Paper 3212-2015

# Using SAS® to Combine Regression and Time Series Analysis on U.S. Financial Data to Predict the Economic Downturn

Avinash Kalwani, Oklahoma State University, Stillwater, Oklahoma

Nishant Vyas, Oklahoma State University, Stillwater, Oklahoma

## ABSTRACT

During the financial crisis of 2007–2009, the U.S. labor market lost 8.4 million jobs, causing the unemployment rate to increase from 5% to 9.5%. One of the indicators for economic recession is negative gross domestic product (GDP) for two consecutive quarters. This paper combines quantitative and qualitative techniques to predict the economic downturn by forecasting recession probabilities. Data was collected from the Board of Governors of the Federal Reserve System and the Federal Reserve Bank of St. Louis, containing 29 variables and quarterly observations from 1976-Q1 to 2013-Q3. Eleven variables were selected as inputs based on their effects on recession and to limit multicollinearity: long-term treasury yield (5-year and 10-year), mortgage rate, CPI inflation rate, prime rate, market volatility index, Better Business Bureau (BBB) corporate yield, house price index, stock market index, commercial real estate price index, and one calculated variable - yield spread (Treasury yield-curve spread). The target variable was a binary variable depicting the economic recession for each quarter (1=Recession). Data was prepared for modeling by applying imputation and transformation on variables. Two-step analysis was used to forecast the recession probabilities for the short-term period. Predicted recession probabilities were first obtained from the Backward Elimination Logistic Regression model that was selected on the basis of misclassification (validation misclassification= 0.115). These probabilities were then forecasted using the Exponential Smoothing method that was selected on the basis of mean average error (MAE= 11.04). Results show the recession periods including the great recession of 2008 and the forecast for eight quarters (up to 2015-Q3).

## INTRODUCTION

The National Bureau of Economic Research (NBER) defines an economic recession as:

"A significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales."

Technical indicators of economic recession are as follows:

- One of the indicators for economic recession is decreasing gross domestic product (GDP) for two consecutive quarters (i.e., negative growth in GDP for two quarters in a row) (Shiskin, 1974).
- The other indicator is 1.5% rise in unemployment within 12 months (Eslake, 2009).

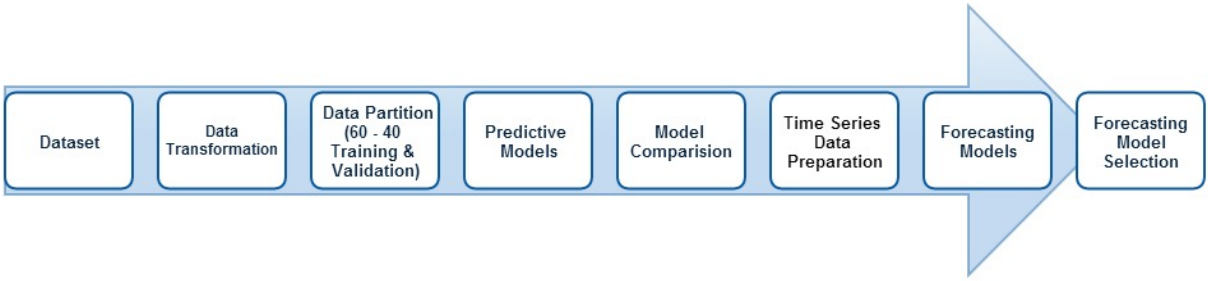



**Figure 1. The modeling process**

## DATA DESCRIPTION

Data was collected from Board of Governors of the Federal Reserve System (Comprehensive Capital Analysis and Review) & Federal Reserve Bank of St. Louis, containing 29 variables and quarterly observations from Q1-1976 to Q3-2013. Out of these 29 variables, 11 variables have been selected for inputs after cleansing variables and limiting multicollinearity.

### DEPENDENT VARIABLE

The target variable is a binary variable depicting the economic recession (1=Recession, 0=No recession). This target variable has observations derived on the basis of GDP i.e. if the GDP is decreasing for two consecutive quarters, then the observation has the value 1; otherwise 0.

### INDEPENDENT VARIABLE

These variables describe the factors affecting the economy and are selected on the basis of limiting multicollinearity, i.e., long term treasury yield (5-year and 10-year), Mortgage rate, CPI inflation rate, Prime rate, Market volatility index, Better Business Bureau (BBB) corporate yield, House price index, Stock market index, Commercial real estate price index. One more calculated input variable was used: Yield spread (Treasury yield - curve spread) has been added (Estrella & Mishkin, 1996).

## DATA TRANSFORMATION

Missing values were imputed using the mean for interval variables. The variables having high kurtosis values were transformed using range standardization. A few other transformations were also applied to the variables as shown in output 1.

Computed Transformations
(maximum 500 observations printed)

| Input Name | Role | Input Level | Name | Level | Formula |
|---|---|---|---|---|---|
| IMP_commercial_real_estate_price | INPUT | INTERVAL | SQRT_IMP_commercial_real_estate_ | INTERVAL | sqrt(max(IMP_commercial_real_estate_price-50.9, 0.0)/209.3) |
| IMP_dow_jones_total_stock_market | INPUT | INTERVAL | RANGE_IMP_dow_jones_total_stock_ | INTERVAL | (IMP_dow_jones_total_stock_market - 2417.1) / (17718.3-2417.1) |
| IMP_house_price_index | INPUT | INTERVAL | RANGE_IMP_house_price_index | INTERVAL | (IMP_house_price_index - 22.1) / (199-22.1) |
| IMP_market_volatility_index__vix | INPUT | INTERVAL | LOG_IMP_market_volatility_index_ | INTERVAL | log(IMP_market_volatility_index__vix + 1) |
| _10_year_treasury_yield | INPUT | INTERVAL | RANGE__10_year_treasury_yield | INTERVAL | (_10_year_treasury_yield - 1.6) / (14.6-1.6) |
| _3_month_treasury_rate | REJECTED | INTERVAL | RANGE__3_month_treasury_rate | INTERVAL | (_3_month_treasury_rate - 0) / (15.1-0) |
| _5_year_treasury_yield | INPUT | INTERVAL | RANGE__5_year_treasury_yield | INTERVAL | (_5_year_treasury_yield - 0.7) / (15-0.7) |
| bbb_corporate_yield | INPUT | INTERVAL | SQRT_bbb_corporate_yield | INTERVAL | sqrt(max(bbb_corporate_yield-3.9, 0.0)/13.7) |
| mortgage_rate | INPUT | INTERVAL | RANGE_mortgage_rate | INTERVAL | (mortgage_rate - 3.4) / (17.7-3.4) |
| prime_rate | INPUT | INTERVAL | SQRT_prime_rate | INTERVAL | sqrt(max(prime_rate-3.3, 0.0)/17) |

**Output 1. Variable Transformation Summary**

## REGRESSSION MODEL

SAS® Enterprise Miner® is used for the modeling process. The dataset is partitioned into training and validation data (60% training & 40% validation) subsets and then the variables are fed to two different logistic regression models (stepwise selection and backward elimination model). Out of these two models, backward elimination logistic regression  the best model for predicting the recession probabilities on the basis of validation misclassification (0.115). The selected backward elimination logistic regression model gives probabilities of recession for the period Q1-1976 to Q3-2013. Variables like GDP and unemployment rate were not considered for input as these directly describe the recession. From the selected regression model:

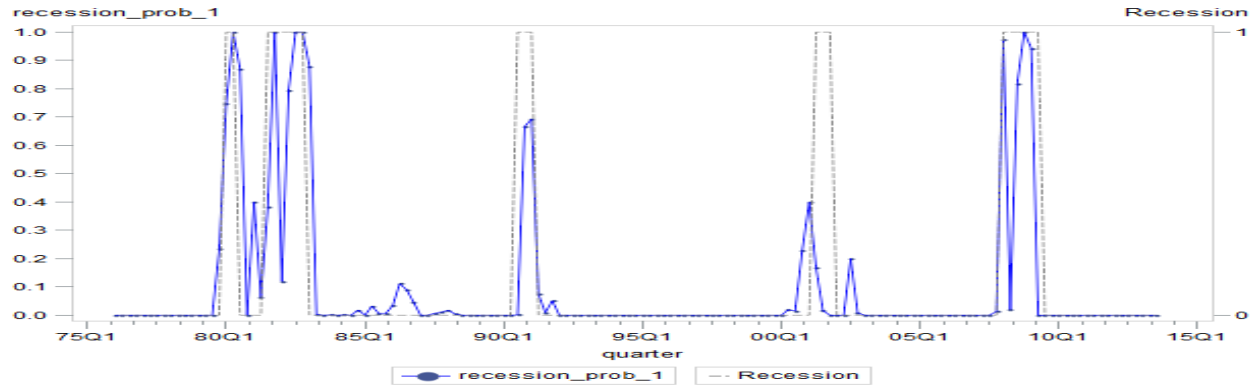
Sensitivity: 60%, Specificity: 98.5%, Overall correct rate: 93.3%

| | | Predicted | | | |
|---|---|---|---|---|---|
| | | Recession=1 | Recession=0 | Total | |
| **Actual** | Recession=1 | **12(TP)** | **8(FN)** | 20(P) | Sensitivity =12/20 **=60%** |
| | Recession=0 | **2(FP)** | **129(TN)** | 131(N) | Specificity =129/131 **=98.5%** |
| | Total | 14 | 137 | 151 | Overall correct rate =(12+129)/151 **=93.30%** |

**Table 1. Confusion matrix for logistic regression model**

| Regression Model | Validation Misclassification |
|---|---|
| Stepwise Logistic Regression | 0.148 |
| Backward Elimination Logistic Regression | 0.115 |

**Table 2. Selection criteria for logistic regression model**



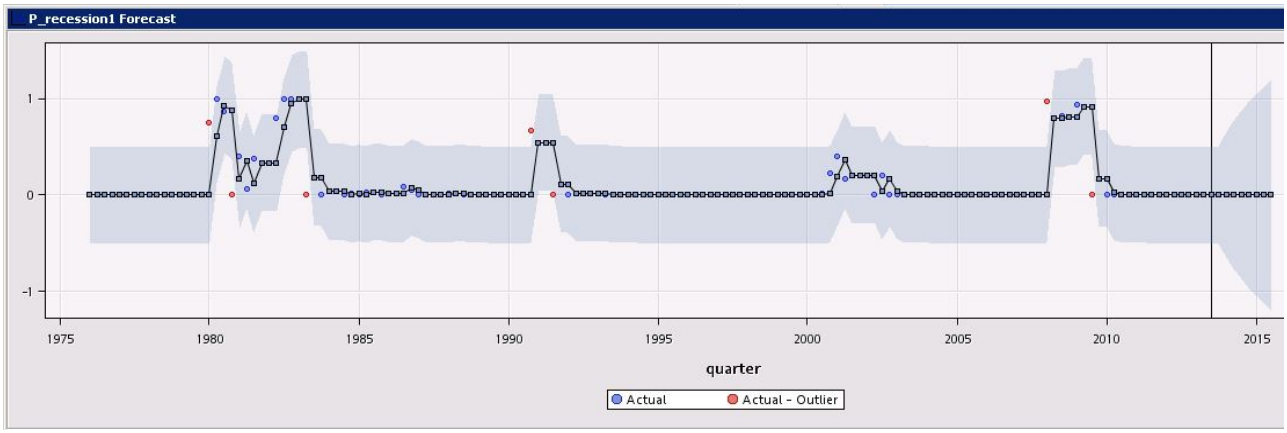

**Output 2. Plot generated by SAS® Enterprise Guide® showing actual probabilities and predicted recession probabilities**

## FORECASTING MODEL

SAS® Enterprise Miner® was used for the modeling process. The Time ID variable was then used to create a time series variable having quarterly intervals. This Time ID and the time series variable P_RECESSION_1 were then fed to different exponential smoothing methods of forecast and the simple exponential smoothing model was selected on the basis of Mean Absolute Error (MAE = 0.1104).

| Exponential Smoothing Model | MAE |
|---|---|
| Damped Trend | 0.1105 |
| Additive Seasonal | 0.1304 |
| Simple | 0.1104 |

**Table 3. Selection criteria for forecasting model**




**Output 3. Forecast generated by SAS® Enterprise Miner® up to 2015-Q3**

The results show the recession periods including the great recession of 2008 and the forecast for 8 QTR i.e. up to 2015-Q3. This model forecasts the economic recession probabilities for 2013-Q4 to 2015-Q3; which shows that probability of recession is relatively low up to 2015-Q3. Looking at the output of the

model, we can say that there will be no economic downturn till 2015-Q3. Result shows recession periods including the great recession of 2008, which justifies the further forecast of probabilities. Lack of some predictors might have resulted in some discrepancies in the forecast which can be enhanced by improving model performance. This model is more suitable for short-term forecast and it can help to identify the risk of forthcoming recession well in advance.

## RESULTS & DISCUSSION

- Selected backward elimination logistic regression model gives probabilities of recession for the period 1976-Q1 to 2013-Q3.
- From selected regression model: Sensitivity: 60%, Specificity: 98.5%, Overall correct rate: 93.3%
- Variables like GDP and unemployment-rate were not considered for input as these directly describe recession.
- The forecasting model selected is the Simple Exponential Smoothing model based on Mean Absolute Error.
- Further research may improve the model performance by including more predictor variables. It can be observed that the presidential elections (Republican Party to Democratic Party shift) also had an impact on the economic downturn of 2007-2009.

## CONCLUSION

- Result shows recession periods that include the great recession of 2008, which justifies the further forecasting of recession probabilities.
- Lack of some predictors might have resulted in some discrepancies in the forecast, which might be enhanced by adding some appropriate inputs like crude oil price index and presidential elections.
- Measurement bias is present in the model due to unbalanced data.
- This model is more suitable for short-term forecasting and it can help in identifying the risk of forthcoming recession well in advance.

## REFERENCES


- Eslake, S. (2009). The difference between a recession and a depression*. *Economic Papers: A journal of applied economics and policy,* 28(2), 75-81.
- Estrella, A., & Mishkin, F. S. (1998). Predicting US recessions: financial variables as leading indicators. *Review of Economics and Statistics,* 80(1), 45-61.
- Lamy, R. (2013). Forecasting the business cycle in the US. *The Forecasting Advisor.*
- Shiskin, J. (1974). The changing business cycle. *New York Times,* 12, 22

## ACKNOWLEDGMENTS

I thank Thomas E. Billings (SAS Developer, MUFG Union Bank) for his constant support and motivation.

I also thank Professor Dr. Goutam Chakraborty (Founder of SAS & OSU Data Mining Program) for his valuable suggestions while developing this poster.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Avinash Kalwani
M.S. MIS, SAS® & OSU Data Mining Program
Phone: 405-412-7591
Email: avinash.kalwani@okstate.edu