# Intro

NJSC consulting was tasked with forecasting monthly bankruptcy rates for Canada, for the period starting on January 2011 and ending in December 2012. Also specified in the instructions was that the forecast should be precise and accurate. No model preference or guidelines were included with the task,

At our disposal were two files:

* “train.csv” contained monthly observations of four variables, from January 1987 to December 2010. These four variables were comprised of the Canadian national bankruptcy rate, the unemployment rate, the house price index, and the country’s population count.
* “test.csv” contained monthly observations of three variables, from January 2011 to December 2012. These three variables were comprised of the Canadian national unemployment rate, the house price index, and the country’s population count.

# VAR

Another multivariate option to forecast the bankrupcy rate is to consider the variables on hand, or a subset of them, as endogenously related. This means that unlike in SARIMAX where we assume no causal effect of the response on any external variables, here we assume the response variable has a causal influence on at least one other variable to be included in the model. In such a situation, we turn to a vector auto-regression (VAR) model to assess the relationship between the response variable and all others which are endogenously related to it.

VAR models can be made more complex in the hopes of better modeling the response variable, and making better predictions. Exogenous variables can be added into the model for instance, as can seasonality indicators.

In our case, we examined a VAR model with the bankruptcy rate and the housing price index as two endogenous variables. Despite further adding the unemployment rate as an exogenous variable to the model, as well as incorporating seasonality, the values of our predictions on the validation set were not as accurate as other models.