CS 229 Machine Learning

Forecasting with Expert Opinions

Khalid El-Awady

Background

In 2003 the Wall Street Journal (WSJ) introduced its Monthly Economic Forecasting Survey. Each month the WSJ polls between 50 and 60 well-known economic experts asking their forecasts of future key economic variables such as GDP, inflation, US treasury rates, unemployment, housing starts, and other data. The forecasts are always for set times of the year, namely the ends of the first and second half of the calendar year, so while the data is collected monthly, the forecast interval varies from one to six months ahead. The forecasts of all participants are made public, but the WSJ also presents a "consensus" view which is simply the sample average forecast of the participants.

The data set suffers from a number of challenges:

- **Small number of observations** – I was able to extract a series of $m = 12$ observations representing six years of 6-month ahead forecasts. Prior to 2005 the data collected and the format used was not consistent with data after 2005.
- **Large number of features (forecasters):** Over the 6 year period, 50 to 60 experts provided forecasts each month, with a total of 79 different forecasters providing inputs over the 6-year period.
- **Missing data:** Only 20 – 30 of the 79 forecasters provided forecasts for the whole time period, depending on indicator. Experts routinely drop out and new ones are added. Also, not all experts polled at a given time chose to provide a forecast for every variable. Of the possible 12×79 data points over one third are missing from each of our variables.

In this paper we seek to explore two things:

1. How much improvement over the average forecast can be made with machine learning techniques discussed in class?
2. Bayesian Networks are often chosen as the preferred way to represent expert opinion [1]. I will attempt to implement one of these on this data and explore the insights it provides.

Data

Three data sets of 6-month ahead forecasts were extracted from the WSJ data for analysis:

1. 10-year US Treasury rate: 12 observations from 79 experts
2. US unemployment rate: 12 observations from 79 experts
3. Oil price: 9 observations from 79 experts (this variable was added to the forecast in 2007)

Linear Regression with PCA

The WSJ data is shown in Figure 1. Each red triangle represents the forecast of a single expert at a point in time 6 months earlier (e.g., the red triangles in Dec08 represent the forecast made in June08 for the value of that variable in Dec08). The blue line is the average or "consensus" forecast of the experts presented by the WSJ, and the black line represents the actual value of the variable on the shown date. As can be seen, the forecast vary widely, and even so, the actual value can deviate far from the “consensus” view. We also notice two issues with the forecast: it tends to lag actuals (i.e., forecasters guess of next period’s value often looks very much like this period’s value), and forecasters underestimate the variance (i.e., they don’t believe it will change as much as it often does).

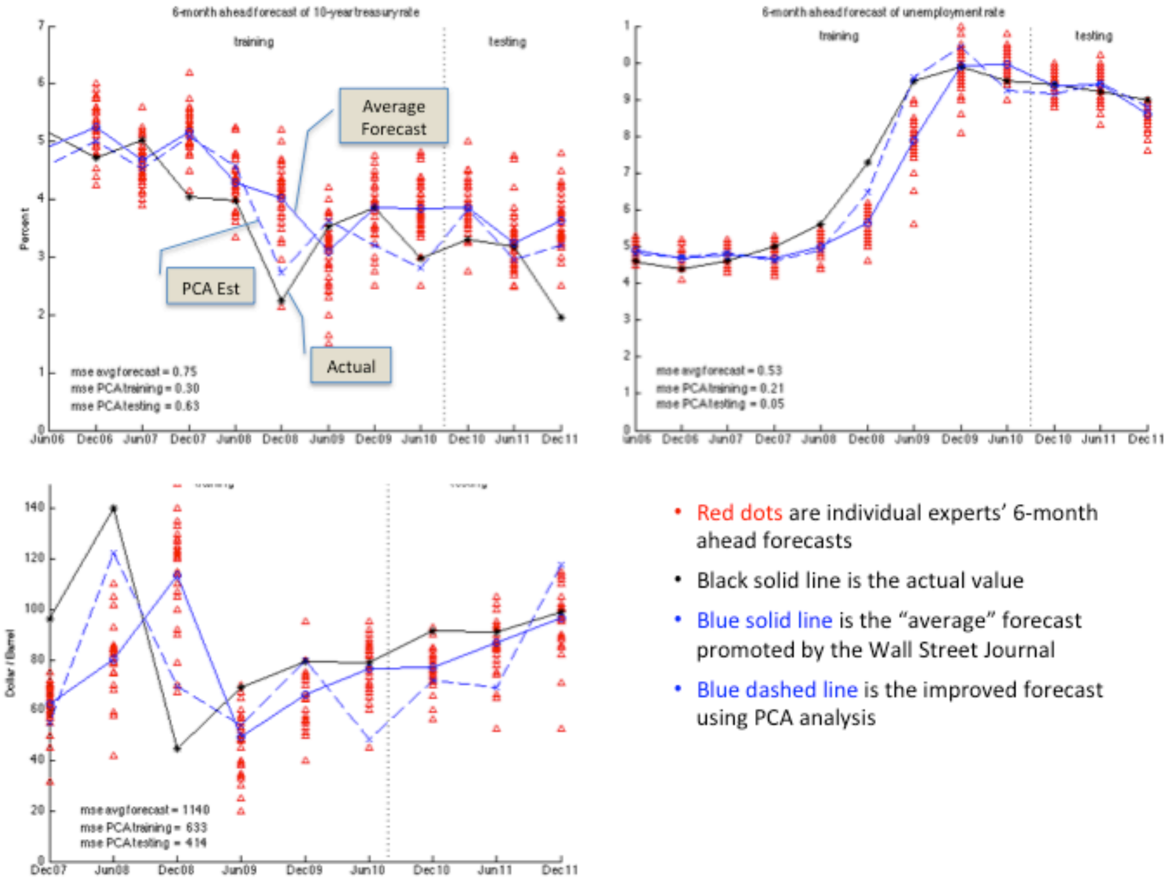

Figure 1. 10-year treasury rate (top left), unemployment rate (top right), and oil price (bottom left).

We first try a linear regression model to see how much improvement can be obtained over the simple average. We do this through the following processing steps on the data:

- Remove forecasters with incomplete data over the whole time period. This leaves 31 forecasters and corresponds to the features of the machine-learning problem.
- Divide the data into a training set of size 9 samples for the rate and unemployment, and 6 samples for the oil price, and a test set of size 3 samples for each indicator.
- Reduce the feature set size. The data set is under-determined with a small number of observations (6 or 9) and a large number of features (> 20). To avoid overfitting, we convert this to a more traditional regression problem using PCA to reduce the feature set to a number below the number of observations. Table 1 below summarizes the chosen PCA parameters and the associated mean square errors (MSEs) of the estimates.

We notice that linear regression with PCA resulted in significant reduction in the mean squared error of the estimate compared to the sample average in the case of unemployment and oil prices. The 10-year rate, though, shows relatively small improvement.

Variable	Samples	Forecasters with Complete Data (Features)	Number of Principal Components in Reduced Feature Set	Portion of Variance in selected Principal Components	MSE using "Consensus" (Average)	Training MSE Using Linear Regression with PCA	Testing MSE Using Linear Regression with PCA
10-Year Rate	12	31	6	95%	0.75	0.30	0.63
Unemployment	12	29	3	79%	0.53	0.21	0.05
Oil Price	9	23	5	100%	1,140	633	414

Table 1. Forecasting with Linear Regression and PCA.

Bayesian Network Estimates

We postulate a Bayesian network inspired by the model presented in Problem Set 4, question 2. For each of the economic indicators we assume there is a latent intrinsic value that is normally distributed with an unknown mean and variance. Each of the forecasters is modeled as a latent node contributing a bias of their own (also normally distributed with unknown mean and variance). The Bayesian network combines these two and adds an independent noise source. This is depicted graphically in figure 2.

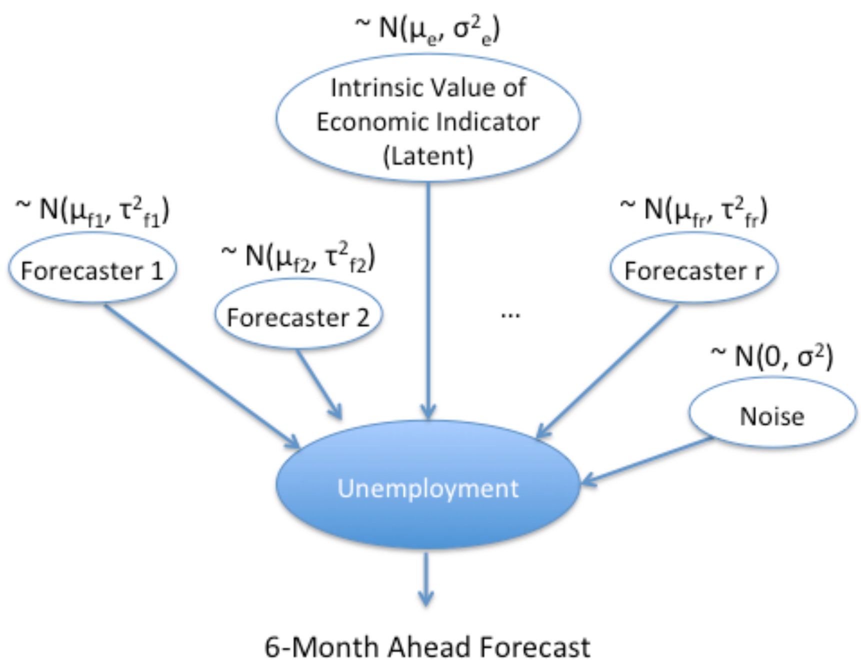

Figure 2. Bayesian network influence diagram for economic indicators.

The parameters $\mu_e, \sigma_e^2, \mu_{fi}$, and σ_{fi}^2 are estimated using the EM algorithm of problem 2 in Problem Set 4, where the subscript 'e' refers to each of the three economic indicators. Because of the different natures and scales of the indicators the parameters for each are estimated from their respective forecasts independently. The parameters σ^2 are assumed known and tuned to accommodate the difference in forecasts from actual value.

This approach allows for a varying number of forecasters each time sample as well as forecasters dropping out and new forecasters coming in. It thus makes maximal use of available information.

Figure 3 shows the estimated intrinsic value from the Bayesian network for the economic indicators.

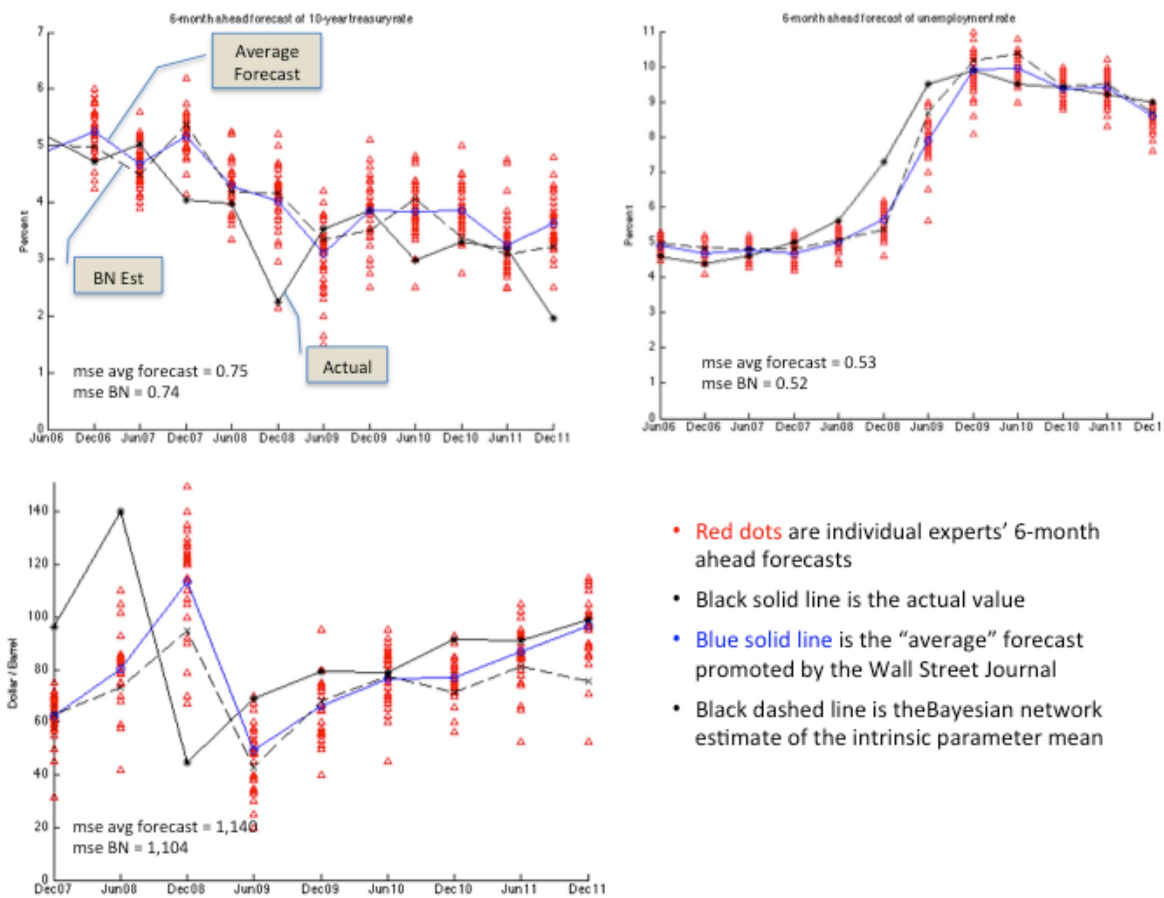

Figure 3. Economic indicator data with Bayesian network intrinsic estimates.

Table 2 shows the performance of the BN intrinsic estimate compared to the sample mean and the PCA / linear regression estimate in terms of mean square error relative to the actual data.

Variable	Observations	Forecasters with Complete Data (Features)	Average Number of Forecasts in any Given Sample	MSE using "Consensus" (Average)	Testing MSE Using Linear Regression with PCA	MSE Using Bayesian Network
10-Year Rate	12	31	53	0.75	0.63	0.74
Unemployment	12	29	45	0.53	0.05	0.52
Oil Price	9	23	53	1,140	414	1,104

Table 2. Comparison of Bayesian Network performance to the sample mean and the PCA estimate.

We notice that the BN estimate of the intrinsic mean differs slightly from the sample mean. But it does not do a better job of forecasting as its mean squared error relative to the actual data is about the same as the sample mean.

Commentary and Conclusions

We investigated a forecast data set based on expert opinions that consists of many more features than samples, and forecasters who drop out and new forecasters who join in at each time sample. The sample mean of the forecasts is nominally put forward as the "consensus" view. Comparison of this consensus estimate to the actual data shows the consensus view is often far from the actual data.

We applied two machine-learning techniques to the data in an attempt to improve the forecast: a supervised approach and an unsupervised one. The supervised technique involves extracting a "complete" subset of the data comprising only forecasters who have provided forecasts for each time sample. This reduces the usable data roughly by over a third. Principal components are used to extract a lower dimensional feature set and a regression is then applied to relate the projected data in the principal component space to the actual values. This PCA-based regression provides a much better fit to the actual data in the case of the unemployment and oil price forecasts, but doesn't improve the 10-year rate forecast much.

The unsupervised learning technique uses a Bayesian network to estimate latent intrinsic values of the indicators. This has the advantage of using all available data at each time sample. The estimate obtained, though, is not better at forecasting the actual value.

We conclude that there are structural deficiencies in the forecasters' estimates that can sometimes be corrected for when actual values are employed to detect these. This would include the tendency of forecasters to rely too heavily on the previous period's values, as well as the fact that they underestimate the volatility of the indicators. Further, the sample mean captures the intrinsic value of the forecasters quite well and a Bayesian network did not provide a meaningfully different estimate.

References

1. D. Heckerman, 1996. "A tutorial on learning with Bayesian networks", Microsoft Research tech. report, MSR-TR-95-06.
2. A. Ng. CS229 Notes. Stanford University, 2011.
3. T. Lai and H. Xing. Statistical Models and Methods for Financial Markets. Springer, 2008.