Modeling & Forecasting Canadian National Bankruptcy Rates

Crystal Xin Sun, Liying Li, Zhi Li, Joy Wenjing Qi

Introduction

Problem Statement

Understanding how national bankruptcy rates change over time is important for risk management and it has always been of interest to national banks, insurance companies, credit-lenders, and politicians etc. The goal of this report is to precisely and accurately forecast monthly bankruptcy rates for Canada. In this report, we will exploit the monthly statistics of Canadian bankruptcy rate, unemployment rate, population as well as house price index, for the period of January 1987 to December 2014, and construct a model for bankruptcy rates. We will use the constructed model to predict the monthly data for bankruptcy rates from January 2015 to December 2017.

Exploratory Data Analysis

For the given dataset, there are 4 variables in the 28-year time series. In this section, we will conduct exploratory data analysis on each of the variables in order to give a preliminary overview for the variables as well as their relationships.

Figure 1 shows four separate time series plots for the variables (Bankruptcy Rate, Unemployment Rate, Population, and House Price Index) from January 1987 to December 2014.

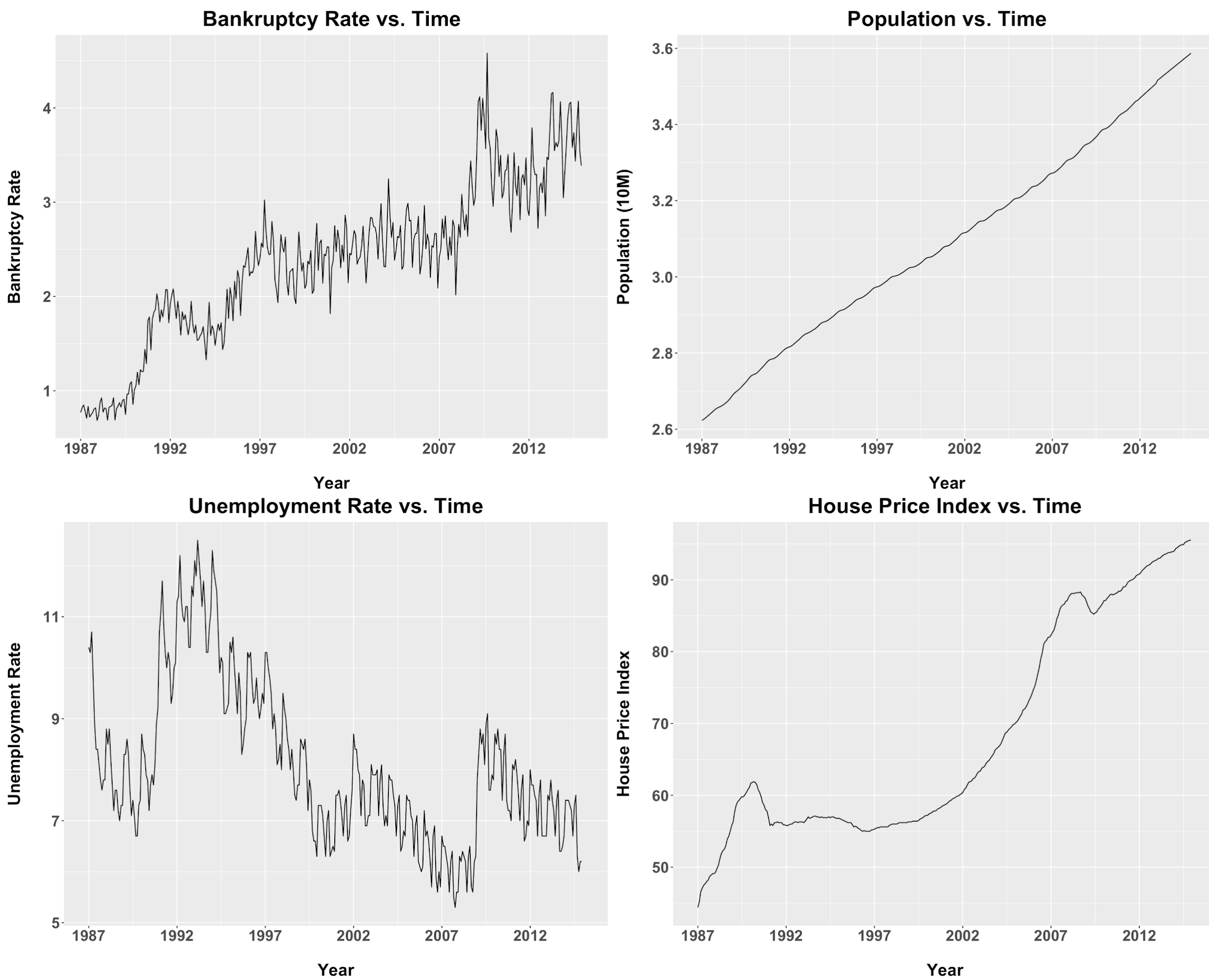

Fig 1. Time Series Plots for Four Variables

- **Bankruptcy Rate**: In general, the bankruptcy rate had a rising trend along the years, it hit certain exceptional peaks in the years of 1992, 1997 and 2009, followed by slight decreasing trends afterward. It also showed a clear seasonality over time.
- **Population**: The population increased in a linear manner over time without any unusual fluctuations.
- **Unemployment Rate**: A decreasing trend resided in the unemployment rate time series plot. However, the unemployment rate had some similarly exceptional jumps and increased around the years of 1991 to 1994, 2001 to 2002 and 2009, and it resumed its declining trend afterward.
- **House Price Rate**: The house price index also had an increasing trend overall. However, at two of the exceptional years of 1990 and 2009, the house price index had an apparent decline shown in the plot.

From the above plots we can see that between Bankruptcy Rate and Unemployment Rate, and between Population and House Price Index, there are some similar trends overall. The unusual fluctuations around the years of 1992 and 2009 happened on three of the variables (Bankruptcy

Rate, Unemployment Rate, House Price Index). This may indicate that there could be some correlations among the variables, meaning that Bankruptcy Rate could be influenced by other variables. In the following sections, while we introduce our modeling methods and constructed our models, we would also explore and reveal the correlations between the variables in order to assist our feature selections.

Methods

As discussed earlier, our goal is to forecast the national bankruptcy rates for Canada. We exploit modeling methods including univariate time series models, (models with a single variable, which is Bankruptcy Rate) and multivariate time series models (models including both target variable and the dependent variables, in this case, some or all of the four variables). In our model selection process, we adopt the comparisons by using values of Log-Likelihood, Akaike Information Criterion(AIC), as well as the root mean square error (RMSE) and identify an optimal model with the minimum RMSE.

Univariate Time Series Models

ARIMA/SARIMA (Box-Jenkins Approach)

SARIMA stands for Seasonal Autoregressive Integrated Moving Average. A SARIMA(p,d,q)(P, D, $Q)_m$ model difference a univariate time series to account for trend and seasonality. By construction, the SARIMA model directly considers trend, seasonality and most recent of the time series. Thus, if the data point is in recent, it has a great impact on making predictions.

Exponential Smoothing (Holt-Winters Approach)

Exponential smoothing is a rule of thumb technique for smoothing time series data using the exponential window function. With the double and triple exponential smoothing, we are able to further smooth the trend and seasonality of the time series.

There are two variations in the Holt-Winter seasonal method that differ in the nature of the seasonal component. One is the additive method, which is appropriate when the seasonal component has constant variance. The other one, the multiplicative method, is appropriate when the variance of the seasonal component is non-constant. Since our bankruptcy data exhibits both trend and seasonality and with non-constant variance, we will fit a multiplicative triple exponential smoothing model to our data.

Multivariate Time Series Models

ARIMAX/SARIMAX

A SARIMAX model is thought of a SARIMA model with explanatory variables. In a SARIMAX model setting, the time series is modeled based on historical data as well as other explanatory variables that may affect the target variable in a unidirectional manner. SARIMAX can make use of the modeling and predicting the power of SARIMA and adopt additional information from the known explanatory variables in order to enhance the model performance.

VAR/VARX

VAR (Vector Autoregression) model is a generalization of the univariate autoregressive model for forecasting a vector of time series. In the VAR framework, the model also takes external variables into considerations. However, it is different from SARIMAX model where it assumes that all the variables are treated symmetrically with no distinction between explanatory and response variables. Unlike SARIMAX model, the VAR model doesn't require data distributional assumptions and predicts all variables simultaneously in a multivariate time series setting. If there exist both unidirectional and bidirectional explanatory variables, a VARX model is commonly used.

Modeling & Model Selection

Feature Selection

For univariate time series models, we are only concerned with the target variable Bankruptcy Rate. As shown below on the left plot of Figure 2, the amplitude of the seasonal fluctuations in the time series data tends to enlarge over the years. In order to stabilize the variance and reduce the non-constant fluctuations of the original time series, we adopt a transformation method and perform on the target variable. On the right plot of Figure 2 is the transformed Bankruptcy Rate, which has approximately constant seasonal fluctuations.

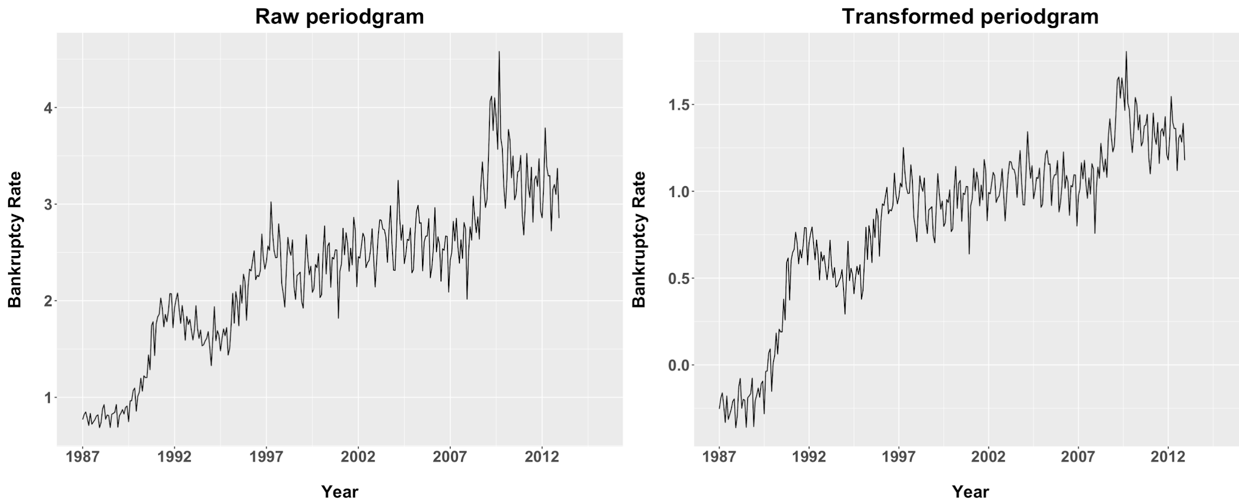

Fig 2. Raw Bankruptcy Rate vs. Transformed Bankruptcy Rate

As we uncovered in the Exploratory Data Analysis section, there might be correlations between the variables. In order to further understand their relationships and correlations, we performed pairwise relationship plots between the variables referenced by the Table 1 below and came up with a correlations matrix as following.

Table 1. Pairwise relationship plots table

Unemployment Rate vs. Bankruptcy Rate	Population vs. House Price Index
Unemployment Rate vs. Population	House Price Index vs. Bankruptcy Rate
Population vs. Bankruptcy Rate	House Price Index vs. Unemployment Rate

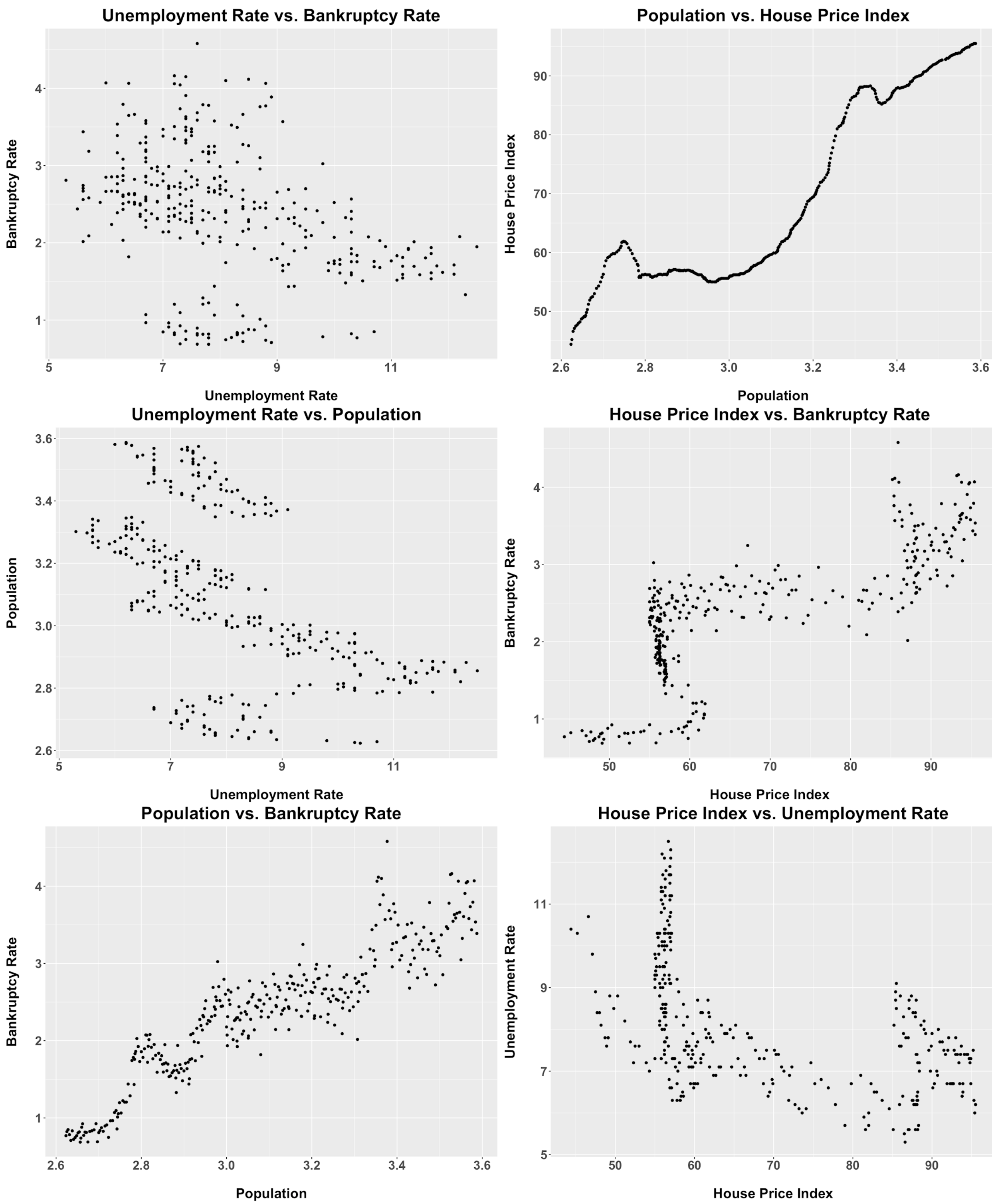

Fig 3. Variables Pairwise Relationship Plots

Fig 4. Variables correlation matrix plot

As shown in both Figure 3 and Figure 4, below are some insights about the correlation relationships:

- Population and House Price Index were almost equivalent on the plot (the upper right plot), indicating that the two of them are highly correlated.
- The patterns between Population and Bankruptcy Rate were similar on the plot (the bottom left plot), indicating there are some correlations between them.
- Similarly for House Price Index and Bankruptcy Rate with similar plot patterns (the middle right plot), indicating there were some correlations between them.
- The patterns on Unemployment Rate vs. Population plot (the middle left plot) seemed less overlapped and more spread out, which means there might be slight correlations between them.
- Similarly, for House Price Index vs Unemployment Rate plot (the bottom right plot), there might be slight correlations between them.

Due to the correlations between the variables, we chose to adopt both univariate and multivariate time series models and take all the four variables into considerations for possible model fitting. Although we are able to identify the existence of correlations between the variables, the influences of the variables on each other could still be either unidirectional or bidirectional. Hence, besides building SARIMA and Holt-Winter models on the target variable (Bankruptcy Rate), we took exhaustive combinations of all four variables and fit the possible options in SARIMAX, VAR, and

VARX models. In the next section, we will further elaborate on our evaluation and model selection processes.

<table>
<tr><td colspan="3">Univariate Models (Target variable)</td></tr>
<tr><td>SARIMA</td><td colspan="2">Holt-Winter</td></tr>
<tr><td colspan="3">Multivariate Models (Combination of variables)</td></tr>
<tr><td>SARIMAX</td><td>VAR</td><td>VARX</td></tr>
</table>

Evaluation Metrics

Since our goal is to have as accurate and precise forecasting results as possible, we choose to use the minimum RMSE as our model selection metrics. The reason is that RMSE is a good measurement of errors that can serve as functions to minimize. In order to test our models, we split the original data (from January 1987 to December 2014) into two parts, one set is the training part (from January 1987 to December 2012) to train our models, and the other set is the validation part (from January 2013 to December 2014) to test the predictive accuracy (RMSE) of the models.

Model Selection & Justification

For each of the five models mentioned above, we iterate through an exhaustive set of possible parameters combinations for univariate models, and through an exhaustive set of possible variables and parameters combinations for multivariate models. We chose an optimal model when it had a minimum RMSE value under each of the five models. Table 2 are the comparisons between the five outstanding optimal models from each of the five iterations selected by its respective minimum RMSE value. We also came up with metrics of Log-likelihood and AIC as a reference besides.

Table 2. Optimal Model for Each Modeling Approach

Model	RMSE	Log-likelihood	AIC
SARIMA$(2,1,1)\times(2,1,4)_{12}$	0.2680	368.1910	-716.3811
Holt-Winters (α=0.6,β=0.2,γ=0.2)	0.2817		
SARIMAX $(1,1,2)\times(3,1,3)_{12}$ Exogenous ➔ Housing Price Index ➔ Population ➔ Unemployment Rate	0.1896 (chosen)	376.2786	-726.5573
VAR(p=2) Endogenous ➔ Housing Price Index ➔ Population ➔ Unemployment Rate	0.2657		
VARX(p=2) Endogenous ➔ Housing Price Index ➔ UNemployment Rate Exogenous ➔ Population	0.2573		

The Optimal Model: SARIMAX$(1,1,2)\times(3,1,3)_{12}$

Based on the RMSE values from Table 2, the SARIMAX$(1,1,2)\times(3,1,3)_{12}$ model has the smallest RMSE (0.1896) across the different modeling approaches, which means the predicted Canadian bankruptcy rate from January 2013 to December 2014 from this model is closest to the real Canadian bankruptcy rate from January 2013 to December 2014. Thus we choose this model as our final optimal model.

The SARIMAX$(1,1,2)\times(3,1,3)_{12}$ model accounts for the trend, seasonality, and the influence of explanatory variables: population, unemployment rate, and house price index. Our optimal model satisfies all required data distributional assumptions, aka normality, zero mean, homogeneity and zero correlation.

Forecasting

The bankruptcy rate predicted by our optimal model can be seen in Table 3 and Figure 4. The forecasting plot also shows the 95% prediction intervals of our prediction values. The interval is the range of values that we are 95% confident that the true forecasted bankruptcy rate lies in.

In order to forecast with a SARIMAX model, the future values of the influential variables (housing price index, population, and unemployment rate) are needed. In other words, the predictions of bankruptcy rate are calculated by using the observed housing price index, population, and unemployment rate from January 2015 to December 2017.

Table 3. Predicted Monthly Canadian Bankruptcy Rate for the year 2015 to 2017

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015	3.356767	3.756709	4.188115	4.251213	3.786493	3.995782	3.659851	3.466036	3.998749	3.715055	3.725223	3.348612
2016	3.124425	3.589487	4.198619	3.901176	3.561479	3.870063	3.251380	3.406103	3.703062	3.427438	3.663245	3.113963
2017	2.999139	3.391720	4.006378	3.537696	3.488355	3.610971	2.994123	3.420986	3.447937	3.409625	3.663413	2.991439

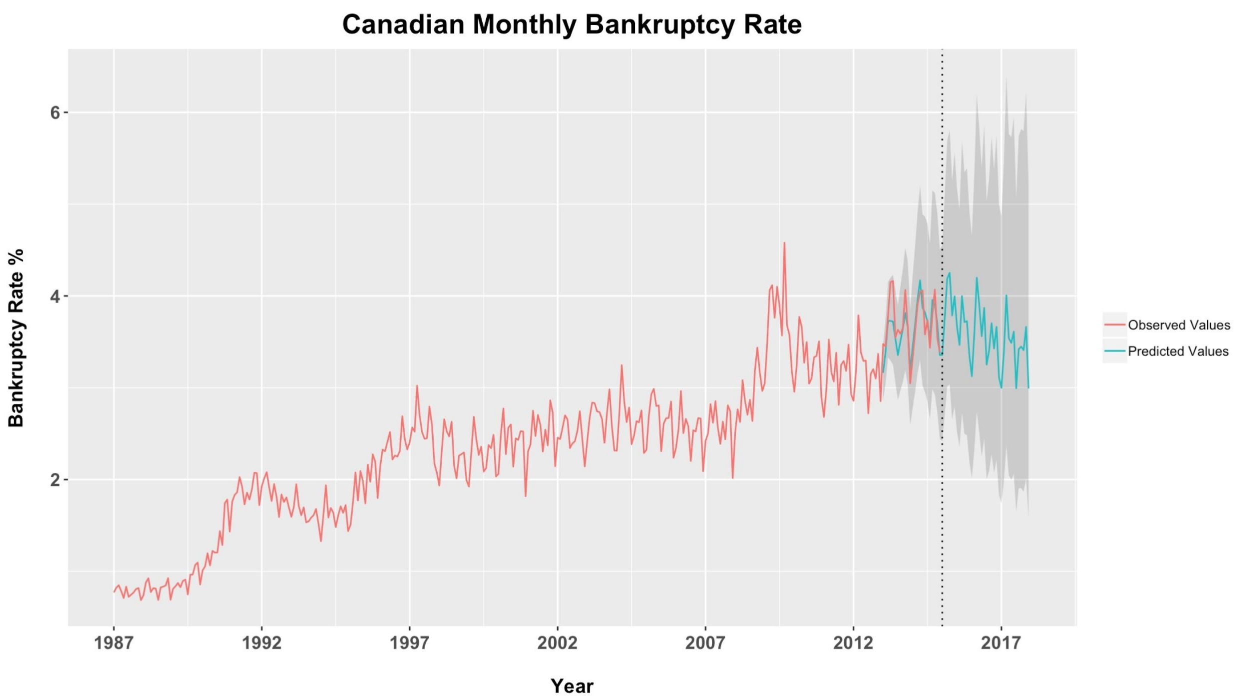

Fig 5. Canadian Monthly Bankruptcy Rate

Our forecasting plot shows that, despite some seasonal fluctuations, there will be a general decreasing trend of the bankruptcy rate for the year 2015 to 2017.

Conclusion

In summary, after we have taken into considerations for the other explanatory variables i.e., Housing Price Index, Population, and Unemployment Rate, we have constructed a model SARIMAX$(1, 1, 2)\times(3, 1, 3)_{12}$ that forecasts the Canadian Bankruptcy Rates during January 2015 to December 2017 to be declining.