Name: Mai Ly
Course: Time Series
Session: Spring 2013 VEE Student Project

Modeling Population in Canada (1946 - 2013)

Introduction

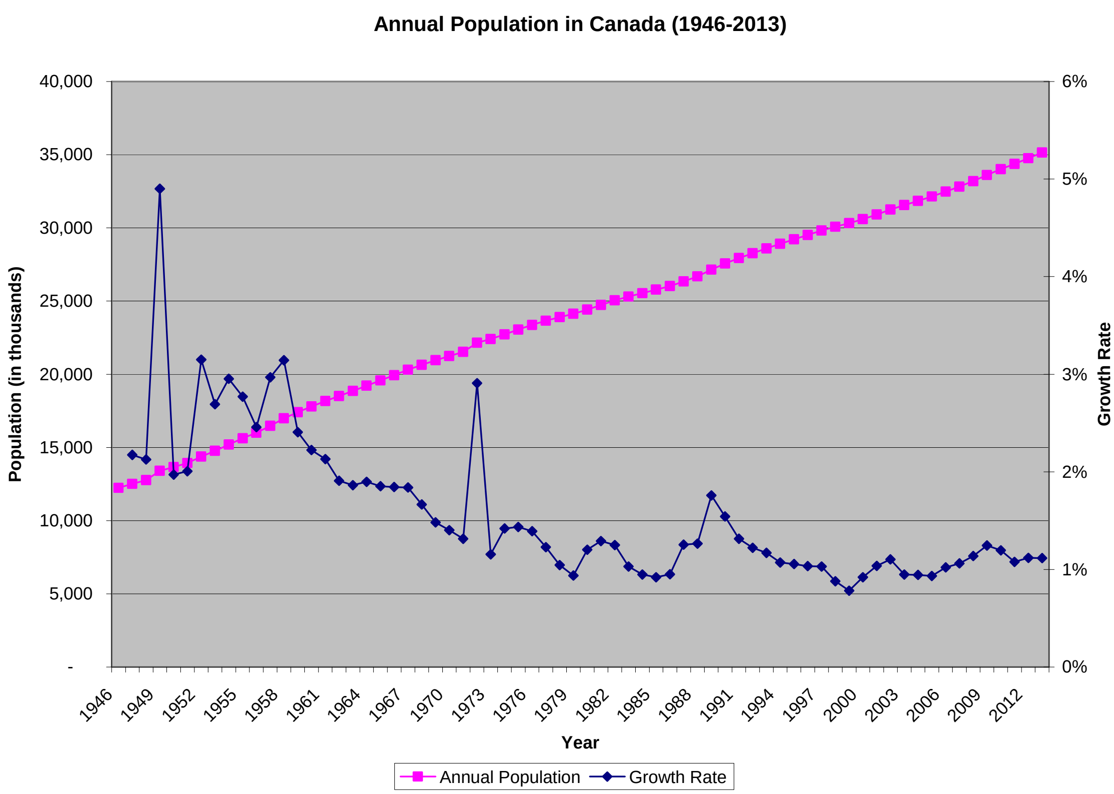
Population growth rates vary greatly among countries. Many countries have experienced rapid growth rates while others have experienced the opposite. My time series project will analyze population in Canada using data from 1946 to 2013 and forecast the expected 2014, 2015 and 2016 population with the resulted model. Canada is a multicultural country, I would expect it has a steady population growth rate as its natural growth rate can be balanced with immigration controls.

Data Source

Data of population in Canada can be extracted from Statistics Canada (www.statcan.gc.ca). These are annual population estimates at each calendar year ending June 30, averaging of four quarterly estimates. Population figures below are all in thousands.

Analysis

Figure 1 shows annual population and growth rate in Canada from 1946 to 2013.



Name: Mai Ly

Course: Time Series

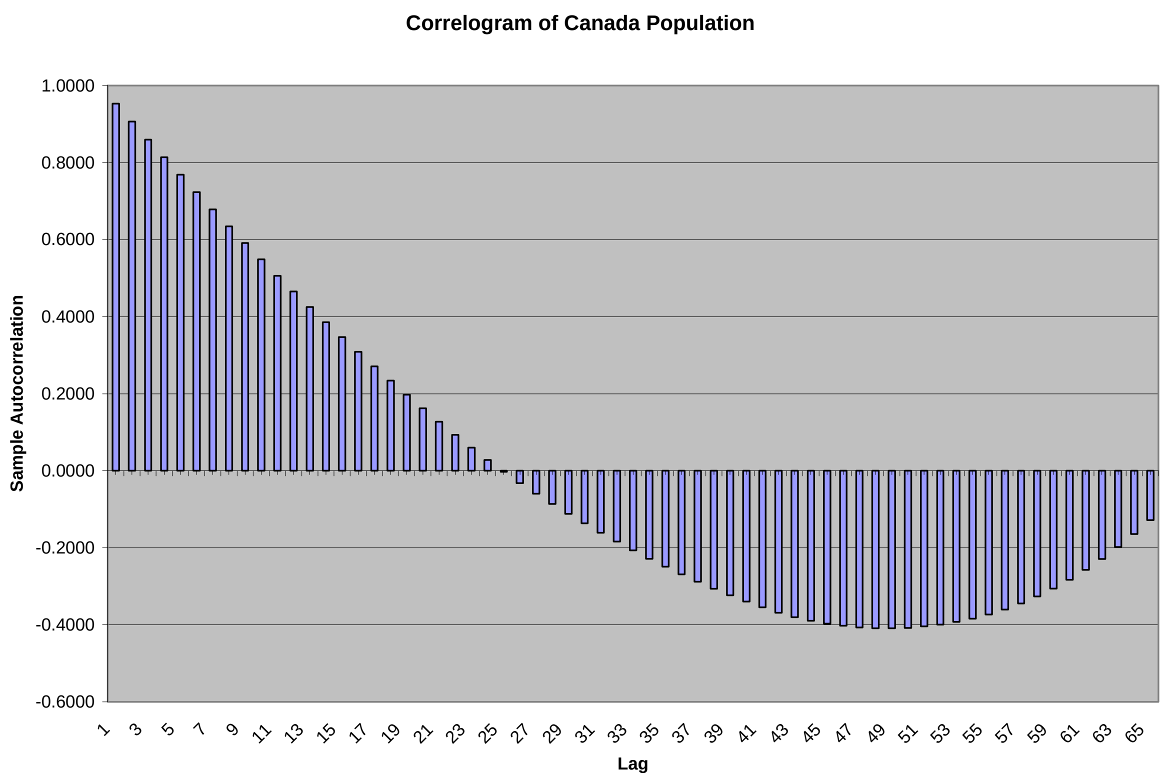
Session: Spring 2013 VEE Student Project

Population remained relatively stable with a period of high population growth from 1946 to 1961 due to strong natural increase and migratory increase. Starting 2001 population grew at an average annual rate of approximately 1.0%.

Before modeling the population data, I will examine autocorrelation to check if this series is stationary and whether the series need to be differenced in order to get a stationary series. The sample autocorrelation is calculated using the following equation where k is from 1 to 65.

$$r_k = \frac{\sum_{t=k+1}^{n}\left(Y_t - \bar{Y}\right)\left(Y_{t-k} - \bar{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \bar{Y}\right)^2}$$

Figure 2 shows sample autocorrelation of Canada population for different lags.



This sample autocorrelation plot shows that the time series is not random, but rather has a high degree of autocorrelation between adjacent and near-adjacent observations. It shows high autocorrelation at lag 1 then slowly and steadily declines. It continues decreasing and becomes negative then starts showing an increasing negative autocorrelation. This indicates an AR(p) time series rather than MA(q) since the autocorrelation shows no sudden drop but gradually decreasing then increasing back up.

Name: Mai Ly

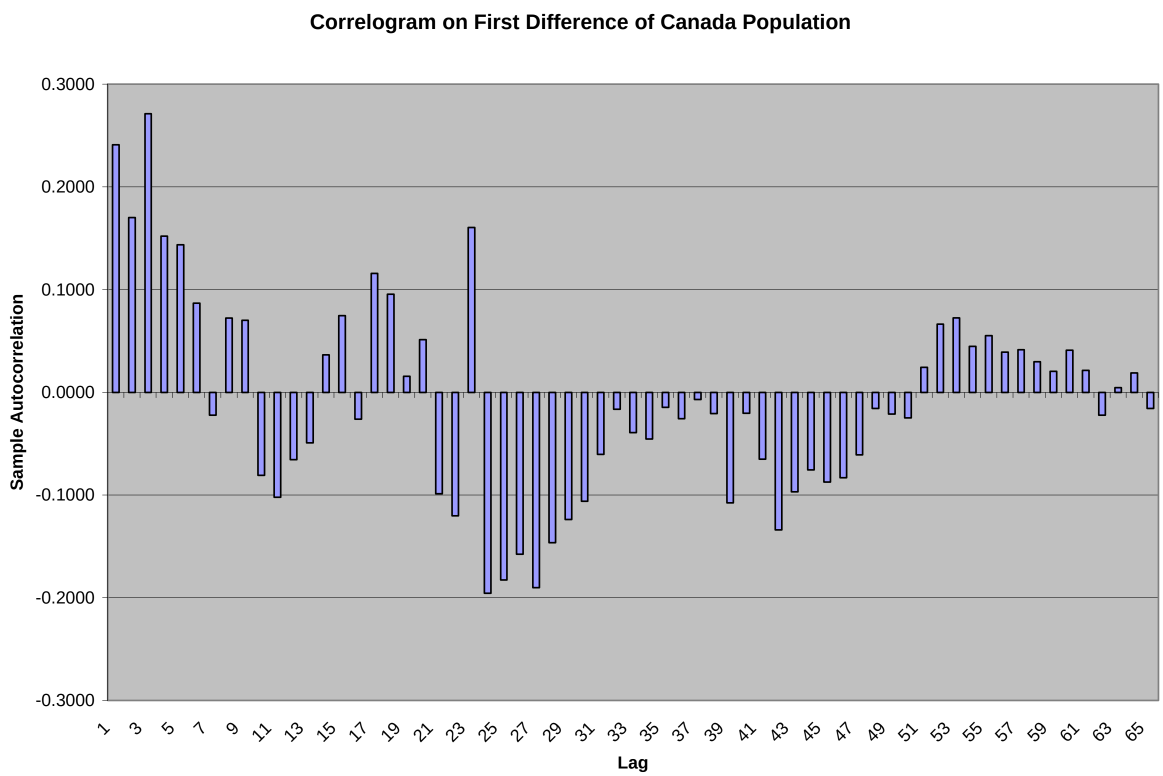
Course: Time Series

Session: Spring 2013 VEE Student Project

Before proceeding with the AR(p) model, I would like to examine the correlogram on the first differences to achieve stationary through differencing which may lead to a simpler model compared to an AR(p) model. The model on first difference of the population is:

$W_t = Y_t – Y_{t-1}$

Figure 3 shows sample autocorrelation on first difference of Canada population for different lags.



For a stationary series, we would expect to see a quick drop off to zero which is not the case shown here. This sample autocorrelation plot indicates non-random patterns and hence non-stationary.

Therefore, we will try to fit two AR(p) models to model the population data.

AR(1) process: $Y_t = \Phi_1 Y_{t-1} + \Theta_0 + e_t$.

AR(2) process: $Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Theta_0 + e_t$.

SAS function was used to fit the population data to AR(1) and AR(2) models and the linear regression results are as follows:

Name: Mai Ly

Course: Time Series

Session: Spring 2013 VEE Student Project

Figure 4 shows SAS Linear Regression results of the AR(1) model

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2756871852	2756871852	4E+05	<.0001
Error	65	404127	6217.34342		
Corrected Total	66	2757275979			

Root MSE	78.85013	**R-Square**	0.9999
Dependent Mean	24196	**Adj R-Sq**	0.9999
Coeff Var	0.32588		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	409.32812	36.99714	11.06	<.0001
Yt-1	1	0.99717	0.0015	665.9	<.0001

Figure 5 shows SAS Linear Regression results of the AR(2) model

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2618201140	1309100570	218829	<.0001
Error	63	376885	5982.29575		
Corrected Total	65	2618578024			

Root MSE	77.3453	**R-Square**	0.9999
Dependent Mean	24373	**Adj R-Sq**	0.9999
Coeff Var	0.31734		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	343.6131	62.89706	5.46	<.0001
Yt-1	1	1.18844	0.12263	9.69	<.0001
yt-2	1	-0.19118	0.12225	-1.56	0.1229

Note that the R^2 and adjusted R^2 statistics for both models are almost equal to 1. This indicates both models would be a good fit to the population data. The residual plots of the AR(1) and AR(2) models are shown in Figures 6 to 9 below.

Name: Mai Ly

Course: Time Series

Session: Spring 2013 VEE Student Project

Figure 6 shows residual plot from AR(1) model

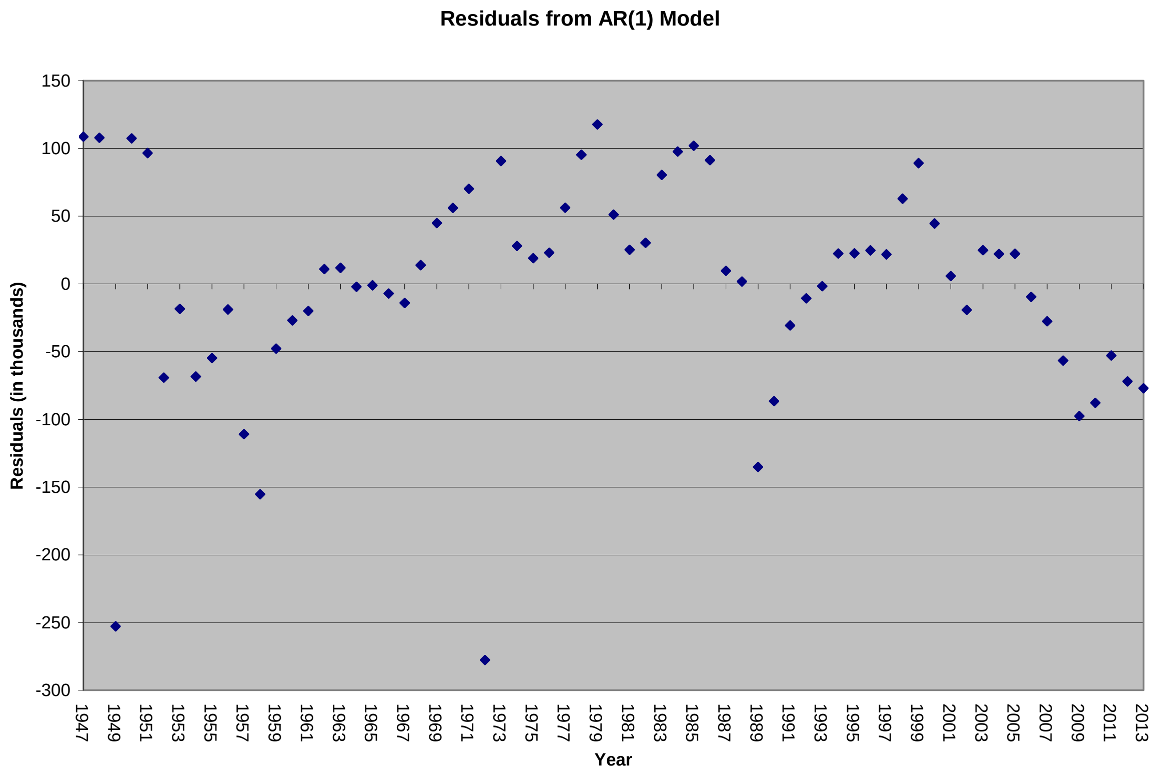
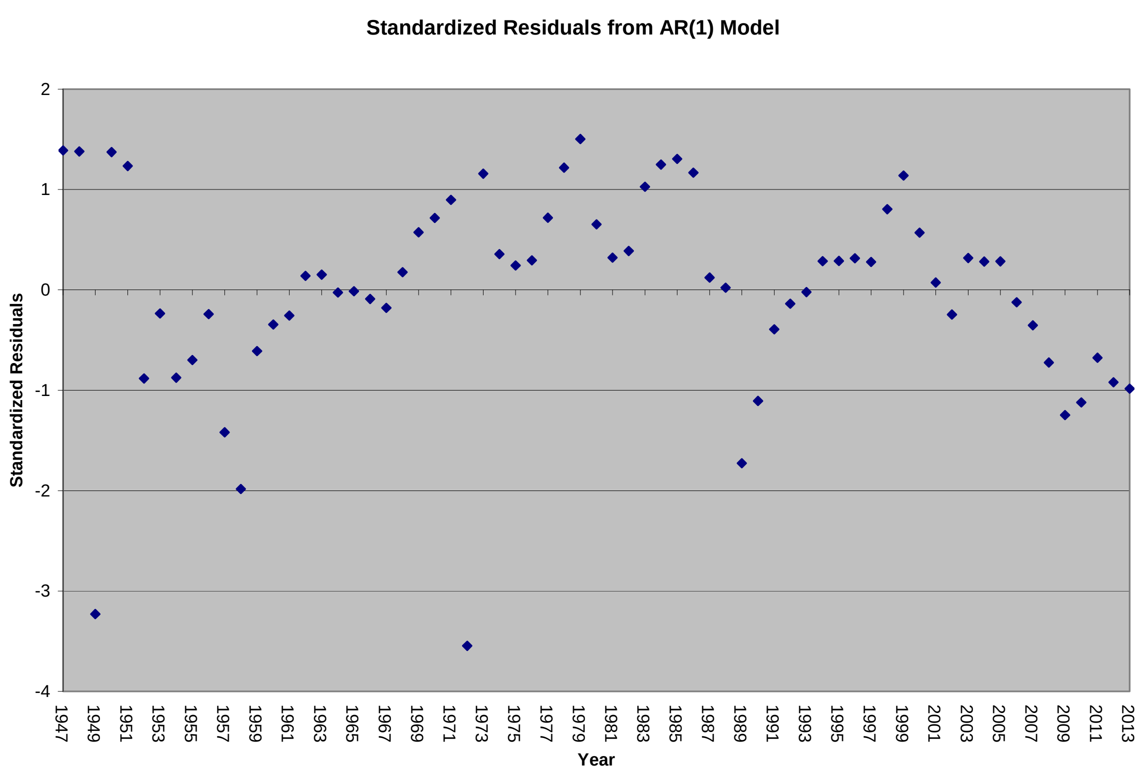


Figure 7 shows standardized residual plot from AR(1) model



Name: Mai Ly

Course: Time Series

Session: Spring 2013 VEE Student Project

Figure 8 shows residual plot from AR(2) model

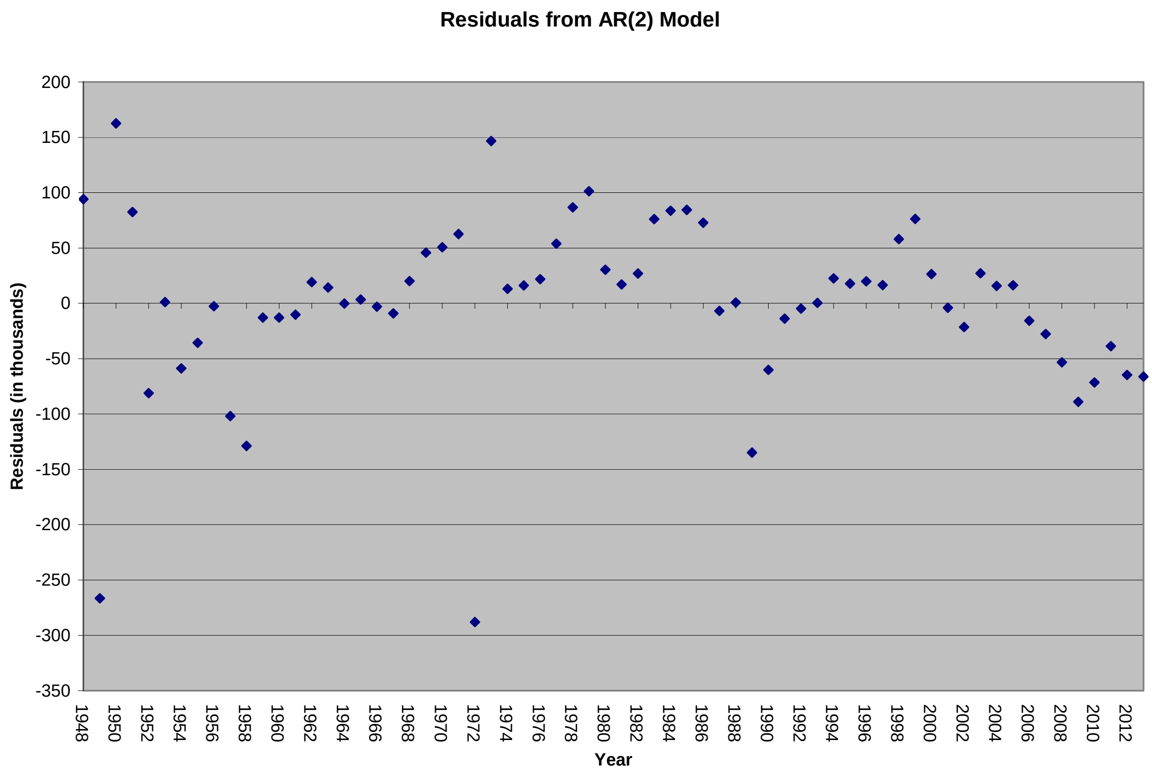
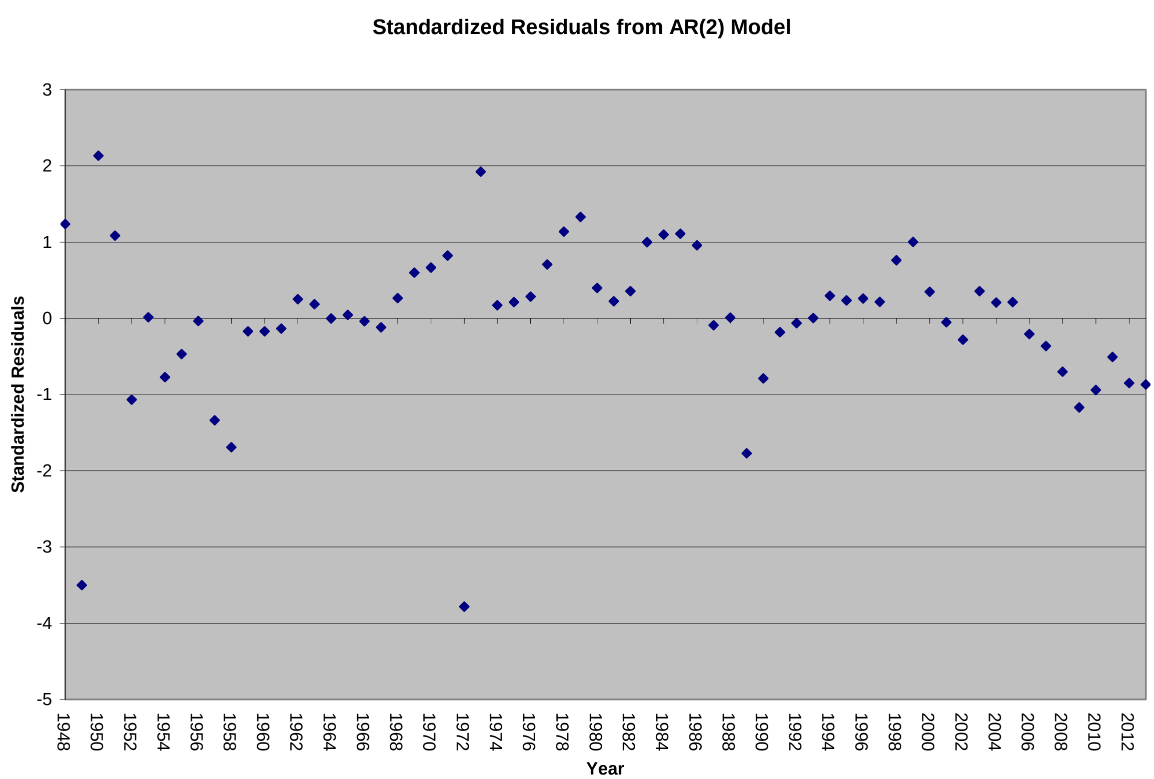


Figure 9 shows standardized residual plot from AR(2) model



We can see that the residuals are very small and almost all the standardized residuals are within 2 standard deviations. Most of the standardized residuals fall within 1 standard deviation with some random fluctuations as expected in the data. This indicates that both models fitted the population data very well.

Conclusion

Based on the analysis provided and the principle of parsimony, the AR(1) model would work best for modeling the population in Canada. The model can be used to forecast the future population using the near-adjacent population data as follows:

$Y_t = 0.99717\ Y_{t-1} + 409.32812 + e_t$

Knowing the 2013 Canada population is 35,142, the expected future populations are:

2014 forecast is 35,142 * 0.99717 + 409.32812 = 35,452

2015 forecast is 35,452 * 0.99717 + 409.32812 = 35,761

2016 forecast is 35,761 * 0.99717 + 409.32812 = 36,069