MSAN 604 Group Project

Arda Aysu, Erin Chinn, Philip Rotella, Zefeng Zhang

12/12/2016

Goal and Data

The goal of this report is to forecast Canadian monthly bankruptcy rates for the years 2011 and 2012 using a time series model. The primary data for constructing this model consisted of historic Canadian monthly bankruptcy rates from January 1987 through December 2010. Three sources of secondary Canadian national data was available from January 1987 through the end of the forecasting period (i.e., December 31, 2012): unemployment rate, population size, and housing price index (HPI). HPI is an economic indicator that measures average price changes in repeat sales or refinancings on the same single-family houses.

The optimal model we propose is the SARIMAX(3,1,5)x(1,0,1)[12] model which includes housing price index as an external variable. Although the details of this model are outside the scope of this report, a brief description of this modeling approach along with other modeling options is included below. The predictions for Canadian monthly bankruptcy rate for 2011 to 2012 made by our optimal model are summarized and plotted at the end of this report.

Available Methods

There are numerous modeling approaches available to forecast bankruptcy rates. This report will summarize some of the most common approaches, highlight the one used for our final model, and discuss what made it the optimal choice. Each model was fit by altering various parameters. These parameters change how the model fits the data and in turn what predictions it makes. They account for the following:
Overall trend - does bankruptcy rate increase or decrease with time?
Seasonal trend - does bankruptcy rate have similar fluctuations within each year?

Exponential smoothing:
The smoothing method uses a set of equations to filter or smooth a model to the data. Specifically this method calculates an average of the observed bankruptcy rates based on the current time in addition to a previous prediction of the rates at an earlier time.

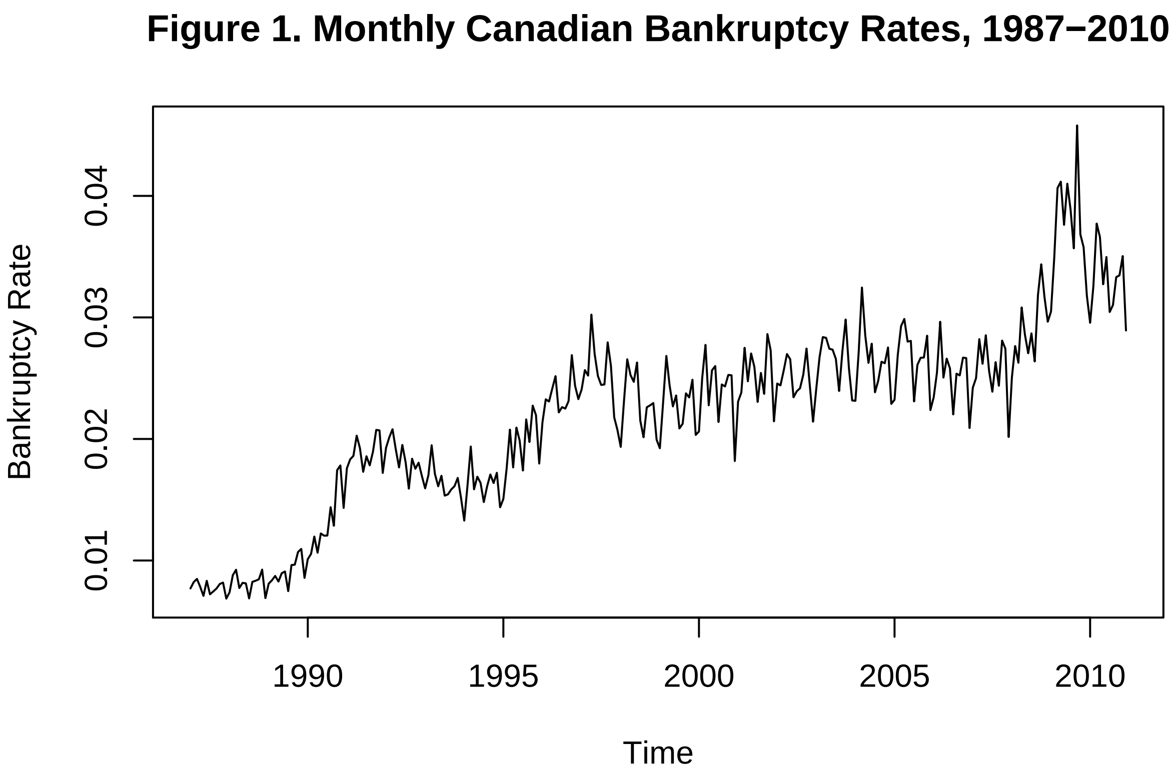
SARIMA:
SARIMA stands for seasonal autoregressive integrated moving average. In essence the model can be broken into two main parts. The first part of this model predicts bankruptcy rate using previous bankruptcy rates. The second part models deviations from the true bankruptcy rate using the deviations from previous bankruptcy rates.

SARIMAX:
A similar modeling approach to SARIMA is the SARIMAX method. The difference here is that SARIMAX allows us to consider external variables that could also affect the prediction variable, bankruptcy rate. In this project these external variables are: unemployment rate, population, and housing price index.

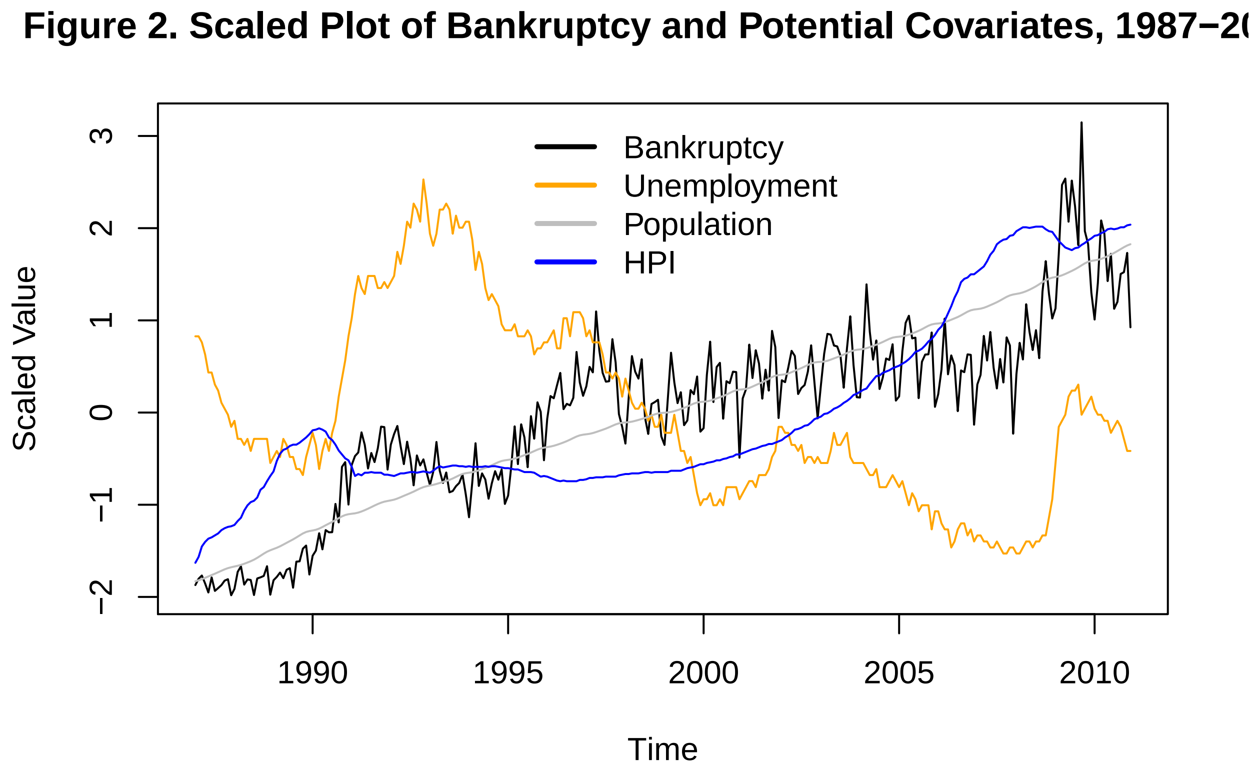
VAR:
The last approach considers external variables as well. However unlike SARIMAX, it treats all variables symmetrically. While SARIMAX only considers the effect the outside variables have on bankruptcy rate, the VAR approach considers all variables as possibly affecting each other. This method simultaneously accounts for these interdependencies.

Model Construction

To select an appropriate modeling approach, we first visualized how bankruptcy data has changed over time from 1987 through 2010. As shown in **Figure 1**, the bankruptcy rates have generally increased since 1987, with marked within-year peaks and valleys and prolonged periods of both growth and decline. The trend over time indicates the need for a model that can handle “non-stationary” data, and cyclical trends indicate that the model should account for seasonality.



Next, we looked at the time trends of our external variables (unemployment, population, and HPI) to assess whether a univariate or multivariate approach would be best. **Figure 2** plots each of these time series curves along with bankruptcy rates, all scaled to fit within the same axis. We judged HPI to be most related to bankruptcy so included it in our model, treating it as a one-directional, non-synergistic relationship. Therefore, we chose SARIMAX as our modeling approach.



In order to tune the model inputs for optimal forecasting accuracy, we fit various models on a subset of years, 1987–2005, and validated each model's performance in predicting bankruptcy rates for 2006–2010. Model fit was determined by calculating the average difference between the predicted bankruptcy rates for our validation time frame and the observed bankruptcy rates for the same period. A smaller difference meant a more accurate prediction and a better fitting model.

Once we found the model parameters that yielded the best fit for the data from January 2006 to December 2010, we re-fit the model using all of our data from January 1987 to December 2010. This was the model used to predict values for bankruptcy rate from January 2011 to December 2012.

While the model was optimized to give us the smallest difference, it was also subject to certain modeling constraints. These constraints involved analysis of the residuals, or the difference between our predicted values and the actual values, from January 2006 to December 2010. They are the following: Normality - does the distribution of residuals follow a bell curve? Zero-mean - is the average value of the residuals zero? Constant variance - do the residuals vary with constant magnitude across all years and months? Zero-correlation - are residuals prone to be similar to the residual from the previous year or month?

Final Model

The final model we chose is a SARIMAX(3,1,5)x(1,0,1)[12] model with housing price index as an external variable. It predicted values of bankruptcy rate closest to the actual values for the validation time frame, January 2006 to December 2010. On average, our point predictions differed from the actual bankruptcy rates by approximately 0.0035, which was just slightly more than the average magnitude of month-to-month fluctuations during that time (0.0030). Although this model uses outside variables for predicting bankruptcy rates, the model was kept conservative. Housing price index was the only external variable included. This model passes three of our four modeling constraints, failing to achieve constant variance. This is likely due to the spike around 2009 which is difficult to properly account for given the limitations of our modeling techniques. As such, we are not overly concerned with not passing that test.

The bankruptcy rates predicted by the model can be seen in **Figure 3** and **Table 1**, below. Included in the plot are 95% prediction intervals. These intervals are the range of values between which we are confident that the true forecasted bankruptcy rate lies.

In order to forecast with a SARIMAX model, future values of the external variable, housing price index, are needed. The prediction interval shown in the plot does not account for the uncertainty of predicting future housing price index values. Predictions of bankruptcy rate were calculated using the observed housing price index for 2011 and 2012.

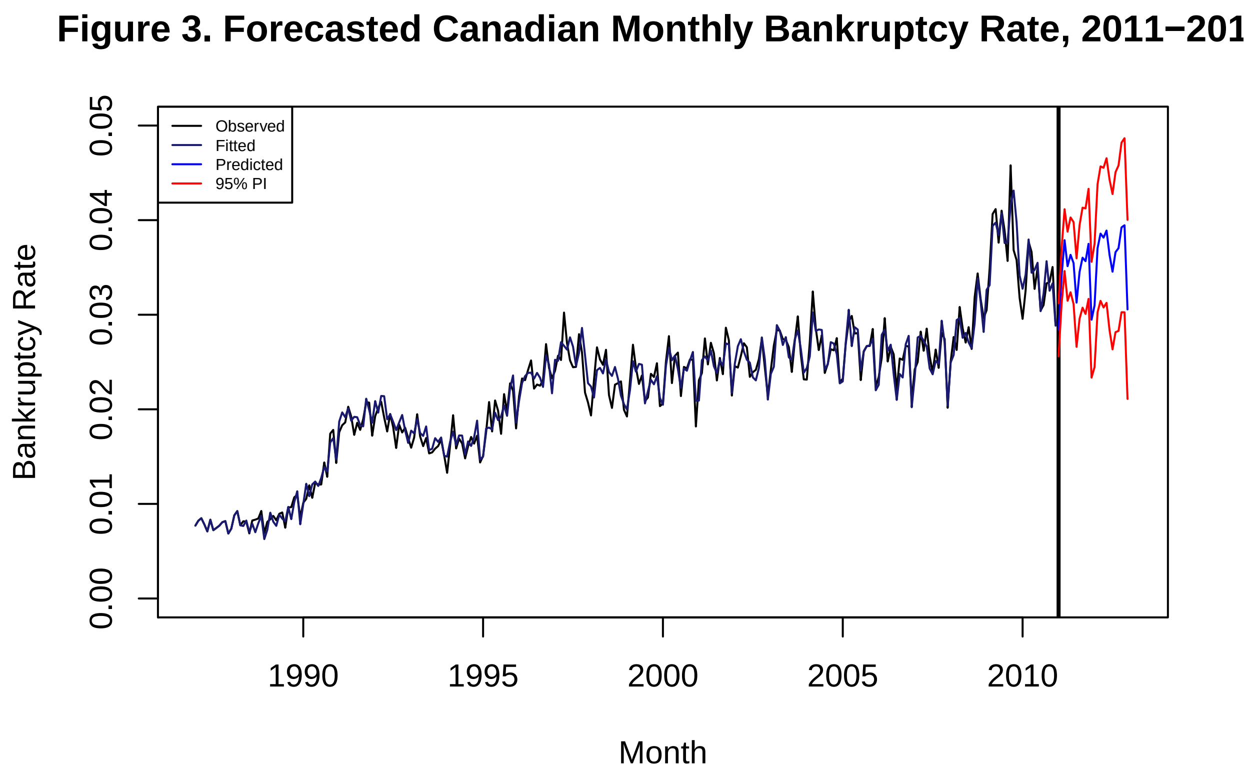


Table 1: Forecasted 2011-2012 Canadian Bankruptcy Rates

	Point Estimate	Lower 95% Bound	Upper 95% Bound
Jan 2011	0.028	0.026	0.031
Feb 2011	0.034	0.031	0.037
Mar 2011	0.038	0.035	0.041
Apr 2011	0.035	0.031	0.039
May 2011	0.036	0.032	0.040
Jun 2011	0.035	0.031	0.040
Jul 2011	0.031	0.027	0.036
Aug 2011	0.035	0.030	0.039
Sep 2011	0.036	0.031	0.041
Oct 2011	0.036	0.030	0.041
Nov 2011	0.037	0.032	0.043
Dec 2011	0.029	0.023	0.036
Jan 2012	0.031	0.024	0.037
Feb 2012	0.037	0.030	0.044
Mar 2012	0.039	0.031	0.046

	Point Estimate	Lower 95% Bound	Upper 95% Bound
Apr 2012	0.038	0.031	0.046
May 2012	0.039	0.031	0.047
Jun 2012	0.036	0.028	0.044
Jul 2012	0.035	0.026	0.043
Aug 2012	0.037	0.028	0.045
Sep 2012	0.037	0.028	0.046
Oct 2012	0.039	0.030	0.048
Nov 2012	0.039	0.030	0.049
Dec 2012	0.031	0.021	0.040