

Introduction:

This project aims to build a predictive model to forecast monthly bankruptcy rates in Canada for the year of 2011 and 2012 with highest possible accuracy, given monthly data from January 1987 to December 2010 on bankruptcy rate, unemployment rate, population, and housing price index in Canada. then selected the optimal model and used it for forecasting.

In this report, we will first explore the data, and discuss the available modeling approaches, including SARIMA, Holt-Winters, SARIMAX and VAR. We will also explain the approach to select our best predictive model and present the forecasting results from our optimal model.

Data Overview:

The available dataset consists of monthly data from January 1987 to December 2010 on the 4 variables: bankruptcy rate, unemployment rate, population, and house price index. Table 1 summarizes the data and Figure 1 and 2 are the plots of each of the variable over time.

Table 1: Summary Table for Training set, 1987 - 2010

Statistic	N	Mean	St. Dev.	Min	Max
Unemployment Rate	288	8.236	1.528	5.900	12.100
Population	288	30,256,218	2,199,282	26,232,423	34,272,214
Bankruptcy Rate	288	0.022	0.008	0.007	0.046
House Price Index	288	75.218	14.124	52.200	104.000

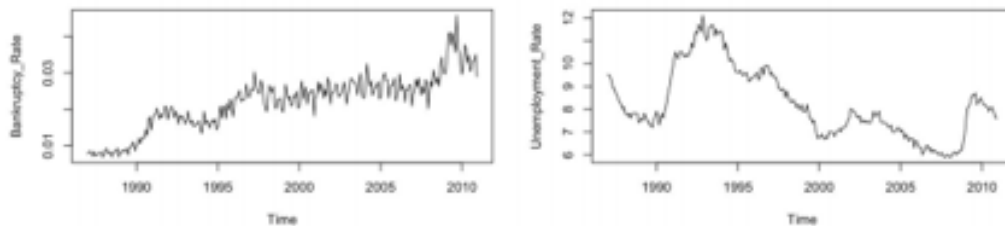


Figure 1: Bankruptcy Rate & Unemployment Rate trends over time (1987 - 2010)

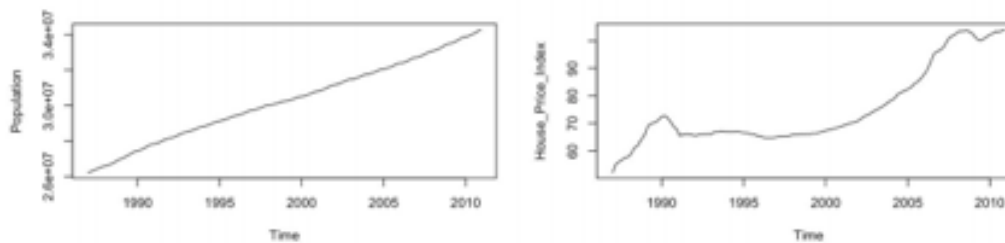


Figure 2: Population & House Price Index trends over time (1987 - 2010)

To explore the relationship between the 4 variables, we constructed a correlation matrix.

	Unemployment Rate	Population	Bankruptcy Rate	House Price Index
Unemployment Rate	1.000	-0.543	-0.317	-0.543
Population	-0.543	1.000	0.898	0.860
Bankruptcy Rate	-0.317	0.898	1.000	0.690
House Price Index	-0.543	0.860	0.690	1.000

We can see that bankruptcy rate is highly correlated with population and is somewhat correlated with house price index. Bankruptcy rate has a smaller, and negative, correlation with unemployment rate. Variables with medium to high correlation are of interest because they can possibly be used as covariates to help accurately predict bankruptcy rates.

Method

In order to find the best model to predict 2011-2012 Canadian bankruptcy rates, we split our training data into training and validation set. The training set consists of observations from January 1987 to December 2008, and is used for constructing the models. The last 2-year of data (48 observations) is held out to determine the predictive accuracy of the model.

Since there is some seasonality and trend, we explored the following methods for modeling the bankruptcy rate.

- SARIMA: this approach works by removing the trend and seasonality through differencing the data and modeling the transformed data with ARIMA.
- Holt-Winters model: this approach works by assigning exponentially decreasing weights to older observations. Considering the trend and seasonality pattern observed in the data, we used triple exponential smoothing for modeling.
- SARIMAX: this approach works by considering external variables in addition to SARIMA, which influence the response series.
- VAR: this approach works by treating the other influential variables as endogenous variable - they influence bankruptcy rate and bankruptcy rate influences them.

For each method, the potential models are mainly compared on the basis of log likelihood, AIC, sigma squared, and an optimal model was selected based on RMSPE on the validation set.