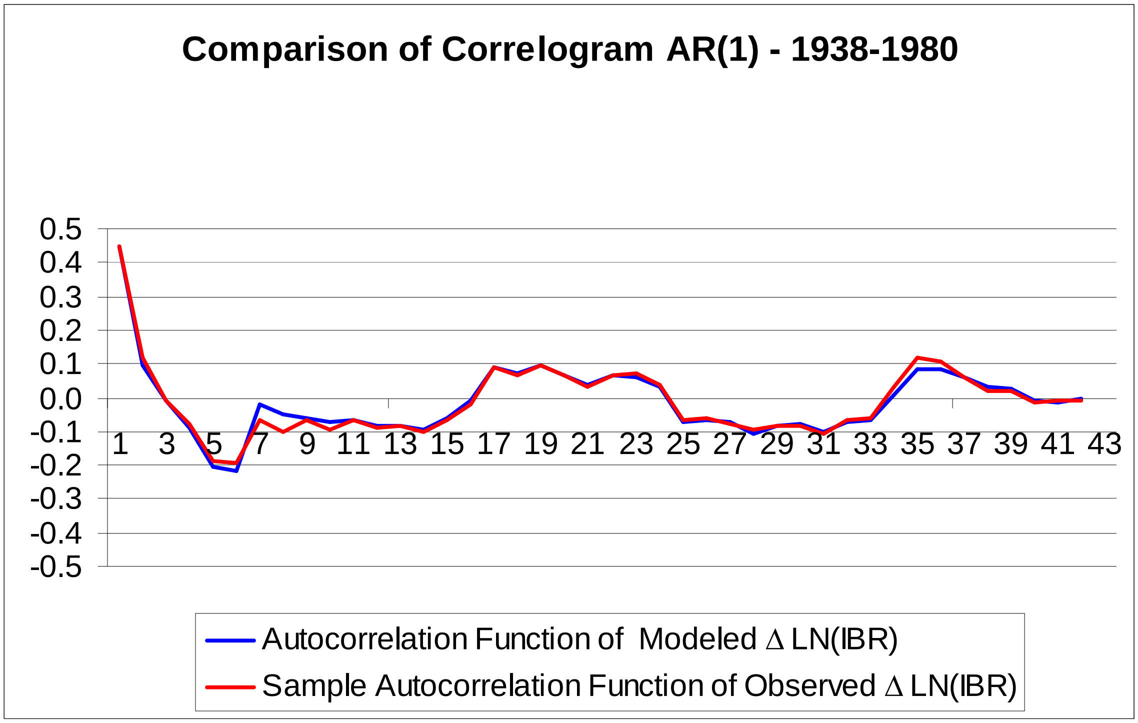
Since the autocorrelation values decline geometrically to zero, this suggests an autoregressive model and not a moving average model. I tested to see what would fit better, an AR(1) or an AR(2) model. This chart shows the results.

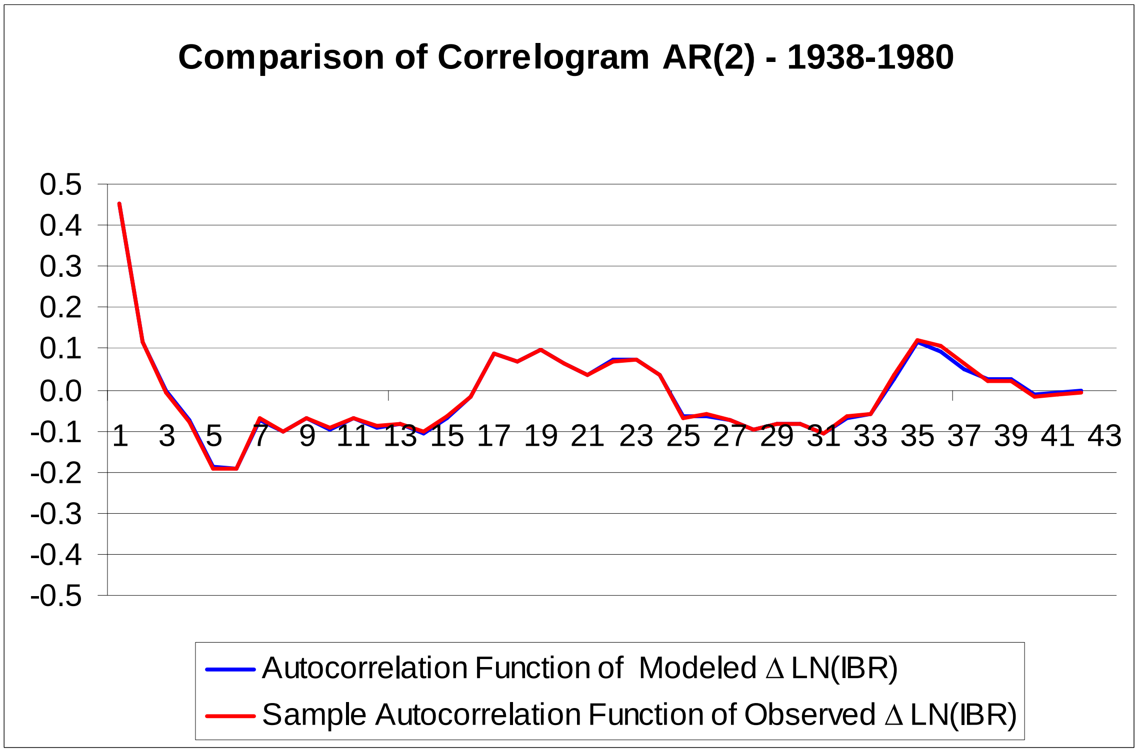
	R Square	Adjusted R Square	P-Value Φ_1
AR(1)	0.00026	-0.02412	0.91825
AR(2)	0.01523	-0.00879	0.43041

The R squared increased as the order of p increased. As was similar to what I looked at in my regression project, we must be careful since R squared is sensitive to the independent variables. We see that the adjusted R squared value goes down as the order of p increases. From the adjusted R squared value, I would say that the AR(3) is the best fit.

Then I compared the correlograms of the two modeled time series to the correlogram of the observed values. The charts below show that AR(2) most closely matches the correllogram of the observed values.

Time Series Project – Illegitimate Birth Rates
Susan Woodward
November 2009



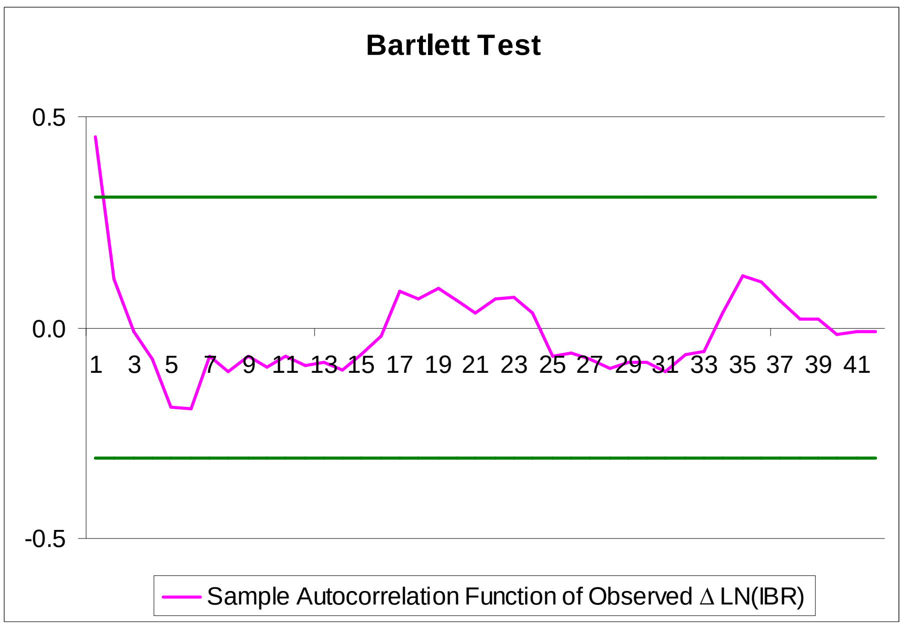


Time Series Project – Illegitimate Birth Rates
Susan Woodward
November 2009

Diagnostic Testing

Now that I have decided that AR(2) is the best fitting model, I want to test to see if the residuals are white noise – this will imply that the model is well fit. I first test for serial correlation in the error terms (residuals). The calculated Durbin-Watson statistic is 1.15. Since this is not close to 2, I am not sure if the lagged residuals are white noise.

The next test I used was Bartlett's test. It tests if the sample autocorrelation coefficients of the residuals have an approximate normal distribution with mean zero and standard deviation 1/√T. We therefore have a standard deviation of 0.1543. If we were to look at the first 20 autocorrelations of the residuals, we would expect (.05*20) = 1 to be outside of +/-(2*0.1543) = +/- .3086. the graph shows only a few values outside of this range, so I feel comfortable saying that the autocorrelations vary normally around zero. See the graphed results of Bartlett's Test below.

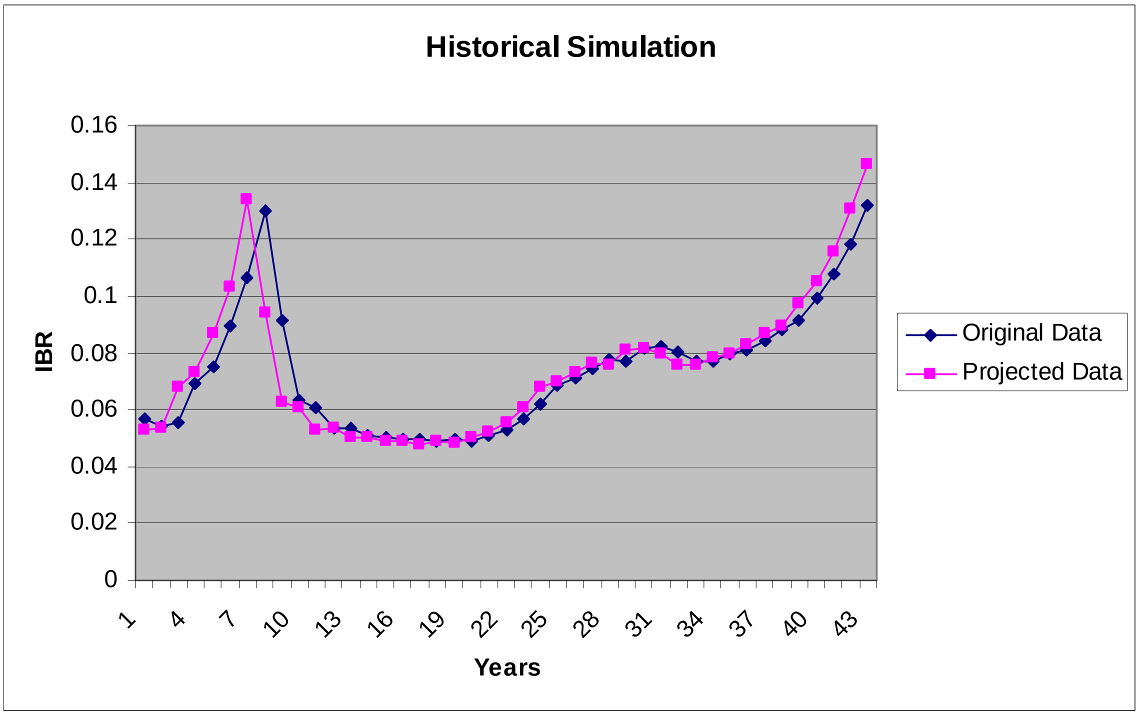


Finally, I calculated the Box-Pierce Q statistic. The Q statistic was calculated for the 42 observations and the autocorrelation function for the first 42 lags of the residuals. I compared the Q statistic of 10.98 with the chi- squared critical value for 41 degrees of freedom which is approximately 51.81. The Q statistic does not exceed this, so I would accept that all the autocorrelations of the residuals are zero with a 90% confidence level. This test implies that the residuals are white noise.

Time Series Project – Illegitimate Birth Rates
Susan Woodward
November 2009

Conclusion

I took historical data and created a time series, which, based on the graph below seems fairly accurate. In 1980, however, how could a time series have predicted the disintegration of reliance on marriage as an appropriate cornerstone of parenthood? Or that young women, seemingly of a mature-enough age to know better, would increasingly make the decision to have unprotected sex?



Please see the attached worksheet for data work.