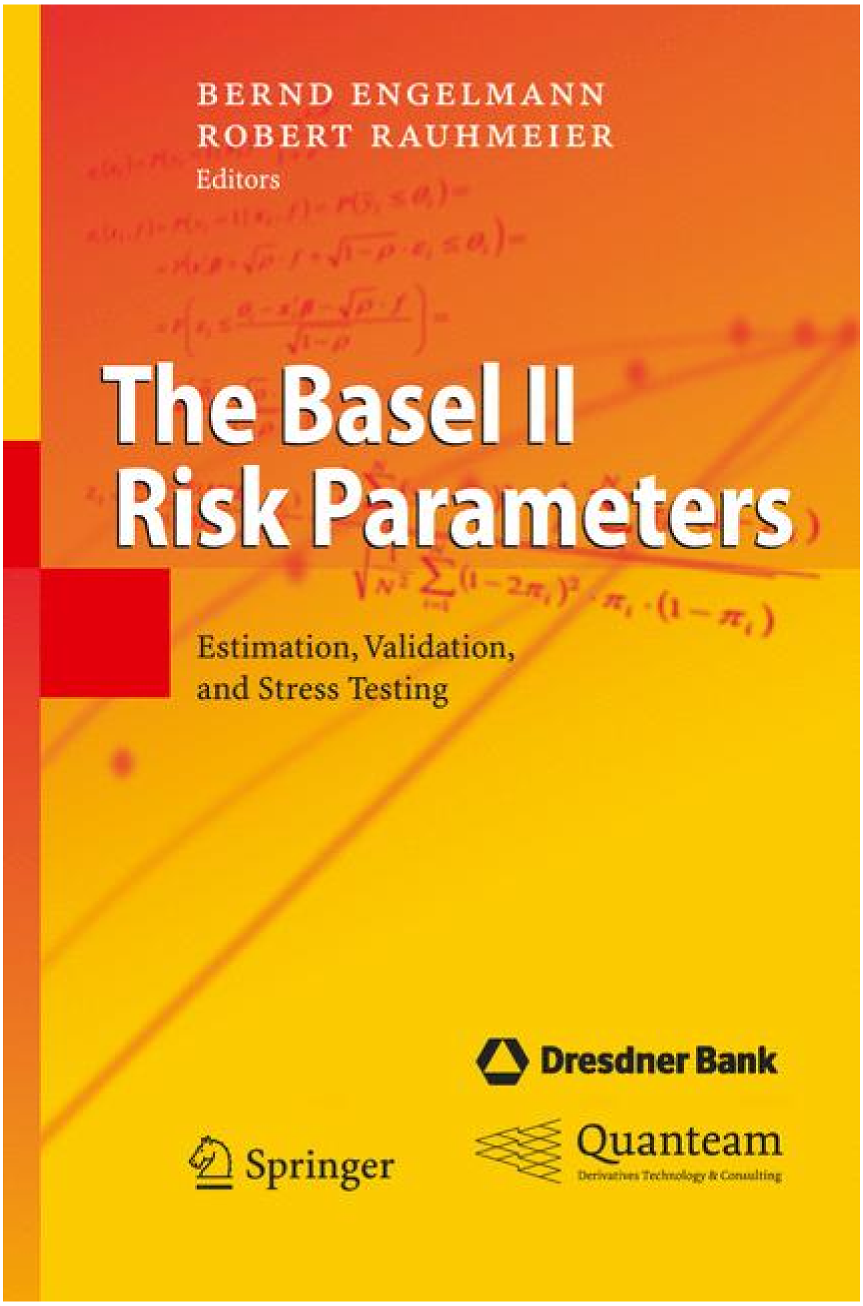
Statistical Methods to Develop Rating Models

[Evelyn Hayden and Daniel Porath,
Österreichische Nationalbank and University of Applied Sciences at Mainz]

Source:

The Basel II Risk Parameters – Estimation, Validation, and Stress Testing
Bernd Engelmann, Robert Rauhmeier, Robert (Editors), 376 p., Springer Verlag, Berlin 2006.



I. Statistical Methods to Develop Rating Models

Evelyn Hayden and Daniel Porath

Österreichische Nationalbank[1] and University of Applied Sciences at Mainz

1. Introduction

The Internal Rating Based Approach (IRBA) of the New Basel Capital Accord allows banks to use their own rating models for the estimation of probabilities of default (PD) as long as the systems meet specified minimum requirements. Statistical theory offers a variety of methods for building and estimation rating models. This chapter gives an overview of these methods. The overview is focused on statistical methods and includes parametric models like linear regression analysis, discriminant analysis, binary response analysis, time-discrete panel methods, hazard models and nonparametric models like neural networks and decision trees. We also highlight the benefits and the drawbacks of the various approaches. We conclude by interpreting the models in light of the minimum requirements of the IRBA.

2. Statistical Methods for Risk Classification

In the following we define statistical models as the class of approach which uses econometric methods to classify borrowers according to their risk. Statistical rating systems primarily involve a search for explanatory variables which provide as sound and reliable a forecast of the deterioration of a borrower's situation as possible. In contrast, structural models explain the threats to a borrower based on an economic model and thus use clear causal connections instead of the mere correlation of variables.

The following sections offer an overview of parametric and nonparametric models generally considered for statistical risk assessment. Furthermore, we discuss the advantages and disadvantages of each approach. Many of the methods are described in more detail in standard econometric textbooks, like Greene (2003).

[1] The opinions expressed in this chapter are those of the author and do not necessarily reflect views of the Österreichische Nationalbank.

In general, a statistical model may be described as follows: As a starting point, every statistical model uses the borrower's characteristic indicators and (possibly) macroeconomic variables which were collected historically and are available for defaulting (or troubled) and non-defaulting borrowers. Let the borrower's characteristics be defined by a vector of n separate variables (also called covariates) $x = x_1,...,x_n$ observed at time $t - L$. The state of default is indicated by a binary performance variable y observed at time t. The variable y is defined as $y = 1$ for a default and $y = 0$ for a non-default.

The sample of borrowers now includes a number of individuals or firms that defaulted in the past, while (typically) the majority did not default. Depending on the statistical application of this data, a variety of methods can be used to predict the performance. A common feature of the methods is that they estimate the correlation between the borrowers' characteristics and the state of default in the past and use this information to build a forecasting model. The forecasting model is designed to assess the creditworthiness of borrowers with unknown performance. This can be done by inputting the characteristics x into the model. The output of the model is the estimated performance. The time lag L between x and y determines the forecast horizon.

3. Regression Analysis

As a starting point we consider the classical regression model. The regression model establishes a linear relationship between the borrowers' characteristics and the default variable:

$$y_i = \boldsymbol{\beta}' \cdot \boldsymbol{x}_i + u_i \tag{1}$$

Again, y_i indicates whether borrower i has defaulted ($y_i = 1$) or not ($y_i = 0$). In period t, $\boldsymbol{x}_i$ is a column vector of the borrowers' characteristics observed in period $t - L$ and $\boldsymbol{\beta}$ is a column vector of parameters which capture the impact of a change in the characteristics on the default variable. Finally, u_i is the residual variable which contains the variation not captured by the characteristics $\boldsymbol{x}_i$.

The standard procedure is to estimate (1) with the ordinary least squares (OLS) estimators of $\boldsymbol{\beta}$ which in the following are denoted by $\boldsymbol{b}$. The estimated result is the borrower's score S_i. This can be calculated by

$$S_i = E(y_i \mid \boldsymbol{x}_i) = \boldsymbol{b}' \cdot \boldsymbol{x}_i \,. \tag{2}$$

Equation (2) shows that a borrower's score represents the expected value of the performance variable when his or her individual characteristics are known. The score can be calculated by inputting the values for the borrower's characteristics into the linear function given in (2).

Note that S_i is continuous (while y_i is a binary variable), hence the output of the model will generally be different from 0 or 1. In addition, the prediction can take on values larger than 1 or smaller than 0. As a consequence, the outcome of the model cannot be interpreted as a probability level. However, the score S_i, can be used for the purpose of comparison between different borrowers, where higher values of S_i correlate with a higher default risk.

The benefits and drawbacks from model (1) and (2) are the following:

- OLS estimators are well-known and easily available.
- The forecasting model is a linear model and therefore easy to compute and to understand.
- The random variable u_i is heteroscedastic (i.e. the variance of u_i is not constant for all i) since

  $$Var(u_i) = Var(y_i) = E(y_i \mid \boldsymbol{x}_i) \cdot [1 - E(y_i \mid \boldsymbol{x}_i)] = \boldsymbol{b}' \cdot \boldsymbol{x}_i (1 - \boldsymbol{b}' \cdot \boldsymbol{x}_i). \quad (3)$$

  As a consequence, the estimation of $\boldsymbol{\beta}$ is inefficient and additionally, the standard errors of the estimated coefficients $\boldsymbol{b}$ are biased. An efficient way to estimate $\boldsymbol{\beta}$ is to apply the Weighted Least Squares (WLS) estimator.
- WLS estimation of $\boldsymbol{\beta}$ is efficient, but the estimation of the standard errors of $\boldsymbol{b}$ still remains biased. This happens due to the fact that the residuals are not normally distributed as they can only take on the values $\boldsymbol{b}'\boldsymbol{x}_i$ (if the borrower does not default and y therefore equals 0) or $(1 - \boldsymbol{b}'\boldsymbol{x}_i)$ (if the borrower does default and y therefore equals 1). This implies that there is no reliable way to assess the significance of the coefficients $\boldsymbol{b}$ and it remains unknown whether the estimated values represent precise estimations of significant relationships or whether they are just caused by spurious correlations. Inputting characteristics which are not significant into the model can seriously harm the model's stability when used to predict borrowers' risk for new data. A way to cope with this problem is to split the sample into two parts, where one part (the training sample) is used to estimate the model and the other part (the hold-out sample) is used to validate the results. The consistency of the results of both samples is then taken as an indicator for the stability of the model.
- The absolute value of S_i cannot be interpreted.

4. Discriminant Analysis

Discriminant analysis is a classification technique applied to corporate bankruptcies by Altman as early as 1968 (see Altman, 1968). Linear discriminant analysis is based on the estimation of a linear discriminant function with the task of separating individual groups (in this case of defaulting and non-defaulting borrowers) according to specific characteristics. The discriminant function is

$$S_i = \boldsymbol{\beta}' \cdot \boldsymbol{x}_i . \quad (4)$$

The Score S_i is also called the discriminant variable. The estimation of the discriminant function adheres to the following principle:

> *Maximization of the spread between the groups (good and bad borrowers) and minimization of the spread within individual groups*

Maximization only determines the optimal proportions among the coefficients of the vector $\boldsymbol{\beta}$. Usually (but arbitrarily), coefficients are normalized by choosing the pooled within-group variance to take the value 1. As a consequence, the absolute level of S_i is arbitrary as well and cannot be interpreted on a stand-alone basis. As in linear regression analysis, S_i can only be used to compare the prediction for different borrowers ("higher score, higher risk").

Discriminant analysis is similar to the linear regression model given in equations (1) and (2). In fact, the proportions among the coefficients of the regression model are equal to the optimal proportion according to the discriminant analysis. The difference between the two methods is a theoretical one: Whereas in the regression model the characteristics are deterministic and the default state is the realization of a random variable, for discriminant analysis the opposite is true. Here the groups (default or non-default) are deterministic and the characteristics of the discriminant function are realizations from a random variable. For practical use this difference is virtually irrelevant.

Therefore, the benefits and drawbacks of discriminant analysis are similar to those of the regression model:

- Discriminant analysis is a widely known method with estimation algorithms that are easily available.
- Once the coefficients are estimated, the scores can be calculated in a straightforward way with a linear function.
- Since the characteristics x_i are assumed to be realizations of random variables, the statistical tests for the significance of the model and the coefficients rely on the assumption of multivariate normality. This is, however, unrealistic for the variables typically used in rating models as for example financial ratios from the balance-sheet. Hence, the methods for analyzing the stability of the model and the plausibility of the coefficients are limited to a comparison between training and hold-out sample.
- The absolute value of the discriminant function cannot be interpreted in levels.

5. Logit and Probit Models

Logit and probit models are econometric techniques designed for analyzing binary dependent variables. There are two alternative theoretical foundations.

The latent-variable approach assumes an unobservable (latent) variable y^* which is related to the borrower's characteristics in the following way:

$$y_i^* = \boldsymbol{\beta}' \cdot \boldsymbol{x}_i + u_i \tag{5}$$

Here $\boldsymbol{\beta}$, $\boldsymbol{x}_i$ and u_i are defined as above. The variable y_i* is metrically scaled and triggers the value of the binary default variable y_i:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

This means that the default event sets in when the latent variable exceeds the threshold zero. Therefore, the probability for the occurrence of the default event equals:

$$P(y_i = 1) = P(u_i > -\boldsymbol{\beta}' \cdot \boldsymbol{x}_i) = 1 - F(-\boldsymbol{\beta}' \cdot \boldsymbol{x}_i) = F(\boldsymbol{\beta}' \cdot \boldsymbol{x}_i). \tag{7}$$

Here $F(.)$ denotes the (unknown) distribution function. The last step in (7) assumes that the distribution function has a symmetric density around zero. The choice of the distribution function $F(.)$ depends on the distributional assumptions about the residuals (u_i). If a normal distribution is assumed, we are faced with the probit model:

$$\mathrm{F}(\boldsymbol{\beta}' \cdot \boldsymbol{x}_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\boldsymbol{\beta}' \cdot \boldsymbol{x}_i} \mathrm{e}^{\frac{-\mathrm{t}^2}{2}} \mathrm{dt} \tag{8}$$

If instead the residuals are assumed to follow a logistic distribution, the result is the logit model:

$$\mathrm{F}(\boldsymbol{\beta}' \cdot \boldsymbol{x}_i) = \frac{\mathrm{e}^{\boldsymbol{\beta}' \cdot \boldsymbol{x}_i}}{1 + \mathrm{e}^{\boldsymbol{\beta}' \cdot \boldsymbol{x}_i}} \tag{9}$$

The second way to motivate logit and probit models starts from the aim of estimating default probabilities. For single borrowers, default probabilities cannot be observed as realizations of default probabilities. However, for groups of borrowers the observed default frequencies can be interpreted as default probabilities. As a starting point consider the OLS estimation of the following regression:

$$p_i = \boldsymbol{b}' \cdot \boldsymbol{x}_i + u_i \tag{10}$$

In (10) the index i denotes the group formed by a number of individuals, p_i is the default frequency observed in group i and $\boldsymbol{x}_i$ are the characteristics observed for group i. The model, however, is inadequate. To see this consider that the outcome (which is $E(y_i|\boldsymbol{x}_i) = \boldsymbol{b}'\boldsymbol{x}_i$) is not bounded to values between zero and one and therefore cannot be interpreted as a probability. As it is generally implausible to assume that a probability can be calculated by a linear function, in a second step the linear expression $\boldsymbol{b}'\boldsymbol{x}_\mathrm{i}$ is transformed by a nonlinear function (link function) F:

$$p_i = F(\boldsymbol{b'} \cdot \boldsymbol{x}_i). \tag{11}$$

An appropriate link function transforms the values of $\boldsymbol{b'x}_i$ to a scale within the interval [0,1]. This can be achieved by any distribution function. The choice of the link function determines the type of model: with a logistic link function equation (11) becomes a logit model, while with the normal distribution (11) results in the probit model.

However, when estimating (10) with OLS, the coefficients will be heteroscedastic, because $Var(u_i) = Var(p_i) = p(\boldsymbol{x}_i)\cdot(1\text{-}p(\boldsymbol{x}_i))$. A possible way to achieve homoscedasticity would be to compute the WLS estimators of $\boldsymbol{b}$ in (10). However, albeit possible, this is not common practice. The reason is that in order to observe default frequencies, the data has to be grouped before estimation. Grouping involves considerable practical problems like defining the size and number of the groups and the treatment of different covariates within the single groups. A better way to estimate logit and probit models, which does not require grouping, is the Maximum-Likelihood (ML) method. For a binary dependent variable the likelihood function looks like:

$$L(\boldsymbol{b}) = \prod_i P(\boldsymbol{b'} \cdot \boldsymbol{x}_i)^{y_i} \left[1 - P(\boldsymbol{b'} \cdot \boldsymbol{x}_i)^{1-y_i}\right]. \tag{12}$$

For the probit model $P(.)$ is the normal density function and for the logit model $P(.)$ is the logistic density function. With equation (12) the estimation of the model is theoretically convincing and also easy to handle. Furthermore, the ML-approach lends itself for a broad set of tests to evaluate the model and its single variables (see Hosmer and Lemeshow (2000) for a comprehensive introduction).

Usually, the choice of the link function is not theoretically driven. Users familiar with the normal distribution will opt for the probit model. Indeed, the differences in the results of both classes of models are often negligible. This is due to the fact that both distribution functions have a similar form except for the tails, which are heavier for the logit model. The logit model is easier to handle, though. First of all, the computation of the estimators is easier. However, today computational complexity is often irrelevant as most users apply statistical software where the estimation algorithms are integrated. What is more important is the fact that the coefficients of the logit model can be more easily interpreted. To see this we transform the logit model given in (9) in the following way:

$$\frac{P_i}{1 - P_i} = e^{\boldsymbol{\beta'} \cdot \boldsymbol{x}_i} \tag{13}$$

The left-hand side of (13) are the odds, i.e. the relation between the default probability and the probability of survival. Now it can be easily seen that a variation of a single variable x_k of one unit has an impact of e^{β_k} on the odds, when β_k denotes the coefficient of the variable x_k. Hence, the transformed coefficients $e^{\boldsymbol{\beta}}$ are called odds-ratios. They represent the multiplicative impact of a borrower's characteristic on the odds. Therefore, for the logit model, the coefficients can be interpreted in a plausible way, which is not possible for the probit model. Indeed, the most important weakness of binary models is the fact that the interpretation of the coefficients is not straightforward.

The strengths of logit and probit models can be summarized as:

- The methods are theoretically sound
- The results generated can be interpreted directly as default probabilities
- The significance of the model and the individual coefficients can be tested. Therefore, the stability of the model can be assessed more effectively than in the previous cases.

6. Panel Models

The methods discussed so far are all cross-sectional methods because all covariates are related to the same period. However, typically banks dispose of a set of covariates for more than one period for each borrower. In this case it is possible to expand the cross-sectional input data to a panel dataset. The main motivation is to enlarge the number of available observations for the estimation and therefore enhance the stability and the precision of the rating model. Additionally, panel models can integrate macroeconomic variables into the model. Macroeconomic variables can improve the model for several reasons. First, many macroeconomic data sources are more up-to-date than the borrowers' characteristics. For example, financial ratios calculated from balance sheet information are usually updated only once a year and are often up to two years old when used for risk assessment. The oil price, instead, is available on a daily frequency. Secondly, by stressing the macroeconomic input factors, the model can be used for a form of stress-testing credit risk. However, as macroeconomic variables primarily affect the absolute value of the default probability, it is only reasonable to incorporate macroeconomic input factors into those classes of models that estimate default probabilities.

In principle, the structure of, for example, a panel logit or probit model remains the same as given in the equations of the previous section. The only difference is that now the covariates are taken from a panel of data and have to be indexed by an additional time series indicator, i.e. we observe $\boldsymbol{x}_{it}$ instead of $\boldsymbol{x}_i$. At first glance panel models seem similar to cross-sectional models. In fact, many developers ignore the dynamic pattern of the covariates and simply fit logit or probit models. However, logit and probit models rely on the assumption of independent observations. Generally, cross-sectional data meets this requirement, but panel data does not. The reason is that observations from the same period and observations from the same borrower should be correlated. Introducing this correlation in the estimation procedure is cumbersome. For example, the fixed-effects estimator known from panel analysis for continuous dependent variables is not available for the probit model. Besides, the modified fixed-effects estimator for logit models proposed by Chamberlain (1980) excludes all non-defaulting borrowers from the analysis and therefore seems inappropriate. Finally, the random-effects estimators proposed in the literature are computationally extensive and can only be computed with specialized software. For an econometric discussion of binary panel analysis, refer to Hosmer and Lemeshow (2000).

7. Hazard Models

All methods discussed so far try to assess the riskiness of borrowers by estimating a certain type of score that indicates whether or not a borrower is likely to default within the specified forecast horizon. However, no prediction about the exact default point in time is made. Besides, these approaches do not allow the evaluation of the borrowers' risk for future time periods given they should not default within the reference time horizon.

These disadvantages can be remedied by means of hazard models, which explicitly take the survival function and thus the time at which a borrower's default occurs into account. Within this class of models, the Cox proportional hazard model (cf. Cox, 1972) is the most general regression model, as it is not based on any assumptions concerning the nature or shape of the underlying survival distribution. The model assumes that the underlying hazard rate (rather than survival time) is a function of the independent variables; no assumptions are made about the nature or shape of the hazard function. Thus, the Cox's regression model is a semi-parametric model. The model can be written as:

$$h_i\left(t \mid \boldsymbol{x}_i\right) = h_0\left(t\right) \cdot e^{\boldsymbol{\beta}' \cdot \boldsymbol{x}_i}, \tag{14}$$

where $h_i(t|\boldsymbol{x}_i)$ denotes the resultant hazard, given the covariates for the respective borrower and the respective survival time t. The term $h_0(t)$ is called the baseline hazard; it is the hazard when all independent variable values are equal to zero. If the covariates are measured as deviations from their respective means, $h_0(t)$ can be interpreted as the hazard rate of the average borrower.

While no assumptions are made about the underlying hazard function, the model equation shown above implies important assumptions. First, it specifies a multiplicative relationship between the hazard function and the log-linear function of the explanatory variables, which implies that the ratio of the hazards of two borrowers does not depend on time, i.e. the relative riskiness of the borrowers is constant, hence the name Cox *proportional* hazard model.

Besides, the model assumes that the default point in time is a continuous random variable. However, often the borrowers' financial conditions are not observed continuously but rather at discrete points in time. What's more, the covariates are treated as if they were constant over time, while typical explanatory variables like financial ratios change with time.

Although there are some advanced models to incorporate the above mentioned features, the estimation of these models becomes complex. The strengths and weaknesses of hazard models can be summarized as follows:

- Hazard models allow for the estimation of a survival function for all borrowers from the time structure of historical defaults, which implies that default probabilities can be calculated for different time horizons.
- Estimating these models under realistic assumptions is not straightforward.

8. Neural Networks

In recent years, neural networks have been discussed extensively as an alternative to the (parametric) models discussed above. They offer a more flexible design to represent the connections between independent and dependent variables. Neural networks belong to the class of non-parametrical methods. Unlike the methods discussed so far they do not estimate parameters of a well-specified model. Instead, they are inspired by the way biological nervous systems, such as the brain, process information. They typically consist of many nodes that send a certain output if they receive a specific input from the other nodes to which they are connected. Like parametric models, neural networks are trained by a training sample to classify borrowers correctly. The final network is found by adjusting the connections between the input, output and any potential intermediary nodes.

The strengths and weaknesses of neural networks can be summarized as:

- Neural networks easily model highly complex, nonlinear relationships between the input and the output variables.
- They are free from any distributional assumptions.
- These models can be quickly adapted to new information (depending on the training algorithm).
- There is no formal procedure to determine the optimum network topology for a specific problem, i.e. the number of the layers of nodes connecting the input with the output variables.
- Neural networks are black boxes, hence they are difficult to interpret.
- Calculating default probabilities is possible only to a limited extent and with considerable extra effort.

In summary, neural networks are particularly suitable when there are no expectations (based on experience or theoretical arguments) on the relationship between the input factors and the default event and the economic interpretation of the resulting models is of inferior importance.

9. Decision Trees

A further category of non-parametric methods comprises decision trees, also called classification trees. Trees are models which consist of a set of if-then split conditions for classifying cases into two (or more) different groups. Under these methods, the base sample is subdivided into groups according to the covariates. In the case of binary classification trees, for example, each tree node is assigned by (usually univariate) decision rules, which describe the sample accordingly and subdivide it into two subgroups each. New observations are processed down the tree in accordance with the decision rules' values until the end node is reached, which then represents the classification of this observation. An example is given in Figure 1.

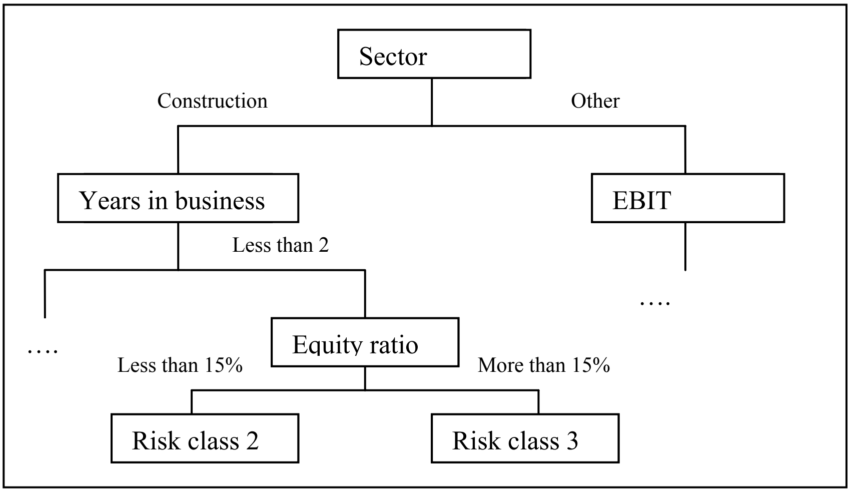


Figure 1. Decision Tree

One of the most striking differences of the parametric models is that all covariates are grouped and treated as categorical variables. Furthermore, whether a specific variable or category becomes relevant depends on the categories of the variables in the upper level. For example, in Figure 1 the variable “years in business” is only relevant for companies which operate in the construction sector. This kind of dependence between variables is called interaction.

The most important algorithms for building decision trees are the Classification and Regression Trees algorithms (C&RT) popularized by Breiman et al. (1984) and the CHAID algorithm (Chi-square Automatic Interaction Detector, see Kass, 1978). Both algorithms use different criteria to identify the best splits in the data and to collapse the categories which are not significantly different in outcome.

The general strengths and weaknesses of trees are:

- Through categorization, nonlinear relationships between the variables and the score can be easily modelled.
- Interactions present in the data can be identified. Parametric methods can model interactions only to a limited extent (by introducing dummy variables).
- As with neural networks, decision trees are free from distributional assumptions.
- The output is easy to understand.
- Probabilities of default have to be calculated in a separate step.
- The output is (a few) risk categories and not a continuous score variable. Consequently, decision trees only calculate default probabilities for the final node in a tree, but not for individual borrowers.
- Compared to other models, trees contain fewer variables and categories. The reason is that in each node the sample is successively partitioned and therefore continuously diminishes.
- The stability of the model cannot be assessed with statistical procedures. The strategy is to work with a training sample and a hold-out sample.

In summary, trees are particularly suited when the data is characterized by a limited number of predictive variables which are known to be interactive.

10. Statistical Models and Basel II

Finally, we ask the question whether the models discussed in this chapter are in line with the IRB Approach of Basel II. Prior to the discussion, it should be mentioned that in the Basel documents, rating systems are defined in a broader sense than in this chapter. Following § 394 of the Revised Framework from June 2004 (cf. BIS, 2004) a rating system "comprises all the methods, processes, controls, and data collection and IT systems that support the assessment of credit risk, the assignment of internal ratings, and the quantification of default and loss estimates". Compared to this definition, these methods provide one component, namely the assignment of internal ratings.

The minimum requirements for internal rating systems are treated in part II, section III, H of the Revised Framework. A few passages of the text concern the assignment of internal ratings, and the requirements are general. They mainly concern the rating structure and the input data, examples being:

- a minimum of 7 rating classes of non-defaulted borrowers (§ 404)
- no undue or excessive concentrations in single rating classes (§§ 403, 406)
- a meaningful differentiation of risk between the classes (§ 410)
- plausible, intuitive and current input data (§§ 410, 411)
- all relevant information must be taken into account (§ 411).

The requirements do not reveal any preference for a certain method. It is indeed one of the central ideas of the IRBA that the banks are free in the choice of the method. Therefore the models discussed here are all possible candidates for the IRB Approach.

The strengths and weaknesses of the single methods concern some of the minimum requirements. For example, hazard rate or logit panel models are especially suited for stress testing (as required by §§ 434, 345) since they contain a time-series dimension. Methods which allow for the statistical testing of the individual input factors (e.g. the logit model) provide a straightforward way to demonstrate the plausibility of the input factors (as required by § 410). When the outcome of the model is a continuous variable, the rating classes can be defined in a more flexible way (§§ 403, 404, 406).

On the other hand, none of the drawbacks of the models considered here excludes a specific method. For example, a bank may have a preference for linear regression analysis. In this case the plausibility of the input factors cannot be verified by statistical tests and as a consequence the bank will have to search for alternative ways to meet the requirements of § 410.

In summary, the minimum requirements are not intended as a guideline for the choice of a specific model. Banks should rather base their choice on their internal aims and restrictions. If necessary, those components that are only needed for the purpose to satisfy the criteria of the IRBA should be added in a second step. All models discussed in this chapter allow for this.

References

Altman EI (1968), Financial Indicators, Discriminant Analysis, and the Prediction of Corporate Bankruptcy, Journal of Finance.

BIS (2004), International Convergence of Capital Measurement and Capital Standards, Basel Committee on Banking Supervision, June 2004.

Breiman L, Friedman JH, Olshen RA, Stone SJ (1984), Classification and Regression Trees, Wadsworth.

Chamberlain G (1980), Analysis of Covariance with Qualitative Data, Review of Economic Studies 47, 225-238.

Cox DR (1972), Regression Models and Life Tables (with Discussion), Journal of Royal Statistical Society, Series B.

Greene W (2003), Econometric Analysis, 5th ed., Prentice-Hall New Jersey.

Hosmer W, Lemeshow S (2000), Applied Logistic Regression, New York, Wiley.

Kass GV (1978), An Exploratory Technique for Investigating Large Quantities of Categorical Data, Applied Statistics 29 (2), pp. 119-127.