Chapter 5
Credit Scoring

Granting credit to both retail and nonretail (e.g., corporate) customers is the core business of a bank. In doing so, banks need to have adequate systems to decide to whom to grant credit. Credit scoring is a key risk assessment technique to analyze and quantify a potential obligor's credit risk. Essentially, credit scoring aims at quantifying the likelihood that an obligor will repay the debt. The outcome of the credit scoring exercise is a score reflecting the creditworthiness of the obligor.

In this chapter, you will learn the fundamentals of credit scoring. We start by introducing the basic idea of credit scoring. We then outline the differences between judgmental and statistical scoring, and discuss the advantages of the latter. Next, we zoom in on credit scoring for both retail and nonretail exposures. This is followed by a discussion of the potential of big data for credit scoring. Another section explains overrides whereby the result of the scorecard is overruled by the credit expert. Various criteria to evaluate scorecard performance are covered. We also review key business applications of credit scores. The chapter concludes by discussing the limitations of credit scoring.

BASIC CONCEPTS

Let's start by defining credit scoring. Throughout the past few decades banks have gathered plenty of information describing the default behavior of their customers. Examples are historical information about a customer's date of birth, gender, income, employment status, and so on. All this data has been nicely stored into huge (e.g., relational) databases or data warehouses. On top of this, banks have accumulated lots of business experience about their credit products. As an example, many credit experts do a pretty good job of discriminating between low-risk and high-risk mortgages using their business expertise only. It is now the aim of credit scoring to analyze both sources of data in more detail and come up with a statistically based decision model that allows scoring future credit applications and ultimately deciding which ones to accept and which to reject.

A key assumption made when building a credit scoring model (and all other credit risk models introduced in later chapters) is that the future resembles the past. By analyzing past repayment behavior of previous customers, it becomes possible to learn how future customers will behave in terms of default risk. More specifically, for the historical customers, we know which ones turned out to be good payers and which ones turned out to be bad payers. This good/bad status is now the binary target variable Y, which we will relate to all information available at scoring time about our obligors. The goal of credit scoring is now to quantify this relationship as precisely as possible to assist credit decisions, monitoring, and management. Banks score borrowers at loan application, as well as at regular times during the term of a financial

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

contract (generally loans, loan commitments, and guarantees).

Once we have our credit scoring model built, we can then use it to decide whether the credit application should be accepted or rejected, or to derive the probability of a future default. To summarize, credit scoring is a key risk management tool for a bank to optimally manage, understand, and model the credit risk it is exposed to.

JUDGMENTAL VERSUS STATISTICAL SCORING

There are basically two main approaches to assessing credit risk: the judgmental approach and the statistical approach. Both rely on historical information, but the type of information they use is different.

The judgmental approach is a qualitative, expert-based approach whereby, based on business experience and common sense, the credit expert or credit committee, which is a group of credit experts, will make a decision about the credit risk. Usually, this is done based on inspecting the five Cs of the applicant and loan:

- **Character** measures the borrower's character and integrity (e.g., reputation, honesty, etc.).
- **Capital** measures the difference between the borrower's assets (e.g., car, house, etc.) and liabilities (e.g., renting expenses, etc.).
- **Collateral** measures the collateral provided in case payment problems occur (e.g., house, car, etc.).
- **Capacity** measures the borrower's ability to pay (e.g., job status, income, etc.).
- **Condition** measures the borrower's circumstances (e.g., market conditions, competitive pressure, seasonal character, etc.).

In analyzing this information, a qualitative or subjective evaluation of the credit risk is made. Although the judgmental approach might seem subjective and thus unsophisticated at first sight, it is still quite commonly used by banks for very specific credit portfolios such as project finance or new credit products.

With the emergence of statistical classification techniques at the beginning of the 1980s, banks became more and more interested in abandoning the judgmental approach and opting for a more formal data-based statistical approach.

The statistical approach is based on statistical analysis of historical data to find the optimal multivariate relationship between a customer's characteristics and the binary good/bad target variable (Baesens et al. 2003). It is less subjective than the judgmental approach since it is not tied to a particular credit expert's background knowledge and experience.

The statistical approach aims at building scorecards, which are based on multivariate correlations between inputs (such as age, marital status, income, savings amount) and a target variable that reflects the risk of default. In other words, a scorecard will assign scores to each of those inputs. In our example, scores will be assigned to age, marital status, income, and

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

savings amount. All those scores will then be added up and compared with the critical threshold, which specifies the minimum level of required credit quality. If the aggregated score exceeds the threshold, then credit will be granted. If it falls below the threshold, then credit will be withheld.

In practice, hybrid approaches may be applied. In a first step, a bank may generate informational values by judgmental scoring. An example may be an expert opinion of a credit analyst on the payment ethics of a borrower (e.g., as a discrete number between 1 and 5). In a second step, the bank may aggregate this judgmental score and other hard information into a statistical score.

ADVANTAGES OF STATISTICAL CREDIT SCORING

Generally speaking, the statistical approach to credit scoring has many advantages compared with the judgmental approach. First, it is better in terms of speed and accuracy. We can now make faster decisions than we were able to do with the judgmental approach. This is especially relevant when working in an online environment where credit decisions need to be made quickly, possibly in real time. Because a credit scorecard is essentially a mathematical formula, it can be easily programmed and evaluated in an automated and fast way.

Another advantage of having statistical credit scoring models is consistency. We no longer have to rely upon the experience, intuition, or common sense of one or multiple business experts. Now it's just a mathematical formula, and the formula will always evaluate in exactly the same way if given the same set of inputs, like age, marital status, income, and so on.

Finally, statistical credit scoring models will typically also be more powerful than judgmental models. This performance boost will allow a reduction of bad debt loss and operating costs, and consequently it will also improve portfolio management.

To summarize, statistical credit scoring models have a lot of advantages compared with judgmental credit scoring models and are thus considered superior.

TECHNIQUES TO BUILD SCORECARDS

In this section, we discuss both logistic regression and decision trees, two classification techniques which are very powerful and popular to build application and/or behavioral scorecards.

Logistic Regression

Basic Model Formulation

Consider a credit scoring data set in panel form as depicted in Exhibit 5.1.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Customer	Age	Income	Employed	…	Default	D
ABC	30	2,000	Yes		No	0
BCD	62	4,600	Yes		No	0
CDE	42	3,200	No		Yes	1
…						
XYZ	56	3,800	No		Yes	1

Exhibit 5.1 Example Credit Scoring Data Set

When modeling the binary default target *D* using linear regression, we get:

$$D = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{ Income} + \beta_3 \text{ Employed}$$

When estimating this using ordinary least squares (OLS), two key problems arise:

1. The errors/target are not normally distributed but follow a Bernoulli distribution with only two outcomes.
2. There is no guarantee that the target is between 0 and 1; it would be handy if it were, because then it could be interpreted as a probability.

Consider now the following bounding function:

$$f(z) = \frac{1}{1 + e^{-z}}$$

which looks like Exhibit 5.2.

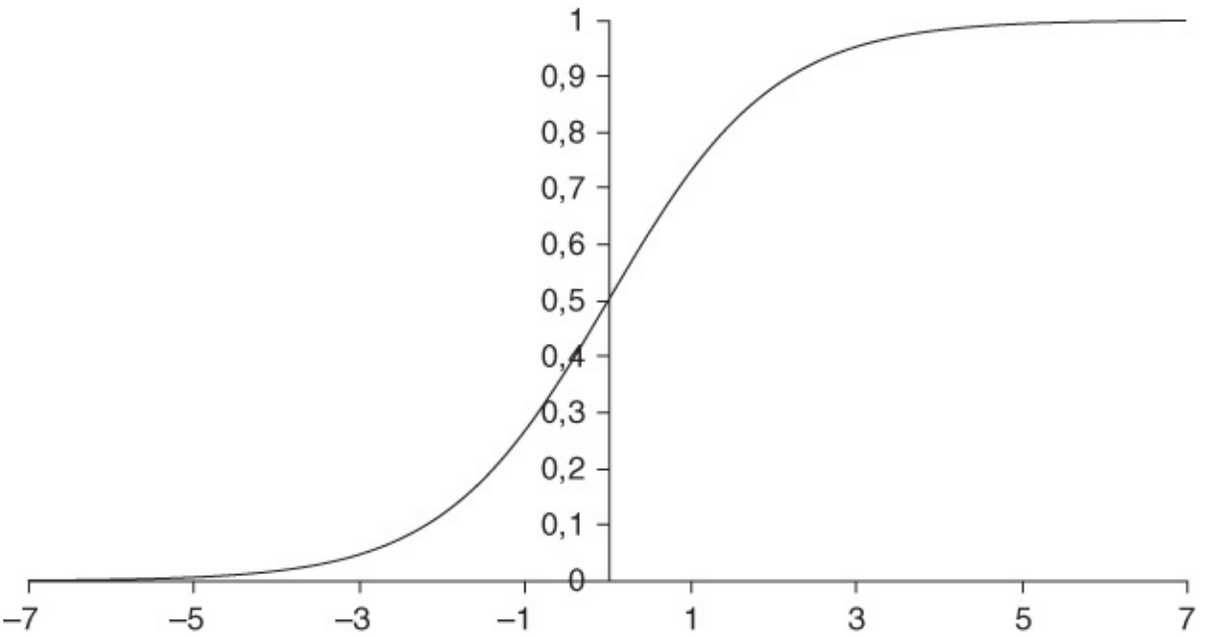

Exhibit 5.2 Bounding Function for Logistic Regression

For every possible value of *z*, the outcome is always between 0 and 1. By combining the linear

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

regression with the bounding function, we get the following logistic regression model:

$$P(D = 1|\text{Age, Income, Employed}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Employed})}}$$

The outcome of this model is always bounded between 0 and 1, no matter which values of age, income, and employed are being used, and can therefore be interpreted as a probability.

The general formulation of the logistic regression model then becomes (Allison 2001):

$$P(D = 1|x_1, \ldots, x_N) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_N x_N)}}$$

or alternatively,

$$\begin{aligned} P(D = 0|x, \ldots, x_N) &= 1 - P(D = 1|x_1, \ldots, x_N) \\ &= 1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_N x_N)}} = \frac{1}{1 + e^{(\beta_0 + \beta_1 x_1 + \ldots + \beta_N x_N)}} \end{aligned}$$

whereby D equals 1 in case of default, and 0 otherwise.

Hence, both $P(D = 1|x_1, \ldots, x_N)$ and $P(D = 0|x_1, \ldots, x_N)$ are bounded between 0 and 1.

Reformulating in terms of the odds, the model becomes:

$$\frac{P(D = 1|x_1, \ldots, x_N)}{P(D = 0|x_1, \ldots, x_N)} = e^{(\beta_0 + \beta_1 x_1 + \ldots + \beta_N x_N)}$$

or in terms of the log odds (logit),

$$\ln\left(\frac{P(D = 1|x_1, \ldots, x_N)}{P(D = 0|x_1, \ldots, x_N)}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_N x_N$$

The β_i parameters of a logistic regression model are then estimated using the idea of maximum likelihood. Maximum likelihood optimization chooses the parameters in such a way as to maximize the probability of getting the sample at hand. First, the likelihood function is constructed. For observation i, the probability of observing either class equals:

$$P(D = 1|x_{1i}, \ldots, x_{Ni})^{D_i}(1 - P(D = 1|x_{1i}, \ldots, x_{Ni})^{1-D_i}$$

whereby D_i represents the target value (either 0 or 1) for observation i. The likelihood function across all n observations then becomes:

$$\prod_{i=1}^{n} P(D = 1|x_{1i}, \ldots, x_{Ni})^{D_i}(1 - P(D = 1|x_{1i}, \ldots, x_{Ni})^{1-D_i}$$

To simplify the optimization, the logarithmic transformation of the likelihood function is taken and the corresponding log-likelihood can then be optimized using, for instance, the iteratively reweighted least squares procedure.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Logistic Regression Properties

Since logistic regression is linear in the log odds (logit), it basically estimates a linear decision boundary to separate both classes. This is illustrated in Exhibit 5.3, whereby G represents the good customers and B the bad customers or defaulters.

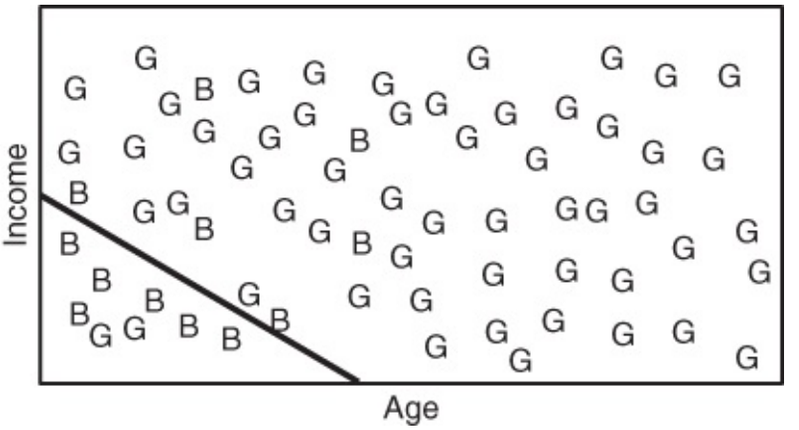

Exhibit 5.3 Linear Decision Boundary of Logistic Regression

To interpret a logistic regression model, one can calculate the odds ratio.

Suppose variable x_i increases with one unit with all other variables being kept constant (ceteris paribus); then the new logit becomes the old logit with β_i added. Likewise, the new odds become the old odds multiplied by $e^{\beta i}$. The latter represents the odds ratio—that is, the multiplicative increase in the odds when x_i increases by 1 (ceteris paribus). Hence,

- $\beta_i > 0$ implies $e^{\beta i} > 1$ and the odds and probability increase with x_i.
- $\beta_i < 0$ implies $e^{\beta i} < 1$ and the odds and probability decrease with x_i.

Another way of interpreting a logistic regression model is by calculating the doubling amount. This represents the amount of change required for doubling the primary outcome odds. It can be easily seen that for a particular variable x_i, the doubling amount equals $\log(2)/\beta_i$.

Variable Selection for Logistic Regressions

Variable selection aims at reducing the number of variables in a model. It will make the model more concise and faster to evaluate. Logistic regression has a built-in procedure to perform variable selection. It is based on a statistical hypothesis test to verify whether the coefficient of a variable i is significantly different from zero:

$$H_0: \ \beta_i = 0$$
$$H_A: \ \beta_i \neq 0$$

In logistic regression, the test statistic is:

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

$$\chi^2 = \left(\frac{\hat{\beta}_i}{s.e.(\hat{\beta}_i)} \right)^2$$

and follows a chi-square distribution with 1 degree of freedom. This test statistic is intuitive in the sense that it will reject the null hypothesis H_0 if the estimated coefficient $\hat{\beta}_i$ is high in absolute value compared to its standard error $s.e.(\hat{\beta}_i)$. The latter can be easily obtained as a by-product of the optimization procedure. Based on the value of the test statistic, we calculate the *p*-value, which is the probability of getting a more extreme value than the one observed. In other words, a low *p*-value represents a significant variable, and a high *p*-value represents an insignificant variable. From a practical viewpoint, the *p*-value can be compared against a significance level. Exhibit 5.4 presents some commonly used values to decide on the degree of variable significance.

Exhibit 5.4 Reference Values for Variable Significance

p-value < 0.01	Highly significant
0.01 < *p*-value < 0.05	Significant
0.05 < *p*-value < 0.10	Weakly significant
p-value > 0.10	Not significant

Various variable selection procedures can now be used based on the *p*-value. Suppose we have four variables, V_1, V_2, V_3, and V_4 (e.g., credit bureau score, income, years employed, and purpose of loan). The number of possible variable subsets equals $2^4 - 1$, or 15, as displayed in Exhibit 5.5.

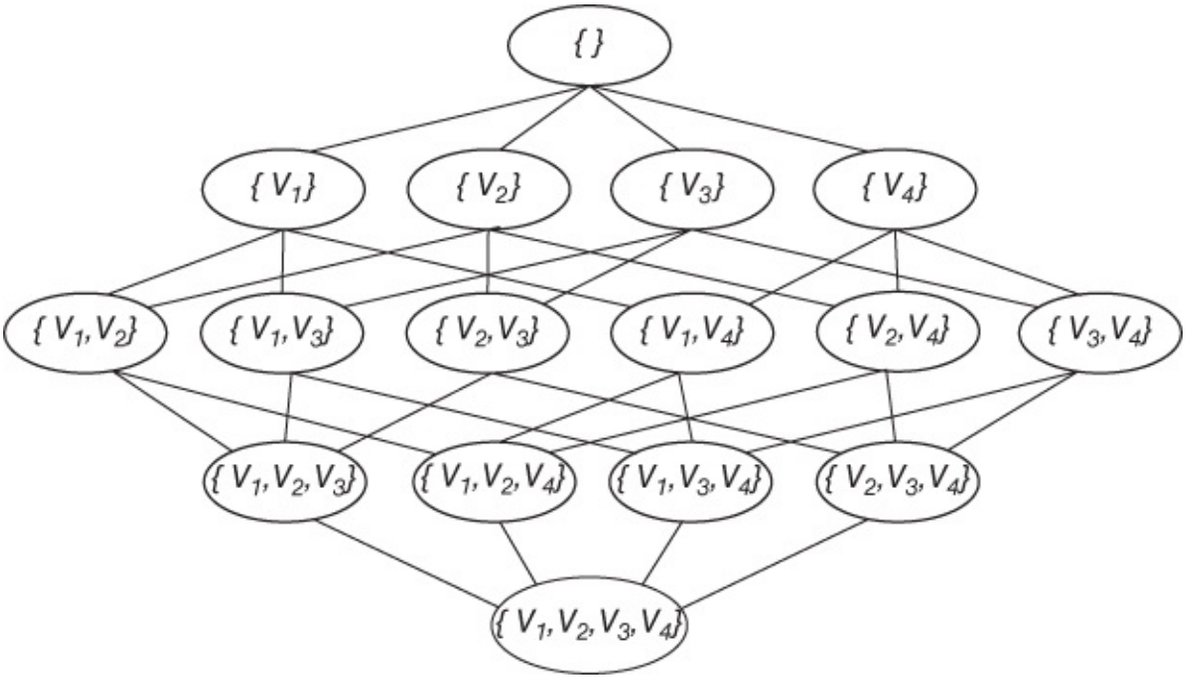

Exhibit 5.5 Variable Subsets for Four Variables, V_1, V_2, V_3, and V_4

When the number of variables is small, an exhaustive search among all variable subsets can be

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

performed. However, as the number of variables increases, the search space grows exponentially and heuristic search procedures are needed. Using the *p*-values, the variable space can be navigated in three possible ways. Forward regression starts from the empty model and always adds variables based on low *p*-values. Backward regression starts from the full model and always removes variables based on high *p*-values. Stepwise regression is a mix of both. It starts off like forward regression, but once the second variable has been added, it will always check the other variables in the model and remove them if they turn out to be insignificant according to their *p*-values. Obviously, all three procedures assume preset significance levels, which should be set by the user before the variable selection procedure starts.

In credit scoring, it is very important to be aware that statistical significance is only one evaluation criterion to consider in doing variable selection. As mentioned before, *interpretability* is also an important criterion (Martens et al. 2007). In logistic regression, this can be easily evaluated by inspecting the sign of the regression coefficient. It is highly preferable that a coefficient has the same sign as anticipated by the credit expert; otherwise he or she will be reluctant to use the model. Coefficients can have unexpected signs due to multicollinearity issues, noise, or small sample effects. Sign restrictions can be easily enforced in a forward regression setup by preventing variables with the wrong sign from entering the model.

Another criterion for variable selection is *operational efficiency*. This refers to the amount of resources needed for the collection and preprocessing of a variable. For example, although trend variables are typically very predictive, they require considerable effort to calculate and thus may not be suitable for use in an online credit scoring environment. The same applies to external data, where the latency might hamper a timely decision. In both cases, it might be worthwhile to look for a correlated, less predictive but easier to collect and calculate variable instead. Finally, *legal issues* also need to be properly taken into account. For example, in the United States, there is the Equal Credit Opportunity Act, which states that no one is allowed to discriminate based on gender, age, ethnic origin, nationality, beliefs, and so on. These variables must not be included in a credit scorecard. Other countries have other regulations, and it is important to be aware of this.

Building Logistic Regression Models in SAS

In Base SAS, a logistic regression model can be estimated using PROC LOGISTIC as follows:

```
PROC LOGISTIC DATA=mydata.hmeq;
CLASS job reason /PARAM=glm;
MODEL bad=clage clno debtinc delinq derog job loan mortdue ninq reason
value yoj/
SELECTION=stepwise SLENTRY=0.05 SLSTAY=0.01;
RUN;
```

The class statement is used to create dummy indicators for the categorical variables. The param=glm option indicates that we want these dummy indicators to be coded as 0/1. For example, since the job variable has five different values, four 0/1 dummy indicators will be

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

created. The selection option indicates that we want to do stepwise logistic regression. The options slentry and slstay indicate that we will consider variables for entering the model at a significance level of 0.05, but in order to stay in the model their significance level should be below 0.01. SAS will report the output for each of the intermediate variable selection steps. The final output then looks like Exhibit 5.6.

The LOGISTIC Procedure

Summary of Stepwise Selection

Step	Effect		DF	Number In	Score Chi-Square	**Wald Chi-Square**	Pr > ChiSq
	Entered	Removed					
1	**DELINQ**		1	1	254.2054		<.0001
2	**DEBTINC**		1	2	142.1980		<.0001
3	**DEROG**		1	3	105.4667		<.0001
4	**CLAGE**		1	4	40.4196		<.0001
5	**JOB**		5	5	23.6862		0.0002
6	**NINQ**		1	6	9.6436		0.0019
7	**CLNO**		1	7	7.2242		0.0072

Type 3 Analysis of Effects

Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
CLAGE	1	30.6003	<.0001
CLNO	1	7.1820	0.0074
DEBTINC	1	94.3510	<.0001
DELINQ	1	121.3538	<.0001
DEROG	1	49.9766	<.0001
JOB	5	24.0728	0.0002
NINQ	1	10.2919	0.0013

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	4.5465	0.5633	65.1386	<.0001
CLAGE		1	0.00574	0.00104	30.6003	<.0001
CLNO		1	0.0204	0.00762	7.1820	0.0074
DEBTINC		1	−0.0996	0.0103	94.3510	<.0001
DELINQ		1	−0.7584	0.0688	121.3538	<.0001
DEROG		1	−0.7213	0.1020	49.9766	<.0001
JOB	**Mgr**	1	0.6078	0.3894	2.4372	0.1185
JOB	**Office**	1	1.1626	0.3983	8.5193	0.0035
JOB	**Other**	1	0.6241	0.3655	2.9163	0.0877
JOB	**ProfExe**	1	0.6390	0.3788	2.8464	0.0916
JOB	**Sales**	1	−0.8330	0.5253	2.5148	0.1128
JOB	**Self**	0	0			
NINQ		1	−0.1199	0.0374	10.2919	0.0013

Odds Ratio Estimates			
Effect	Point Estimate	95% WaId Confidence Limits	
CLAGE	1.006	1.004	1.008
CLNO	1.021	1.006	1.036
DEBTINC	0.905	0.887	0.924
DELINQ	0.468	0.409	0.536
DEROG	0.486	0.398	0.594
JOB Mgr vs Self	1.836	0.856	3.939
JOB Office vs Self	3.198	1.465	6.981
JOB Other vs Self	1.867	0.912	3.821
JOB ProfExe vs Self	1.895	0.902	3.980
JOB Sales vs Self	0.435	0.155	1.217
NINQ	0.887	0.824	0.954

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	79.3	**Somers' D**	0.593
Percent Discordant	20.0	**Gamma**	0.597
Percent Tied	0.7	**Tau-a**	0.096
Pairs	919200	**c**	0.796

Exhibit 5.6 Output of PROC LOGISTIC

From this output, we can decide the following:

- The final model consists of the variables CLAGE, CLNO, DEBTINC, DELINQ, DEROG, JOB, and NINQ. The LOAN, MORTDUE, VALUE, REASON, and YOJ variables have no impact on the default risk.
- The performance of the model is reported in the final table. The c-coefficient corresponds to the area under the receiver operating characteristic (ROC) curve. We will discuss this measure in more detail in the chapter on validation. For the moment, it suffices to say that the number 0.796 indicates a very good performance.

In SAS Enterprise Miner, the Regression node from the Model tab can be used to estimate a logistic regression model (see Exhibit 5.7). If this node is connected to a data set with a binary variable, it will perform logistic regression by default. The Model Selection section of the property panel can be used to specify the variable selection options.

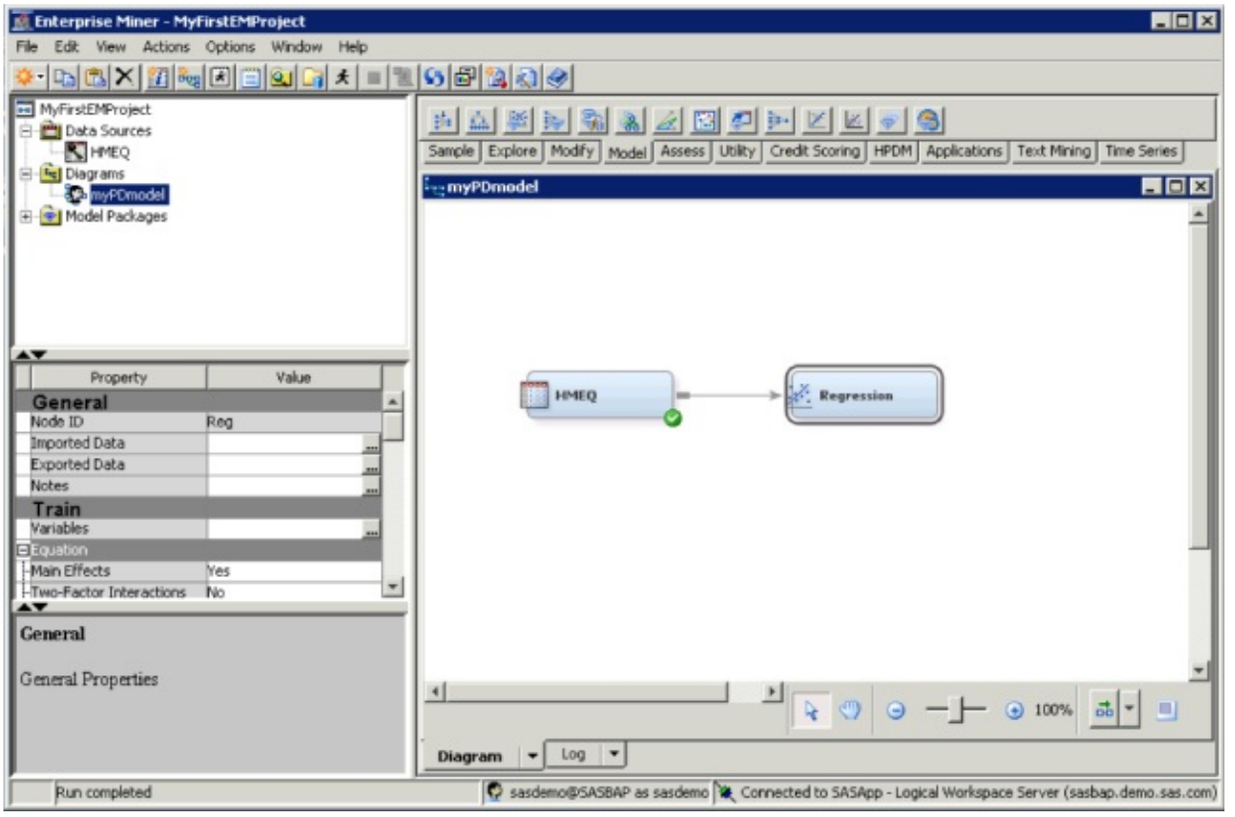

Exhibit 5.7 Logistic Regression in SAS Enterprise Miner

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Using Logistic Regression for Credit Scoring

Logistic regression is a very popular credit scoring classification technique due to its simplicity and good performance. Just as with linear regression, once the parameters have been estimated, the regression can be evaluated in a straightforward way, contributing to its operational efficiency. From an interpretability viewpoint, it can be easily transformed into an interpretable, user-friendly, points-based credit scorecard. Let's assume we start from the following logistic regression model whereby the explanatory variables have been coded using weights of evidence coding (see the chapter on data preprocessing):

$$P(default = yes|Age,\ Income,\ Employed,\ \dots) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 WOE_{Age} + \beta_2 WOE_{Income} + \beta_3 WOE_{Employed} + \dots)}}$$

As discussed earlier, this model can be easily reexpressed in a linear way in terms of the log odds as follows:

$$\log\left(\frac{P(default = yes|Age,\ Income,\ Employed,\ \dots)}{P(default = no|Age,\ Income,\ Employed,\ \dots)}\right) = \beta_0 + \beta_1 WOE_{Age} + \beta_2 WOE_{Income} + \beta_3 WOE_{Employed} + \dots$$

A scaling can then be introduced by calculating a credit score, which is linearly related to the log odds as follows:

$$\text{Credit Score} = \text{offset} + \text{factor} * \log(\text{odds})$$

Assume that we want a credit score of 100 for odds of 50:1, and a credit score of 120 for odds of 100:1. This gives the following:

$$100 = \text{offset} + \text{factor} * \log 50$$
$$120 = \text{offset} + \text{factor} * \log 100$$

The offset and factor then become:

$$\text{Factor} = 20/\ln(2) = 28.85$$
$$\text{Offset} = 100 - \text{factor} * \ln(50) = -12.87$$

Once these values are known, the credit score becomes:

$$\text{Credit Score} = \left(\sum_{i=1}^{N}(WOE_i * \beta_i) + \beta_0\right) * factor + offset$$
$$\text{Credit Score} = \left(\sum_{i=1}^{N}\left(WOE_i * \beta_i + \frac{\beta_0}{N}\right)\right) * factor + offset$$
$$\text{Credit Score} = \left(\sum_{i=1}^{N}\left(WOE_i * \beta_i + \frac{\beta_0}{N}\right) * factor + \frac{offset}{N}\right)$$

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

The points for each attribute are calculated by multiplying the weight of evidence of the attribute with the regression coefficient of the characteristic, then adding a fraction of the regression intercept, multiplying the result by the factor, and finally adding a fraction of the offset. The corresponding credit scorecard can then be visualized as depicted in Exhibit 5.8.

Exhibit 5.8 Example Credit Scorecard

Characteristic Name	Attribute	Points
Age 1	Up to 30	80
Age 2	30–45	120
Age 3	45–60	160
Age 4	65+	240
Income 1	Up to $2,000	5
Income 2	$2,000–$3,500	20
Income 3	$3,500+	80
Employed	No	100
Employed	Yes	140
…		

The credit scorecard is very easy to work with. Suppose a new customer with the following characteristics needs to be scored:

$$\text{Age} = 48,\ \text{Income} = \$2{,}500,\ \text{Employed} = \text{Yes},\ \ldots$$

The score for this customer can then be calculated as follows: 160 + 20 + 140 + …. This score can then be compared with a critical cutoff to help decide whether the customer is a defaulter. A key advantage of this credit scoring model is its interpretability. We can clearly see which are the most risky categories and how they contribute to the overall credit score. This is a very useful technique in credit scoring settings where interpretability is a key concern (Martens et al. 2007).

Logistic regression models can also be estimated in SAS/STAT. We will discuss this and other nonlinear regressions such as probit and cloglog model in the next chapter (Probabilities of Default: Discrete-Time Hazard Models).

Building Logistic Regression Scorecards in SAS Enterprise Miner

As part of the Credit Scoring tab, SAS offers a Scorecard node, which allows building scorecards using the procedure outlined in the previous section (see Exhibit 5.9). First, we added an Interactive Grouping node from the Credit Scoring tab to do the categorization and weights of evidence coding. We accepted the default settings for this node. For the Scorecard node, we set the Odds to 50, the Scorecard Points to 600, and the Points to Double Odds to 20. This will ensure that if the odds are 50, 600 points will be assigned, and 620 in case the odds

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

double to 100. We also set the Scorecard Type property to Detailed. The output is given in Exhibit 5.10, and the corresponding scorecard is shown in Exhibit 5.11.

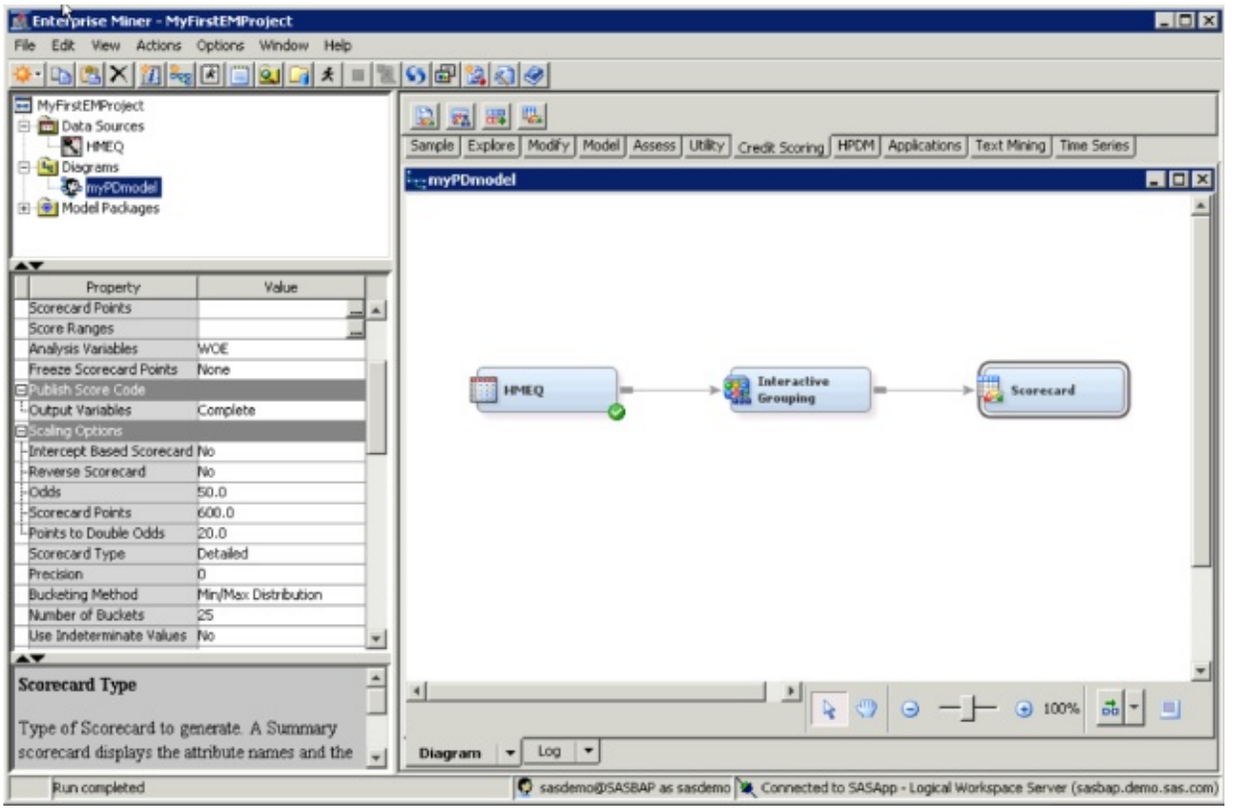

Exhibit 5.9 The Scorecard Node in SAS Enterprise Miner

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

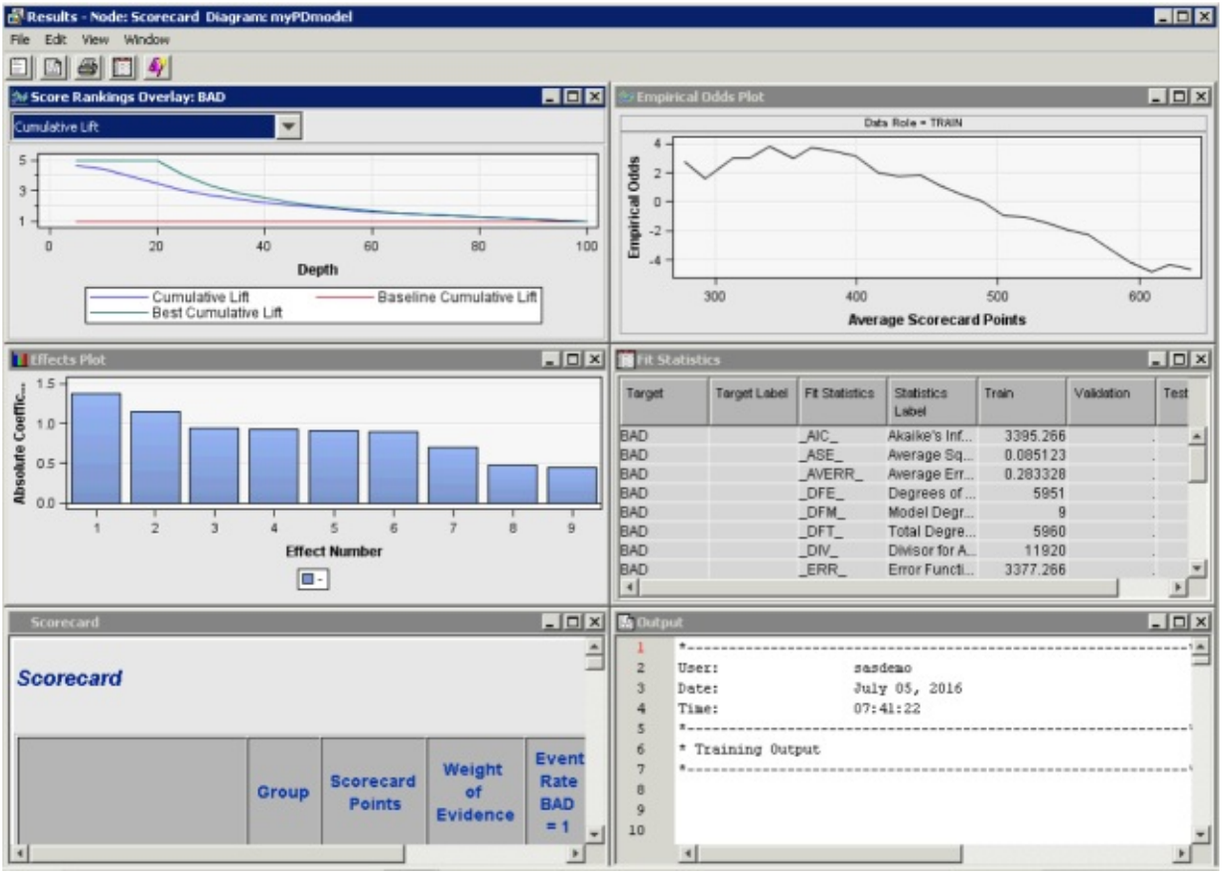

Exhibit 5.10 Output of the Scorecard Node in SAS Enterprise Miner

	Group	Scorecard Points	Weight of Evidence	Event Rate BAD = 1	Percentage of Population	Coefficient	
CLAGE	CLAGE< 84.55	1.00	41	–0.75	34.51	9.48	–1.15
	84.55<= CLAGE< 173.47	2.00	57	–0.27	24.54	37.95	–1.15
	173.47<= CLAGE< 247.1	3.00	79	0.39	14.46	28.42	–1.15
	247.1<= CLAGE	4.00	92	0.78	10.26	18.98	–1.15
	MISSING	5.00	56	–0.31	25.32	5.17	–1.15
DEBTINC	_MISSING_	5.00	16	–1.88	62.04	21.26	–0.91
	DEBTINC< 23.77	1.00	96	1.13	7.48	7.85	–0.91

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

	23.77<= DEBTINC< 30.31	2.00	111	1.72	4.26	15.77	– 0.91
	30.31<= DEBTINC< 41.44	3.00	98	1.20	6.96	47.23	– 0.91
	41.44<= DEBTINC	4.00	54	–0.45	28.09	7.89	– 0.91
DELINQ	_MISSING_	4.00	80	0.56	12.41	9.73	– 0.89
	DELINQ< 1	1.00	77	0.43	13.95	70.12	– 0.89
	1<= DELINQ< 2	2.00	47	–0.72	33.94	10.97	– 0.89
	2<= DELINQ	3.00	23	–1.67	57.04	9.18	– 0.89
DEROG	_MISSING_	3.00	77	0.58	12.29	11.88	– 0.70
	DEROG< 1	1.00	70	0.22	16.66	75.96	– 0.70
	1<= DEROG	2.00	39	–1.31	48.00	12.16	– 0.70
JOB	OFFICE	1.00	79	0.50	13.19	15.91	– 0.94
	PROFEXE	2.00	72	0.22	16.61	21.41	– 0.94
	MGR, OTHER	3.00	61	–0.19	23.23	52.94	– 0.94
	SALES, SELF	4.00	49	–0.63	31.79	5.07	– 0.94
	MISSING, _UNKNOWN_	5.00	94	1.02	8.24	4.68	– 0.94
LOAN	LOAN< 7600	1.00	54	–0.92	38.49	9.98	– 0.44
	7600<= LOAN< 10000	2.00	68	0.17	17.38	8.98	– 0.44
	10000<= LOAN< 15300	3.00	64	–0.12	21.87	26.01	– 0.44

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

	15300<= LOAN< 40000, _MISSING_	4.00	70	0.31	15.39	49.93	–0.44
	40000<= LOAN	5.00	63	–0.18	23.03	5.10	–0.44
NINQ	_MISSING_	5.00	71	0.37	14.71	8.56	–0.48
	NINQ< 1	1.00	70	0.30	15.65	42.47	–0.48
	1<= NINQ< 2	2.00	67	0.06	18.97	22.47	–0.48
	2<= NINQ< 4	3.00	62	–0.27	24.57	19.66	–0.48
	4<= NINQ	4.00	51	–1.11	43.14	6.85	–0.48
VALUE	_MISSING_	5.00	–43	–4.10	93.75	1.88	–0.92
	VALUE< 48800	1.00	49	–0.65	32.25	9.78	–0.92
	48800<= VALUE< 89235.5	2.00	68	0.08	18.75	39.28	–0.92
	89235.5<= VALUE< 132297	3.00	77	0.41	14.20	29.43	–0.92
	132297<= VALUE	4.00	70	0.14	17.78	19.63	–0.92

Exhibit 5.11 Credit Scorecard for HMEQ Data Set

Decision Trees

Basic Concepts

Decision trees are recursive partitioning algorithms (RPAs) that develop a tree-like structure representing patterns in an underlying data set (Duda, Hart, and Stork 2001). Exhibit 5.12 provides an example of a decision tree for credit scoring.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

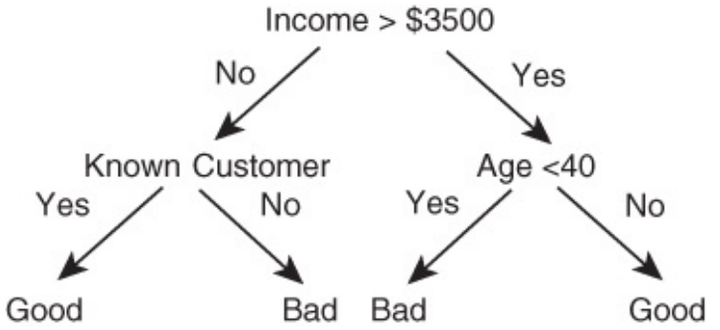

Exhibit 5.12 Example Decision Tree for Credit Scoring

The top node is the root node specifying a testing condition of which the outcome corresponds to a branch leading up to an internal node. The terminal nodes of the tree assign the classifications (in our case good or bad) and are also referred to as the leaf nodes. Many algorithms have been suggested in the literature to construct decision trees. Among the most popular are: C4.5 (See5) (Quinlan 1993), CART (Breiman et al. 1984), and CHAID (Hartigan 1975). These algorithms differ in their way of answering the key decisions to build a tree, which are:

- **Splitting decision**: Which variable to split and at what value (e.g., Income is > $3,500 or not, Known Customer is yes or no, etc.)
- **Stopping decision**: When to stop adding nodes to the tree
- **Assignment decision**: What class (e.g., good or bad) to assign to a leaf node

Assignment Decision

Usually, the assignment decision is the most straightforward to make since we typically look at the majority class within the leaf node to make the decision. This idea is also referred to as winner-take-all learning. The other two decisions are less straightforward and are elaborated on next.

Splitting Decision

In order to answer the splitting decision, one needs to define the concept of impurity or chaos. Consider, for example, the three data sets of Exhibit 5.13, each containing good customers (unfilled circles) and bad customers (filled circles). Quite obviously the good customers are nondefaulters, whereas the bad customers are defaulters. Minimal impurity occurs when all customers are either good or bad. Maximal impurity occurs when one has the same number of good and bad customers (i.e., the data set in the middle).

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

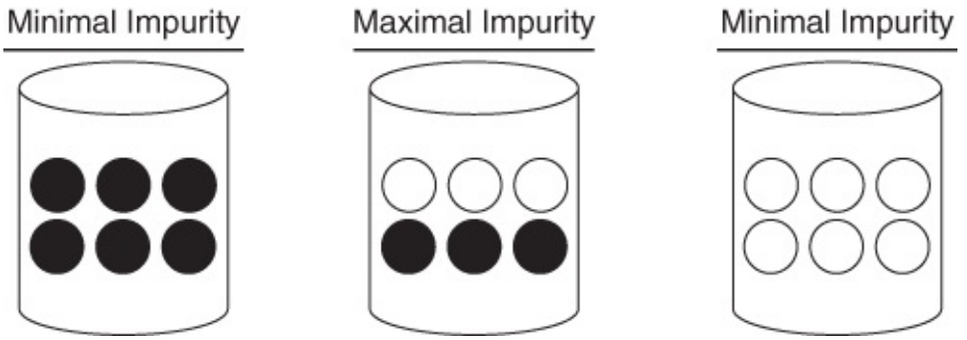

Exhibit 5.13 Example Data Sets for Calculating Impurity

Decision trees will now aim at minimizing the impurity in the data. In order to do so appropriately, one needs a measure to quantify impurity. Various measures have been introduced in the literature, and the most popular are:

- Entropy: $E(S) = -p_G \log_2(p_G) - p_B \log_2(p_B)$ (C4.5/See5)
- Gini: $\text{Gini}(S) = 2p_G p_B$ (CART)
- Chi-square analysis (CHAID)

with p_G (p_B) being the proportions of good and bad, respectively. Both measures are depicted in Exhibit 5.14, where it can be clearly seen that the entropy (Gini) is minimal when all customers are either good or bad, and maximal in cases of the same number of good and bad customers.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

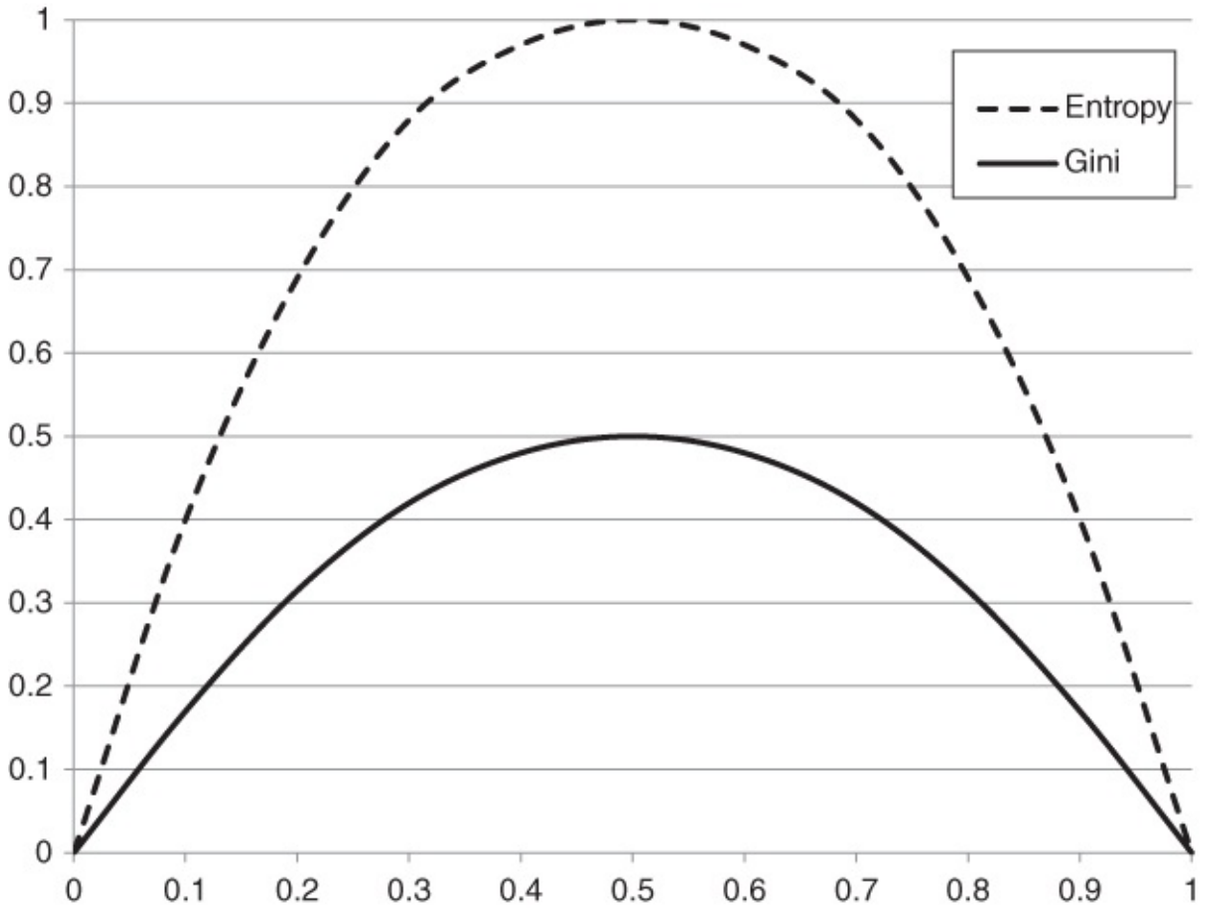

Exhibit 5.14 Entropy versus Gini

In order to answer the splitting decision, various candidate splits will now be evaluated in terms of their decrease in impurity. Consider, for example, a split on age as depicted in Exhibit 5.15.

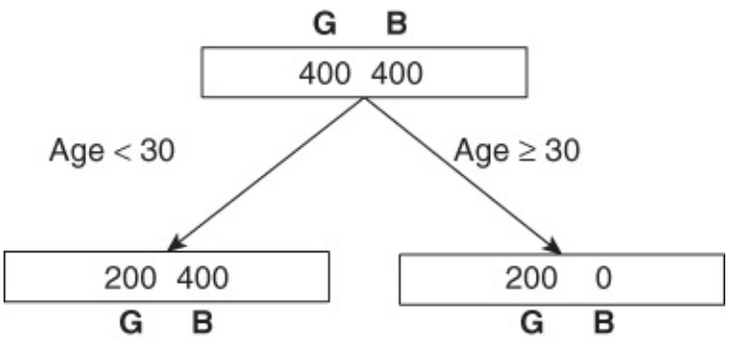

Exhibit 5.15 Calculating the Entropy for Age Split

The original data set had maximum entropy since the amount of goods and bads were the same. The entropy calculations now become:

- Entropy top node $= -1/2 \times \log_2(1/2) - 1/2 \times \log_2(1/2) = 1$
- Entropy left node $= -1/3 \times \log_2(1/3) - 2/3 \times \log_2(2/3) = 0.91$
- Entropy right node $= -1 \times \log_2(1) - 0 \times \log_2(0) = 0$

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

The weighted decrease in entropy, also known as the gain, can then be calculated as follows:

$$\text{Gain} = 1 - (600/800) \times 0.91 - (200/800) \times 0 = 0.32$$

The gain measures the weighted decrease in entropy thanks to the split. It speaks for itself that a higher gain is to be preferred. The decision tree algorithm will now consider different candidate splits for its root node and adopt a greedy strategy by picking the one with the biggest gain. Once the root node has been decided upon, the procedure continues in a recursive way, each time adding splits with the biggest gain. In fact, this can be perfectly parallelized and both sides of the tree can grow in parallel, hereby increasing the efficiency of the tree construction algorithm.

Stopping Decision

The third decision relates to the stopping criterion. Obviously, if the tree continues to split, it will become very detailed with leaf nodes containing only a few observations. In the most extreme case, the tree will have one leaf node per observation and as such perfectly fit the data. However, by doing so, the tree will start to fit the specificities or noise in the data, which is also referred to as *overfitting*. In other words, the tree has become too complex and fails to correctly model the noise-free pattern or trend in the data. As such, it will generalize poorly to new unseen data. In order to prevent this from happening, the data will be split into a training sample and a validation sample. The training sample will be used to make the splitting decision. The validation sample is an independent sample, set aside to monitor the misclassification error (or any other performance metric such as a profit-based measure) as the tree grows. A commonly used split is a 70 percent training sample and a 30 percent validation sample. We then typically observe a pattern as depicted in Exhibit 5.16.

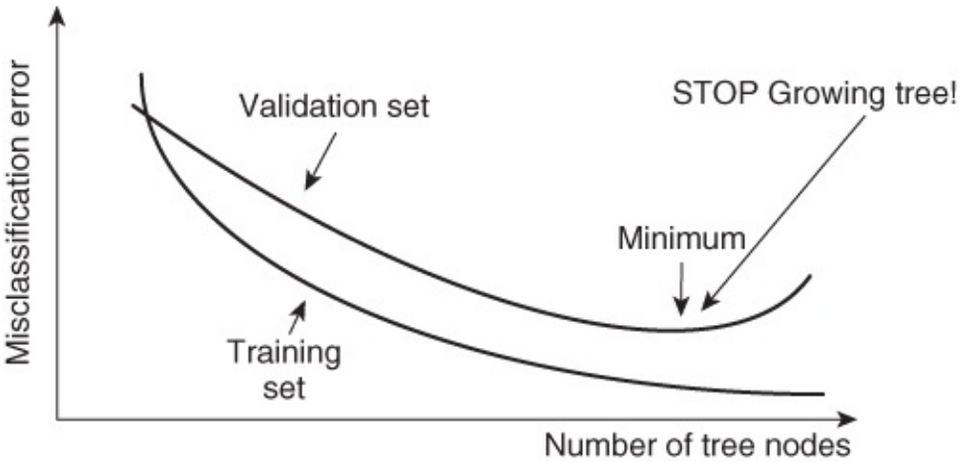

Exhibit 5.16 Using a Validation Set to Stop Growing a Decision Tree

The error on the training sample keeps on decreasing as the splits become more and more specific and tailored toward it. On the validation sample, the error will initially decrease, which indicates that the tree splits generalize well. However, at some point the error will increase since the splits become too specific for the training sample as the tree starts to memorize it. Where the validation set curve reaches its minimum, the procedure should be

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

stopped, as otherwise overfitting will occur. Note that, as already mentioned, besides classification error, we might also use accuracy or profit-based measures on the *y*-axis to make the stopping decision. Also note that sometimes simplicity is preferred above accuracy, and one can select a tree that does not necessarily have minimum validation set error, but a lower number of nodes.

Decision Tree Properties

In the example of Exhibit 5.12, every node had only two branches. The advantage of this is that the testing condition can be implemented as a simple yes/no question. Multiway splits allow for more than two branches and can provide trees that are wider but less deep. In a read-once decision tree, a particular attribute can be used only once in a certain tree path.

Every tree can also be represented as a rule set since every path from a root node to a leaf node makes up a simple "If–Then" rule. For the tree depicted in Exhibit 5.17, the corresponding rules are:

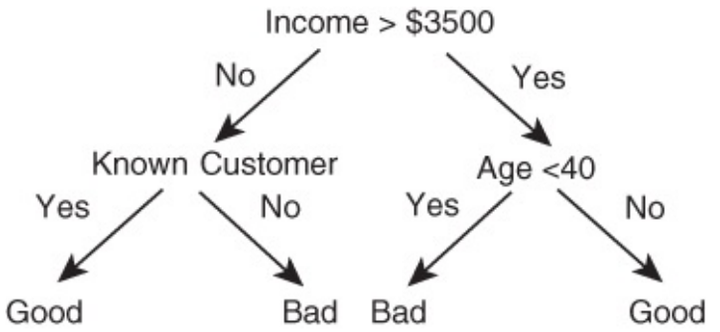

Exhibit 5.17 Example Decision Tree.

If Income > $3,500 = Yes **And** Age < 40 = No **Then** Good

If Income > $3,500 = Yes **And** Age < 40 = Yes **Then** Bad

If Income > $3,500 = No **And** Known Customer = No **Then** Bad

If Income > $3,500 = No **And** Known Customer = Yes **Then** Good

These rules can then be easily implemented in all kinds of software packages (e.g., Microsoft Excel).

Decision trees essentially model decision boundaries orthogonally to the axes. This is illustrated in Exhibit 5.18 for an example decision tree.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

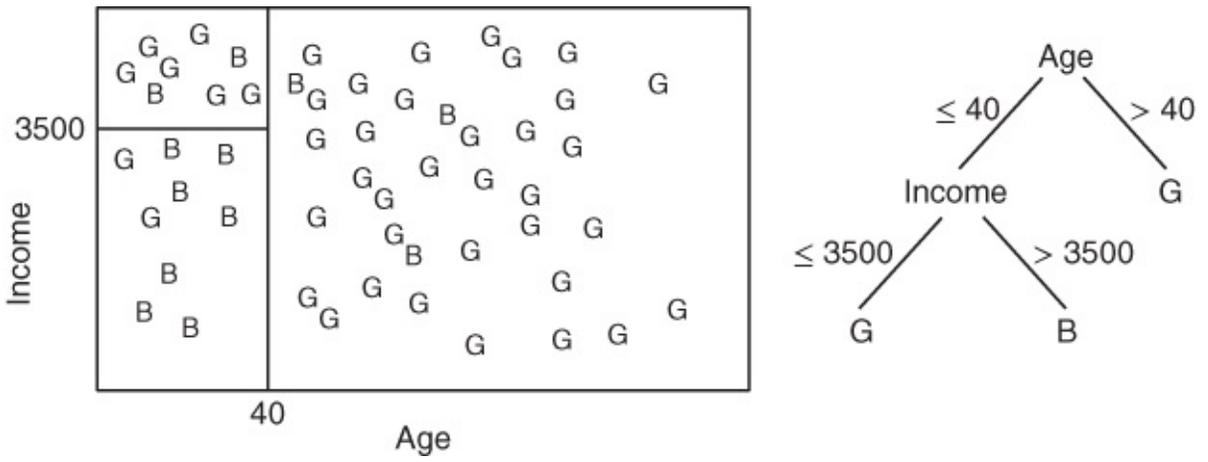

Exhibit 5.18 Decision Boundary of a Decision Tree

Building Decision Trees in SAS Enterprise Miner

There are no readily available procedures to build decision trees in Base SAS. In SAS Enterprise Miner, decision trees can be built using the Decision Tree node from the Model tab. This is illustrated in Exhibit 5.19. Note that we also added a Data Partition node between the HMEQ data source and the Decision Tree node. This node will split the data into a training set (40 percent of the observations), a validation set (30 percent of the observations), and a test set (30 percent of the observations). The validation set will be used to make the stopping decision as we discussed earlier. The output of the node is given in Exhibit 5.20, whereas the tree is represented in Exhibit 5.21. Note that the nodes are colored according to the impurity, where a darker node corresponds to greater impurity. The thickness of the branches is proportional to the amount of training observations that follow them.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

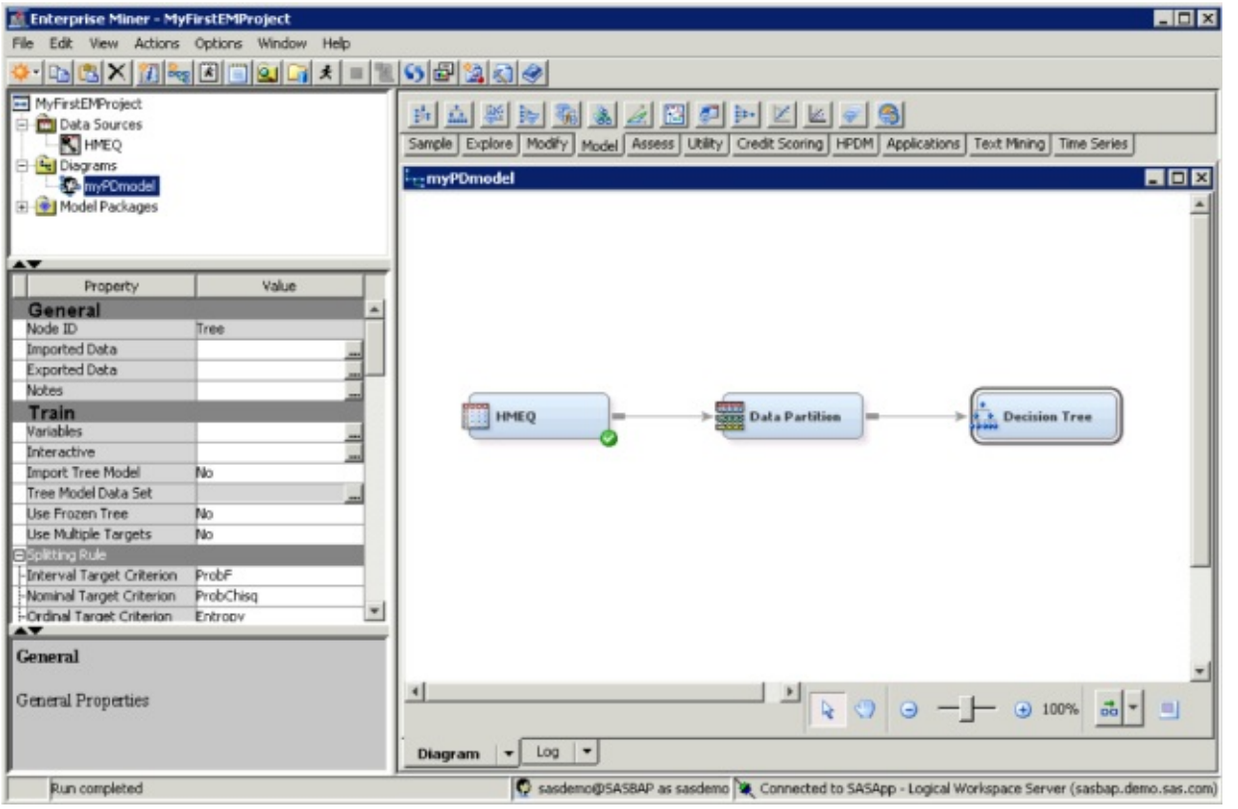

Exhibit 5.19 Decision Tree Node in SAS Enterprise Miner

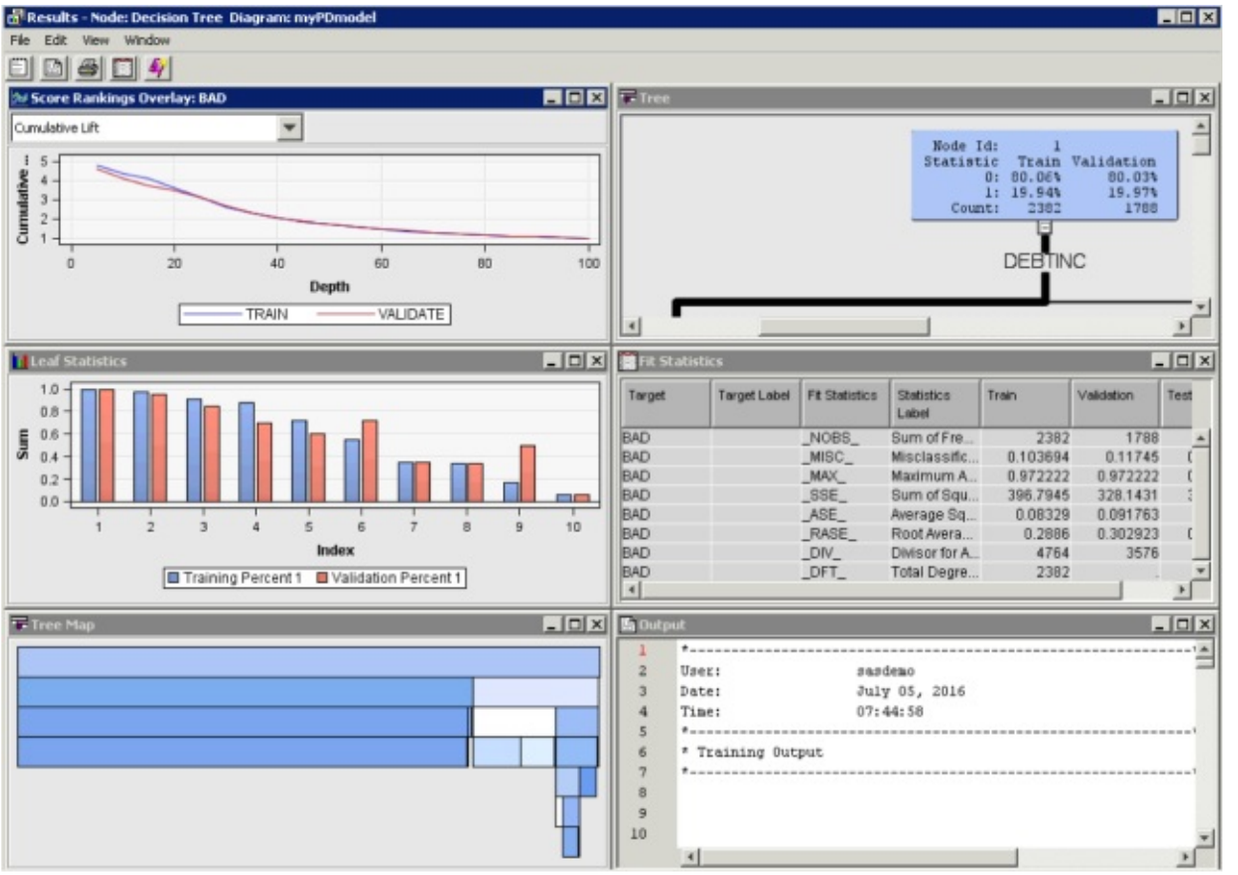

Exhibit 5.20 Output of the Decision Tree Node

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

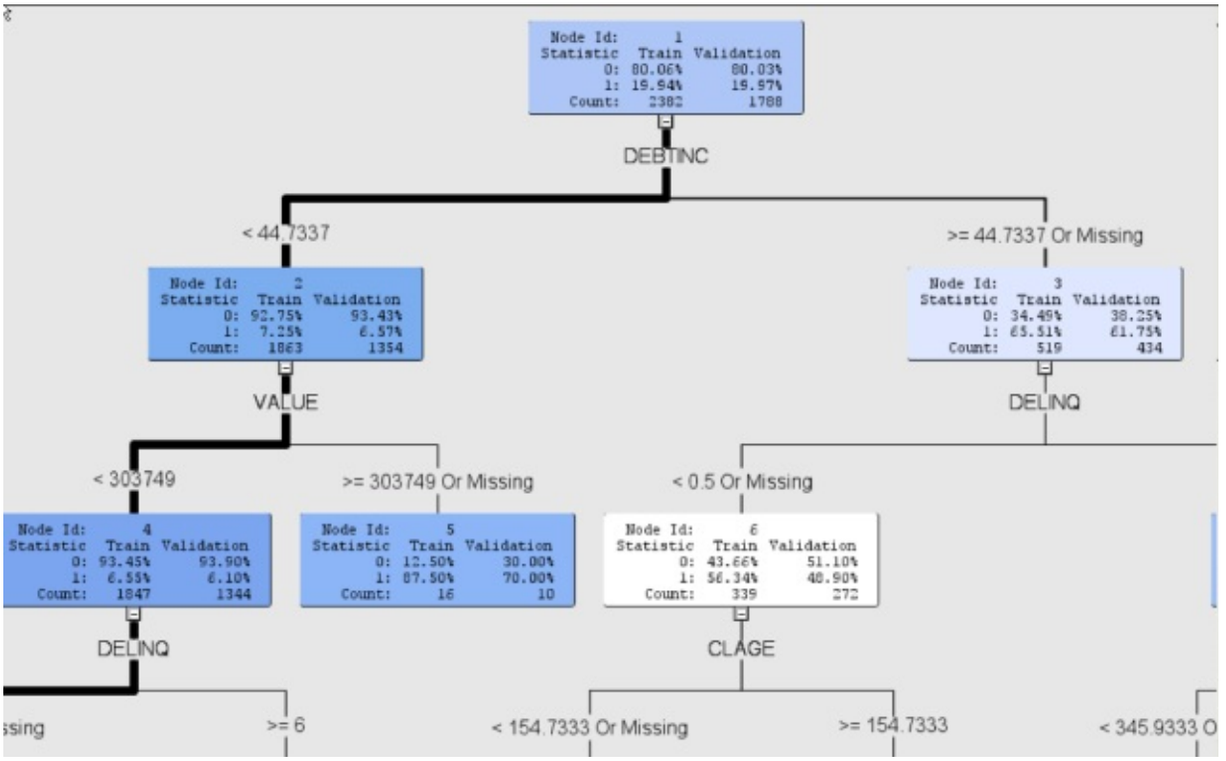

Exhibit 5.21 Decision Tree for HMEQ Data Set

Logistic Regression versus Decision Trees

Logistic regression is the most popular scorecard construction technique used in industry. Its key advantage, when compared to decision trees, is that a continuous range of scores is provided between 0 and 1. For decision trees, every leaf node corresponds to a particular score (i.e., the proportion of goods in the leaf node). Hence, only a limited set of score values is provided, which may not be sufficient to provide a fine, granular distinction between obligors in terms of default risk. Decision trees are, however, often used during the data preprocessing step for variable selection, categorization, or segmentation (as discussed in the chapter on data preprocessing).

Other Classification Techniques

Other classification techniques have been developed to build scorecards, such as discriminant analysis, neural networks, and support vector machines (SVMs), as well as ensemble methods such as bagging, boosting, and random forests (Baesens 2014). Especially the last-named methods have proven to be very powerful in terms of classification performance. However, despite their potential, these techniques yield very complex models that are hard to understand and hence not useful for building credit scoring models, where model interpretability is a key concern (Martens et al. 2007).

CREDIT SCORING FOR RETAIL EXPOSURES

Popular examples of retail portfolios are mortgages, revolving exposures such as credit cards

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

or overdraft accounts, and installment loans such as car loans. Three key statistical approaches for credit scoring in retail portfolios are application scoring, behavioral scoring, and dynamic scoring. All three approaches rely on historical data to build scorecards. A scorecard then provides a score whereby a higher score typically indicates less credit risk and a lower score a riskier obligor. The three approaches differ in the way they construct their historical data and set their prediction horizons. Let's go into these three approaches in more detail.

Application Scoring

Application scoring is the first important statistical credit scoring approach. The purpose of application scoring is to come up with a credit score that reflects the default risk of a customer at the moment of loan application. This is a very important scoring mechanism, as it will help the lender decide whether the credit application should be accepted or rejected.

In order to build an application scorecard, one first needs to define the concept of default. Multiple definitions of default can be adopted. It could be based on profit, amount owed, negative net present value, or number of months in payment arrears. A popular definition of default in the earlier days of credit scoring was that a customer was considered to be a defaulter if he or she ran into more than three months of payment arrears. With the introduction of the Basel Capital Accords, the default definition has now been set to 90 days in payment arrears, which is similar (Van Gestel and Baesens 2009). Note, however, that in some countries this definition has been overruled. In the United States, for example, in retail credit for residential mortgages the default definition is 180 days, for qualifying revolving exposures it's also 180 days, and for other retail exposures it's 120 days.

Let's assume now that we have our definition of default set. The next step is then to identify the information that can be used to predict default. Two different types of information can be distinguished: application variables and bureau variables.

Let's first discuss the application variables. This is the information provided to the bank by the applicant upon loan application. Popular examples are age, gender, marital status, income, time at residence, time at employment, time in industry, first digit of postal code, geographical (urban/rural/regional/provincial), residential status, employment status, lifestyle code, existing client (Y/N), number of years as client, number of products internally, total liabilities, total debt, total debt service ratio, gross debt service ratio, revolving debt/total debt, and number of credit cards. All these variables are internally available to the bank. They can be complemented by bureau variables.

Bureau variables are obtained from credit bureaus (also called credit reference agencies), which are external to the bank. A credit bureau is an organization that assembles and aggregates credit information from various financial institutions or banks. It can collect both positive and negative credit information, depending upon the country in which it operates. Usually, credit bureaus provide two sources of information. A first example is raw bureau data such as number of previous delinquencies, total amount of credit outstanding, previous delinquency history, time at credit bureau, total credit bureau inquiries, time since last credit bureau inquiry, inquiries in the past 3/6/12 months, inquiries in the past 3/6/12 months as

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

percentage of total, and so on.

Using this raw bureau data, credit bureaus can now build bureau credit scores. These bureau scores can then be sold to banks, which can then use them in their application scoring models. Credit bureaus are all around these days. In the United States, popular bureaus are Experian, Equifax, and TransUnion, each of which covers its own geographical region. All three provide a FICO score, which ranges between 300 to 850 with higher scores reflecting better credit quality. A FICO score essentially relies on the following five data sources to determine creditworthiness:

1. Payment history: Has the customer any delinquency history? This accounts for 35 percent of the FICO score.
2. Amount of current debt: How many credits does the customer have in total? This accounts for 30 percent of the FICO score.
3. Length of credit history: How long has the customer been using credit? This accounts for 15 percent of the FICO score.
4. Types of credit in use: What kind of loans does the customer currently have (e.g., credit cards, installment loans, mortgage, etc.)? This accounts for 10 percent of the FICO score.
5. Pursuit of new credit: How many new credits is the customer applying for? This accounts for 10 percent of the FICO score.

These FICO scores are commonly used in the United States, not only by banks, but also by insurance providers, telecommunications firms, utilities companies, and others. Other countries obviously also have their own credit bureaus. In Australia, there is Baycorp Advantage, Germany has the Schufa, Netherlands BKR, and Belgium CKP. Dun & Bradstreet is a popular credit bureau targeting the midsize corporation and small to medium-sized enterprise market.

Exhibit 5.22 shows an example of an application scorecard. It includes three characteristics: age, known customer, and salary. Each of these characteristics has been categorized or coarse classified into several categories. For example, age has been categorized into four categories: the first is up to 26 years, the second between 26 and 35 years, and so on. Each of these categories has points assigned to it: The more points, the higher the credit quality. We can see that within the age characteristic the points are monotonically increasing from 100 to 225. We can also see that unknown customers are considered more risky than known customers since they receive only 90 points instead of 180. For salary, we again see a monotonic increase.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Characteristic Name	Attribute	Scorecard Points
Age 1 (years)	Up to 26	100
Age 2	26–35	120
Age 3	35–37	185
Age 4	37+	225
Known customer	No	90
Known customer	Yes	180
Salary 1	Up to $500	120
Salary 2	$501–$1,000	140
Salary 3	$1,001–$1,500	160
Salary 4	$1,501–$2,000	200
Salary 5	$2,000+	240

Exhibit 5.22 Example Application Scorecard

Let's now imagine that a new application for credit is submitted by a known customer whose age is 32 and salary is $1,150. We can now look up the points for each of these three characteristics. For age, the customer gets 120 points, for known customer 180 points, and for salary 160 points. This all adds up to 460 points, which represents the total credit quality of this particular customer. This now needs to be compared against a cutoff of 500, which represents the minimum required credit quality set by the bank. We can see that this particular customer falls below the cutoff and will thus be rejected. Imagine now that we consider the same customer, but with a salary of $2,500 instead of $1,150. Do you think this customer will be accepted or rejected? Well, let's verify. The new score now becomes 120 + 180 + 240, which equals 540. This is above the cutoff of 500, so the customer will be accepted.

The purpose of application scoring is now to build a scorecard like the one you see in Exhibit 5.22. Note that this encompasses many decisions that need to be made during the scorecard development process. For example:

- Why do we select the characteristics age, known customer, and salary? Why don't we include other ones like employment status or number of years living at current address?
- Why do we categorize age into four categories?
- How do we decide on the points assigned to each category?
- Why do we set the cutoff at 500? If it had been set at 400, the customer would have been accepted with the lower salary.

When building an application scorecard, you are actually taking two snapshots of customer behavior (see Exhibit 5.23).

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

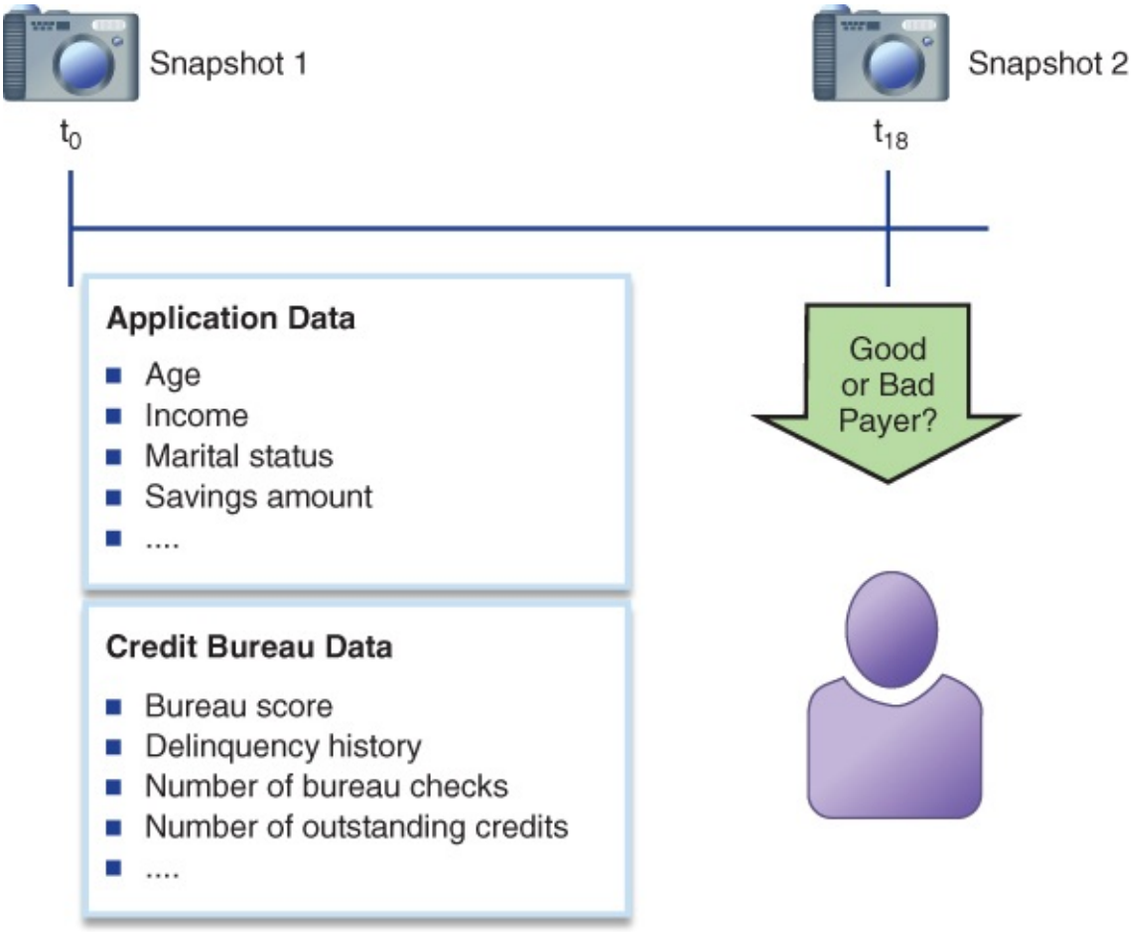

Exhibit 5.23 Application Scoring: Snapshot to Snapshot

The first snapshot is taken at loan origination where you will gather both the application and credit bureau data. This is the information that will then be used to predict default. The second snapshot is taken at some later point during the loan at which the default behavior will be determined. Ideally, you should wait until the end of the loan to be absolutely certain about whether a customer defaulted. However, for mortgages this would imply having to wait 15, 20, or even more years to do this, and obviously this is not feasible. Empirical analysis has shown that the majority of customers who default do so in the first 18 months. Hence, many firms take the second snapshot 18 months after loan origination to see whether the customer defaulted. In this way, they construct a data set and build an application scorecard from it.

Application scorecards provide scores. A score is a measure that allows lenders to rank customers from high risk (low score) to low risk (high score) and as such provides a relative measure of credit risk. Scores are unlimited and can be measured within any range; they can even be negative. A score is not the same as a probability. A probability also allows us to rank, but on top of that, since it is limited between 0 and 1, it also gives an absolute interpretation of credit risk. Hence, probabilities provide more information than scores do. For application scoring, one does not need well-calibrated probabilities of default. However, for other application areas such as regulatory capital calculation in a Basel setting, as we will discuss later, calibrated default probabilities are needed (Van Gestel and Baesens 2009). In later sections, we will discuss how to transform scores to probabilities.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Behavioral Scoring

Behavioral scoring is another statistical credit scoring approach, in this case one that analyzes the behavior of existing credit customers. Imagine that customer Bart applies for credit at your bank. First you are going to put this borrower through an application scoring model, and let's say that you decide to accept his application. At some point, you want to reassess the borrower's credit risk, taking into account all his recent behavior. That recent information could be his checking account behavior summarized by the average of the checking account balance, the maximum or minimum thereof, or the trend during the previous 12 months. Other interesting information could be delinquency information like whether the borrower had already incurred payment delays. Also, changes in his job status or home address could be considered. All this behavioral information can then be combined into a behavioral credit score, which provides the bank with a new and better assessment of the credit risk for the already existing obligor.

Behavioral scoring models are typically constructed using a 24-month time frame. Twelve months are taken to measure and quantify all the information that will be used as predictors, and the subsequent 12 months to determine the default status.

Behavioral scoring is dynamic since it summarizes the behavior into various dynamic variables such as average checking account balance, maximum checking account balance, trend in checking account balance, and more. As such, it is as if we construct a video clip of customer behavior during a period of 12 months. This video clip is then again summarized using a snapshot to determine the good/bad status 12 months after the observation point. Hence, behavioral scoring is often referred to as a video clip to snapshot problem (see Exhibit 5.24).

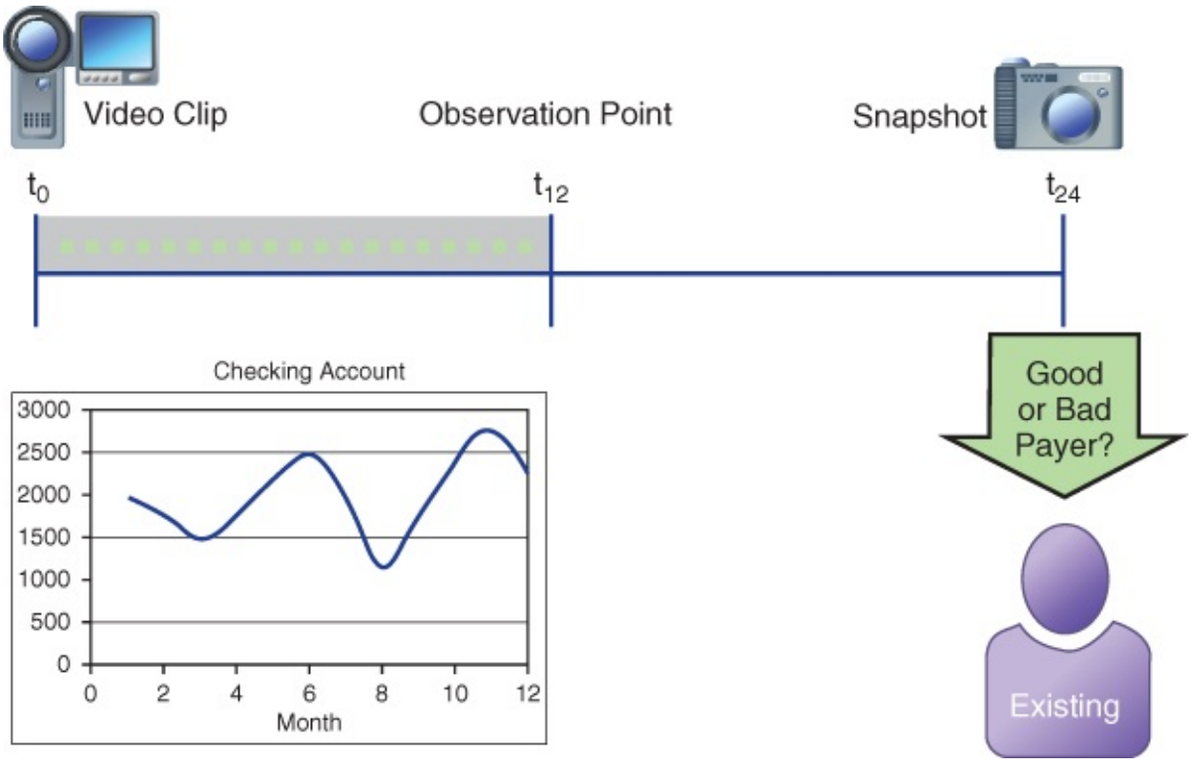

Exhibit 5.24 Behavioral Scoring: Video Clip to Snapshot

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

When compared to application scoring, behavioral scoring starts from a much bigger data set in terms of the number of variables. Behavioral scoring data sets typically have a few hundred variables to consider. Some examples are maximum and minimum levels of balance, credit turnover, trend in payments, trend in balance, number of missed payments, times exceeded credit limit, times changed home address, and so forth. Hence, during scorecard construction, it will be very important to carefully select the variables that contribute to predicting default risk.

When creating a behavioral scoring data set, it is really important to carefully think about the definitions of the variables. These variables should as much as possible be related to the financial solvency of the customer. In doing so, a first issue to thoroughly consider is the roles that a customer can take for a particular product. A customer can be the primary debtor, secondary debtor, or guarantor of the credit. When defining a behavioral variable such as the number of credits of a customer, all these roles should be taken into account. Furthermore, a customer can have both private and professional products. Also, the behavior can be measured at various levels in a product taxonomy. Think about a behavioral variable such as average, maximum, or minimum savings amount. When quantifying this variable, various savings-related products could be considered simultaneously such as checking accounts, saving accounts, term accounts, and so on.

When defining behavior, the same variable can be measured multiple times during the observation period. Think about variables such as checking account, credit balance, or bureau score. During a 12-month observation period, multiple values for these variables will be available. Hence, aggregate functions need to be used to aggregate these variables. Popular aggregate functions are the mean, which is sensitive to outliers; the median, which is robust to outliers; the minimum; and the maximum. Think of variables such as the worst delinquency status during the past six months, or the highest credit utilization during the past 12 months. Also, trend variables can be computed, such as the absolute or relative trend. The latter takes into account the starting value of the variable. These trends can be computed during the previous six or 12 months, as illustrated. Trend variables are usually very predictive but require more time to be computed. It may also be worthwhile to add variables such as the most recent value, the value one month ago, and so on. The aggregate functions can also be used for ratio variables, which are very popular in behavioral scoring. One example is the obligation/income ratio, which is calculated by adding the monthly house payment and any regular monthly installment or revolving debt and dividing by the monthly income. Another example could be the ratio of the current credit balance to the credit limit.

Just as with application scoring, the aim of behavioral scoring is to provide a score that is, as explained earlier for application scores, a relative credit assessment allowing banks to rank order customers from low risk to high risk in terms of their default likelihood.

A final issue concerns the migration from an application to a behavioral score. A sudden move from an application to a behavioral score can cause big fluctuations (e.g., in terms of expected or unexpected losses) and is thus discouraged. Many banks will work with a transition period of about six months during which a weighted combination of both scores is used. The transition

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

score then becomes $\alpha \times \mathrm{AS} + (1-\alpha) \times \mathrm{BS}$. During the transition period, the weight α is gradually decreased such that at the end of the period the customer has fully migrated to the behavioral score.

Dynamic Scoring

We already discussed application scoring as a snapshot to snapshot problem, and behavioral scoring as a video clip to snapshot problem, and in what follows we will now discuss dynamic scoring as a video clip to video clip statistical credit scoring approach.

In dynamic scoring, a risk assessment is provided for any future moment in time (see Exhibit 5.25). Contrary to application and behavioral scoring, where the risk assessment is provided for a fixed time horizon (e.g., 18 or 12 months), in dynamic scoring risk assessments will be provided for 3, 4, 5, 6, 12, or 18 months (or more) into the future. As such it gives a lot more information than an application scorecard or a behavioral scorecard. However, it is a lot harder to construct since we need to use special statistical techniques such as survival analysis, which can be difficult. Basically, these techniques allow us to model not just if a borrower will default, but also when. The key advantages of using survival analysis techniques for credit scoring are:

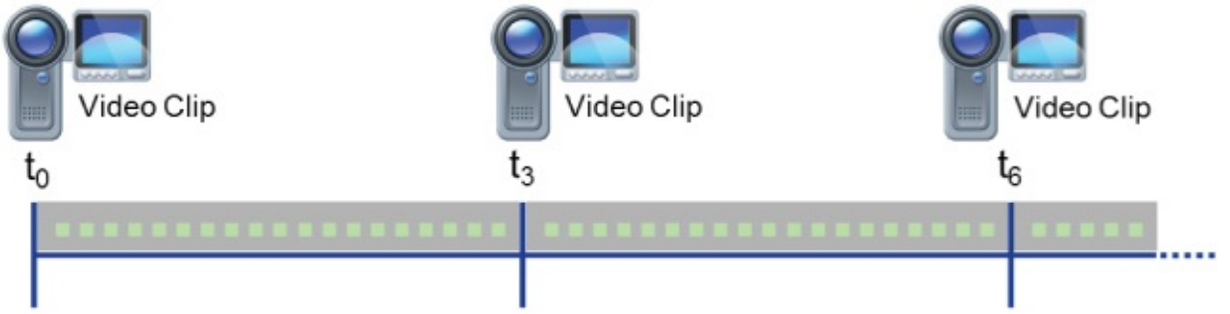

Exhibit 5.25 Dynamic Scoring: Video Clip to Video Clip

- They provide a natural way to model the loan default process and also incorporate obligors who have not defaulted during the observation period.
- They do not necessitate the definition of a fixed performance window during which default is measured.
- They can easily accommodate time-varying behavioral and economic factors.
- The survival probabilities can be easily used for profit scoring.

We do see a growing interest in dynamic credit scoring models in the industry, and some banks have already started to experiment with them. The outcomes of these models not only are useful for credit risk assessment but also can be used to model customer lifetime value (CLV), for example.

REJECT INFERENCE

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

The problem of reject inference refers to the fact that the data available to build application scorecards concerns only the past accepts and not the rejects since we don't know their default or nondefault behavior. Not including the rejects in the application scorecard development process will create a bias. Various methods for handling reject inference have been suggested, as we will discuss in what follows.

Classifying the Rejects as Bads

A first method to perform reject inference is to classify all rejects as bad payers. Obviously, this method will reinforce the credit scoring policy and prejudices of the past. It is also a conservative approach since it is plausible to assume that not all rejects would have turned out to be bad payers. Hence, by classifying all rejects as bads, the bad rate in the sample will be too high. A softer version can also be adopted whereby only a subsample of the rejects is classified as bads based on expert knowledge.

Hard Cutoff Augmentation

The hard cutoff augmentation method starts by building a scorecard on the accepts only. It then uses this scorecard to calculate scores for the rejects. An assumption is then made about the bad rate in the rejects. Consider, for example, Exhibit 5.26 with four rejects and their scores.

Reject	Score	Inferred G/B
R1	0.46	B
R2	0.22	B
R3	0.58	G
R4	0.04	B

Exhibit 5.26 Hard Cutoff Augmentation

If the assumed bad rate equaled 75 percent, that would mean that there were three bads among the four rejects. The three rejects with the lowest scores would then be considered as bad payers. In our case, this would mean rejects R1, R2, and R4 would be considered as bad payers. Once all the rejects have been labeled as good or bad payers, they are added to the accepts data set. The final scorecard can then be built on this combined data set. Although this method looks quite appealing at first sight, a key input parameter is the bad rate in the rejects, which is usually not known.

Parceling

Parceling is another method that works similarly to hard cutoff augmentation. It starts by building a scorecard on the accepts only. This model is then also used to score the rejects. The score range is then categorized, and for the accepts the percentages of goods and bads within each category are calculated. These percentages are then used to allocate the rejects to the good and bad classes.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

In the example depicted in Exhibit 5.27, this would mean that for the 654 rejects in the score range 100 to 199, 21.6 percent or 141 will be classified as bad, and 78.4 percent or 513 will be classified as good. This allocation can then be done randomly. Obviously, the proportion of bads for the rejects and accepts cannot be the same. Hence, an alternative could be to use a scaling factor to allocate a higher proportion of the rejects to the bad class. As an example rule of thumb when determining this scaling factor, the bad rate for the rejects could be 2 to 4 times higher than the bad rate for the accepts.

	Accepts				Rejects		
Score	**# Bad**	**# Good**	**% Bad**	**% Good**	**# Rejects**	**# Bad**	**# Good**
0–99	24	10	70.3%	29.7%	342	240	102
100–199	54	196	21.6%	78.4%	654	141	513
200–299	43	331	11.5%	88.5%	345	40	305
300–399	32	510	5.9%	94.1%	471	28	443
400+	29	1,232	2.3%	97.7%	778	18	760

Exhibit 5.27 Parceling

Fuzzy Augmentation

Fuzzy augmentation also starts by first creating a model based on the accepts, and then using that model to score the rejects. Every reject now gets a probability of being good, *p*(Good), and a probability of being bad, *p*(Bad). Every reject will now be duplicated into two observations with target good and bad and corresponding weights *p*(Good) and *p*(Bad), respectively. In Exhibit 5.28, you can see that the reject customer Dan has been duplicated into a first observation with target good and weight 0.80, and a second observation with target bad and weight 0.20. The rejects can then be combined with the accepts and a new model can be estimated on this combined data set. Ideally, the classification technique that is being used to build the model should take into account the weights set for the rejects.

Accepts/Rejects	**Customer ID**	**Age**	**Income**	**…**	**Weight**	**Good/Bad**
Accepts	John	56	$3,400		1	Bad
Accepts	Sara	38	$2,200		1	Good
					…	
Rejects	Dan	44	$2,500		0.80	Good
Rejects	Dan	44	$2,500		0.20	Bad

Exhibit 5.28 Fuzzy Augmentation

Nearest Neighbor Methods

Nearest neighbor methods can also be used to do reject inference. They work very intuitively

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

as follows. For every reject, look at the *k* most similar accepts in the Euclidean sense. These are also called the *k* nearest neighbors. *k* can be set to 1, 10, 100, or even more. You can then assign the most common class among those nearest neighbors to the reject. For example, if you consider the 100 nearest neighbors of a reject and it turns out that 80 of them are good payers and 20 are bad payers, then the reject will be considered as a good payer. Once all rejects have been classified as good or bad, they can be added to the accepts and a new model on the combined data set can be estimated. Although the nearest neighbor method to do reject inference looks quite appealing, instability problems can occur when the rejects are too far away from the accepts. In this case, the classes assigned to the rejects will be very uncertain and unstable.

Grant Credit to All Customers

A controversial approach to do reject inference is granting credit to all customers, including the rejects, during a specific period of time. This is motivated by a statement made by professor David Hand (2001):

> … There is no unique best method of universal applicability, unless extra information is obtained.
>
> That is, the best solution is to obtain more information (perhaps by granting loans to some potential rejects) about those applicants who fall in the reject region.

The idea here is to grant credit to some but not all rejects by evaluating the trade-off between the cost of granting credit to rejects and the benefit of obtaining a better scoring model. Banasik, Crook, and Thomas (2001) were in the exceptional situation of being able to observe the repayment behavior of customers who would normally have been rejected. They were able to contrast the effect on scorecard performance of both using and not using reject inference procedures. They concluded that the scope for improving scorecard performance by including the rejected applicants in the model development process is present but modest. However, note that this was studied on only one data set and can thus not be generalized.

Credit Bureau Based Inference

Another way to get more information about the rejects is via the credit bureau. Remember that a credit bureau or credit reference agency gathers information from various financial institutions about the delinquency behavior of their customers. So what you could do is also ask the bureau if some past rejects received credit elsewhere, and also inquire after their performance at these other financial institutions. Actually, you can ask the credit bureau for two pieces of information. You can give the credit bureau a sample of your past rejects and ask the bureau to classify them as good or bad. Or, if privacy regulations would not allow you to do so, you could also provide the credit bureau with a sample of your past rejects and ask the bureau about the bad rate in that sample. This bad percentage can then be used in the hard cutoff augmentation method, for example. One problem with bureau-based reject inference is

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

that customers that were previously rejected by your bank may have received credit elsewhere but under other conditions, making the comparison not strictly apples to apples.

Other Methods

Other methods to do reject inference have also been developed but are less popular. Some examples are a three-group approach, iterative reclassification, and mixture distribution modeling (Thomas, Edelman, and Crook 2002). We will not consider those any further since they are less frequently used in the industry.

Withdrawal Inference

In this section, we have discussed the reject inference problem. Remember, the reject inference problem is essentially a sampling problem because no information about the target good/bad class is available for the previously rejected customers. However, when you think a bit more closely about the historical through-the-door population, you will see that you also have withdrawals in there. These are customers who decided themselves to not take up the offer because they found a better offer elsewhere. In other words, these are the shoppers. Also for these withdrawals, we may not know the true good/bad class, and procedures for withdrawal inference should be adopted. One easy way to do withdrawal inference could again be via the credit bureau, whereby a sample of past withdrawals is given to the credit bureau so as to obtain information about their good/bad status at other financial institutions. However, it needs to be noted that not many firms consider the withdrawals in their scorecard development.

Reject Inference in SAS Enterprise Miner

As part of the Credit Scoring tab, SAS Enterprise Miner offers the Reject Inference node to perform reject inference using hard cutoff augmentation, parceling, or fuzzy augmentation.

CREDIT SCORING FOR NONRETAIL EXPOSURES

In this section, we will discuss four credit scoring approaches for nonretail credit portfolios such as corporate exposures, sovereign exposures, and bank exposures. As opposed to retail portfolios, a key issue in these portfolios is the availability of data. The four approaches covered differ in the type of data that they use. Let's continue and discuss these approaches in more detail.

Prediction Approach

The prediction approach assumes that there is historical data available about the obligors (e.g., firms) and their default (e.g., bankruptcy) status. This data can then be analyzed by statistical techniques to predict default behavior. As an example, popular data to be used for bankruptcy prediction are accounting information such as balance sheet and financial statement ratios, and stock price behavior if the firm is publicly listed.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

A well-known model for bankruptcy prediction is the Altman z-model for manufacturing firms. It was built in the late 1960s using a statistical technique called linear discriminant analysis. Separate versions exist for public and private industrial companies:

- For public industrial companies: $z = 1.2 \times 1 + 1.4 \times 2 + 3.3 \times 3 + 0.6 \times 4 + 1.0 \times 5$ (healthy if z > 2.99, unhealthy if z < 1.81)
- For private industrial companies: $z = 6.56 \times 1 + 3.26 \times 2 + 6.72 \times 3 + 1.05 \times 4$ (healthy if z > 2.60, unhealthy if z < 1.1)

The ×1,…×5 variables are the following accounting ratios: ×1: working capital/total assets; ×2: retained earnings/total assets; ×3: earnings before interest and taxes (EBIT)/total assets; ×4: market (book) value of equity/total liabilities; ×5: net sales/total assets.

Essentially, the z-score is a linear combination of these five accounting ratios. A higher z score reflects a healthier firm and thus a lower bankruptcy risk. Extensions of the original z-score model have been provided for privately held and nonmanufacturing firms. The z-score can be used by a bank as its internal bankruptcy prediction model. It can also be used to benchmark other bankruptcy prediction models. Note, however, that the z-score model has been built using U.S. corporates, and care should be taken when applying it to other countries.

Expert-Based Approach

As discussed previously, the prediction approach necessitates the availability of data, and you may not always have data available in a nonretail setting. So in the absence of data, the expert-based approach is also used in nonretail credit risk modeling. Basically, the expert-based approach builds a scorecard in a qualitative way using the business experience, intuition, and common sense of one or more credit experts.

Exhibit 5.29 shows an example of an expert-based scorecard for corporate credit risk modeling. You can see that it considers various characteristics, such as industry position, market share trends and prospects, and so on. Each of these characteristics has been defined in a qualitative way. They also have scores assigned to them. These scores have not been estimated from historical data, since this is not available here, but have been determined by the business experts themselves in a subjective way. Expert-based scorecards are often written down as a set of “If–Then” business rules. Although they might seem inferior to statistically based scorecards at first sight, they are still quite commonly used in the industry for specific corporate portfolios where no historical data is available.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Business Risk	Score
Industry Position	6
Market Share Trends and Prospects	2
Geographical Diversity of Operations	2
Diversity Product and Services	6
Customer Mix	1
Management Quality and Depth	4
Executive Board Oversight	2
…	…

Exhibit 5.29 Expert-Based Scorecard (Ozdemir and Miu 2009)

Agency Ratings Approach

We already discussed the prediction approach, which assumes the availability of historical data, and the expert-based approach, which assumes the availability of expert-based knowledge or experience. The agency ratings approach is an approach that can be adopted if none of these is available. In other words, if the bank cannot come up with an internal approach to do credit risk assessment, it needs to look externally. Rating agencies are interesting partners to collaborate with in this case since they provide credit ratings for almost any type of nonretail exposure. These ratings typically vary from AAA, which represents excellent credit quality, to AA, A, … and down to D, which represents the default status. The ratings also come with default rates measured across different time horizons such as one, two, three, or even five years. Banks can then purchase these credit ratings to score their nonretail exposures. Popular rating agencies are Moody's, Standard & Poor's, and Fitch. They provide ratings to almost any type of debt or fixed income securities. Examples are ratings for companies (both private and public), countries and governments (sovereign ratings), local authorities, and banks. Retail exposures are typically not covered by the rating agencies. The methodology behind the rating assignment is obviously not disclosed, but it is based on a combination of both quantitative and qualitative modeling. Exhibit 5.30 shows the list of ratings adopted by these three agencies.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Moody's	S&P	Fitch	Credit Quality
Aaa	AAA	AAA	Extremely strong
Aa1	AA+	AA+	
Aa2	AA	AA	Very strong
Aa3	AA–	AA–	
A1	A+	A+	
A2	A	A	Strong
A3	A–	A–	
Baa1	BBB+	BBB+	
Baa2	BBB	BBB	Adequate
Baa3	BBB–	BBB–	
Ba1	BB+	BB+	
Ba2	BB	BB	Speculative
Ba3	BB–	BB–	
B1	B+	B+	
B2	B	B	Highly speculative
B3	B–	B–	
Caa1	CCC+	CCC+	
Caa2	CCC	CCC	Vulnerable
Caa3	CCC–	CCC–	
Ca	CC	CC	Highly vulnerable
C	C	C	Extremely vulnerable
RD	SD	RD	Selective, restrictive default
D	D	D	Default

Exhibit 5.30 Credit Ratings by Moody's, S&P, and Fitch (Van Gestel and Baesens 2009)

Shadow Ratings Approach

The shadow ratings approach starts from a data set with ratings for a particular set of obligors. In a next step, information will be collected for each obligor that might have an influence on the rating. Example data that can be considered in a corporate setting are accounting ratios, firm characteristics, and stock price behavior. The aim is then to combine all this information in one data set (see Exhibit 5.31) and build an analytical model to predict the ratings (Van Gestel et al. 2005, 2007). Exhibit 5.32 shows an example of a decision tree predicting ratings. The advantage of this approach is that we obtain a “white box” understandable model that clearly indicates how the various characteristics of an obligor contribute to the rating. It will

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

also provide clear advice to corporates on how to improve their ratings. Furthermore, in the long term, this approach allows the bank to become independent from the rating agency, since the internal statistical model can now be used to rate any obligor given its characteristics.

Company	Solvency	Liquidity	Stock Price	…	Rating
ABC	10%	66%	100		B
CDE	5%	90%	16		A
DEF	78%	12%	225		A
FGH	24%	58%	88		C
…					

Exhibit 5.31 Example Data Set for the Shadow Rating Approach

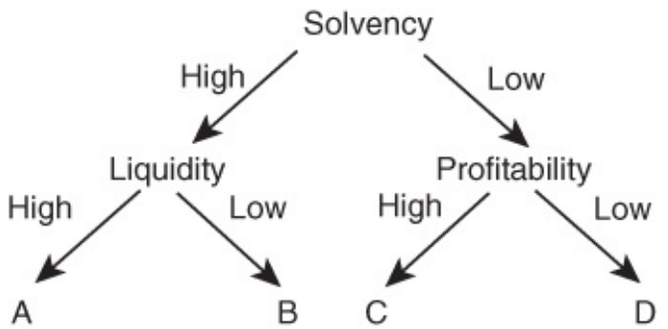

Exhibit 5.32 Example Shadow Rating Model

BIG DATA FOR CREDIT SCORING

Data are everywhere these days. IBM projects that every day we generate 2.5 quintillion bytes of data. In relative terms, this means 90 percent of the data in the world has been created in the past two years. These massive amounts of data yield an unprecedented treasure of information, ready to be analyzed using state of the art analytical techniques to build better credit scorecards. Before we illustrate its potential in a credit scoring context, let's first get a closer look at big data.

Big data is often characterized in terms of its four Vs: volume, variety, velocity, and veracity. To illustrate this, let's briefly zoom in on some key sources or processes generating big data. Traditional sources are large-scale transactional enterprise systems such as online transaction processing (OLTP), enterprise resource planning (ERP), and customer relationship management (CRM) applications. Companies have been deploying these systems for about two decades now. Classic credit scorecards are typically constructed using data extracted from these transactional systems.

The online social graph is a more recent example. Think about the major social networks such as Facebook, Twitter, LinkedIn, Weibo, and WeChat. All together, these networks capture information on close to two billion people about their friends, preferences, and other behavior,

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

leaving a massive digital trail of data. With close to five billion handsets worldwide and with the mobile channel serving as the primary gateway to the Internet in many developed and developing countries, this is another source of big data, as every action taken by the user can be tracked and potentially geo-tagged. Also think about the Internet of Things (IoT) or the emerging sensor-enabled ecosystem that is going to connect various objects (e.g., homes, cars) with each other and with humans. Finally, we see more and more open or public data such as data about weather, traffic, maps, and the macroeconomy.

All these data-generating processes can be characterized in terms of the sheer *volume* of data that is being generated. Clearly, this poses serious challenges in terms of setting up scalable storage architectures combined with a distributed approach to data manipulation and querying.

Big data usually comes in a great *variety* of various formats. Traditional data types or structured data such as customer name, customer birth date, and the like are more and more complemented with unstructured data such as images, fingerprints, tweets, e-mails, Facebook pages, sensor data, and GPS data. Although the former can be easily stored in traditional (e.g., relational) databases, the latter needs to be accommodated using the appropriate database technology facilitating the storage, querying, and manipulation of each of these types of unstructured data. This also requires a substantial effort since it is claimed that at least 80 percent of all data is unstructured.

Velocity refers to the speed at which the data is generated and needs to be stored and analyzed. Think about streaming applications such as online trading platforms, YouTube, SMS messages, credit card swipes, phone calls, and the like, which are all examples where high velocity is a key concern.

Veracity indicates the quality or trustworthiness of the data. Unfortunately, more data does not automatically imply better data, so the quality of the data-generating process must be closely monitored and guaranteed.

As the volume, variety, velocity, and veracity of data continue to grow, so do the new opportunities for building better credit scoring models. Think about Facebook or Twitter as examples. It is quite obvious that knowing a credit applicant's hobbies, followers, friends, likes, education, and workplace could be very beneficial to better quantify his or her creditworthiness. In other words, a customer's social standing, online reputation, and professional connections are likely to be related to his or her credit quality. Another useful data source concerns call detail records (CDR) data, which capture the mobile phone usage of an applicant. Also, surfing behavior could be a nice add-on. Obviously, any privacy concerns surrounding the usage of this data should be properly addressed.

Clearly, the availability of these big data sources creates both opportunities as well as challenges for credit scoring. For example, the availability of social network and CDR data may be beneficial in various settings. First, it may be useful to score customers who lack borrowing experience (e.g., because it's their first loan or they recently moved to a new country) and would be automatically perceived as risky according to traditional credit scoring models that rely on historical information. By using these alternative data sources, a lender can

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

make a better assessment of the credit risk, which can then be translated into a more favorable interest rate. This obviously gives an incentive to the customer to disclose his or her social network, CDR, or other relevant data to the bank. Another example is developing countries. In these countries, banks often lack historical credit information and no local credit bureaus may be available. Other data sources should be used to optimize access to credit. Given the widespread use of social networks and mobile phones (even in developing countries), the data gathered might be an interesting alternative to undertaking credit scoring.

Obviously, using the aforementioned data sources also comes with various challenges. The first one concerns privacy. It is important that customers are properly informed about what data is used to calculate their credit score. An opt-out option should always be provided. Furthermore, using social network data for credit scoring can trigger new default behavior whereby customers strategically construct their social network to artificially and maliciously brush up their credit quality. One example is that customers can easily buy Twitter followers to boost their credit scores. Finally, regulatory compliance might become an important issue. Many countries prohibit the use of gender, age, marital status, national origin, ethnicity, and beliefs for credit scoring. Much of this information can be easily scraped from social networks. It may be harder to oversee regulatory compliance when using social network or other big data for credit scoring.

OVERRIDES

Decisions made by a scorecard may be overruled by human judgment when extra information is present that has not been captured by the scorecard, or because of specific bank policies or strategies. A low-side override or upgrade override occurs when a customer is rejected by the scorecard but accepted anyway because recent information indicates that the customer has improved (or is expected to improve) his solvency status. The default status of the low-side override can then be subsequently tracked in order to determine whether it was the right decision to accept the customer. A high-side override or downgrade override occurs when a customer is accepted by the scorecard but rejected by the credit officer because new information shows, for example, that this customer is expected to change his or her employment status in the near future. Since credit is rejected, the true default status of the customer will never be known unless the customer receives credit elsewhere and his or her default status can be tracked via the credit bureau.

Exhibit 5.33 provides an example of an override report wherein the italic numbers indicate overrides. It is important to note that an excessive number of overrides is a sign that there is no longer confidence in the scorecard, and rebuilding should be considered. Financial regulators discourage financial institutions from doing ad hoc overrides, but instead insist on having clear, well-articulated override policies. Note that an override is sometimes also referred to as an overruling.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Score Range	Accepts	Rejects
<400	*2*	10
400–425	*5*	50
425–450	*10*	80
450–475	*20*	100
475–500	*25*	120
500–525	200	*5*
525–550	500	*5*
550–575	400	*4*
575–600	300	*2*
>600	200	*2*

Exhibit 5.33 Example Override Report for an Application Scorecard with Cutoff Equal to 500

EVALUATING SCORECARD PERFORMANCE

Before bringing a scorecard into production, it needs to be thoroughly evaluated. Depending on the exact setting and usage of the model, different aspects may need to be assessed during evaluation in order to ensure the model is acceptable for implementation. A number of key characteristics of successful scorecards, which may or may not apply depending on the exact application, are defined and explained in Exhibit 5.34.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Statistical accuracy	Refers to the detection power and the correctness of the scorecard in labeling customers as defaulters. Several statistical evaluation criteria exist and may be applied to evaluate this aspect, such as the hit rate, lift curves, area under the curve (AUC), and so on. Statistical accuracy may also refer to statistical significance, meaning that the patterns that have been found in the data have to be valid and not the consequence of noise. In other words, we need to make sure that the model generalizes well and is not overfitted to the historical data set.
Interpretability	A scorecard needs to be interpretable. In other words, a deeper understanding of the detected default behavior is required, for instance to validate the scorecard before it can be used. This aspect involves a certain degree of subjectivism, since interpretability may depend on the credit expert's knowledge. The interpretability of a model depends on its format, which in turn is determined by the adopted analytical technique. Models that allow the user to understand the underlying reasons why the model signals a customer to be a defaulter are called white box models, whereas complex, incomprehensible, mathematical models are often referred to as black box models.
Operational efficiency	Operational efficiency refers to the time that is required to evaluate the scorecard, or in other words the time required to evaluate whether a customer is a defaulter. When customers need to be scored in real time, operational efficiency is crucial and is a main concern during model performance assessment. Operational efficiency also entails the efforts needed to collect and preprocess the data, evaluate the scorecard, monitor and back-test the scorecard, and reestimate it when necessary.
Economical cost	Developing and implementing a scorecard involves a significant cost to an organization. The total cost includes the costs to gather, preprocess, and analyze the data, and the costs to put the resulting scorecards into production. In addition, the software costs as well as human and computing resources should be taken into account. Possibly also external (e.g., credit bureau) data has to be bought to enrich the available in-house data. Clearly it is important to perform a thorough cost-benefit analysis at the start of the credit scoring project, and to gain insight into the constituent factors of the return on investment of building a scorecard system.
Regulatory compliance	A scorecard should be in line and compliant with all applicable regulations and legislation. In a credit scoring setting, the Basel Accords specify what information can or cannot be used and how the target (i.e., default) should be defined. Other regulations (e.g., with respect to privacy and/or discrimination) should also be respected.

Exhibit 5.34 Key Characteristics of Successful Scorecards

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

BUSINESS APPLICATIONS OF CREDIT SCORING

The most important usage of application scores is to decide on loan approval. The scores can also be used for pricing purposes. Risk-based pricing (sometimes also referred to as risk-adjusted pricing) sets the price or other characteristics (e.g., loan term, collateral) of the loan based on the perceived risk as measured by the application score. A lower score will imply a higher interest rate. Hence, subprime loans (e.g., having a FICO score of less than 620) will come with higher rates and fees.

Behavioral scores can be used for various business purposes. First, they can be used for marketing applications. The behavioral scores can be segmented and each of the segments can then be individually approached with targeted mailings. Another usage is for up-, down-, or cross-selling. Up-selling means that you want to sell more of the same product. Think about credit cards or lines of credit, for example. In the case of a good behavioral score and thus low credit risk, the bank may consider increasing the credit limit, thereby generating more revenue. Down-selling means selling less of the same product. So, in case of a bad behavioral score, the bank may consider mitigating its potential loss by lowering the credit limit. Finally, cross-selling means selling other products. For example, if the customer has a good behavioral score on his or her mortgage, the bank may try to sell some additional insurance products.

Although the idea of using behavioral scores to set credit limits sounds reasonable, there has been some debate in the literature about whether this is appropriate. In their book *Credit Scoring and Its Applications* (2002), Thomas et al. argue that one should be careful when using behavioral scores for limit setting. Their reasoning goes as follows. A behavioral credit score is calculated using a given operating policy and credit limit. Hence, using the behavioral score to change the credit limit basically invalidates the effectiveness of the score. To further illustrate this, they came up with an analogy. Suppose it is proposed that only those people who have no or few accidents when they drive a car at 30 miles per hour in town be allowed to drive at 70 miles per hour on the highways. Clearly, this is not a good reasoning since other skills may be required to drive faster. Similarly, it's not because a customer has a good behavioral score with a low credit limit that his or her behavioral score will remain good with a high credit limit, since other characteristics or skills might be needed to manage accounts with large credit limits. However, despite this argument, behavioral scores are commonly used in the industry to manage credit limits. Typically, the behavioral score will be categorized into bands, whereby each band will correspond to a specific credit limit.

Behavioral scores can also be used to authorize accounts to go in excess of their credit limit A gradually decreasing behavioral score could be an early warning signal for looming credit problems, which can be very useful information from a proactive debt collection perspective. It allows time to develop a collection strategy by working out actions that might prevent default.

Both application and behavioral scores will also be used for risk management in a Basel II/III context (Van Gestel and Baesens 2009). More specifically, they will be used as key inputs to estimate the default rate on a loan portfolio, which will then be used to calculate the expected

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

losses (covered by provisions) and unexpected losses (covered by capital). They can also be helpful for securitization purposes by slicing and dicing a credit portfolio into tranches with similar risk.

Besides financial institutions, other organizations can also use credit scores to support their business decisions. For example, electricity and telecommunications companies can use credit scores in their pricing or contracting policies. Employers can use them to get a better idea of the profiles of job applicants, while landlords can get a better idea about the solvency of their future renters. Insurance companies can use credit scores to set insurance premiums or decide for whom to accept the insurance policy. Note that some of these applications are controversial and subject to debate.

The widespread use of both application and behavioral scorecards has made them a key decision support tool in modern risk measurement and management.

LIMITATIONS

Although credit scoring systems are being implemented and used by most banks nowadays, they do face a number of limitations. A first limitation concerns the data that is used to estimate credit scoring models. Since data is the major, and in most cases the only, ingredient to build these models, its quality and predictive ability is key to the models' success. The quality of the data refers, for example, to the number of missing values and outliers, and to the recency and representativity of the data. Data quality issues can be difficult to detect without specific domain knowledge, but have an important impact on the scorecard development and resulting risk measures. The availability of high-quality data is a very important prerequisite for building good credit scoring models. However, not only does the data need to be of high quality, but it should be predictive as well, in the sense that the captured characteristics are related to the customer's likelihood of defaulting. Before constructing a scorecard, we need to thoroughly reflect on why a customer defaults and which characteristics could potentially be related to this. Customers may default because of unknown reasons or information not available to the financial institution, thereby posing another limitation to the performance of credit scoring models. The statistical techniques used in developing credit scoring models typically assume a data set of sufficient size containing enough defaults. This may not always be the case for specific types of portfolios where only limited data is available, or only a low number of defaults is observed. For these types of portfolios, one may have to rely on alternative risk assessment methods using, for example, expert judgment based on the five Cs, as we discussed earlier.

Financial institutions should also be aware that scorecards have only a limited lifetime. The populations on which they were estimated will typically vary throughout time because of changing economic conditions or new strategic actions (e.g., new customer segments targeted, new credit products introduced) undertaken by the bank. This is often referred to as population drift and will necessitate the financial institution rebuilding its scorecards if the default risk in the new population is totally different from the one present in the population that was used to

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

build the old scorecards.

Many credit bureaus nowadays start disclosing how their bureau scores (e.g., FICO scores) are computed in order to encourage customers to improve their financial profiles, and hence increase their success in getting credit. Since this gives customers the tools to polish up their scores and make them look good in future credit applications, this may trigger new types of default risk (and fraud), thereby invalidating the original scorecard and necessitating more frequent rebuilds.

Introducing credit scoring into an organization requires serious investments in information and communications technology (ICT) hardware and software, personnel training, and support facilities. The total cost needs to be carefully considered beforehand and compared with future benefits, which may be hard to quantify.

Finally, a last criticism concerns the fact that most credit scoring systems model only default risk (i.e., the risk that a customer runs into payment arrears on one of his or her financial obligations). Default risk is, however, only one type of credit risk. Besides default risk, credit risk also entails recovery risk and exposure risk.

PRACTICE QUESTIONS

1. Contrast the judgmental approach and the statistical approach to credit scoring. Give examples of situations where one is to be preferred over the other.
2. Contrast logistic regression to decision trees in terms of model formulation, model representation, and decision boundary. Give examples of situations where one is to be preferred over the other.
3. Discuss the criteria that should be considered when performing variable selection for credit scoring.
4. What decisions should be made when building a decision tree?
5. Discuss the approaches available to do credit scoring for retail portfolios.
6. Give some examples of modeling decisions that need to be made when building an application scorecard.
7. Discuss some issues that arise when defining variables for behavioral scoring.
8. What is reject inference? How can it be dealt with?
9. Discuss the approaches available to do credit scoring for nonretail portfolios.
10. Discuss the potential and risks of using social media data (e.g., Facebook, Twitter) for credit scoring.
11. What are the key characteristics of successful scorecards? Discuss which ones are important in what situations.
12. Discuss who is allowed to use credit scores in your country (e.g., utility companies,

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

telecommunications firms, employers, landlords, etc.).

13. What is the importance of data quality in the context of credit scoring?

REFERENCES

Allison, P. D. 2001. *Logistic Regression Using the SAS System: Theory and Application*. New York: John Wiley & Sons–SAS.

Baesens, B., T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. 2003. "Benchmarking State of the Art Classification Algorithms for Credit Scoring." *Journal of the Operational Research Society* 54 (6): 627–635.

Baesens, B. 2014. *Analytics in a Big Data World*, Wiley.

Banasik, J., J. N. Crook, and L. C. Thomas. 2001. "Sample Selection Bias in Credit Scoring Models." In *Proceedings of the Seventh Conference on Credit Scoring and Credit Control (CSCCVII'2001)*, Edinburgh, Scotland.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Duda, R. O., P. E. Hart, and D. G. Stork. 2001. *Pattern Classification*. New York: John Wiley & Sons.

Hand, D. J. 2001. "Reject Inference in Credit Operations: Theory and Methods." In *Handbook of Credit Scoring*, edited by Elizabeth Mays. Chicago: Glenlake Publishing Company.

Hartigan, J. A. 1975. *Clustering Algorithms*. New York: John Wiley & Sons.

Martens, D., B. Baesens, T. Van Gestel, and J. Vanthienen. 2007. "Comprehensible Credit Scoring Models Using Rule Extraction from Support Vector Machines." *European Journal of Operational Research* 183:1466–1476.

Ozdemir, B., and P. Miu. 2009. *Basel II Implementation: A Guide to Developing and Validating a Compliant, Internal Risk Rating System*. New York: McGraw-Hill.

Quinlan, J. R. 1993. *C4.5 Programs for Machine Learning*. San Francisco: Morgan Kaufmann.

Thomas, L. C., D. B. Edelman, and J. N. Crook. 2002. *Credit Scoring and Its Applications*. Monographs on Mathematical Modeling and Computation. Philadelphia: Society for Industrial and Applied Mathematics.

Van Gestel, T., and B. Baesens. 2009. *Credit Risk Management: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. Oxford: Oxford University Press.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Van Gestel, T., B. Baesens, P. Van Dijcke, J. Suykens, J. Garcia, and T. Alderweireld. 2005. “Linear and Nonlinear Credit Scoring by Combining Logistic Regression and Support Vector Machines.” *Journal of Credit Risk* 1, no. 4.

Van Gestel, T., D. Martens, B. Baesens, D. Feremans, J. Huysmans, and J. Vanthienen. 2007. “Forecasting and Analyzing Insurance Companies' Ratings.” *International Journal of Forecasting* 23 (3): 513–529.

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.

Chapter 6
Probabilities of Default (PD): Discrete-Time Hazard Models

INTRODUCTION

In the previous chapter, we saw that credit scores are indicators of the credit risk of borrowers. Default probabilities or probabilities of default (PDs) are in essence credit scores that are standardized likelihood measures with a range between zero and one, whereby zero implies that an event is impossible to occur and one implies certainty. Realistic models generally assign PDs between zero and 30 percent to loans.

The PD is the most scrutinized parameter in credit risk analytics and subject to minimum standards imposed by prudential regulators. For example, banks are required to include and exclude specific risk factors. Furthermore, minimum floors such as three basis points are imposed (often to have a nonzero PD value for low default portfolios, which is a subject that we explore later). Banks are also required to validate their PD estimates with rigorous tests, which we discuss in the validation chapter.

Default Events

A PD describes the likelihood of a default event. Banks observe whether borrowers default, and generally indicate this with a default indicator:

$$D_{it} = \begin{cases} 1 & \text{borrower } i \text{ defaults at time } t \\ 0 & \text{otherwise} \end{cases}$$

with $i = 1, \ldots, I$ and $t = 1, \ldots, T$. We assume that the default event is random and use an uppercase letter D as the random variable and a lowercase letter d as its realization. A default event may be defined by any of the following events:

- Payment delinquency of a number of days or more; popular thresholds are 30, 60, and 90 days
- Bankruptcy of the borrower
- Collateral owned by a bank (e.g., real estate owned after an unsuccessful sale at a foreclosure auction)
- Foreclosure of loan
- Short sale of loan
- Loss/write-down amount

Copyright © 2016. John Wiley & Sons, Incorporated. All rights reserved.

Baesens, Bart, et al. Credit Risk Analytics : Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, Incorporated, 2016. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/nmims-ebooks/detail.action?docID=4694230.
Created from nmims-ebooks on 2023-07-05 05:50:51.