Time Series Final Project

Forecasting Canadian Bankruptcy Rate

April Wang, Zainab Danish, Yixin Zhang

I. Problem Statement

Accurately forecasting national bankruptcy rates has always been a primary interest to national banks, insurance companies, credit-lenders, politicians, etc. With the development of global economy, the bankruptcy rate in Canada saw a steady rise. From "Bankruptcy Canada"[1], one in six adults in Canada face bankruptcy. The most important reason behind this problem is they borrow money from bank but cannot pay off the debt. Since loans are very important earning asset for banks, an accurate forecast of the national bankruptcy rate is desired. The goal of this project is to develop a model that accurately predicts monthly bankruptcy rate in Canada from January 2011 to December 2012, based on historical data from January 1987 to December 2010.

II. Data Description

Data used in the report:

- Bankruptcy rate in Canada from January 1987 to December 2010.
- Unemployment rate in Canada from January 1987 to December 2010.
- Housing price index in Canada from January 1987 to December 2010.

[1] Hoyes, J. (2012, March 12). Why 1 in 6 Canadians will go bankrupt. Retrieved December 10, 2016, from http://bankruptcy-canada.com/bankruptcy-blog/1-in-6-canadians-will-go-bankrupt/

III. Forecast Summary

Four different time series models were employed in order to forecast the Canadian national bankruptcy rate. As a brief summary, we chose Vector Autoregressive (VAR) model as our "best" model. Forecast Results are shown in Table 1 and Figure 1:

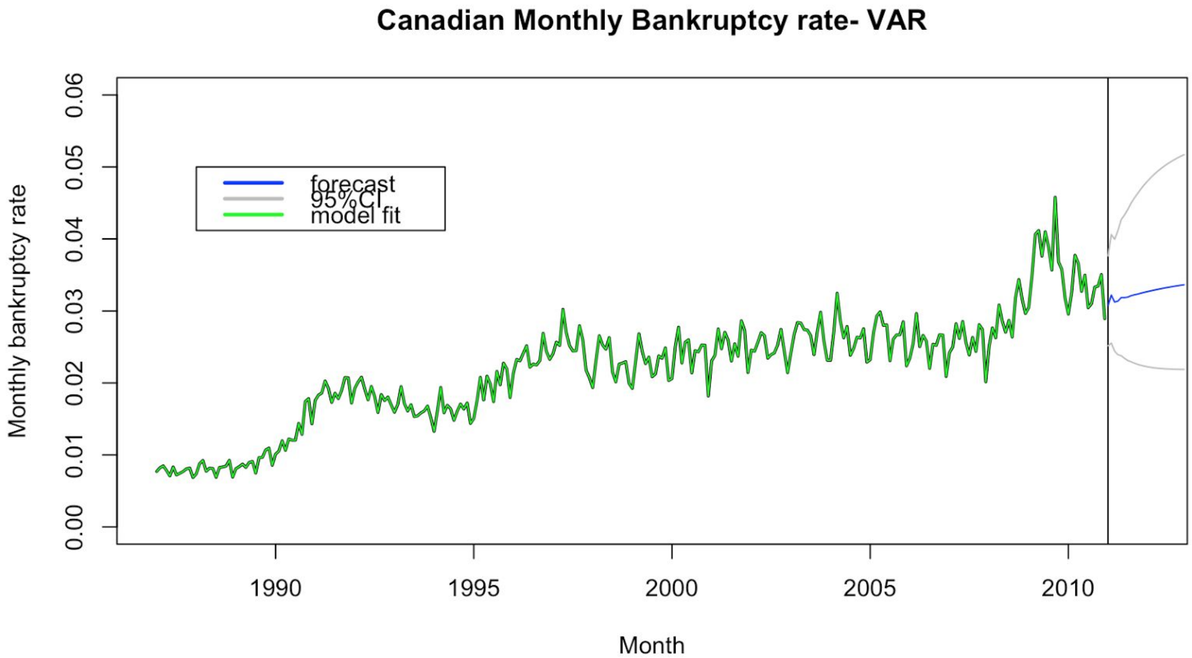
Table1: Forecast Summary

Forecast Time Period: January 2011 - December 2012

Date	Forecast	[2]Lower Bound	[2]Upper Bound
Jan 2011	3.08%	2.52%	3.76%
Feb 2011	3.22%	2.55%	4.06%
Mar 2011	3.12%	2.44%	4.00%
Apr 2011	3.14%	2.39%	4.12%
May 2011	3.19%	2.38%	4.27%
Jun 2011	3.19%	2.34%	4.33%
Jul 2011	3.19%	2.31%	4.41%
Aug 2011	3.21%	2.29%	4.50%
Sep 2011	3.22%	2.27%	4.57%
Oct 2011	3.23%	2.26%	4.63%
Nov 2011	3.25%	2.24%	4.70%
Dec 2011	3.26%	2.23%	4.75%
Jan 2012	3.27%	2.22%	4.81%
[2]Feb 2012	3.28%	2.22%	4.85%
Mar 2012	3.29%	2.21%	4.90%
Apr 2012	3.30%	2.20%	4.94%
May 2012	3.31%	2.20%	4.98%
Jun 2012	3.32%	2.20%	5.01%
Jul 2012	3.33%	2.19%	5.04%
Aug 2012	3.33%	2.19%	5.07%
Sep 2012	3.34%	2.19%	5.10%
Oct 2012	3.35%	2.19%	5.13%
Nov 2012	3.36%	2.19%	5.15%
Dec 2012	3.36%	2.19%	5.17%

[2]Lower and Upper Bound are based on 95% confidence interval

Figure 1: Forecast Time Period: January 2011 - December 2012



As is forecasted in Table 1 and Figure 1, the expected insolvency rate will continue to slowly climb up and reach 3.36% by the end of year 2012. There is 95% confidence that the insolvency rate will reach between 2.19% to 5.17% by the end of year 2012.

IV. Method Applied - Vector Autoregressive Model (VAR)

A. Model Description

For our final model we decided to treat unemployment rate, bankruptcy rate and housing prices as endogenous variables and ran a Vector Autoregression. We believe that bankruptcy rate, unemployment rate and housing price are highly correlated with each other and influence each other symmetrically. Therefore, the model performance can be improved by adding unemployment rate and housing price as additional information, in order for the model to best forecast the 2011 and 2012 bankruptcy rate.

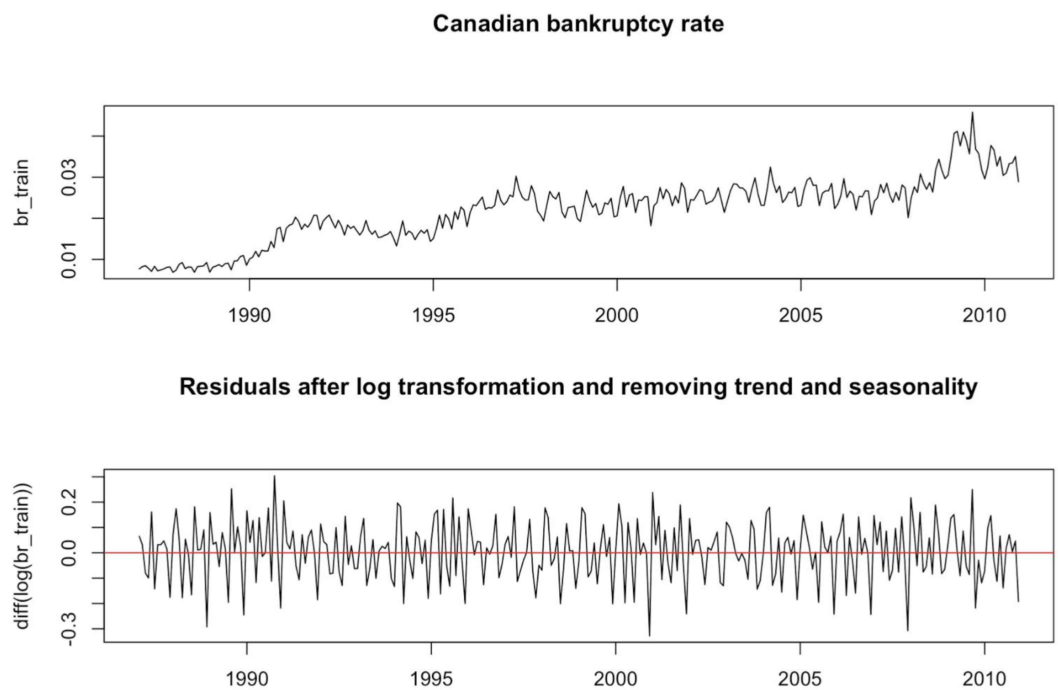
For qualifying our models in modeling part, so we are able to calculate root mean square errors (RMSE) as part of the model selection process, we took the split of 80, 20 of the overall set to create our training set and test set. More specifically, we used data from

January 1987 to December 2006 for modeling and compared the prediction results with real data from January 2007 to December 2010.

B. Model Preprocessing and Model Selection

The plot below of training data showed the Canadian bankruptcy rate between 1987 and 2010. As we can see, the bankruptcy rate increased gradually with some fluctuations from 1987 to 2010 with moderate volatilities. In order to make the bankruptcy rate stationary, one of the assumptions we need to make is constant variance. Therefore, with current data showing slight volatilities, we performed log-transformation of the data, so that the residuals after log transformation meet the underlying assumptions of model fitting.

Figure 2. Bankruptcy Rate in Canada from 1987 to 2010



Next, in order to select the order for VAR we used the VARselect function in R which gave us the following output:

AIC(n)	HQ(n)	SC(n)	FPE(n)
6	5	3	6

We chose order p equal to three because we want to keep the lag of the model as small as possible to maintain the simplicity of the model and prevent the model from overfitting. The root mean square error (RMSE) using the hold out 20% sample came to 0.0045.

C. Model Assumptions Verification

There are four underlying assumptions about the model residuals that need to be met. These assumptions include zero mean, constant variance, uncorrelatedness and normality distribution. All these four assumptions need to be met in order for this model to be valid and appropriate. In other words, after model fitting, the residuals of the model should preserve all the statistical properties after the time shift, in order for us to use historical data to forecast into future. Therefore, diagnostics for residuals have been conducted to verify the underlying assumptions of the model. Below, scatter plot of the residuals, autocorrelation function ACF plot and quantile - quantile(QQ) plot have been examined in details. Residual Diagnostic Plots are shown in Appendix II.

a) **Zero Mean**

From the residual versus time plot in Appendix II, the mean appears to be zero. Hence, this assumption is met.

b) **Constant variance**

According to the residual plot, constant variance assumption is reasonable as we do not see any change in variance with the passage of time.

c) **Uncorrelatedness**

From the ACF plot and partial autocorrelation function plot (PACF), we can see that there is no significant spikes that is outside the confidence interval after the lag zero, indicating that all spikes are relatively negligible. Therefore, the assumption of no serial correlation between residuals is largely maintained.

d) **Normality distribution**

Finally, normality distribution is verified through QQ plot, indicating the residuals of the model approximately normally distributed and therefore, meets the normality assumption.

V. Other Model Applied and Performance Comparison

Besides vector autoregressive model, we also performed other models including ARIMA, ARIMAX and Holt Winters. Model Performance of each model has been performed and is shown in Table2. Similar residual diagnostics have also been performed to ensure the validity of each of the following models. More technical details and residual diagnostics for each model are available upon request.

Table 2. Goodness of Fit Metrics for VAR, ARIMA, ARIMAX, Holt Winters

Models	σ^2	Part Log Likelihood	RMSE	Fitting Method
VAR	n/a	n/a	0.0045	Bankruptcy rate, housing price index and unemployment rate are treated as equally important in the model fitting
ARIMA	0.0035	385.75	0.0075	Least Square
ARIMAX	0.0029	411.8	0.0036	Least Square
Holt Winters	0.0028	434.35	0.0037	3rd exponential smoothing

A. ARIMA

We performed differencing techniques and log transformation to remove any trend or seasonal component of the data to make it stationary.

The “optimal” ARIMA model was $ARIMA(3,\ 1,\ 1) \times (6,\ 0,\ 0)_{12}$.

B. ARIMAX

Appendix I shows the plots of bankruptcy rate, unemployment rate and housing price index. Overall, house pricing index climbed slowly over decades, which had similar trends as the bankruptcy rate. The unemployment rate decreased from 1987 to 1990, and then recovered but dropped again after 1993 with lots of volatilities. These rates might be externally influencing factors to the bankruptcy rate; consequently, we want to know if some of these factors could help us to better predict the bankruptcy rate. By adding unemployment rate and housing price index, we fit an ARIMAX model. The "optimal" model was $ARIMAX(2, 1, 1) \times (5, 1, 0)_{12}$, with explanatory variables unemployment rate and housing price index.

C. Holt Winters

Holt Winters uses exponential smoothing techniques to forecast into the future using the historical data. We applied Triple Exponential Smoothing to model the Canadian bankruptcy rate with trend and seasonality. To take care of increasing variance we used a multiplicative rather than an additive model. Moreover, we fine tuned our parameters in order to obtain a small training error. Finally, our Holt Winters model has the following parameters:

- alpha = 0.2
- beta = 0.6
- gamma = 0.1

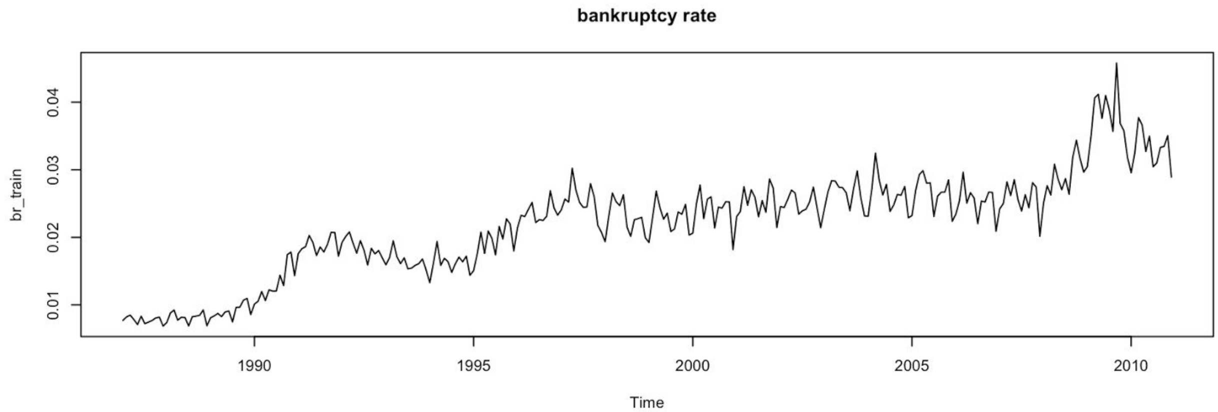
As we can see from table 3 and forecast plot of Holt Winters(shown in Appendix III), the forecast for the 20% hold out set appears to follow the actual bankruptcy rates very closely. This is verified by the fact that the RMSE of this model is the lowest of all. However, this model can potentially suffer from high variance problem as we can see from the forecast plot. In other words, due to the fact that the confidence interval is too wide, it is hard to make precise predictions based on the model result.

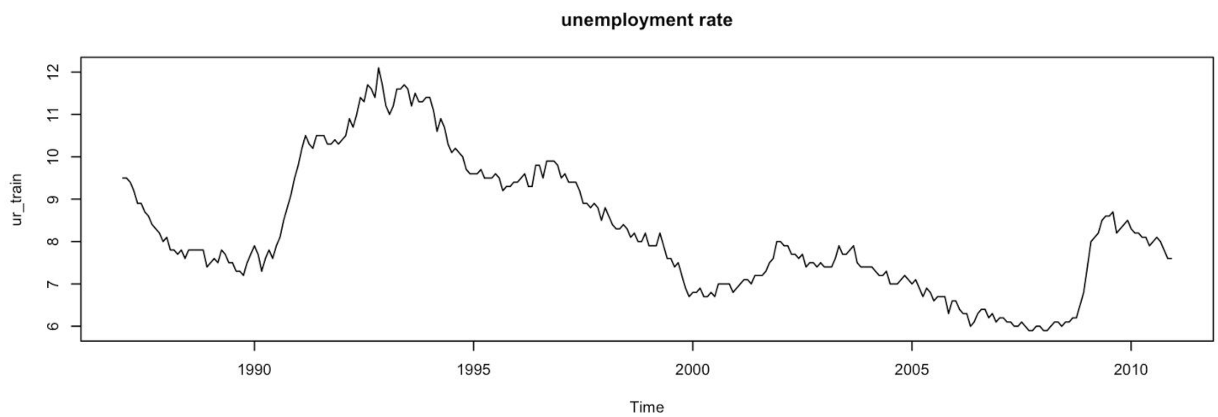
VI. Conclusion

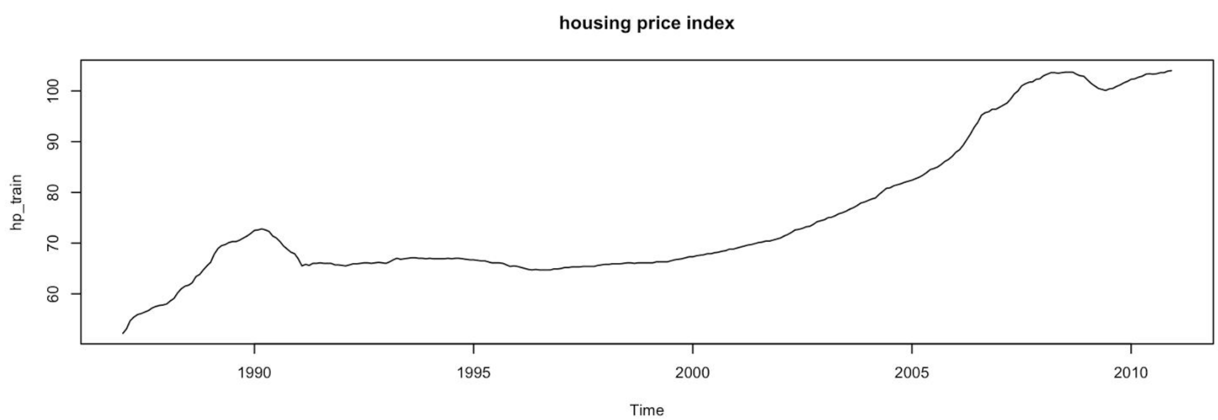
To sum up, all underlying assumptions of the model have been verified and overall, the autoregressive model has a good model performance. In addition, this model has been compared with other models (see Section V) and has shown a solid performance as it produces a reasonably small root mean squared errors (RMSE) between the training set and hold out set. Therefore, we are confident to use Vector autoregressive model as our primary model in forecasting.

Appendix I

Plots of Bankruptcy Rate, Unemployment Rate, Population, Housing Price Index



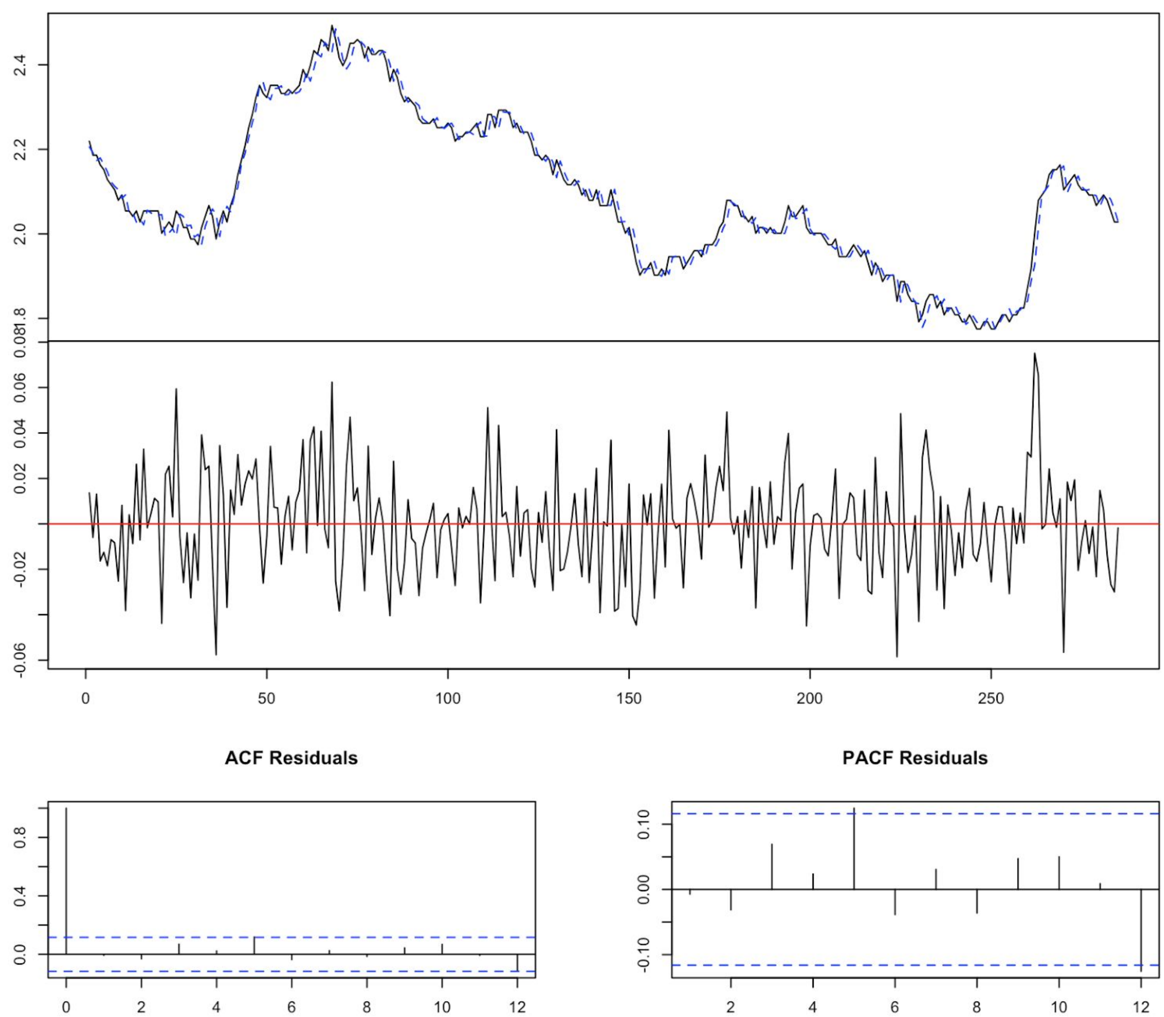




Appendix II

Residual Diagnostics Plots for VAR model

Diagram of fit and residuals for log.ur_train.



Appendix III

Forecast Plots for ARIMA, ARIMAX, Holt Winters