Forecasting in Business

Alexandru Constangioara [*], Simona Aurelia Bodog [**],
Florian Gyula Laszlo*, Dana Petrica*

University of Oradea,

* Faculty of Economic Studies

**Faculty of Electrical Engineering and Information Technology

sbodog@uoradea.ro, aconstangioara@uoradea.ro

***Abstract* – There is little management literature on forecasting at national level and scarce practical issues. We are told only that there are several forecasting methods and that forecasting is a tool of strategic management. After addressing several claims about forecasting made in management literature, this paper presents several forecasting methods used in operational management**

***Keywords:* operational vs. strategic forecasting, decision tress, parametric forecasting**

I. INTRODUCTION

There is little management literature on forecasting at national level. However there is general agreement about the using time series in business forecasting. It is also taking for granted that forecasting is a tool of strategic management. While strategic forecasting might exist, the fact is that in most businesses operational forecasting is most prevalent.

The fundamental problem in prediction is the correct determination of an unknown quantity in the presence of supplementary facts. The unknown quantity is called a target and the supplementary facts are called inputs. To build a successful predictive model one must first define an analytic objective. The predictive model serves as a means of fulfilling the analytic objective. Analytic objectives that involve predictive modeling generally fall in one of a limited number of categories (Par Rud, 2001).

- Response models attempt to identify individuals likely to respond to an offer or solicitation.
- Up-sell and cross-sell models are used to predict the likelihood of existing customers wanting additional products from the same company.
- Risk assessment models quantify the likelihood events that will adversely affect a business.
- Attrition models gauge the probability of a customer taking his or her business elsewhere.
- Lifetime value models evaluate the overall profitability of a customer over a predetermined length of time.

There are many business applications of predictive modeling**{** XE "predictive model:business applications" **}**. Database marketing uses customer databases to improve sales promotions and product loyalty. In target marketing, the cases are customers, the inputs are attributes such as previous purchase history and demographics, and the target is often a binary variable indicating a response to a past promotion. The aim is to find segments of customers that are likely to respond to some offer so they can be targeted. Historic customer databases can also be used to predict who is likely to switch brands or cancel services (churn). Loyalty promotions can then be targeted at new cases that are at risk. **{** XE "database marketing" **}{** XE "target marketing" **}{** XE "attrition predicting" **}**Credit scoring (Hand and Henley 1997) is used to decide whether to extend credit to applicants. The cases are past applicants. Most input variables come from the credit application or credit reports. A relevant binary target is whether the case defaulted. The aim is to reduce defaults and serious delinquencies on new applicants for credit. **{** XE "credit scoring" **}**In fraud detection**{** XE "fraud detection" **}**, the cases are transactions (for example, telephone calls, and credit card purchases) or insurance claims. The inputs are the particulars and circumstances of the transaction. The binary target is whether that case was fraudulent. The aim is to anticipate fraud or abuse on new transactions or claims so that they can be investigated or impeded. Supervised classification also has less business-oriented uses. Image classification has applications in areas such as astronomy, nuclear medicine, and molecular genetics (McLachlan 1992; Ripley 1996; Hand 1997).

All these are operational forecasts, used in conducting day to day selection of customers and risk management. It is hard to identify a strategic forecast. Generally strategic forecasts are highly specialized analysis performed in macroeconomics. While in macroeconomics necessary data for strategic forecasts is easily available, this is hardly the case in business world. Moreover, there are economic mechanisms which might render strategic forecast inefficient. Using forecast to predict future prices and performance is a central issue in Financial Economics. Prices, rather then reflecting performance they reflect expectations about future performance. Changes in prices reflect new information about future performance. As information arrives randomly, prices do follow a random walk. Consequently we cannot predict future performance using past information. If our estimates are statistically significant we have found evidence that the market is not efficient. Of course we can conduct fundamental analysis that this is by no means strategic forecasting. Even specialists might not trust the usefulness of strategic forecasting. Kenneth Arrows, Nobel price winner for his work on economic contracts, while he was working in the Army's weather forecast department during World War II, wrote a letter to his commander asking to be discharged from his duties because long term weather forecast is useless. He got a surprising answer, explaining that the army is well aware that long term weather forecast is useless but it needs it for strategic planning.

II. FORECASTING METHODS

Having said that forecasting is mainly operational and time series analysis is rare in business practice, we will briefly present main methods of forecasting.

To solve the fundamental problem in prediction, a mathematical relationship between the inputs and the target is constructed. This mathematical relation is known as a predictive model. Once established, the predictive model can be used to produce an estimate of an unknown target value given a set of input measurements. The observations used in analysis are assumed to be representative of future (unobserved) input and target measurements. An extremely simplistic predictive model assumes all possible input and target combinations are recorded in the training data. Given a set of input measurements, you need only to scan the training data for identical measurements and note the corresponding target measurement (EViews Tutorial). Often in a real set of training data, a particular set of inputs corresponds to a range of target measurements. Because of this noise, predictive models usually provide the expected (average) value of the target for a given set of input measurements. With a qualitative target, (ordinal, nominal, or binary) the expected target value may be interpreted as the probability of each qualitative level. Both situations suggest that there are limits to the accuracy achievable by any predictive model (EViews Tutorial).

Usually, a given set of input measurements does not yield an exact match in the training data. How you compensate for this fact distinguishes various predictive modeling methods. Perhaps the most intuitive way to predict cases lacking an exact match in the training data is to look for a nearly matching case and note the corresponding target measurement. This is the philosophy behind nearest-neighbor prediction and other local smoothing methods (EViews Tutorial).

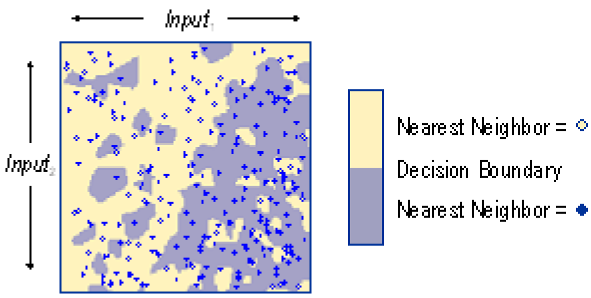

Figure 1. Nearest Neighbor Prediction
Source: EViews Tutorial

Nearest neighbor prediction' roots date back to the taxonomists of the 19^{th} Century. In its simplest form, the predicted target value equals the target value of the nearest data case. Basically the input space is divides into cells of distinct target value. The edge of these cells, where the predicted value changes, is known as the decision boundary. A nearest neighbor model has a very complex decision boundary (EViews Tutorial).

Recursive partitioning models, commonly called decision trees are the most ubiquitous of predictive modeling tools, although may not yield the largest generalization profit (EViews Tutorial). They do not assume a particular functional relationship which allows them to find complex relationships between target and input variables. It also allows them, if not carefully tuned, to find complex input and target associations that do not really exist. Tree algorithm is very complex. First part searches for the best split for each variable, using for example a Pearson chi-squared statistic to quantify the independence of counts resulted. Large values for the chi-squared statistic suggest the proportion of 0's and 1's in the left branch is different than the proportion in the right branch. A large difference in target level proportions indicates a good split.

After determining the best split for every input, the tree algorithm compares each best split's corresponding log worth. The process repeats in each leaf until there are no more allowed splits whose adjusted log worth exceeds the depth-adjusted thresholds. This completes the split search portion of the tree algorithm. The resulting partition of the input space is known as the *maximal tree*. Development of the maximal tree was based exclusively on statistical measures of split worth on the training data. It is likely that the maximal will fail to generalize well on an independent set of validation data. The second part of the Tree algorithm, called *pruning*, attempts to improve generalization by determining the optimal tree, i.e. the one which maximizes profits (EViews Tutorial).

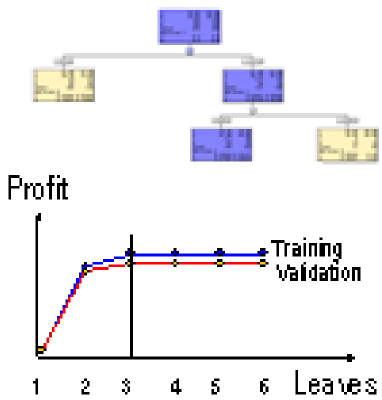

Figure 2. Selecting optimal Tree
Source: EViews Tutorial

Figure 3 shows an example of bi-dimensinal partitioning of input space.

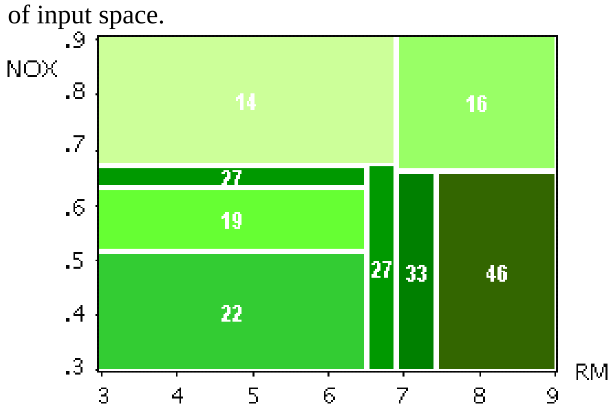

Figure 3. Bi-dimensional segmentation of input space

As figure 3 shows, leafs or terminal nodes are rectangular areas based on segmentation rules. They group observaton with same estimated value of target variable. In the case where target is a binary variable, leafs group observations with same estimated probability of target variable. An example is given in Figure 4.

Leaf	Pr(1|x)	Pr(7|x)	Pr(9|x)	Decision
1	.03	.96	.01	7
2	.09	.91	.00	7
3	.56	.44	.00	1
4	.95	.05	.00	1
5	.80	.10	.10	1
6	.64	.09	.27	1
7	.00	.13	.87	9
8	.10	.73	.17	7
9	.78	.01	.21	1
10	.01	.00	.99	9

Figure 4. Leafs segmentation

The second class of parametric models is given by parametric prediction models. They do assume a functional relationship between target variable and independent variable. For example a standard logistic regression model assumes the logit (p) is a linear combination of the inputs.

$$\log \frac{p}{1-p} = \beta_0 + \beta x' + \varepsilon$$

This is a non-linear model. Non-linearity concerns coefficients not variables, as basic econometrics textbooks say (Wooldridge, 1999). Consequently it might include squared or interaction terms as is often the case. One can see that in this case the probability of the event of interest is given by:

$$p = \frac{e^{\beta_0 + \beta x' + \varepsilon}}{1 + e^{\beta_0 + \beta x' + \varepsilon}}$$

The main advantage of logit modelling is that the estimated probabilities are constrained to take values between zero and one due to logistic function (Figure 5).

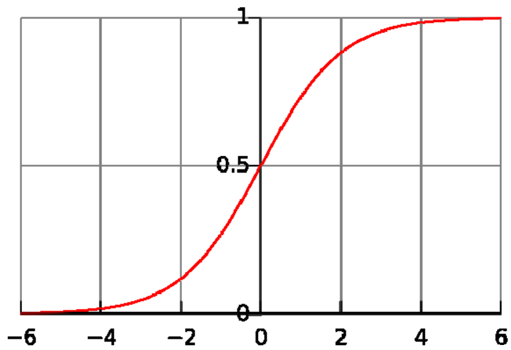

Figure. 5 Logistic Function

Neural models (multi-layer perceptrons) are in fact a natural extension of a regression model. A neural network can be thought of as a generalized linear model on a set of derived inputs. These derived inputs are themselves a generalized linear model on the original inputs. The usual link for the derived input's model is inverse hyperbolic tangent, a shift and rescaling of the logit function (EViews Tutorial). Same tutorial shows that neural networks have the ability to approximate virtually any continuous association between the inputs and the target by simply specifying the correct number of derived inputs.

Multi-layer perceptron models were originally inspired by neurophysiology and the interconnections between neurons. The basic model form arranges neurons in layers. The first layer, called the input layer connects to a layer of neurons called a hidden layer, which, in turn, connects to a final layer called the target, or output, layer. The structure of a multi-layer perceptron lends itself to a graphically representation called a network diagram. Each element in the diagram has a counterpart in the network equation. The model parameters are given random initial values, and predictions of the target are computed. These predictions are compared to the actual values of the target via an objective function which minimizes the difference between the actual and predicted values of the target. Training proceeds by updating the parameter estimates in a manner that decreases the value of the objective function. Training concludes when small changes in the parameter values no longer decrease the value of the objective function. The network is said to have reached a local minimum in the objective. At convergence, however, the model is likely to be highly over generalized. The overall average profit is examined to compensate for overgeneralization. The final parameter estimates for the model are taken from the iteration with the maximum validation profit (EViews Tutorial). Note that the coefficients of a nonlinear model are not directly interpretable (Wooldridge, 1999) not to mention that in the case of a neural network to interpret them is futile.

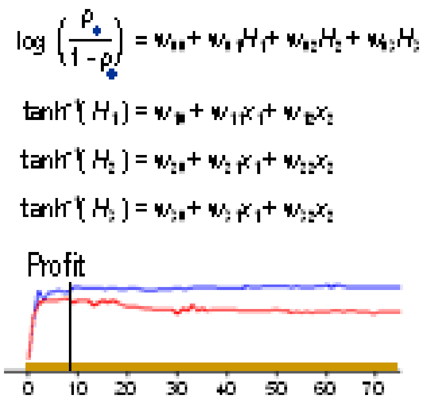

Figure 6. Neural Network Final Model

III. CONCLUSIONS

We have examined several claims about forecasting found in Romanian management literature. We have present argument that dismiss strategic forecasting and time series analysis in day to day business. As shown, forecasting is above all operational. Then we have presented the characteristics of main forecasting methods. Decision Trees are the most prevalent forecasting method. They tend to find complex relationships between target and inputs which can easily lead to over generalization. Decision trees are also used as a preliminary analysis for more sophisticated parametric estimations. Among parametric models logistic regressions are most used for modeling probabilities although neural network offer an exotic alternative, which can approximate virtually any type of relation between target and inputs, at the cost of employing increasingly complex multilayer hidden structures. We have also shown that forecasting does not stand for itself. The final forecasting algorithm of any forecasting method employs decision theory to determine the maximum profit or loss involved. Once again, the fact that firms maximize profits not utilities might come as a surprise for Romanian management literature which agrees the other way.

REFERENCES

[1] Bernstein, P., „*Against the Gods. The remarkable story of risk*”, John Wiley & Sons, Inc., 1998;

[2] Hand D., Henley W., “*Statistical Classification Methods in Consumer Credit Scoring: A Review*”, Journal of the Royal Statistical Society, 1997;

[3] Niu Jack, 2004. “*Managing Risks in Consumer Credit Industry*”, Policy Conference on Chinese Consumer Credit;

[4] Par Rud, O., „Data mining cookbook set. Modeling data for marketing, risk and customer relationship management”, John Wiley & Sons, Inc., 2001;

[5] Wiginton John, 1980. “*A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior*”, The Journal of Financial and Quantitative analysis;

[6] Wooldridge, J., „*Introductory Econometrics*”, South Western College Publishing, 1999;

[7] EViews 5 Tutorials, 2004