A Systematic Analysis of Potential Leading Indicators in the United States through Vector Autoregression

Nicholas D. Holschuh (Holschun@Carleton.edu)

Carleton College
Senior Integrative Exercise
Adviser: Pavel Kapinos
May 26, 2011

Abstract

The business cycle has been a subject of great economic interest over the past century. Decision making in both the public and private sector is influenced by the phase of the business cycle, and as a result, our ability to understand and model real economic activity is incredibly important. This study presents a group of linear models that attempt to explain the evolution of real economic activity, in an effort to determine how the inclusion of leading indicators affects out-of-sample predictive power. I focus on 10 potential leading indicators: interest rate spread, producer price index, hours worked, corporate profits, M1, M2, the Federal Funds Rate, the S&P 500, and the Dow Jones industrial average. Using a rolling vector autoregressive structure and two different forecasting methods, all possible combinations of these leading indicators were analyzed. I found that including any of the viable leading indicator candidates in the model improves performance, however interest rate spread, the producer price index, and M1 yield the best results. With every additional variable beyond two included in the regression, the loss in degrees of freedom results in worse forecasts despite better in-sample fit.

Potential Leading Indicators

Corporate Profits
Business Loans
M1 and M2
Federal Funds Rate
Interest Rate Term Spread
Producer Price Index
S&P 500
Dow Jones
Hours Worked

Preliminary Data Analysis: Unit Root Testing

Unit root testing was performed using the Augmented Dickey Fuller test, based on a regression of the form:

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \sum_{i=1}^{l} \gamma_i * \Delta y_{t-i} + \epsilon_t$$

Non-stationary sequences were differenced until they became stationary.

Preliminary Data Analysis: Static Granger Causality

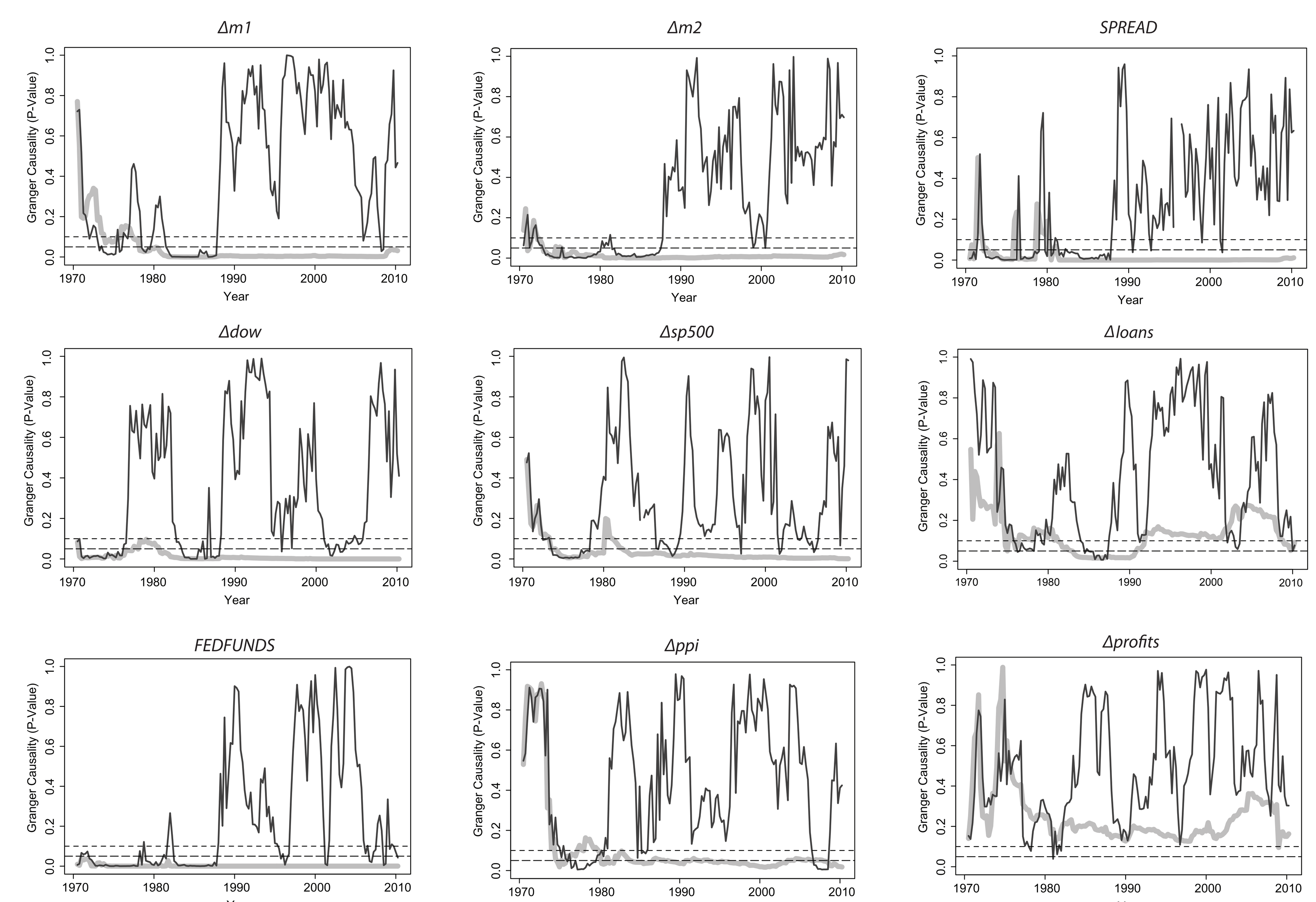
Granger causality tests were performed to determine the temporal relationship between variables. The test statistic is calculated by the estimation of:

$$y_t = a_0 + \sum_{i=1}^{lmax} a_i y_{t-i} + \sum_{j=1}^{lmax} b_j z_{t-j} + e_{yt}$$

This test eliminated Hours Worked as a potential leading indicator.

Preliminary Data Analysis: Rolling Granger Causality

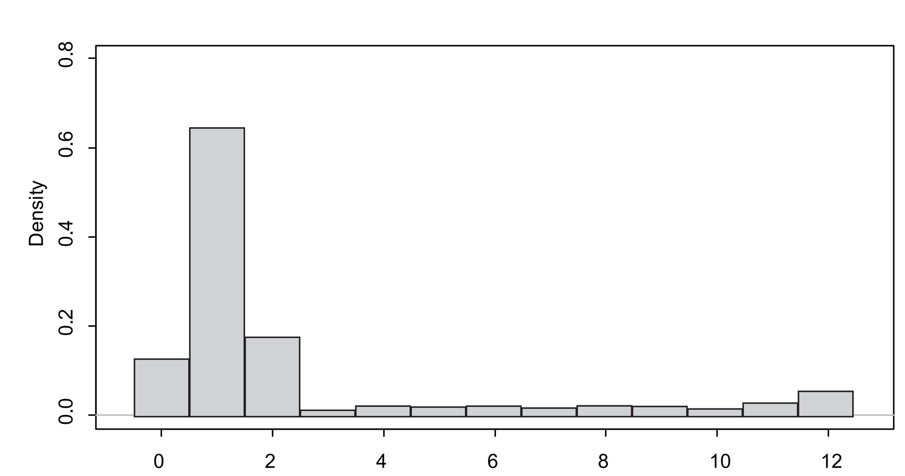
Rolling Granger causality tests were performed to qualitatively determine periods of strong leading behavior for each of the different variables in question. This was done two ways, with a fixed start date (grey line) and with a rolling start date (black line). 5% and 10% p-values are plotted as dashed lines, to indicate periods of strong Granger causality. Further study is required to determine how Granger causality and forecast accuracy are related.



VAR Specification and Lag Selection

Forecasting was performed based on a series of multivariate vector autoregressions. These took the standard form:

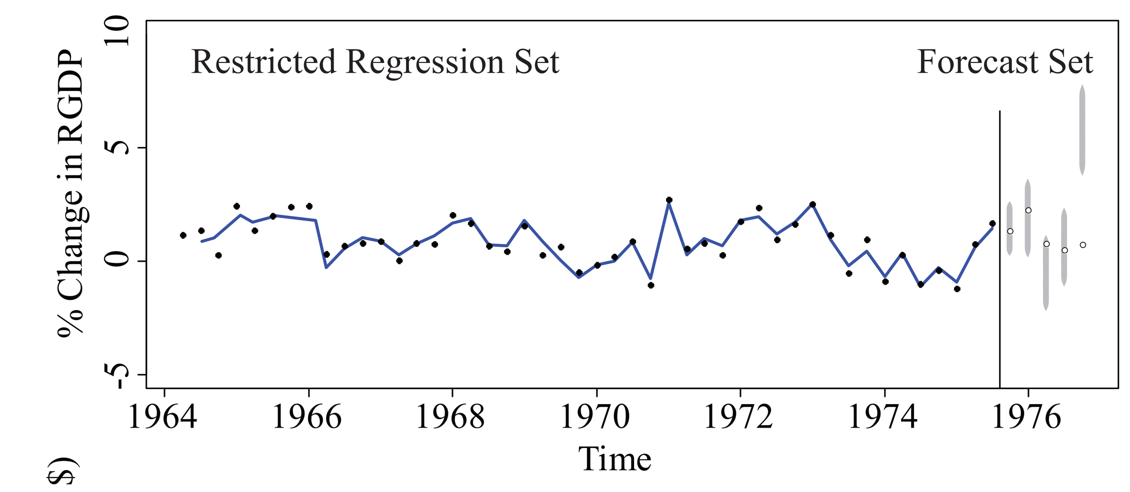
$$x_t^i = \alpha + \sum_{l=1}^{lmax} \boldsymbol{\gamma_l} \mathbf{X}_{t-l} + \mathrm{u_t}$$

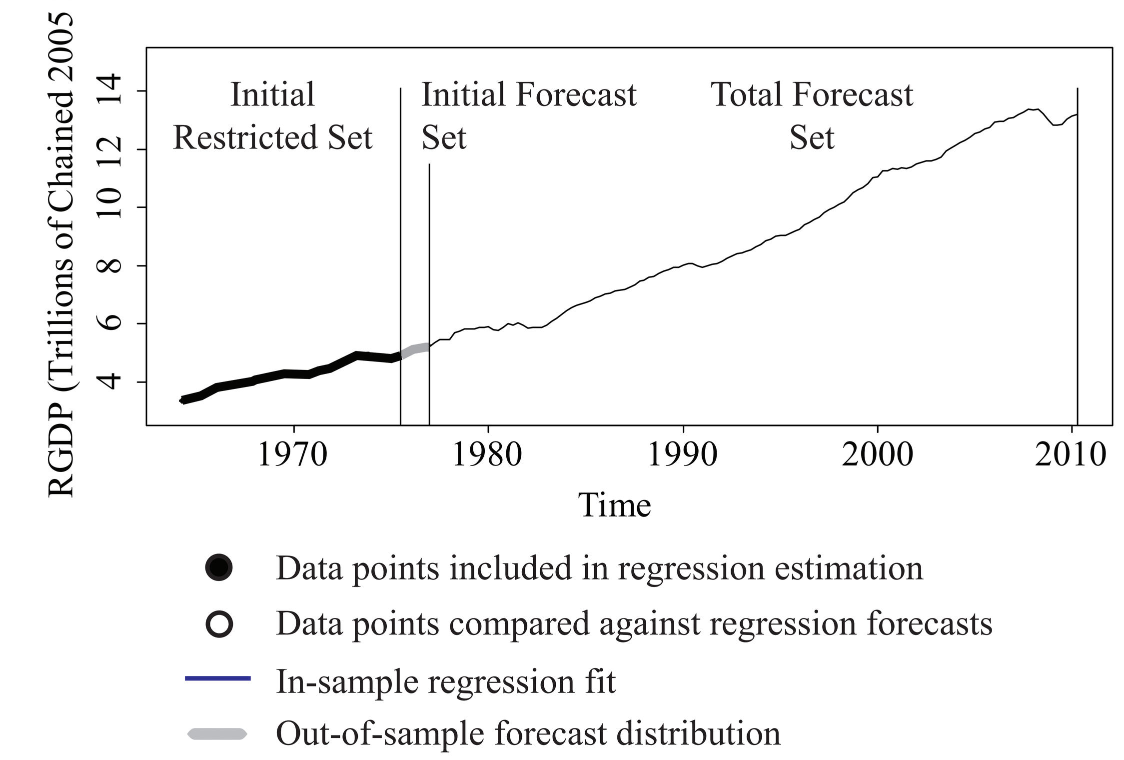


For consistency in the number of required parameter estimates, I chose to use a fixed number of lags across all regressions. To the left is a plot of selected lag orders for a random draw of regressions used in the forecasting process, calculated using the Hannon-Quinn criterion. On average I found that one lag was optimal, regardless of the number of time periods included in the rolling regression. Therefor, all regressions were calculated using an lmax=1.

Forecasting and Evaluation

Regressions were estimated from a restricted set of the data, and parameter estimates were used to forecast RGDP values 5 periods out from the restricted set (shown below). Forecasts distributions were found using two methods, and model performance was evaluated using a quadratic probability score (QPS), comparing predicted business cycle turning points to observed ones.





(Individual model performance can be found in the companion material

The QPS is defined as: $\frac{2}{T}\sum_{t=1}^{T}(P_t - D_t)^2$

where P is the percentage of turning points predicted at each time (t) and D is an index indicating whether or not a turning point actually occurred at time (t).

Conclusions

- Leading indicators provide valuable forecasting information
- Interest Rate Spread, M1, and PPI produce the best bivariate VAR forecasts.
- Including more than 1 leading indicator worsens model performance