

Predicting U.S. Industrial Production with Oil and Natural Gas Prices

Matthew L. Higgins
Department of Economics
Western Michigan University

Prediction is very important in economic analysis. The prediction of a variable that will be determined in the future is called a forecast. Business enterprises require forecasts of sales, input costs and interest rates to make capital allocation decisions. Government agencies, such as the U.S. Federal Reserve, the International Monetary Fund and the World Bank, require forecasts of economic growth, unemployment and inflation to better design economic policy. Firms and government agencies employ large numbers of economists, statisticians and information technology specialist to build sophisticated mathematical models to do this forecasting.

This article demonstrates how economists forecast by building a small illustrative model to predict future levels of the U.S. monthly index of industrial production. The level of industrial production is greatly affected by the price of oil and natural gas because these commodities are a major source of energy and refined products that are used in industrial production. Oil and natural gas prices are highly volatile due to supply shocks from technological innovations, such as hydraulic fracturing, and political turmoil in the regions of the world where petroleum is produced. The demand for these products is also very inelastic, and therefore, large price fluctuations are necessary to affect the quantity demanded. Incorporating information on oil and natural gas prices should improve the forecasting accuracy of a statistical model to forecast industrial production.

Determining the appropriate measure of prices for oil and gas productions is difficult because of the large number of fuel types and refined products that are derived from oil and gas. For example, a barrel of oil is typically refined into over 40 different commercial products. Although historical data is not readily available on all of these products, Table 1 shows 20 price series that will be used in the forecasting model. The data was obtained from the U.S. Energy Information Administration and are monthly from January of 1995 through January of 2016. The variables include the spot price of crude oil and natural gas. The variables also include futures prices of these commodities. Industrial firms use futures markets to hedge the risk of volatile prices. Also included are price series for ten different refined products. Twenty variables are too many to explicitly include in the forecasting model. Including too many variables can increase the sampling variability of the statistical estimator of the model and cause the model to predict poorly.

When confronted with large data, economists often use modern machine learning algorithms to reduce the dimensionality of the data. We assume that the information in the oil and gas price variables can be represented by a smaller number of latent factors. More formally, let X_t be a 20×1 vector containing the price variables from Table 1 observed in time period t . We assume the vector of prices are related to the vector of factors by the model

$$X_t = \Lambda F_t + u_t, \quad (1)$$

where F_t is the $k \times 1$ vector of latent factors. In (1), Λ is a $20 \times k$ matrix of unknown parameters. Each row of Λ contains the coefficients that relate each factor to the corresponding price variable. The 20×1 vector u_t contains the idiosyncratic errors that affect the prices that are not accounted for by the factors.

The latent factors can be estimated for the observed price data by the method of principal components. Principal components is a data reduction technique which estimates the latent factors with linear combinations of the observed data that have the greatest variance. Applying this methodology, we find that the majority of the information in the prices can be represented by two factors.

The estimated factors can then be entered into a forecasting model. Let y_t denote the observed value of industrial production at time t . We assume industrial production can be represented by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \alpha' F_t + \epsilon_t, \quad (2)$$

where the model includes the factors and lagged values of industrial production to account for a trend. The β_i 's and the vector α are parameters and ϵ_t is a random error. The model is estimated and forecasted in two steps. First, the latent factors in (2) are replaced by their estimates $\hat{F}_t$ from (1) and the model is estimated by the method of least squares. In the second step, the trends in the factors are extrapolated and equation (2) is iterated to obtain forecasts of future values of industrial production.

Figure 1 shows the observed index of industrial production over the sample period and monthly forecasts of the index through January of 2018. The blue line in the figure represent the forecasts. The shaded gray regions are 80 percent and 95 percent prediction regions. The prediction regions are obtained from the estimated probability distribution of the error ϵ_t in model (2). The forecasts show a slight downward trend with a strong seasonal component.

To show the effect of including the oil and gas price factors, we also estimate model (2) without the factors. The forecasts and prediction regions for this simplified model are shown in Figure 2. The forecasts in the simplified model show a slightly more pronounced downward trend. The most prominent difference between the two sets of forecasts is the size of the prediction regions. The prediction regions for the model which includes the factors are significantly smaller. The smaller forecast regions indicate that the oil and gas price factors contain useful information, and reduce the uncertainty in the predictions about the future path of industrial production.

Economic forecasting has progressed rapidly in recent years. The increasing amount of online real-time economic data has made “big data” available to economic forecasters. Economic forecasters are just beginning to adopt advances in machine learning to exploit this data. Economic forecasting should be an appealing field for young mathematicians interested in economics, statistics and data algorithms.

Table 1. Oil and Natural Gas Prices used to Forecast Industrial Production.

Commodity Spot Price:

Crude Oil Prices: West Texas Intermediate Cushing, Oklahoma

Natural Gas Price: Henry Hub

Commodity Futures Price:

Cushing OK Crude Oil Future Contract 1 Month

Cushing OK Crude Oil Future Contract 2 Month

Cushing OK Crude Oil Future Contract 3 Month

Cushing OK Crude Oil Future Contract 4 Month

Natural Gas Futures Contract 1 Month

Natural Gas Futures Contract 2 Month

Natural Gas Futures Contract 3 Month

Natural Gas Futures Contract 4 Month

Refined Products Spot Price:

No. 2 Heating Oil Prices: New York Harbor

US Diesel Sales Price

Kerosene-Type Jet Fuel Prices: U.S. Gulf Coast

Propane Prices: Mont Belvieu, Texas

US Premium Conventional Gas Price

US Midgrade Conventional Gas Price

US Regular Conventional Gas Price

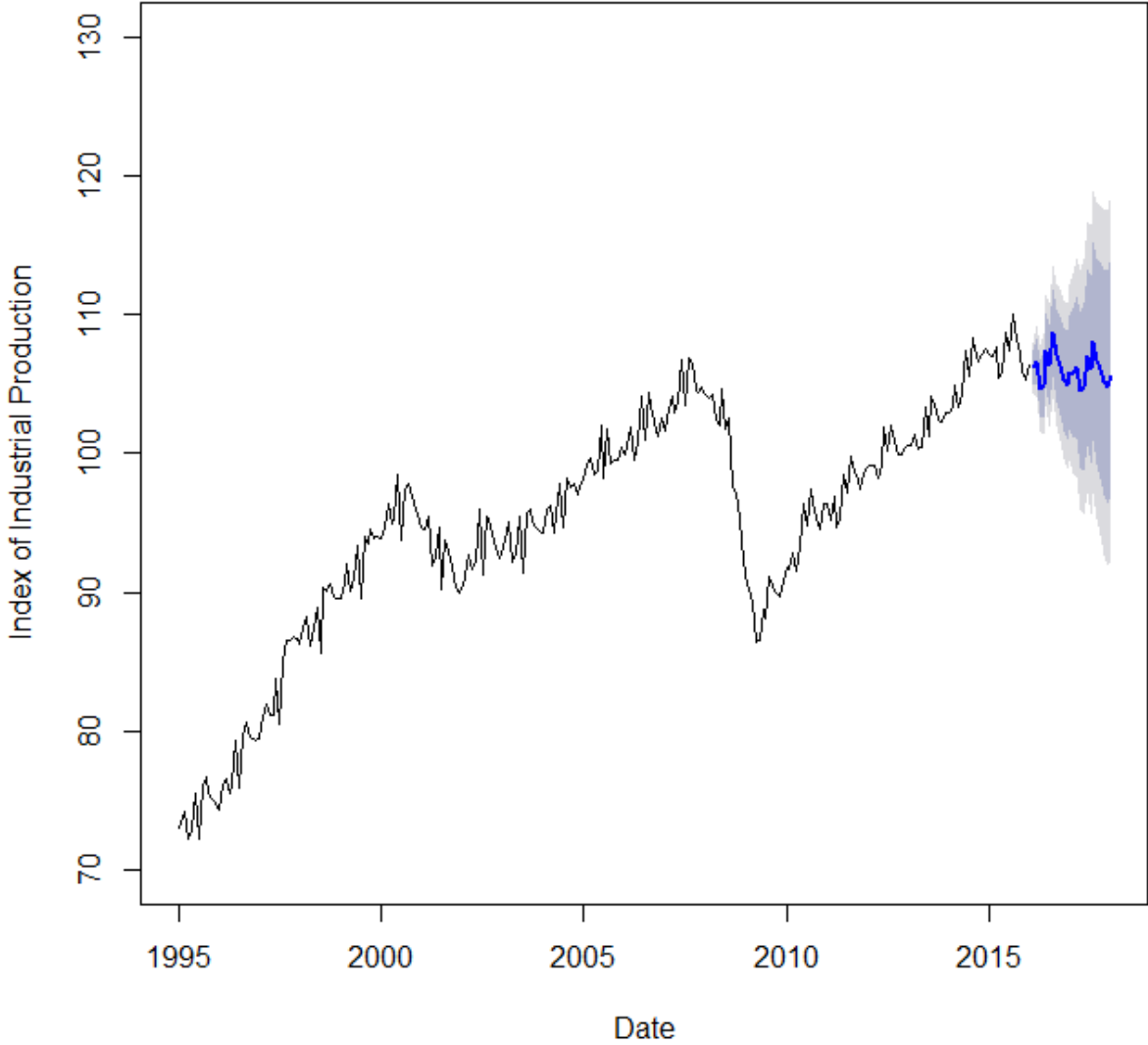
US Premium Reformulated Gas Price

US Midgrade Reformulated Gas Price

US Regular Reformulated Gas Price

Source: U.S. Energy Information Administration

**Figure 1. Forecast of Industrial Production
Using Oil and Gas Factors**



**Figure 2. Forecast of Industrial Production
Without Oil and Gas Factors**

