OeNB

FMA

Guidelines on Credit Risk Management

Rating Models and Validation

These guidelines were prepared by the Oesterreichische Nationalbank (OeNB) in cooperation with the Financial Market Authority (FMA)

Published by:

Oesterreichische Nationalbank (OeNB)
Otto Wagner Platz 3, 1090 Vienna, Austria
Austrian Financial Market Authority (FMA)
Praterstrasse 23, 1020 Vienna, Austria

Produced by:

Oesterreichische Nationalbank

Editor in chief:

Günther Thonabauer, Secretariat of the Governing Board and Public Relations (OeNB)
Barbara Nösslinger, Staff Department for Executive Board Affairs and Public Relations (FMA)

Editorial processing:

Doris Datschetzky, Yi-Der Kuo, Alexander Tscherteu, (all OeNB)
Thomas Hudetz, Ursula Hauser-Rethaller (all FMA)

Design:

Peter Buchegger, Secretariat of the Governing Board and Public Relations (OeNB)

Typesetting, printing, and production:

OeNB Printing Office

Published and produced at:

Otto Wagner Platz 3, 1090 Vienna, Austria

Inquiries:

Oesterreichische Nationalbank
Secretariat of the Governing Board and Public Relations
Otto Wagner Platz 3, 1090 Vienna, Austria
Postal address: PO Box 61, 1011 Vienna, Austria
Phone: (+43-1) 40 420-6666
Fax: (+43-1) 404 20-6696

Orders:

Oesterreichische Nationalbank
Documentation Management and Communication Systems
Otto Wagner Platz 3, 1090 Vienna, Austria
Postal address: PO Box 61, 1011 Vienna, Austria
Phone: (+43-1) 404 20-2345
Fax: (+43-1) 404 20-2398

Internet:

http://www.oenb.at
http://www.fma.gv.at

Paper:

Salzer Demeter, 100% woodpulp paper, bleached without chlorine, acid-free, without optical whiteners

DVR 0031577

Preface

The ongoing development of contemporary risk management methods and the increased use of innovative financial products such as securitization and credit derivatives have brought about substantial changes in the business environment faced by credit institutions today. Especially in the field of lending, these changes and innovations are now forcing banks to adapt their in-house software systems and the relevant business processes to meet these new requirements.

The **OeNB Guidelines on Credit Risk Management** are intended to assist practitioners in redesigning a bank's systems and processes in the course of implementing the Basel II framework.

Throughout 2004 and 2005, OeNB guidelines will appear on the subjects of securitization, rating and validation, credit approval processes and management, as well as credit risk mitigation techniques. The content of these guidelines is based on current international developments in the banking field and is meant to provide readers with best practices which banks would be well advised to implement regardless of the emergence of new regulatory capital requirements.

The purpose of these publications is to develop mutual understanding between regulatory authorities and banks with regard to the upcoming changes in banking. In this context, the Oesterreichische Nationalbank (OeNB), Austria's central bank, and the Austrian Financial Market Authority (FMA) see themselves as partners to Austria's credit industry.

It is our sincere hope that the OeNB Guidelines on Credit Risk Management provide interesting reading as well as a basis for efficient discussions of the current changes in Austrian banking.

Vienna, November 2004

Univ. Doz. Mag. Dr. Josef Christl
Member of the Governing Board
of the Oesterreichische Nationalbank

Dr. Kurt Pribil,
Dr. Heinrich Traumüller
FMA Executive Board

Contents

I	**INTRODUCTION**	7
II	**ESTIMATING AND VALIDATING PROBABILITY OF DEFAULT (PD)**	8
1	**Defining Segments for Credit Assessment**	8
2	**Best-Practice Data Requirements for Credit Assessment**	11
2.1	Governments and the Public Sector	12
2.2	Financial Service Providers	15
2.3	Corporate Customers – Enterprises/Business Owners	17
2.4	Corporate Customers – Specialized Lending	22
2.4.1	*Project Finance*	24
2.4.2	*Object Finance*	25
2.4.3	*Commodities Finance*	26
2.4.4	*Income-Producing Real Estate Financing*	26
2.5	Retail Customers	28
2.5.1	*Mass-Market Banking*	28
2.5.2	*Private Banking*	31
3	**Commonly Used Credit Assessment Models**	32
3.1	Heuristic Models	33
3.1.1	*"Classic" Rating Questionnaires*	33
3.1.2	*Qualitative Systems*	34
3.1.3	*Expert Systems*	36
3.1.4	*Fuzzy Logic Systems*	38
3.2	Statistical Models	40
3.2.1	*Multivariate Discriminant Analysis*	41
3.2.2	*Regression Models*	43
3.2.3	*Artificial Neural Networks*	45
3.3	Causal Models	48
3.3.1	*Option Pricing Models*	48
3.3.2	*Cash Flow (Simulation) Models*	49
3.4	Hybrid Forms	50
3.4.1	*Horizontal Linking of Model Types*	51
3.4.2	*Vertical Linking of Model Types Using Overrides*	52
3.4.3	*Upstream Inclusion of Heuristic Knock-Out Criteria*	53
4	**Assessing the Models' Suitability for Various Rating Segments**	54
4.1	Fulfillment of Essential Requirements	54
4.1.1	*PD as Target Value*	54
4.1.2	*Completeness*	55
4.1.3	*Objectivity*	55
4.1.4	*Acceptance*	56
4.1.5	*Consistency*	57
4.2	Suitability of Individual Model Types	57
4.2.1	*Heuristic Models*	57
4.2.2	*Statistical Models*	58
4.2.3	*Causal Models*	60

5	**Developing a Rating Model**	60
5.1	Generating the Data Set	62
5.1.1	*Data Requirements and Sources*	62
5.1.2	*Data Collection and Cleansing*	64
5.1.3	*Definition of the Sample*	72
5.2	Developing the Scoring Function	82
5.2.1	*Univariate Analyses*	75
5.2.2	*Multivariate Analysis*	80
5.2.3	*Overall Scoring Function*	82
5.3	Calibrating the Rating Model	84
5.3.1	*Calibration for Logistic Regression*	85
5.3.2	*Calibration in Standard Cases*	86
5.4	Transition Matrices	88
5.4.1	*The One-Year Transition Matrix*	88
5.4.2	*Multi-Year Transition Matrices*	91
6	**Validating Rating Models**	94
6.1	Qualitative Validation	96
6.2	Quantitative Validation	98
6.2.1	*Discriminatory Power*	98
6.2.2	*Back-Testing the Calibration*	115
6.2.3	*Back-Testing Transition Matrices*	132
6.2.4	*Stability*	134
6.3	Benchmarking	128
6.4	Stress Tests	130
6.4.1	*Definition and Necessity of Stress Tests*	130
6.4.2	*Essential Factors in Stress Tests*	131
6.4.3	*Developing Stress Tests*	133
6.4.4	*Performing and Evaluating Stress Tests*	137
III	**ESTIMATING AND VALIDATING LGD/EAD AS RISK COMPONENTS**	139
7	**Estimating Loss Given Default (LGD)**	139
7.1	Definition of Loss	140
7.2	Parameters for LGD Calculation	140
7.2.1	*LGD-Specific Loss Components in Non-Retail Transactions*	140
7.2.2	*LGD-Specific Loss Components in Retail Transactions*	143
7.3	Identifying Information Carriers for Loss Parameters	144
7.3.1	*Information Carriers for Specific Loss Parameters*	144
7.3.2	*Customer Types*	146
7.3.3	*Types of Collateral*	148
7.3.4	*Types of Transaction*	149
7.3.5	*Linking of Collateral Types and Customer Types*	150
7.4	Methods of Estimating LGD Parameters	151
7.4.1	*Top-Down Approaches*	151
7.4.2	*Bottom-Up Approaches*	153
7.5	Developing an LGD Estimation Model	157

8	**Estimating Exposure at Default (EAD)**	162
8.1	Transaction Types	162
8.2	Customer Types	163
8.3	EAD Estimation Methods	165
IV	**REFERENCES**	167
V	**FURTHER READING**	170

I INTRODUCTION

The OeNB Guideline on Rating Models and Validation was created within a series of publications produced jointly by the Austrian Financial Markets Authority and the Oesterreichische Nationalbank on the topic of credit risk identification and analysis. This set of guidelines was created in response to two important developments: First, banks are becoming increasingly interested in the continued development and improvement of their risk measurement methods and procedures. Second, the Basel Committee on Banking Supervision as well as the European Commission have devised regulatory standards under the heading "Basel II" for banks' in-house estimation of the loss parameters probability of default (PD), loss given default (LGD), and exposure at default (EAD). Once implemented appropriately, these new regulatory standards should enable banks to use IRB approaches to calculate their regulatory capital requirements, presumably from the end of 2006 onward. Therefore, these guidelines are intended not only for credit institutions which plan to use an IRB approach but also for all banks which aim to use their own PD, LGD, and/or EAD estimates in order to improve assessments of their risk situation.

The objective of this document is to assist banks in developing their own estimation procedures by providing an overview of current best-practice approaches in the field. In particular, the guidelines provide answers to the following questions:

- Which segments (business areas/customers) should be defined?
- Which input parameters/data are required to estimate these parameters in a given segment?
- Which models/methods are best suited to a given segment?
- Which procedures should be applied in order to validate and calibrate models?

In part II, we present the special requirements involved in PD estimation procedures. First, we discuss the customer segments relevant to credit assessment in chapter 1. On this basis, chapter 2 covers the resulting data requirements for credit assessment. Chapter 3 then briefly presents credit assessment models which are commonly used in the market. In Chapter 4, we evaluate these models in terms of their suitability for the segments identified in chapter 1. Chapter 5 discusses how rating models are developed, and part II concludes with chapter 6, which presents information relevant to validating estimation procedures. Part III provides a supplement to Part II by presenting the specific requirements for estimating LGD (chapter 7) and EAD (chapter 8). Additional literature and references are provided at the end of the document.

Finally, we would like to point out that these guidelines are only intended to be descriptive and informative in nature. They cannot (and are not meant to) make any statements on the regulatory requirements imposed on credit institutions dealing with rating models and their validation, nor are they meant to prejudice the regulatory activities of the competent authorities. References to the draft EU directive on regulatory capital requirements are based on the latest version available when these guidelines were written (i.e. the draft released on July 1, 2003) and are intended for information purposes only. Although this document has been prepared with the utmost care, the publishers cannot assume any responsibility or liability for its content.

II ESTIMATING AND VALIDATING PROBABILITY OF DEFAULT (PD)

1 Defining Segments for Credit Assessment

Credit assessments are meant to help a bank measure whether potential borrowers will be able to meet their loan obligations in accordance with contractual agreements. However, a credit institution cannot perform credit assessments in the same way for all of its borrowers.

This point is supported by three main arguments, which will be explained in greater detail below:

1. The factors relevant to creditworthiness vary for different borrower types.
2. The available data sources vary for different borrower types.
3. Credit risk levels vary for different borrower types.

Ad 1.
Wherever possible, credit assessment procedures must include all *data and information relevant to creditworthiness.* However, the factors determining creditworthiness will vary according to the type of borrower concerned, which means that it would not make sense to define a uniform data set for a bank's entire credit portfolio. For example, the credit quality of a government depends largely on macroeconomic indicators, while a company will be assessed on the basis of the quality of its management, among other things.

Ad 2.
Completely *different data sources* are available for various types of borrowers. For example, the bank can use the annual financial statements of companies which prepare balance sheets in order to assess their credit quality, whereas this is not possible in the case of retail customers. In the latter case, it is necessary to gather analogous data, for example by requesting information on assets and liabilities from the customers themselves.

Ad 3.
Empirical evidence shows that average default rates vary widely for different types of borrowers. For example, governments exhibit far lower default rates than business enterprises. Therefore, banks should account for these *varying levels of risk* in credit assessment by segmenting their credit portfolios accordingly. This also makes it possible to adapt the intensity of credit assessment according to the risk involved in each segment.

Segmenting the credit portfolio is thus a basic prerequisite for assessing the creditworthiness of all a bank's borrowers based on the specific risk involved.

On the basis of *business considerations*, we distinguish between the following general segments in practice:

- Governments and the public sector
- Financial service providers
- Corporate customers
  - Enterprises/business owners
  - Specialized lending
- Retail customers

This segmentation from the business perspective is generally congruent with the *regulatory categorization of assets in the IRB approach under Basel II and the draft EU directive*:[1]

- Sovereigns/central governments
- Banks/institutions
- Corporates
  - Subsegment: Specialized lending
- Retail customers
- Equity

Due to its highly specific characteristics, the equity segment is not discussed in detail in this document.

However, as the above-mentioned general segments themselves are generally not homogeneous, a more *specific segmentation* is necessary (see chart 1).

One conspicuous feature of our best-practice segmentation is its inclusion of product elements in the retail customer segment. In addition to borrower-specific creditworthiness factors, transaction-specific factors are also attributed importance in this segment. Further information on this special feature can be found in Section 2.5, Retail Customers, where in particular its relationship to Basel II and the draft EU directive is discussed.

[1] *EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 47, No. 1–9.*

Governments and the Public Sector	Central Governments
	Public Sector
	Regional Governments
	Local Authorities
	Public Sector Entities
Financial Service Providers	Credit Institutions
	Insurance Companies
	Misc. Financial Institutions
Corporate Customers	Enterprises / Business Owners
	Capital-Market-Oriented / International Companies
	Other enterprises which prepare balance sheets
	Businesses and Independent Professionals (not preparing balance-sheets)
	Small Businesses
	Start-ups
	Pre-launch Stage
	Post-launch Stage
	NPOs (Non-profit Organizations)
	Specialized Lending
	Project Finance
	Before the Project
	During the Project
	Object Finance
	Before the Project
	During the Project
	Commodities Finance
	Financing of income-producing real estate
	Before the Project
	During the Project
Retail Customers	Mass-Market Banking
	Current Accounts
	Upon Credit Application
	During the Credit Term
	Consumer Loans
	Upon Credit Application
	During the Credit Term
	Credit Cards
	Upon Credit Application
	During the Credit Term
	Residential Construction Loans
	Upon Credit Application
	During the Credit Term
	Private Banking

Chart 1: Best-Practice Segmentation

The best-practice segmentation presented here on the basis of individual loans and credit facilities for retail customers reflects customary practice in banks, that is, scoring procedures for calculating the PD of individual customers usually already exist in the retail customer segment.

The draft EU directive contains provisions which ease the burden of risk measurement in the retail customer segment. For instance, retail customers do not have to be assessed individually using rating procedures; they can be assigned to pools according to specific borrower and product characteristics. The risk components PD, LGD, and EAD are estimated separately for these pools and then assigned to the individual borrowers in the pools.

Although the approach provided for in Basel II is not discussed in greater detail in this document, this is not intended to restrict a bank's alternative courses of action in any way. A pool approach can serve as an alternative or a supplement to best practices in the retail segment.

2 Best-Practice Data Requirements for Credit Assessment

The previous chapter pointed out the necessity of defining segments for credit assessment and presented a segmentation approach which is commonly used in practice. Two essential reasons for segmentation are the different factors relevant to creditworthiness and the varying availability of data in individual segments.

The relevant data and information categories are presented below with attention to their actual availability in the defined segments. In this context, the data categories indicated for individual segments are to be understood as part of a best-practice approach, as is the case throughout this document. They are intended not as compulsory or minimum requirements, but as an orientation aid to indicate which data categories would ideally be included in rating development. In our discussion of these information categories, we deliberately confine ourselves to a highly aggregated level. We do not attempt to present individual rating criteria. Such a presentation could never be complete due to the huge variety of possibilities in individual data categories. Furthermore, these guidelines are meant to provide credit institutions with as much latitude as possible in developing their own rating models.

The data necessary for all segments can first be subdivided into three data types:

Quantitative Data/Information

This type of data generally refers to objectively measurable numerical values. The values themselves are categorized as quantitative data related to the past/present or future. Past and present quantitative data refer to actual recorded values; examples include annual financial statements, bank account activity data, or credit card transactions.

Future quantitative data refer to values projected on the basis of actual numerical values. Examples of these data include cash flow forecasts or budget calculations.

Qualitative Data/Information
This type is likewise subdivided into qualitative data related to the past/present or to the future.

Past or present qualitative data are subjective estimates for certain data fields expressed in ordinal (as opposed to metric) terms. These estimates are based on knowledge gained in the past. Examples of these data include assessments of business policies, of the business owner's personality, or of the industry in which a business operates.

Future qualitative data are projected values which cannot currently be expressed in concrete figures. Examples of these data include business strategies, assessments of future business development or appraisals of a business idea.

Within the bank, possible *sources* of quantitative and qualitative data include:
– Operational systems
– IT centers
– Miscellaneous IT applications (including those used locally at individual workstations)
– Files and archives

External Data/Information
In contrast to the two categories discussed above, this data type refers to information which the bank cannot gather internally on the basis of customer relationships but which has to be acquired from external information providers. Possible sources of external data include:
– Public agencies (e.g. statistics offices)
– Commercial data providers (e.g. external rating agencies, credit reporting agencies)
– Other data sources (e.g. freely available capital market information, exchange prices, or other published information)

The information categories which are generally relevant to rating development are defined on the basis of these three data types. However, as the data are not always completely available for all segments, and as they are not equally relevant to creditworthiness, the relevant data categories are identified for each segment and shown in the tables below.

These tables are presented in succession for the four general segments mentioned above (governments and the public sector, financial service providers, corporate customers, and retail customers), after which the individual data categories are explained for each subsegment.

2.1 Governments and the Public Sector

In general, banks do not have internal information on central governments, central banks, and regional governments as borrowers. Therefore, it is necessary to extract creditworthiness-related information from external data sources. In contrast, the availability of data on local authorities and public sector entities certainly allows banks to consider them individually using in-house data.

Central Governments
Central governments are subjected to credit assessment by external rating agencies and are thus assigned external country ratings. As external rating agencies

			Central governments	Regional governments	Local authorities	Public sector entities
Quantitative Data / Information	Based on past or present development	Annual financial statements				
		Customer account management				
		Income & expense accounts				
		Tax returns				
		Assets / liabilities (information from customer)				
		Equity/grant funding/financing needs				
		Collateral value of project assets (realization proceeds)				
		Collateral value of object financed (realization proceeds)				
		Fair market value of real estate				
		Rent levels for comparable properties at same location				
		Development of fair market value and rent levels (historical/expected)				
		Liquidity and revenue development				
		Target/actual comparison of project progress				
		Target/actual comparison of cash flows				
		Target/actual comparison of construction progress				
		Credit stage / time to maturity				
		Credit card transactions				
		Purchasing behavior				
	Based on future development	Cash flow forecasts				
		Debt service capacity				
		Budget calculation				
		Expected sale proceeds (collateral value)				
Qualitative Daten/ Informationen	Based on past or present development	Business areas (customers/products)				
		Risk structure / risk management.				
		Business policies				
		Industry information				
		Enterprise/management				
		Market (customers/products)				
		Entrepreneur's personal traits				
		Type of project		✓	✓	✓
		Ownership structure / owners' creditworthiness				
		Liability relationships				
		Type of object				
		Type of commodity				
		Structure of financing transaction				
		Type of real estate				
		Socio-demographic data				
		Type and history of customer relationship to date				
		Information on living environment				
		Distribution channel				
		Purpose				
		Fulfillment of requirements				
		Occupancy rate				
		Further development of customer relationship				
		Payment reminder behavior				
		Payment behavior			✓	✓
		Special agreements			✓	✓
		General legal conditions (group ties)				✓
		Distribution of tasks at local government level		✓	✓	✓
	Based on future development	Planning and future company development				
		Business idea / business plan				
External Data / Information	Capital market information	Stock price / price development	✓	✓	✓	
		Public relations/public information	✓	✓	✓	
		Current exchange price				
		Price development (historical/expected)				
	Miscellaneous external information	Country rating (sovereigns)	✓	✓	✓	✓
		Form of government	✓	✓	✓	✓
		General legal and political conditions	✓	✓	✓	✓
		Macroeconomic indicators	✓	✓	✓	✓
		External rating	✓	✓	✓	✓
		Infrastructure	✓	✓	✓	
		Credit reporting information				
		Creditworthiness of potential tenant(s)/lessee(s)/buyer(s				
		Budget situation	✓	✓	✓	✓

Chart 2: Data Requirements for Governments and the Public Sector

perform comprehensive analyses in this process with due attention to the essential factors relevant to creditworthiness, we can regard country ratings as the primary source of information for credit assessment. This external credit assessment should be supplemented by observations and assessments of macroeconomic indicators (e.g. GDP and unemployment figures as well as business cycles) for each country. Experience on the capital markets over the last few decades has shown that the repayment of loans to governments and the redemption of government bonds depend heavily on the legal and political stability of the country in question. Therefore, it is also important to consider the form of government as well as its general legal and political situation. Additional external data which can be used include the development of government bond prices and published capital market information.

Regional Governments
This category refers to the individual political units within a country (e.g. states, provinces, etc.). Regional governments and their respective federal governments often have a close liability relationship, which means that if a regional government is threatened with insolvency the federal government will step in to repay the debt. In this way, the credit quality of the federal government also plays a significant role in credit assessments for regional governments, meaning that the country rating of the government to which a regional government belongs is an essential criterion in its credit assessment. However, when the creditworthiness of a regional government is assessed, its own external rating (if available) also has to be taken into account. A supplementary analysis of macroeconomic indicators for the regional government is also necessary in this context. The financial and economic strength of a regional government can be measured on the basis of its budget situation and infrastructure. As the general legal and political circumstances in a regional government can sometimes differ substantially from those of the country to which it belongs, lending institutions should also perform a separate assessment in this area.

Local Authorities
The information categories relevant to the creditworthiness of local authorities do not diverge substantially from those applying to regional governments. However, it is entirely possible that individual criteria within these categories will be different for regional governments and local authorities due to the different scales of their economies.

Public Sector Entities
As public sector entities are also part of the "Other public agencies" sector, their credit assessment should also rely on a data set similar to the one used for regional governments and local authorities. However, such assessments should also take any possible group interdependences into account, as such relationships may have a substantial impact on the repayment of loans in the "Public sector entities" segment. In some cases, data which is generally typical of business enterprises will contain relevant information and should be used accordingly.

2.2 Financial Service Providers

In this context, *financial service providers* include credit institutions (e.g. banks, building and loan associations, investment fund management companies), insurance companies and financial institutions (e.g. leasing companies, asset management companies).

For the purpose of rating financial service providers, credit institutions will generally have more in-house quantitative and qualitative data at their disposal than in the case of borrowers in the "Governments and the public sector" segment. In order to gain a complete picture of a financial service provider's creditworthiness, however, lenders should also include external information in their credit assessments.

In practice, separate in-house rating models are rarely developed specifically for insurance companies and financial institutions. Instead, the rating models developed for credit institutions or corporate customers can be modified and employed accordingly.

Credit institutions

One essential source of quantitative information for the assessment of a credit institution is its annual financial statements. However, financial statements only provide information on the organization's past business success. For the purpose of credit assessment, however, the organization's future ability and willingness to pay are decisive factors which means that credit assessments should be supplemented with cash flow forecasts. Only on the basis of these forecasts is it possible to establish whether the credit institution will be able to meet its future payment obligations arising from loans. Cash flow forecasts should be accompanied by a qualitative assessment of the credit institution's future development and planning. This will enable the lending institution to review how realistic its cash flow forecasts are.

Another essential qualitative information category is the credit institution's risk structure and risk management. In recent years, credit institutions have mainly experienced payment difficulties due to deficiencies in risk management. This is one of the main reasons why the Basel II Committee decided to develop new regulatory requirements for the treatment of credit risk. In this context, it is also important to take group interdependences and any resulting liability obligations into account.

In addition to the risk side, however, the income side also has to be examined in qualitative terms. In this context, analysts should assess whether the credit institution's specific policies in each business area will also enable the institution to satisfy customer needs and to generate revenue streams in the future.

Finally, lenders should also include external information (if available) in their credit assessments in order to obtain a complete picture of a credit institution's creditworthiness. This information may include external ratings of the credit institution, the development of its stock price, or other published information (e.g. ad hoc reports). The rating of the country in which the credit institution is domiciled deserves special consideration in the case of credit institutions for which the government has assumed liability.

			Credit institutions	Insurance companies	Misc. financial institutions
Quantitative Data / Information	Based on past or present development	Annual financial statements	✓	✓	✓
		Customer account management			
		Income & expense accounts			
		Tax returns			
		Assets / liabilities (information from customer)			
		Equity/grant funding/financing needs			
		Collateral value of project assets (realization proceeds)			
		Collateral value of object financed (realization proceeds)			
		Fair market value of real estate			
		Rent levels for comparable properties at same location			
		Development of fair market value and rent levels (historical/expected)			
		Liquidity and revenue development			
		Target/actual comparison of project progress			
		Target/actual comparison of cash flows			
		Target/actual comparison of construction progress			
		Credit stage / time to maturity			
		Credit card transactions			
		Purchasing behavior			
	Based on future development	Cash flow forecasts	✓	✓	✓
		Debt service capacity			
		Budget calculation			
		Expected sale proceeds (collateral value)			
Qualitative Daten/ Informationen	Based on past or present development	Business areas (customers/products)	✓	✓	✓
		Risk structure / risk management.	✓	✓	✓
		Business policies	✓	✓	✓
		Industry information			
		Enterprise/management	✓	✓	✓
		Market (customers/products)	✓	✓	✓
		Entrepreneur's personal traits			
		Type of project			
		Ownership structure / owners' creditworthiness	✓	✓	✓
		Liability relationships	✓	✓	✓
		Type of object			
		Type of commodity			
		Structure of financing transaction			
		Type of real estate			
		Socio-demographic data			
		Type and history of customer relationship to date			
		Information on living environment			
		Distribution channel			
		Purpose			
		Fulfillment of requirements			
		Occupancy rate			
		Further development of customer relationship			
		Payment reminder behavior			
		Payment behavior			
		Special agreements			
		General legal conditions (group ties)	✓	✓	✓
		Distribution of tasks at local government level			
	Based on future development	Planning and future company development	✓	✓	✓
		Business idea / business plan			
External Data / Information	Capital market information	Stock price / price development	✓	✓	✓
		Public relations/public information	✓	✓	✓
		Current exchange price			
		Price development (historical/expected)			
	Miscellaneous external information	Country rating (sovereigns)	✓	✓	✓
		Form of government			
		General legal and political conditions			
		Macroeconomic indicators			
		External rating	✓	✓	✓
		Infrastructure			
		Credit reporting information			
		Creditworthiness of potential tenant(s)/lessee(s)/buyer(s)			
		Budget situation			

Chart 3: Data Requirements for Financial Service Providers

Insurance Companies
Due to their different business orientation, insurance companies have to be assessed using different creditworthiness criteria from those used for credit institutions. However, the existing similarities between these institutions mean that many of the same information categories also apply to insurers.

Financial institutions
Financial institutions, or "other financial service providers," are similar to credit institutions. However, the specific credit assessment criteria taken into consideration may be different for financial institutions. For example, asset management companies which only act as advisors and intermediaries but to do not grant loans themselves will have an entirely different risk structure to that of credit institutions. Such differences should be taken into consideration in the different credit assessment procedures for the subsegments within the "financial service providers" segment.

However, it is not absolutely necessary to develop an entirely new rating procedure for financial institutions. Instead, it may be sufficient to use an adapted version of the rating model applied to credit institutions. It may also be possible to assess certain financial institutions with a modified corporate customer rating model, which would change the data requirements accordingly.

2.3 Corporate Customers – Enterprises/Business Owners

The general segment "Corporate Customers – Enterprises/Business Owners" can be subdivided into the following subsegments:

- Capital market-oriented[2]/international companies
- Other companies which prepare balance sheets
- Businesses and independent professionals (not preparing balance sheets)
- Small businesses
- Start-ups
- NPOs (non-profit organizations)

The first four subsegments consist of enterprises which have already been on the market for some time. These enterprises differ in size and thus also in terms of the available data categories.

In the case of start-ups, the information available will be very depending on the enterprise's current stage of development and should be taken into account accordingly.

The main differentiating criterion in the case of NPOs is the fact that they are not operated for the purpose of making a profit.

Moreover, it is common practice in the corporate segment to develop separate rating models for various countries and regions (e.g. for enterprises in CEE countries). Among other things, these models take the accounting standards applicable in individual countries into consideration.

[2] *Capital market-oriented means that the company funds itself (at least in part) by means of capital market instruments (stocks, bonds, securitization).*

			Corporate customers	Enterprises / Business Owners	Capital market-oriented / international companie	Other companies which prepare balance sheets	Enterprises not preparing balance sheets	Small businesses	Start-ups	Pre-launch stage	Post-launch stage	NPOs
Quantitative Data / Information	Based on past or present development	Annual financial statements			✓	✓					✓	✓
		Customer account management				✓	✓	✓			✓	✓
		Income & expense accounts					✓				✓	✓
		Tax returns					✓	✓				
		Assets / liabilities (information from customer)					✓	✓				
		Equity/grant funding/financing needs								✓	✓	
		Collateral value of project assets (realization proceeds)										
		Collateral value of object financed (realization proceeds)										
		Fair market value of real estate										
		Rent levels for comparable properties at same location										
		Development of fair market value and rent levels (historical/expected)										
		Liquidity and revenue development									✓	
		Target/actual comparison of project progress										
		Target/actual comparison of cash flows										
		Target/actual comparison of construction progress										
		Credit stage / time to maturity										
		Credit card transactions										
		Purchasing behavior										
	Based on future development	Cash flow forecasts			✓							
		Debt service capacity				✓	✓			✓	✓	✓
		Budget calculation						✓				
		Expected sale proceeds (collateral value)										
Qualitative Daten/ Informationen	Based on past or present development	Business areas (customers/products)			✓							
		Risk structure / risk management										
		Business policies										
		Industry information			✓	✓	✓	✓		✓	✓	✓
		Enterprise/management			✓	✓						✓
		Market (customers/products)				✓	✓			✓	✓	
		Entrepreneur's personal traits					✓	✓		✓	✓	
		Type of project										
		Ownership structure / owners' creditworthiness										
		Liability relationships			✓	✓						
		Type of object										
		Type of commodity										
		Structure of financing transaction										
		Type of real estate										
		Socio-demographic data										
		Type and history of customer relationship to date										
		Information on living environment										
		Distribution channel										
		Purpose										
		Fulfillment of requirements										
		Occupancy rate										
		Further development of customer relationship										
		Payment reminder behavior										
		Payment behavior										
		Special agreements										
		General legal conditions (group ties)			✓	✓						
		Distribution of tasks at local government level										
	Based on future development	Planning and future company development			✓	✓	✓				✓	✓
		Business idea / business plan								✓	✓	
External Data / Information	Capital market information	Stock price / price development			✓							
		Public relations/public information			✓							
		Current exchange price										
		Price development (historical/expected)										
	Miscellaneous external information	Country rating (sovereigns)										
		Form of government										
		General legal and political conditions										✓
		Macroeconomic indicators										
		External rating			✓	✓						
		Infrastructure										
		Credit reporting information				✓	✓	✓		✓	✓	
		Creditworthiness of potential tenant(s)/lessee(s)/buyer(s)										
		Budget situation										

Chart 4: Data Requirements for Corporate Customers – Enterprises/Business Owners

Capital Market-Oriented/International Companies

The main source of credit assessment data on capital market-oriented/international companies is their annual financial statements. However, financial statement analyses are based solely on the past and therefore cannot fully depict a company's ability to meet future payment obligations. To supplement these analyses, cash flow forecasts can also be included in the assessment process. This requires a qualitative assessment of the company's future development and planning in order to assess how realistic these cash flow forecasts are.

Additional qualitative information to be assessed includes the management, the company's orientation toward specific customers and products in individual business areas, and the industry in which the company operates. The core objective of analyzing these information categories should always be an appraisal of an enterprise's ability to meet its future payment obligations. As capital market-oriented/international companies are often broad, complex groups of companies, legal issues – especially those related to liability – should be examined carefully in the area of qualitative information.

One essential difference between capital market-oriented/international companies and other types of enterprises is the availability of external information. The capital market information available may include the stock price and its development (for exchange-listed companies), other published information (e.g. ad hoc reports), and external ratings.

Other enterprises which prepare balance sheets (not capital market-oriented/international)

Credit assessment for other companies which prepare balance sheets is largely similar to the assessment of capital market-oriented/international companies. However, there are some differences in the available information and the focuses of assessment.

In this context, analyses also focus on the company's annual financial statements. In contrast to the assessment of capital market-oriented/international companies, however, these analyses are not generally supplemented with cash-flow forecasts, but usually with an analysis of the borrower's debt service capacity. This analysis gives a simplified presentation of whether the borrower can meet the future payment obligations arising from a loan on the basis of income and expenses expected in the future. In this context, therefore, it is also necessary to assess the company's future development and planning in qualitative terms.

In addition, bank account activity data can also provide a source of quantitative information. This might include the analysis of long-term overdrafts as well as debit or credit balances. This type of analysis is not feasible for capital market-oriented/international companies due to their large number of bank accounts, which are generally distributed among multiple (national and international) credit institutions.

On the qualitative level, the management and the respective industry of these companies also have to be assessed. As the organizational structure of these companies is substantially less complex than that of capital market-oriented/international companies, the orientation of business areas is less important in this context. Rather, the success of a company which prepares balance sheets hinges on its strength and presence on the relevant market. This means

that it is necessary to analyze whether the company's orientation in terms of customers and products also indicates future success on its specific market.

In individual cases, external ratings can also be used as an additional source of information. If such ratings are not available, credit reporting information on companies which prepare balance sheets is generally also available from independent credit reporting agencies.

Businesses and Independent Professionals (not preparing balance sheets)

The main difference between this subsegment and the enterprise types discussed in the previous sections is the fact that the annual financial statements mentioned above are not available. Therefore, lenders should use other sources of quantitative data – such as income and expense accounts – in order to ensure as objective a credit assessment as possible. These accounts are not standardized to the extent that annual financial statements are, but they can yield reliable indicators of creditworthiness.

Due to the personal liability of business owners, it is often difficult to separate their professional and private activities clearly in this segment. Therefore, it is also advisable to request information on assets and liabilities as well as tax returns and income tax assessments provided by the business owners themselves.

Information derived from bank account activity data can also serve as a complement to the quantitative analysis of data from the past.

In this segment, data related to the past also have to be accompanied by a forward-looking analysis of the borrower's debt service capacity.

On the qualitative level, it is necessary to assess the same data categories as in the case of companies which prepare balance sheets (market, industry, etc.). However, the success of a business owner or independent professional depends far more on his/her personal characteristics than on the management of a complex organization. Therefore, assessment focuses on the personal characteristics of the business owners – not the management of the organization – in the case of these businesses and independent professionals. As regards external data, it is advisable to obtain credit reporting information (e.g. from the consumer loans register) on the business owner or independent professional.

Small Businesses

In some cases, it is sensible to use a separate rating procedure for small businesses. Compared to other businesses which do not prepare balance sheets, these businesses are mainly characterized by the smaller scale of their business activities and therefore by lower capital needs. In practice, analysts often apply simplified credit assessment procedures to small businesses, thereby reducing the data requirements and thus also the process costs involved.

The resulting simplifications compared to the previous segment (business owners and independent professionals who do not prepare balance sheets) are as follows:

- Income and expense accounts are not evaluated.
- The analysis of the borrower's debt service capacity is replaced with a simplified budget calculation.

– Market prospects are not assessed due to the smaller scale of business activities.

Aside from these simplifications, the procedure applied is analogous to the one used for business owners and independent professionals who do not prepare balance sheets.

Start-Ups

In practice, separate rating models are not often developed for start-ups. Instead, they adapt the existing models used for corporate customers. These adaptations might involve the inclusion of a qualitative "start-up criterion" which adds a (usually heuristically defined) negative input to the rating model. It is also possible to include other soft facts or to limit the maximum rating class attained in this segment.

If a separate rating model is developed for the start-up segment, it is necessary to distinguish between the pre-launch and post-launch stages, as different information will be available during these two phases.

Pre-Launch Stage

As quantitative data on start-ups (e.g. balance sheet and profit and loss accounts) are not yet available in the pre-launch stage, it is necessary to rely on other – mainly qualitative – data categories.

The decisive factors in the future success of a start-up are the business idea and its realization in a business plan. Accordingly, assessment in this context focuses on the business idea's prospects of success and the feasibility of the business plan. This also involves a qualitative assessment of market opportunities as well as a review of the prospects of the industry in which the start-up founder plans to operate. Practical experience has shown that a start-up's prospects of success are heavily dependent on the personal characteristics of the business owner. In order to obtain a complete picture of the business owner's personal characteristics, credit reporting information (e.g. from the consumer loans register) should also be retrieved.

On the quantitative level, the financing structure of the start-up project should be evaluated. This includes an analysis of the equity contributed, potential grant funding and the resulting residual financing needs. In addition, an analysis of the organization's debt service capacity should be performed in order to assess whether the start-up will be able to meet future payment obligations on the basis of expected income and expenses.

Post-Launch Stage

As more data on the newly established enterprise are available in the post-launch stage, credit assessments should also include this information.

In addition to the data requirements described for the pre-launch stage, it is necessary to analyze the following data categories:

– Annual financial statements or income and expense accounts (as available)
– Bank account activity data
– Liquidity and revenue development
– Future planning and company development

This will make it possible to evaluate the start-up's business success to date on the basis of quantitative data and to compare this information with the business plan and future planning information, thus providing a more complete picture of the start-up's creditworthiness.

NPOs (Non-Profit Organizations)
Although NPOs do not operate for the purpose of making a profit, it is still necessary to review the economic sustainability of these organizations by analyzing their annual financial statements. In comparison to those of conventional profit-oriented companies, the individual balance sheet indicators of NPOs have to be interpreted differently. However, these indicators still enable reliable statements as to the organization's economic efficiency. In order to allow forward-looking assessments of whether the organization will be able to meet its payment obligations, it is also necessary to analyze the organization's debt service capacity. This debt service capacity analysis is to be reviewed in a critical light by assessing the organization's planning and future development. It is also important to analyze bank account activity data in order to detect payment disruptions at an early stage. The viability of an NPO also depends on qualitative factors such as its management and the prospects of the industry.

As external information, the general legal and political circumstances in which the NPO operates should be taken into account, as NPOs are often dependent on current legislation and government grants (e.g. in organizations funded by donations).

2.4 Corporate Customers – Specialized Lending

Specialized lending operations can be characterized as follows:[3]

- *The exposure is typically to an entity (often a special purpose entity (SPE)) which was created specifically to finance and/or operate physical assets;*
- *The borrowing entity has little or no other material assets or activities, and therefore little or no independent capacity to repay the obligation, apart from the income that it receives from the asset(s) being financed;*
- *The terms of the obligation give the lender a substantial degree of control over the asset(s) and the income that it generates; and*
- *As a result of the preceding factors, the primary source of repayment of the obligation is the income generated by the asset(s), rather than the independent capacity of a broader commercial enterprise.*

On the basis of the characteristics mentioned above, specialized lending operations have to be assessed differently from conventional companies and are therefore subject to different data requirements. In contrast to that of conventional companies, credit assessment in this context focuses not on the borrower but on the assets financed and the cash flows expected from those assets.

[3] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 47, No. 8.*

			Specialized lending					Project finance				Financing of income-producing real estate		
					Before the project	During the project	Object finance	Before the project	During the project	Commodities finance		Before the project	During the project	
Quantitative Data / Information	Based on past or present development	Annual financial statements												
		Customer account management												
		Income & expense accounts												
		Tax returns												
		Assets / liabilities (information from customer)												
		Equity/grant funding/financing needs			✓	✓		✓	✓					
		Collateral value of project assets (realization proceeds)			✓	✓								
		Collateral value of object financed (realization proceeds)						✓	✓					
		Fair market value of real estate										✓	✓	
		Rent levels for comparable properties at same location										✓	✓	
		Development of fair market value and rent levels (historical/expected)										✓	✓	
		Liquidity and revenue development												
		Target/actual comparison of project progress				✓								
		Target/actual comparison of cash flows				✓			✓				✓	
		Target/actual comparison of construction progress							✓				✓	
		Credit stage / time to maturity												
		Credit card transactions												
		Purchasing behavior												
	Based on future development	Cash flow forecasts												
		Debt service capacity												
		Budget calculation												
		Expected sale proceeds (collateral value)								✓		✓	✓	
Qualitative Daten/ Informationen	Based on past or present development	Business areas (customers/products)												
		Risk structure / risk management.												
		Business policies												
		Industry information												
		Enterprise/management												
		Market (customers/products)												
		Entrepreneur's personal traits												
		Type of project			✓	✓								
		Ownership structure / owners' creditworthiness			✓	✓		✓	✓	✓		✓	✓	
		Liability relationships			✓	✓								
		Type of object						✓	✓					
		Type of commodity								✓				
		Structure of financing transaction								✓		✓	✓	
		Type of real estate										✓	✓	
		Socio-demographic data												
		Type and history of customer relationship to date												
		Information on living environment												
		Distribution channel												
		Purpose												
		Fulfillment of requirements				✓			✓				✓	
		Occupancy rate											✓	
		Further development of customer relationship												
		Payment reminder behavior												
		Payment behavior												
		Special agreements												
		General legal conditions (group ties)												
		Distribution of tasks at local government level												
	Based on future development	Planning and future company development												
		Business idea / business plan			✓	✓		✓	✓					
External Data / Information	Capital market information	Stock price / price development												
		Public relations/public information												
		Current exchange price								✓				
		Price development (historical/expected)												
	Miscellaneous external information	Country rating (sovereigns)			✓	✓		✓	✓	✓		✓	✓	
		Form of government												
		General legal and political conditions			✓	✓		✓	✓	✓		✓	✓	
		Macroeconomic indicators												
		External rating												
		Infrastructure												
		Credit reporting information												
		Creditworthiness of potential tenant(s)/lessee(s)/buyer(s)										✓	✓	
		Budget situation												

Chart 5: Data Requirements for Corporate Customers – Specialized Lending

In general, four different types of specialized lending can be distinguished on the basis of the assets financed:[4]

– Project finance
– Object finance
– Commodities finance
– Financing of income-producing real estate

For project finance, object finance and the financing of income-producing real estate, different data will be available for credit assessment purposes depending on the stage to which the project has progressed. For these three types of specialized lending operations, it is necessary to differentiate between credit assessment before and during the project. In commodities finance, this differentiation of stages is not necessary as these transactions generally involve only short-term loans.

2.4.1 Project Finance

This type of financing is generally used for large, complex and expensive projects such as power plants, chemical factories, mining projects, transport infrastructure projects, environmental protection measures and telecommunications projects. The loan is repaid exclusively (or almost exclusively) using the proceeds of contracts signed for the facility's products. Therefore, repayment essentially depends on the project's cash flows and the collateral value of project assets.[5]

Before the Project

On the basis of the dependences described above, it is necessary to assess the expected cash flow generated by the project in order to estimate the probability of repayment for the loan. This requires a detailed analysis of the business plan underlying the project. In particular, it is necessary to assess the extent to which the figures presented in the plan can be considered realistic. This analysis can be supplemented by a credit institution's own cash flow forecasts. This is common practice in real estate finance transactions, for example, in which the bank can estimate expected cash flows quite accurately in-house.

In this segment, the lender must compare the expected cash flow to the project's financing requirements, with due attention to equity contributions and grant funding. This will show whether the borrower is likely to be in a position to meet future payment obligations. The risk involved in project finance also depends heavily on the specific type of project involved. If the planned project does not meet the needs of the respective market (e.g. the construction of a chemical factory during a crisis in the industry), this may cause repayment problems later.

Should payment difficulties arise, the collateral value of project assets and the estimated resulting sale proceeds will be decisive for the credit institution.

Besides project-specific information, data on the borrowers also have to be analyzed. This includes the ownership structure as well as the respective credit standing of each stakeholder in the project. Depending on the specific liability

[4] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 47, No. 8.*

[5] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-1, No. 8.*

relationships in the project, these credit ratings will affect the assessment of the project finance transaction in various ways.

One external factor which deserves attention is the country in which the project is to be carried out. Unstable legal and political circumstances can cause project delays and can thus result in payment difficulties. Country ratings can be used as indicators for assessing specific countries.

During the Project

In addition to the information available at the beginning of the project, additional data categories can be assessed during the project due to improved data availability. At this stage, it is also possible to compare target figures with actual data. Such comparisons can first be performed for the general progress of the project by checking the current project status against the status scheduled in the business plan. The results will reveal any potential dangers to the progress of the project.

Second, assessment may also involve comparing cash flow forecasts with the cash flows realized to date. If large deviations arise, this has to be taken into account in credit assessment.

Another qualitative factor to be assessed is the fulfillment of specific covenants or requirements, such as construction requirements, environmental protection requirements and the like. Failure to fulfill these requirements can delay or even endanger the project.

2.4.2 Object Finance

Object finance (OF) refers to a method of funding the acquisition of physical assets (e.g. ships, aircraft, satellites, railcars, and fleets) where the repayment of the exposure is dependent on the cash flows generated by the specific assets that have been financed and pledged or assigned to the lender.[6] Rental or leasing agreements with one or more contract partners can be a primary source of these cash flows.

Before the Project

In this context, the procedure to be applied is analogous to the one used for project finance, that is, analysis should focus on expected cash flow and a simultaneous assessment of the business plan. Expected cash flow is to be compared to financing requirements with due attention to equity contributions and grant funding.

The type of assets financed can serve as an indicator of the general risk involved in the object finance transaction. Should payment difficulties arise, the collateral value of the assets financed and the estimated resulting sale proceeds will be decisive factors for the credit institution.

In addition to object-specific data, it is also important to review the creditworthiness of the parties involved (e.g. by means of external ratings). One external factor to be taken into account is the country in which the object is to be constructed. Unstable legal and political circumstances can cause project delays and can thus result in payment difficulties. The relevant country rating can serve as an additional indicator in the assessment of a specific country.

[6] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-1, No. 12.*

Although the data categories for project and object finance transactions are identical, the evaluation criteria can still differ in specific data categories.

During the Project
In addition to the information available at the beginning of the project, it is possible to assess additional data categories during the project due to improved data availability. The procedure to be applied here is analogous to the one used for project finance transactions (during the project), which means that the essential new credit assessment areas are as follows:
– Target/actual comparison of cash flows
– Target/actual comparison of construction progress
– Fulfillment of requirements

2.4.3 Commodities Finance

Commodities finance refers to structured short-term lending to finance reserves, inventories or receivables of exchange-traded commodities (e.g. crude oil, metals, or grains), where the exposure will be repaid from the proceeds of the sale of the commodity and the borrower has no independent capacity to repay the exposure.[7]

Due to the short-term nature of the loans (as mentioned above), it is not necessary to distinguish various project stages in commodities finance.

One essential characteristic of a commodities finance transaction is the fact that the proceeds from the sale of the commodity are used to repay the loan. Therefore, the primary information to be taken into account is related to the commodity itself. If possible, credit assessments should also include the current exchange price of the commodity as well as historical and expected price developments. The expected price development can be used to derive the expected sale proceeds as the collateral value. By contrast, the creditworthiness of the parties involved plays a less important role in commodities finance.

External factors which should not be neglected in the rating process include the legal and political circumstances at the place of fulfillment for the commodities finance transaction. A lack of clarity in the legal situation at the place of fulfillment could cause problems with the sale – and thus payment difficulties. The country rating can also serve as an indicator in the assessment of specific countries.

2.4.4 Income-Producing Real Estate Financing

The term "Income-producing real estate (IPRE)" refers to a method of providing funding to real estate (such as, office buildings to let, retail space, multifamily residential buildings, industrial or warehouse space, and hotels) where the prospects for repayment and recovery on the exposure depend primarily on the cash flows generated by the asset.[8] The main source of these cash flows is rental and leasing income or the sale of the asset.

[7] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-1, No. 13.*
[8] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-1, No. 14.*

Before the Project

As the repayment of the loan mainly depends on the income generated by the real estate, the main data category used in credit assessment is the cash flow forecast for proceeds from rentals and/or sales. In order to assess whether this cash flow forecast is realistic, it is important to assess the rent levels of comparable properties at the respective location as well as the fair market value of the real estate. For this purpose, historical time series should be observed in particular in order to derive estimates of future developments in rent levels and real estate prices. These expected developments can be used to derive the expected sale proceeds as the collateral value in the case of default. The lender should compare a plausible cash flow forecast with the financing structure of the transaction in order to assess whether the borrower will be able to meet future payment obligations.

Furthermore, it is necessary to consider the type of property financed and whether it is generally possible to rent out or sell such properties on the current market.

Even if the borrower's creditworthiness is not considered crucial in a commercial real estate financing transaction, it is also necessary to examine the ownership structure and the credit standing of each stakeholder involved. The future income produced by the real estate depends heavily on the creditworthiness of the future tenant or lessee, and therefore credit assessments for the real estate financing transaction should also include this information whenever possible.

Another external factor which plays an important role in credit assessment is the country in which the real estate project is to be constructed. It is only possible to ensure timely completion of the project under stable general legal and political conditions. The external country rating can serve as a measure of a country's stability.

During the Project

Aside from the information available at the beginning of the project, a number of additional data categories can be assessed during the project. These include the following:

– Target/actual comparison of construction progress
– Target/actual comparison of cash flows
– Fulfillment of covenants/requirements
– Occupancy rate

With the help of target/actual comparisons, the project's construction progress can be checked against its planned status. In this context, substantial deviations can serve as early signs of danger in the real estate project.

Second, the assessment can also involve comparing the planned cash flows from previous forecasts with the cash flows realized to date. If considerable deviations arise, it is important to take them into account in credit assessment.

Another qualitative factor to be assessed is the fulfillment of specific requirements, such as construction requirements, environmental protection requirements and the like. In cases where these requirements are not fulfilled, the project may be delayed or even endangered.

As the loan is repaid using the proceeds of the property financed, the occupancy rate will be of particular interest to the lender in cases where the property in question is rented out.

2.5 Retail Customers

In the retail segment, we make a general distinction between *mass-market banking* and *private banking*. In contrast to the Basel II segmentation approach, our discussion of the retail segment only includes loans to private individuals, not to SMEs.

Mass-market banking refers to general (high-volume) business transacted with retail customers. For the purpose of credit assessment, we can differentiate the following standardized products in this context:

– Current accounts
– Consumer loans
– Credit cards
– Residential construction loans

Private banking involves transactions with high-net-worth retail customers and goes beyond the standardized products used in mass-market banking. Private banking thus differs from mass-market banking due to the special financing needs of individual customers.

Unlike in the general segments described above, we have also included a *product component* in the retail customer segment. This approach complies with the future requirements arising from the Basel II regulatory framework. For example, this approach makes it possible to define retail loan defaults on the level of specific exposures instead of specific borrowers.[9] Rating systems for retail credit facilities have to be based on risks specific to borrowers as well as those specific to transactions, and these systems should also include all relevant characteristics of borrowers and transactions.[10]

In our presentation of the information categories to be assessed, we distinguish between assessment *upon credit application* and ongoing risk assessment *during the credit term*.

Credit card business is quite similar to current account business in terms of its risk level and the factors to be assessed. For this reason, it is not entirely necessary to define a separate segment for credit card business.

2.5.1 Mass-Market Banking

Current Accounts

Upon Credit Application

As standardized documents (such as annual financial statements in the corporate customer segment) are not available for the evaluation of a retail customer's financial situation, it is necessary to assess these customers on the basis of information they provide regarding their assets and liabilities. In order to evaluate whether the borrower is likely to be able to meet future payment obligations, lenders should also calculate a budget for the borrower.

[9] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 1, No. 46.*

[10] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 7.*

			Retail customers	Mass-market banking	Current accounts	Upon credit application	During credit term	Consumer loans	Upon credit application	During credit term	Credit cards	Upon credit application	During credit term	Residential construction loans	Upon credit application	During credit term	Private banking
Quantitative Data / Information	Based on past or present development	Annual financial statements															
		Customer account management					✓			✓			✓			✓	✓
		Income & expense accounts															
		Tax returns															✓
		Assets/ liabilities (information from customer)				✓			✓			✓			✓		✓
		Equity/grant funding/financing needs															
		Collateral value of project assets (realization proceeds)															
		Collateral value of object financed (realization proceeds)															
		Fair market value of real estate													✓	✓	
		Rent levels for comparable properties at same location													✓	✓	
		Development of fair market value and rent levels (historical/expected)													✓	✓	
		Liquidity and revenue development															
		Target/actual comparison of project progress															
		Target/actual comparison of cash flows															
		Target/actual comparison of construction progress															
		Credit stage / time to maturity								✓						✓	
		Credit card transactions											✓				
		Purchasing behavior											✓				
	Based on future development	Cash flow forecasts															
		Debt service capacity															
		Budget calculation				✓			✓			✓			✓		✓
		Expected sale proceeds (collateral value)													✓	✓	✓
Qualitative Daten/ Informationen	Based on past or present development	Business areas (customers/products)															
		Risk structure / risk management.															
		Business policies															
		Industry information															
		Enterprise/management															
		Market (customers/products)															
		Entrepreneur's personal traits															
		Type of project															
		Ownership structure / owners' creditworthiness															
		Liability relationships															
		Type of object															
		Type of commodity															
		Structure of financing transaction															
		Type of real estate															
		Socio-demographic data				✓			✓			✓			✓		
		Type and history of customer relationship to date				✓			✓			✓			✓		
		Information on living environment				✓			✓			✓					
		Distribution channel							✓			✓			✓		
		Purpose							✓						✓		✓
		Fulfillment of requirements															
		Occupancy rate															
		Further development of customer relationship					✓			✓			✓			✓	✓
		Payment reminder behavior					✓			✓			✓			✓	✓
		Payment behavior					✓			✓			✓			✓	✓
		Special agreements					✓			✓			✓			✓	✓
		General legal conditions (group ties)															
		Distribution of tasks at local government level															
	Based on future development	Planning and future company development															
		Business idea / business plan															
External Data / Information	Capital market information	Stock price / price development															
		Public relations/public information															
		Current exchange price															
		Price development (historical/expected)															
	Miscellaneous external information	Country rating (sovereigns)															
		Form of government															
		General legal and political conditions															
		Macroeconomic indicators															
		External rating															
		Infrastructure															
		Credit reporting information				✓	✓		✓	✓		✓	✓		✓	✓	✓
		Creditworthiness of potential tenant(s)/lessee(s)/buyer(s)															
		Budget situation															

Chart 6: Data Requirements for Retail Customers

An essential qualitative element in retail credit assessment at the time of credit application is socio-demographic data (age, profession, etc.). If the customer relationship has existed for some time, it is advisable to assess the type and history of the relationship.

Finally, the credit institution should also evaluate external data in the form of credit reporting information (e.g. from the consumer loans register).

During the Credit Term

During the term of the credit transaction, the lender should evaluate activity patterns in the customer's current account on the quantitative level. This will require historical records of the corresponding account data. Examples of the information to be derived from these data include overdraft days as well as debit and credit balances, which make it possible to detect payment disruptions at an early stage.

In addition, the general development of the customer relationship as well as reminder and payment behavior should be observed on the qualitative level. Credit assessments should also take account of any special agreements (e.g. troubled loan restructuring, deferral) made with the borrower. If possible, the lender should retrieve current credit reporting information from external agencies on a regular basis.

Consumer Loans

Upon Credit Application

The procedure applied to consumer loans is analogous to the one used for current accounts. In addition, the purpose of the loan (e.g. financing of household appliances, automobiles, etc.) can also be included in credit assessments.

During the Credit Term

In this context, the procedure applied is analogous to the one used for current accounts. The additional information to be taken into account includes the credit stage and the residual term of the transaction. Practical experience has shown that consumer loans are especially prone to default in the initial stage of a transaction, which means that the default risk associated with a consumer loan tends to decrease over time.

Credit Cards

Credit card business is quite similar to current accounts in terms of its risk level and the factors to be assessed. For this reason, it is not entirely necessary to define a separate segment for credit card business.

Upon Credit Application

In general, banks do not offer credit cards themselves but serve as distribution outlets for credit card companies. However, as the credit institution usually also bears liability if the borrower defaults, credit assessment should generally follow the same approach used for current accounts.

During the Credit Term
Instead of observing the customer's bank account activity patterns, the credit institution should assess the customer's credit card transactions and purchasing behavior in this context. As in the case of current accounts, this will make it possible to detect payment disruptions at an early stage.

The qualitative data categories assessed in this segment are no different from those evaluated for current accounts.

Residential Construction Loans

Upon Credit Application
In addition to the borrower's current financial situation (as indicated by the customer him/herself) and the customer's probable future ability to meet payment obligations (based on budget calculations), the (residential) property financed also plays a decisive role in credit assessment for this segment, as this property will serve as collateral in the case of default. For this reason, the fair market value and probable sale proceeds should be calculated for the property. In order to facilitate assessments of how the fair market value of the property will develop in the future, it is necessary to consider its historical price development. If the property financed includes more than one residential unit and part of it is to be rented out, it is also advisable to assess the current and expected rent levels of comparable properties.

The relevant qualitative and external sources of information in this context are analogous to the other subsegments in mass-market banking: socio-demographic data, the type and history of the customer relationship to date, and credit reporting information.

During the Credit Term
During the term of the loan, bank account activity data can also provide essential information. In addition, the property-specific data assessed at the time of the credit application should also be kept up to date. As in the case of consumer loans, the credit stage and residual term of residential construction loans are also significant with regard to the probability of default. Likewise, the general development of the customer relationship, reminder and payment behavior, as well as special agreements also deserve special consideration. The lender should retrieve updated credit reporting information immediately upon the first signs of deterioration in the customer's creditworthiness.

2.5.2 Private Banking

Credit assessment in private banking mainly differs from assessment in mass-market banking in that it requires a greater amount of quantitative information in order to ensure as objective a credit decision as possible. This is necessary due to the increased level of credit risk in private banking. Therefore, in addition to bank account activity data, information provided by the borrower on assets and liabilities, as well as budget calculations, it is also necessary to collect data from tax declarations and income tax returns. The lender should also take the borrower's credit reports into account and valuate collateral wherever necessary.

3 Commonly Used Credit Assessment Models

In chapter 2, we described a best-practice approach to segmentation and defined the data requirements for credit assessment in each segment. Besides the creation of a complete, high-quality data set, the method selected for processing data and generating credit assessments has an especially significant effect on the quality of a rating system.

This chapter begins with a presentation of the credit assessment models commonly used in the market, with attention to the general way in which they function and to their application in practice. This presentation is not meant to imply that all of the models presented can be considered best-practice approaches. The next chapter discusses the suitability of the various models presented. The models discussed further below are shown in chart 7.

In addition to these "pure" models, we frequently encounter combinations of heuristic methods and the other two model types in practice. The models as well as the corresponding hybrid forms are described in the sections below.

The models described here are primarily used to rate borrowers. In principle, however, the architectures described can also be used to generate transaction ratings.

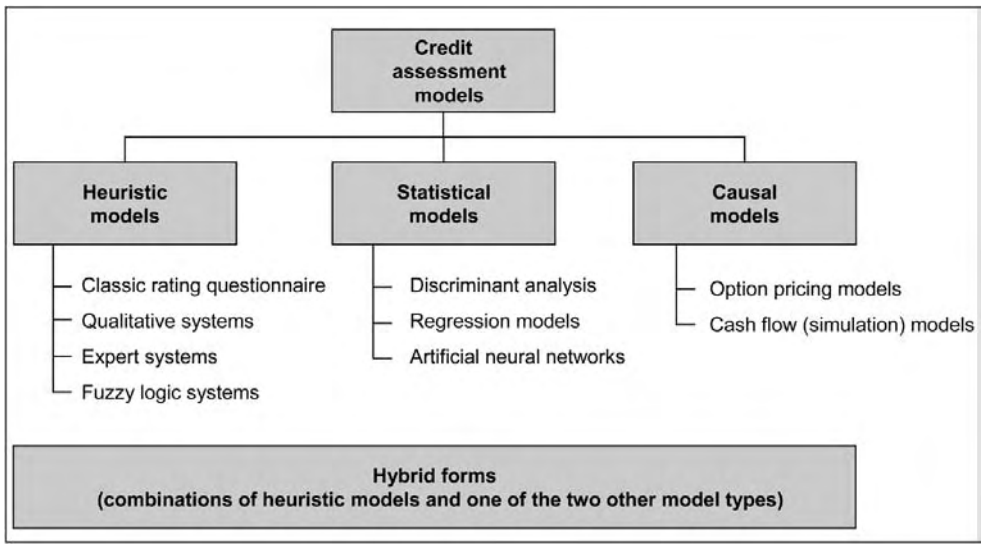

Chart 7: Systematic Overview of Credit Assessment Models

In this document, we use the term "rating models" consistently in the context of credit assessment. "Scoring" is understood as a component of a rating model, for example in Section 5.2., "Developing the Scoring Function."

On the other hand, "scoring" – as a common term for credit assessment models (e.g. application scoring, behavior scoring in retail business) – is not differentiated from "rating" in this document because the terms "rating" and "scoring" are not clearly delineated in general usage.

3.1 Heuristic Models

Heuristic models attempt to gain insights methodically on the basis of previous experience. This experience is rooted in:

- subjective practical experience and observations
- conjectured business interrelationships
- business theories related to specific aspects.

In credit assessment, therefore, these models constitute an attempt to use experience in the lending business to make statements as to the future creditworthiness of a borrower. The quality of heuristic models thus depends on how accurately they depict the subjective experience of credit experts. Therefore, not only the factors relevant to creditworthiness are determined heuristically, but their influence and weight in overall assessments are also based on subjective experience.

In the development of these rating models, the factors used do not undergo statistical validation and optimization.

In practice, heuristic models are often grouped under the heading of *expert systems*. In this document, however, the term is only used for a specific class of heuristic systems (see section 3.1.3).

3.1.1 "Classic" Rating Questionnaires

"Classic" rating questionnaires are designed on the basis of credit experts' experience. For this purpose, the lender defines clearly answerable questions regarding factors relevant to creditworthiness and assigns fixed numbers of points to specific factor values (i.e. answers). This is an essential difference between classic rating questionnaires and qualitative systems, which allow the user some degree of discretion in assessment. Neither the factors nor the points assigned are optimized using statistical procedures; rather, they reflect the subjective appraisals of the experts involved in developing these systems.

For the purpose of credit assessment, the individual questions regarding factors are to be answered by the relevant customer service representative or clerk at the bank. The resulting points for each answer are added up to yield the total number of points, which in turn sheds light on the customer's creditworthiness.

Chart 8 shows a sample excerpt from a classic rating questionnaire used in the retail segment.

In this example, the credit experts who developed the system defined the borrower's sex, age, region of origin, income, marital status, and profession as factors relevant to creditworthiness. Each specific factor value is assigned a fixed number of points. The number of points assigned depends on the presumed impact on creditworthiness. In this example, practical experience has shown that male borrowers demonstrate a higher risk of default than female borrowers. Male borrowers are therefore assigned a lower number of points. Analogous considerations can be applied to the other factors.

The higher the total number of points is, the better the credit rating will be.

In practice, classic rating questionnaires are common both in the retail and corporate segments. However, lending institutions are increasingly replacing these questionnaires with statistical rating procedures.

Characteristic \ Value				
Sex	Male **-20**	Female **+5**		
Age	18 - 23 **-30**	24 - 30 **-15**	31 - 40 **0**	> 40 **+20**
Postal code	Region A **-5**	Region B **+20**	Region C **0**	Region D **+5**
Monthly income	<TEUR 1 **-40**	<TEUR 2 **-10**	<TEUR 3 **+10**	>=TEUR 3 **+30**
Marital status	Single **-10**	Married **+15**	Divorced **-15**	Widowed **+5**
Profession	Wage earner **+10**	Salaried empl. **-20**	Self-employed **0**	Unemployed **-50**

Chart 8: Excerpt from a Classic Rating Questionnaire

3.1.2 Qualitative Systems

In qualitative systems,[11] the information categories relevant to creditworthiness are also defined on the basis of credit experts' experience. However, in contrast to classic rating questionnaires, qualitative systems do not assign a fixed number of points to each specific factor value. Instead, the individual information categories have to be evaluated in qualitative terms by the customer service representative or clerk using a predefined scale. This is possible with the help of a grading system or ordinal values (e.g. "good," "medium," "poor"). The individual grades or assessments are combined to yield an overall assessment. These individual assessment components are also weighted on the basis of subjective experience. Frequently, these systems also use equal weighting.

In order to ensure that all of the users have the same understanding of assessments in individual areas, a qualitative system must be accompanied by a user's manual. Such manuals contain verbal descriptions for each information category relevant to creditworthiness and for each category in the rating scale in order to explain the requirements a borrower has to fulfill in order to receive a certain rating.

In practice, credit institutions have used these procedures frequently, especially in the corporate customer segment. In recent years, however, qualitative systems have been replaced more and more by statistical procedures due to improved data availability and the continued development of statistical methods.

One example of a qualitative system is the BVR-I rating system used by the Federal Association of German Cooperative Banks (shown below). This system, however, is currently being replaced by the statistical BVR-II rating procedure.

The BVR-I rating uses 5 information categories relevant to creditworthiness, and these categories are subdivided into a total of 17 subcriteria (see chart 9).

[11] *In contrast to the usage in this guide, qualitative systems are also frequently referred to as expert systems in practice.*

Criterion Type		BVR-I Rating Criterion
Quantitative Criteria	Economic situation	Overall assets situation Annual financial statements
	Cust. relationship / account mgmt.	Customer account management Customer transparency/information behavior
Qualitative Criteria	Future company development	Development since last financial statements Enterprise planning Income planning/future debt service capacity Special company risks
	Market/industry	Development of market/industry Cyclical dependence Customer/supplier structure Export/import risks Intensity of competition Products/range Service level
	Management	Management Accounting/management accounting

Chart 9: Information Categories for BVR-I Ratings[12]

All 17 sub-areas use the grading system used in German schools (1 to 6, with 1 being the best possible grade), and the arithmetic mean of the grades assigned is calculated to yield the average grade.

When carrying out these assessments, users are required to adhere to specific rating guidelines which explain the individual creditworthiness factors and define the information sources and perspectives to be considered. Each specific grade which can be assigned is also described verbally.

In the "Management" information category, for example, the grades are described as follows:[13]

The key difference between qualitative models and classic rating questionnaires lies in the user's discretion in assessment and interpretation when assigning ratings to the individual factors.

[12] *See KIRMSSE, S. / JANSEN, S., BVR-II-Rating.*

[13] *Cf. EIGERMANN, J., Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale, p. 120.*

Assessment	Grade
High-level professional and personal qualifications in all areas	1
Good qualifications with strengths in specific areas	2
Satisfactory qualifications with strengths in specific areas	3
Average qualifications with weaknesses in specific areas	4
Pronounced weaknesses in specific areas/latent succession problems	5
Insufficient qualifications/acute succession problems	6

Chart 10: Rating Scale in the "Management" Information Category for BVR-I Ratings

3.1.3 Expert Systems

Expert systems are software solutions which aim to recreate human problem-solving abilities in a specific area of application. In other words, expert systems attempt to solve complex, poorly structured problems by making conclusions on the basis of "intelligent behavior." For this reason, they belong to the research field of artificial intelligence and are also often referred to as "knowledge-based systems."

The essential components of an expert system are the knowledge base and the inference engine.[14]

The knowledge base in these systems contains the knowledge acquired with regard to a specific problem. This knowledge is based on numbers, dates, facts and rules as well as "fuzzy" expert experience, and it is frequently represented using "production rules" (if/then rules). These rules are intended to recreate the analytical behavior of credit experts as accurately as possible.

The inference engine links the production rules in order to generate conclusions and thus find a solution to the problem.

The expert system outputs partial assessments and the overall assessment in the form of verbal explanations or point values.

Additional elements of expert systems include:[15]

Knowledge Acquisition Component

As the results of an expert system depend heavily on the proper and up-to-date storage of expert knowledge, it must be possible to expand the knowledge base with new insights at all times. This is achieved by means of the knowledge acquisition component.

Dialog Component

The dialog component includes elements such as standardized dialog boxes, graphic presentations of content, help functions and easy-to-understand menu structures. This component is decisive in enabling users to operate the system effectively.

[14] *Cf. HEITMANN, C., Neuro-Fuzzy, p. 20ff.*

[15] *Cf. BRUCKNER, B., Expertensysteme, p. 391.*

Explanatory Component

The explanatory component makes the problem-solving process easier to comprehend. This component describes the specific facts and rules the system uses to solve problems. In this way, the explanatory component creates the necessary transparency and promotes acceptance among the users.

Applied example:

One example of an expert system used in banking practice is the system at Commerzbank:[16] The CODEX (*Commerzbank Debitoren Experten System*) model is applied to domestic small and medium-sized businesses.

The knowledge base for CODEX was compiled by conducting surveys with credit experts. CODEX assesses the following factors for all borrowers:

– *Financial situation* (using figures on the business' financial, liquidity and income situation from annual financial statements)
– *Development potential* (compiled from the areas of market potential, management potential and production potential)
– *Industry prospects.*

In all three areas, the relevant customer service representative or clerk at the bank is required to answer questions on defined creditworthiness factors. In this process, the user selects ratings from a predefined scale. Each rating option is linked to a risk value and a corresponding grade. These rating options, risk values and grades were defined on the basis of surveys conducted during the development of the system.

A schematic diagram of how this expert system functions is provided in chart 11.

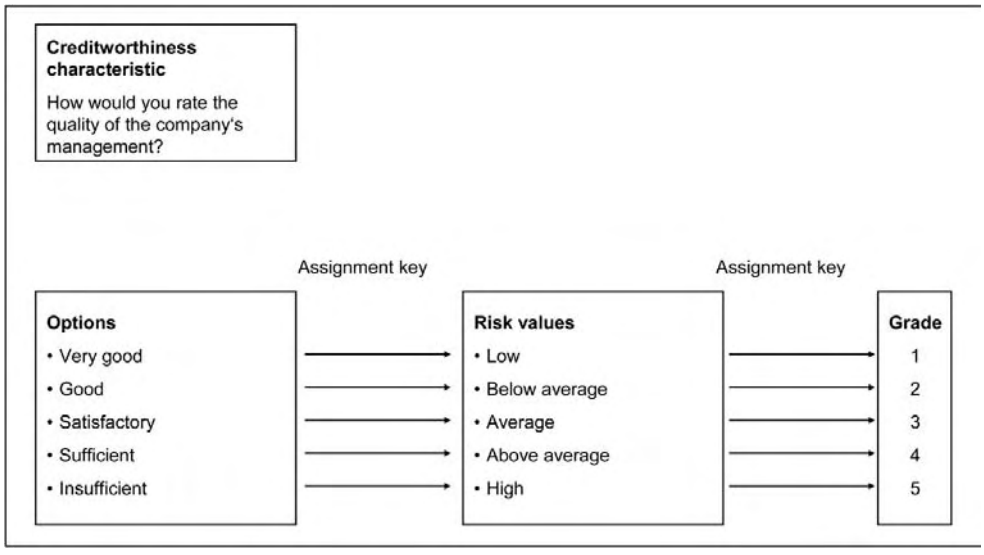

Chart 11: How the CODEX Expert System Works[17]

[16] *Cf. EIGERMANN, J., Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale, p. 104ff.*

[17] *Adapted from EIGERMANN, J., Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale, p. 107.*

The system transforms all of the individual creditworthiness characteristics into grades and then combines them to yield an overall grade. This involves two steps: First the system compresses grades from an individual information category into a partial grade by calculating a weighted average. The weights used here were determined on the basis of expert surveys. Then the system aggregates individual assessments to generate an overall assessment. The aggregation process uses the expert system's hierarchical aggregation rules, which the credit analyst cannot influence.

3.1.4 Fuzzy Logic Systems

Fuzzy logic systems can be seen as a special case among the classic expert systems described above, as they have the additional ability to evaluate data using fuzzy logic. In a fuzzy logic system, specific values entered for creditworthiness criteria are no longer allocated to a single linguistic term (e.g. "high," "low"); rather they can be assigned to multiple terms using various degrees of membership.

For example, in a classic expert system the credit analyst could be required to rate a return on equity of 20% or more as good and a return on equity of less than 20% as "poor." However, such dual assignments are not in line with human assessment behavior. A human decision maker would never rate a return on equity of 19.9% as "low" and a return on equity of 20.0% as "high" at the same time.

Fuzzy logic systems thus enable a finer gradation which bears more similarity to human decision-making behavior by introducing linguistic variables. The basic manner in which these linguistic variables are used is shown in chart 12.

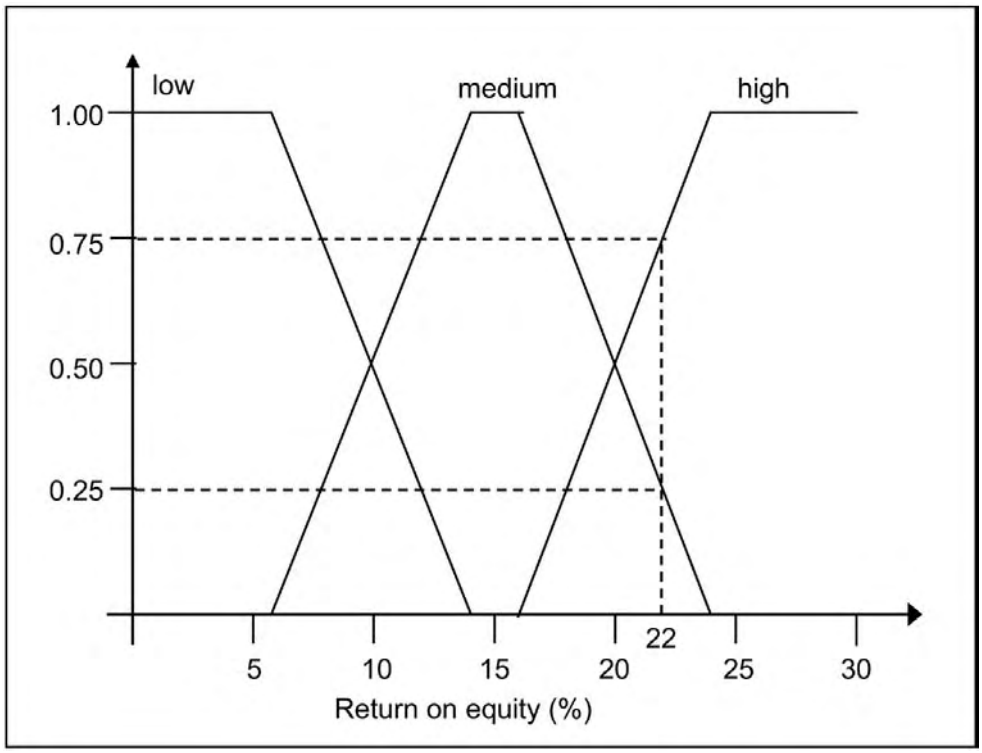

Chart 12: Example of a Linguistic Variable[18]

[18] *Adapted from HEITMANN, C., Neuro-Fuzzy, p. 47.*

This example defines linguistic terms for the evaluation of return on equity ("low," "medium," and "high") and describes membership functions for each of these terms. The membership functions make it possible to determine the degree to which these linguistic terms apply to a given level of return on equity. In the diagram above, for example, a return on equity of 22% would be rated "high" to a degree of 0.75, "medium" to a degree of 0.25, and "low" to a degree of 0.

In a fuzzy logic system, multiple distinct input values are transformed using linguistic variables, after which they undergo further processing and are then compressed into a clear, distinct output value. The rules applied in this compression process stem from the underlying knowledge base, which models the experience of credit experts. The architecture of a fuzzy logic system is shown in chart 13.

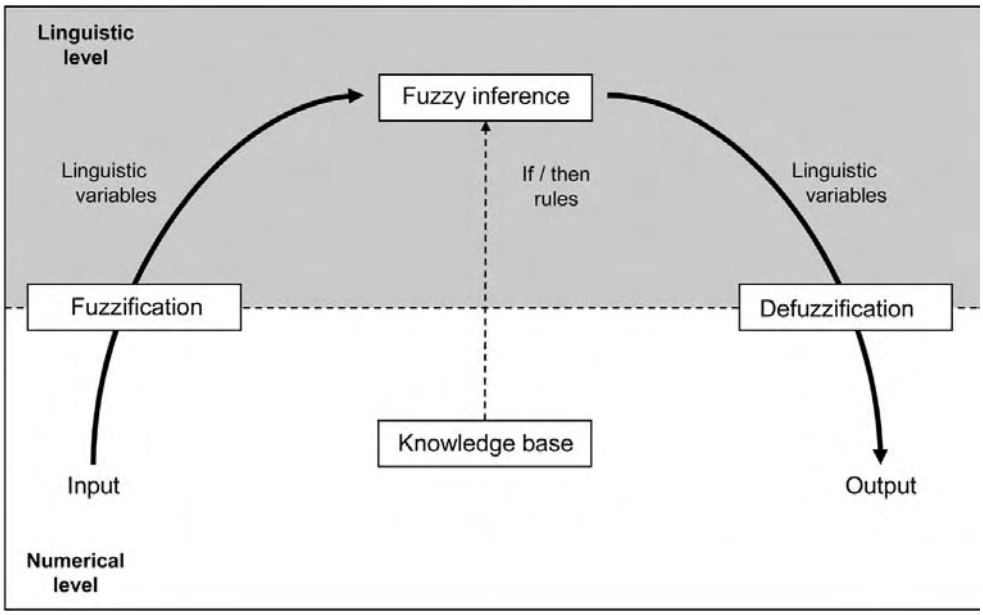

Chart 13: Architecture of a Fuzzy Logic System[19]

In the course of *fuzzification*, degrees of membership in linguistic terms are determined for the input values using linguistic variables. The data then undergo further processing in the fuzzy logic system solely on the basis of these linguistic terms.

The if/then rules in the *knowledge base* model the links between input values and the output value and represent the experience of credit experts. One simple example of an if/then rule might be: "IF return on equity is high AND debt-to-equity ratio is low, THEN creditworthiness is good."

The fuzzy inference engine is responsible for the computer-based evaluation of the if/then rules in the knowledge base.

The result output by the fuzzy inference engine is an overall assessment based on a linguistic variable. At this point, degrees of membership are still used, meaning that the resulting statement is still expressed in fuzzy terms.

[19] *Adapted from HEITMANN, C., Neuro-Fuzzy, p. 53.*

The result from the inference engine is therefore transformed into a clear and distinct credit rating in the process of *defuzzification*.

Applied example:
The Deutsche Bundesbank uses a fuzzy logic system as a module in its credit assessment procedure.[20]

The Bundesbank's credit assessment procedure for corporate borrowers first uses industry-specific discriminant analysis to process figures from annual financial statements and qualitative characteristics of the borrower's accounting practices. The resulting overall indicator is adapted using a fuzzy logic system which processes additional qualitative data (see chart 14).

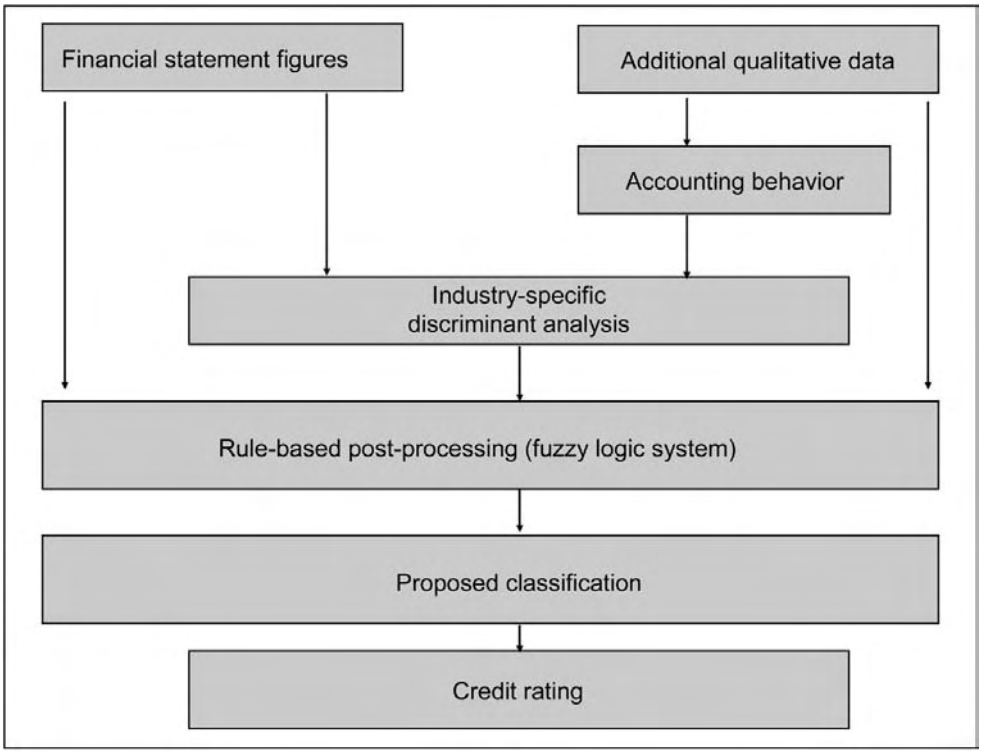

Chart 14: Architecture of Deutsche Bundesbank's Credit Assessment Procedure[21]

In this context, the classification results for the sample showed that the error rate dropped from 18.7% after discriminant analysis to 16% after processing with the fuzzy logic system.

3.2 Statistical Models

While heuristic credit assessment models rely on the subjective experience of credit experts, statistical models attempt to verify hypotheses using statistical procedures on an empirical database.

For credit assessment procedures, this involves formulating hypotheses concerning potential creditworthiness criteria: These hypotheses contain state-

[20] *Cf. BLOCHWITZ, STEFAN/EIGERMANN, JUDITH, Bonitätsbeurteilungsverfahren der Deutschen Bundesbank.*

[21] *Adapted from BLOCHWITZ, STEFAN/EIGERMANN, JUDITH, Bonitätsbeurteilungsverfahren der Deutschen Bundesbank.*

ments as to whether higher or lower values can be expected on average for solvent borrowers compared to insolvent borrowers. As the solvency status of each borrower is known from the empirical data set, these hypotheses can be verified or rejected as appropriate.

Statistical procedures can be used to derive an objective selection and weighting of creditworthiness factors from the available solvency status information. In this process, selection and weighting are carried out with a view to optimizing accuracy in the classification of solvent and insolvent borrowers in the empirical data set.

The goodness of fit of any statistical model thus depends heavily on the quality of the empirical data set used in its development. First, it is necessary to ensure that the data set is large enough to enable statistically significant statements. Second, it is also important to ensure that the data used accurately reflect the field in which the credit institution plans to use the model. If this is not the case, the statistical rating models developed will show sound classification accuracy for the empirical data set used but will not be able to make reliable statements on other types of new business.

The statistical models most frequently used in practice – discriminant analysis and regression models – are presented below. A different type of statistical rating model is presented in the ensuing discussion of artificial neural networks.

3.2.1 Multivariate Discriminant Analysis

The general objective of multivariate discriminant analysis (MDA) within a credit assessment procedure is to distinguish solvent and insolvent borrowers as accurately as possible using a function which contains several independent creditworthiness criteria (e.g. figures from annual financial statements). Multivariate discriminant analysis is explained here on the basis of a linear discriminant function, which is the approach predominantly used in practice. In principal, however, these explanations also apply to nonlinear functions.

In linear multivariate discriminant analysis, a weighted linear combination of indicators is created in order to enable good and bad cases to be classified with as much discriminatory power as possible on the basis of the calculated result (i.e. the discriminant score D):

$$D = a_0 + a_1 \cdot K_1 + a_2 \cdot K_2 + ... + a_n \cdot K_n.$$

In this equation, n refers to the number of financial indicators included in the scoring function, I_i refers to the specific indicator value, and a_i stands for each indicator's coefficient within the scoring function.

The chart below illustrates the principle behind linear discriminant analysis on the basis of a two-criterion example. The optimum cutoff line represents a linear combination of the two criteria. The line was determined with a view to discriminating between solvent and insolvent borrowers as accurately as possible (i.e. with a minimum of misclassifications).

One advantage of using MDA compared to other classification procedures is that the linear function and the individual coefficients can be interpreted directly in economic terms.

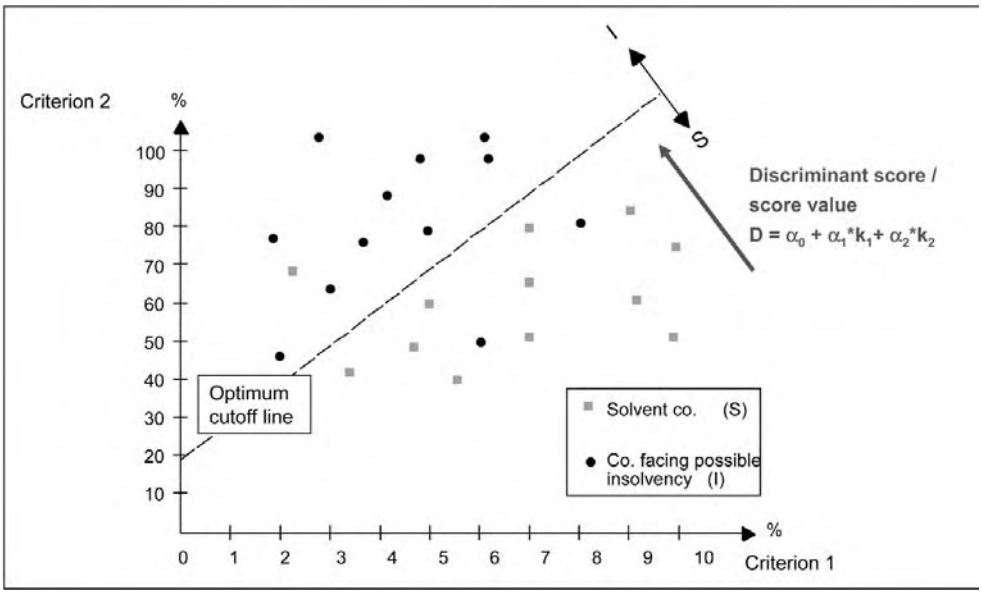

Chart 15: How Linear Discriminant Analysis Works

Linear multivariate discriminant analysis requires normal distribution (in the strict mathematical sense of the term) in the indicators examined. Therefore, the assumption of normal distribution has to be tested for the input indicators. In cases where the indicators used in analysis are not normally distributed, the MDA results may be compromised.

In particular, practitioners should bear this in mind when using qualitative creditworthiness criteria, which generally come in the form of ordinal values and are therefore not normally distributed. However, studies have shown that rescaling the qualitative creditworthiness criteria in a suitable manner can also fulfill the theoretical prerequisites of MDA.[22] For example, Lancaster scaling can be used.[23]

In addition to the assumption of normal distribution, linear discriminant analysis also requires the same variance/covariance matrices for the groups to be discriminated. In practice, however, this prerequisite is attributed less significance.[24]

Applied example:
In practice, linear multivariate discriminant analysis is used quite frequently for the purpose of credit assessment. One example is Bayerische Hypo- und Vereinsbank AG's "Crebon" rating system, which applies linear multivariate discriminant analysis to annual financial statements. A total of ten indicators are processed in the discriminant function (see chart 16). The resulting discriminant score is referred to as the MAJA value at Bayerische Hypo- und Vereinsbank. MAJA is the German acronym for automated financial statement analysis (***MA****schinelle* ***J****ahresabschluß****A****nalyse).*

[22] *Cf. BLOCHWITZ, S./EIGERMANN, J., Unternehmensbeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen.*

[23] *HARTUNG, J./ELPELT, B., Multivariate Statistik, p. 282ff.*

[24] *Cf. BLOCHWITZ, S./EIGERMANN, J., Effiziente Kreditrisikobeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen, p. 10.*

Area of Analysis	Name	Explanation
Capital structure	IND 1	(Economic equity – intangible assets + long-term provisions) / (Total assets – provisions)
Assets structure	IND 2	Accounts receivable / Current assets
Liquidity	IND 3	(Liquid funds + trade accounts receivable) / Short-term debt
Asset turnover	IND 4	Short-term debt / Revenues
	IND 5	(Economic equity + provisions) / Revenues
	IND 6	Total assets / Revenues
	IND 7	Liability relationships / Revenues
Profitability	IND 8	Annual profit / Total assets
Financial strength	IND 9	Financial cash flow / Revenues
Miscellaneous	IND 10	Interest expenses / Liabilities

Chart 16: Indicators in the "Crebon" Rating System at Bayerische Hypo- und Vereinsbank[25]

3.2.2 Regression Models

Like discriminant analysis, regression models serve to model the dependence of a binary variable on other independent variables. If we apply this general definition of regression models to credit assessment procedures, the objective is to use certain creditworthiness characteristics (independent variables) to determine whether borrowers are classified as solvent or insolvent (dependent binary variable). The use of nonlinear model functions as well as the maximum likelihood method to optimize those functions means that regression models also make it possible to calculate membership probabilities and thus to determine default probabilities directly from the model function. This characteristic is relevant in rating model calibration (see section 5.3).

In this context, we distinguish between logit and probit regression models. The curves of the model functions and their mathematical representation are shown in chart 17. In this chart, the function Φ denotes the cumulative standard normal distribution, and the term $(\sum)$ stands for a linear combination of the factors input into the rating model; this combination can also contain a constant term. By rescaling the linear term, both model functions can be adjusted to yield almost identical results. The results of the two model types are therefore not substantially different.

Due to their relative ease of mathematical representation, logit models are used more frequently for rating modeling in practice. The general manner in which regression models work is therefore only discussed here using the logistic regression model (logit model) as an example.

In (binary) logistic regression, the probability p that a given case is to be classified as solvent (or insolvent) is calculated using the following formula:[26]

$$p = \frac{1}{1 + \exp[-(b_0 + b_1 \cdot K_1 + b_2 \cdot K_2 + ... + b_n \cdot K_n)]}.$$

[25] *See EIGERMANN, J., Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale, p. 102.*

[26] *In the probit model, the function used is $p = \Phi(\sum)$, where $N(.)$ stands for standard normal distribution and the following applies for $\sum$: $\sum = b_0 + b_1 \cdot x_1 + ... + b_n \cdot x_n$.*

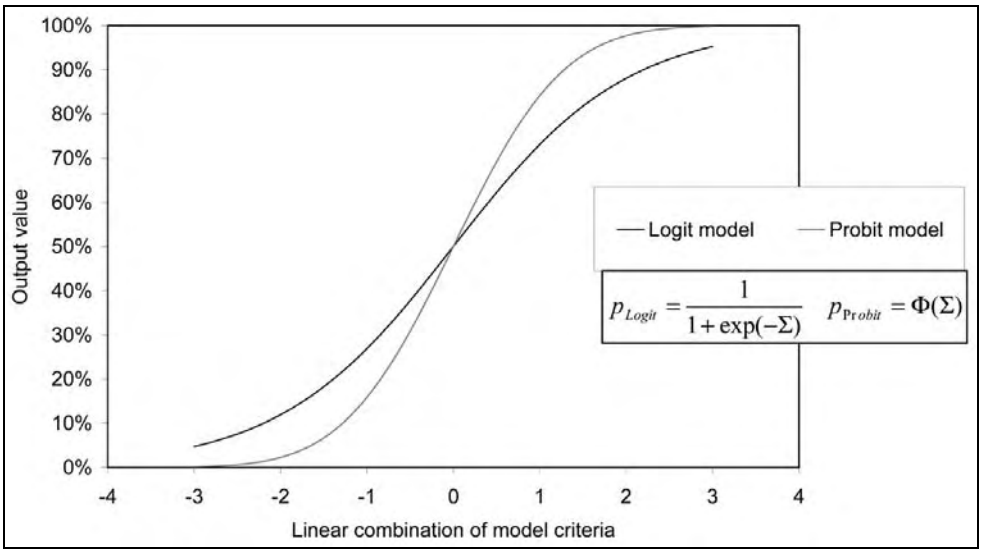

Chart 17: Functional Forms for Logit and Probit Models

In this formula, n refers to the number of financial indicators included in the scoring function, K_i refers to the specific value of the creditworthiness criterion, and b_i stands for each indicator's coefficient within the scoring function (for $i = 1, ...n$). The constant b_0 has a decisive impact on the value of p (i.e. the probability of membership).

Selecting an S-shaped logistic function curve ensures that the p values fall between 0 and 1 and can thus be interpreted as actual probabilities. The typical curve of a logit function is shown again in relation to the result of the exponential function (score) in chart 18.

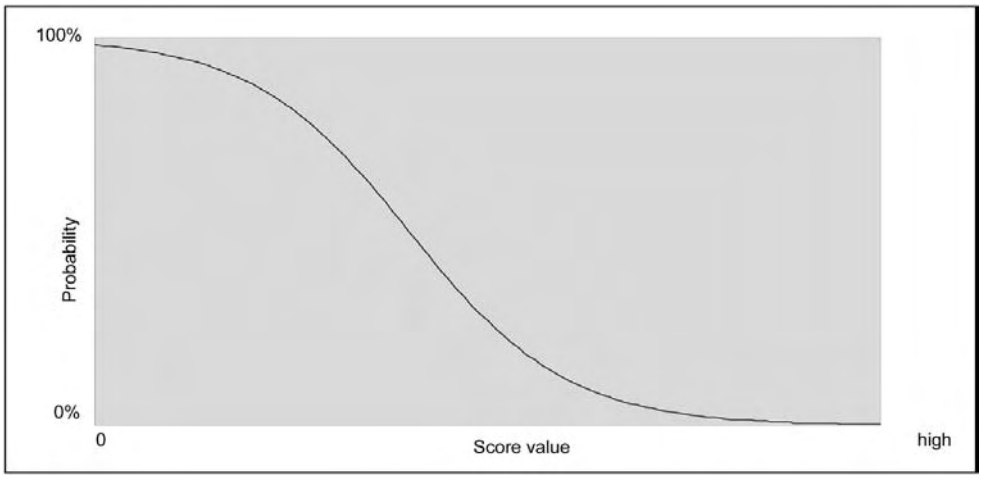

Chart 18: Logit Function Curve

The maximum likelihood method is used to estimate the coefficients. The maximum likelihood function describes how frequently the actual defaults observed match the model forecasts in the development sample.[27]

[27] *Cf. KALTOFEN, D./MÖLLENBECK, M./STEIN, S., Risikofrüherkennung im Kreditgeschäft mit kleinen und mittleren Unternehmen, p. 14.*

Logistic regression has a number of advantages over MDA. First, logistic regression does not require normal distribution in input indicators. This allows logistic regression models to process qualitative creditworthiness characteristics without previous transformation. Second, the result of logistic regression can be interpreted directly as the probability of group membership. This makes it possible to assign a one-year default rate to each result, for example by rescaling the value p.

Logistic regression models are often characterized by more robust[28] and more accurate results than those generated by discriminant analysis.

In recent years, logistic regression has seen more widespread use both in academic research and in practice. This can be attributed to the lower demands it makes on data material as well as its more robust results compared to discriminant analysis.

One example of the practical use of logistic regression in banks is the BVR-II rating model used by the Federal Association of German Cooperative Banks to rate small and medium-sized enterprises.[29]

3.2.3 Artificial Neural Networks

Structure of Artificial Neural Networks

Artificial neural networks use information technology in an attempt to simulate the way in which the human brain processes information. In simplified terms, the human brain consists of a large number of nerve cells (neurons) connected to one another by a network of synapses. Neurons receive signals through these synapses, process the information, and pass new signals on through other neurons. The significance of a particular piece of information is determined by the type and strength of the links between neurons. In this way, information can be distributed and processed in parallel across the entire network of neurons. The human brain is able to learn due to its capacity to adjust the weighting of links between neurons.

Artificial neural networks attempt to model this biological process. An artificial neural network consists of an input layer, the inner layers and an output layer (see chart 19).

The input layer serves the purpose of taking in information (e.g. specific indicator values) and passing it on to the downstream neurons via the connections shown in the diagram below. These links are assigned weights in an artificial neural network and thus control the flow of information.

In the neurons, all incoming information i_j is first linked with a value v. This is done by means of a simple sum function. Each piece of information is then assigned a connection weight w. The compressed value v is transformed into value o by a nonlinear function. The function used for this purpose depends on the specific model. One example is the following logistic function:

$$o = \frac{1}{1 + e^{-v}}.$$

[28] *Cf. KALTOFEN, D./MÖLLENBECK, M./STEIN, S., Risikofrüherkennung im Kreditgeschäft mit kleinen und mittleren Unternehmen, p. 14.*

[29] *Cf. STUHLINGER, MATTHIAS, Rolle von Ratings in der Firmenkundenbeziehung von Kreditgenossenschaften, p. 72.*

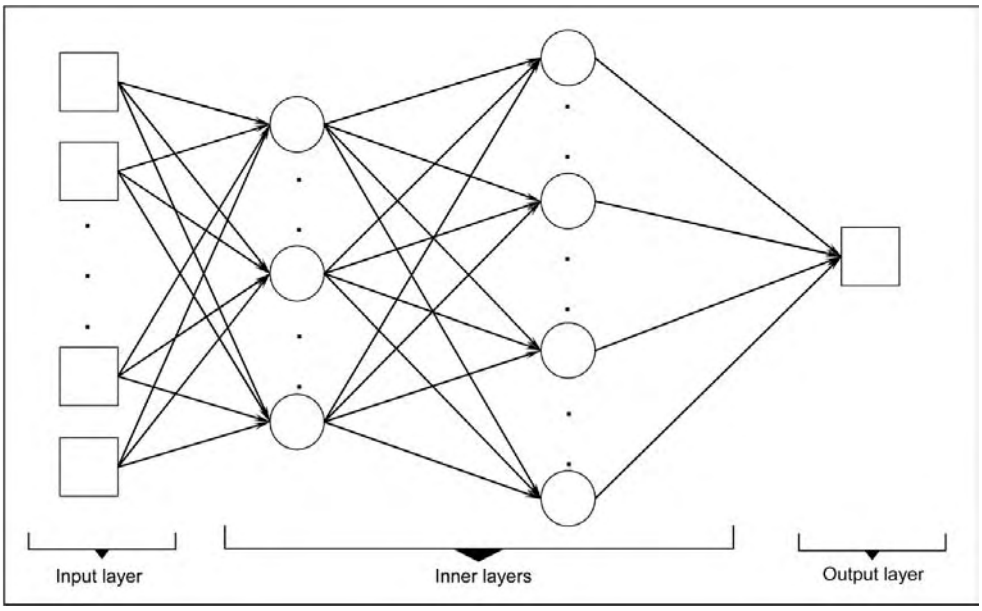

Chart 19: Architecture of an Artificial Neural Network

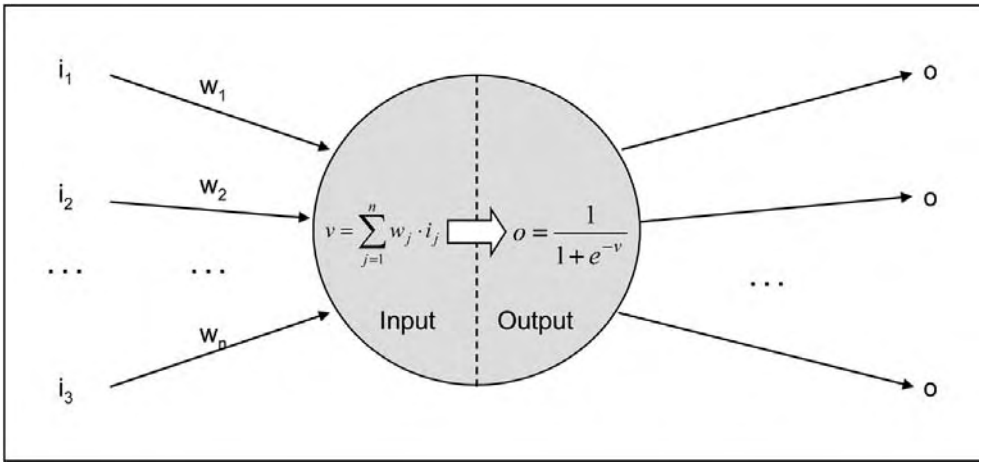

Chart 20: How Neurons Work

This transformed value is passed on to all downstream neurons, which in turn carry out the same procedure with the output factors from the upstream neurons. Chart 20 gives a schematic depiction of how a neuron works. In general, other nonlinear functions can be used instead of the logistic function.

Once the information has passed through the inner layers, it is delivered to the neuron in the output layer. This information represents the network's output, or the result generated by the artificial neural network. The inner layers are also referred to as hidden layers because the state of the neurons in these layers is not visible from the outside.

Training of Artificial Neural Networks

An artificial neural network learns on the basis of training data sets for which the actual correct output is already known. In the training process, the artificial neural network compares the output generated with the actual output and

adapts the network according to any deviations it finds. Probably the most commonly used method of making such changes in networks is the adjustment of weights between neurons. These weights indicate how important a piece of information is considered to be for the network's output. In extreme cases, the link between two neurons will be deleted by setting the corresponding weight to zero.

A classic learning algorithm which defines the procedure for adjusting weights is the back-propagation algorithm. This term refers to "a gradient descent method which calculates changes in weights according to the errors made by the neural network."[30]

In the first step, output results are generated for a number of data records.

The deviation of the calculated output o_d from the actual output t_d is measured using an error function. The sum-of-squares error function is frequently used in this context:

$$e = \frac{1}{2}\sum_d (t_d - o_d)^2$$

The calculated error can be back-propagated and used to adjust the relevant weights. This process begins at the output layer and ends at the input layer.[31]

When training an artificial neural network, it is important to avoid what is referred to as overfitting. Overfitting refers to a situation in which an artificial neural network processes the same learning data records again and again until it begins to recognize and "memorize" specific data structures within the sample. This results in high discriminatory power in the learning sample used, but low discriminatory power in unknown samples. Therefore, the overall sample used in developing such networks should definitely be divided into a learning, testing and a validation sample in order to review the network's learning success using "unknown" samples and to stop the training procedure in time. This need to divide up the sample also increases the quantity of data required.

Application of Artificial Neural Networks

Neural networks are able to process both quantitative and qualitative data directly, which makes them especially suitable for the depiction of complex rating models which have to take various information categories into account. Although artificial neural networks regularly demonstrate high discriminatory power and do not involve special requirements regarding input data, these rating models are still not very prevalent in practice. The reasons for this lie in the complex network modeling procedures involved and the "black box" nature of these networks. As the inner workings of artificial neural networks are not transparent to the user, they are especially susceptible to acceptance problems.

One example of an artificial neural network used in practice is the BBR (Baetge-Bilanz-Rating® BP-14 used for companies which prepare balance sheets. This artificial neural network uses 14 different figures from annual financial statements as input parameters and compresses them into an "N-score," on the basis of which companies are assigned to rating classes.

[30] *See HEITMANN, C., Neuro-Fuzzy, p. 85.*

[31] *Cf. HEITMANN, C., Neuro-Fuzzy, p. 86ff.*

In practice, rating models which use artificial neural networks generally attain high to very high levels of discriminatory power.[32] However, when validating artificial neural networks it is advisable to perform rigorous tests in order to ensure that the high discriminatory power of individual models is not due to overfitting.

3.3 Causal Models

Causal models in credit assessment procedures derive direct analytical links to creditworthiness on the basis of financial theory. In the development of such models, this means that statistical methods are not used to test hypotheses against an empirical data set.

3.3.1 Option Pricing Models

The option pricing theory approach supports the valuation of default risk on the basis of individual transactions without using a comprehensive default history. Therefore, this approach can generally also be used in cases where a sufficient data set of bad cases is not available for statistical model development (e.g. discriminant analysis or logit regression). However, this approach does require data on the economic value of debt and equity, and especially volatilities.

The main idea underlying option pricing models is that a credit default will occur when the economic value of the borrower's assets falls below the economic value of its debt.[33]

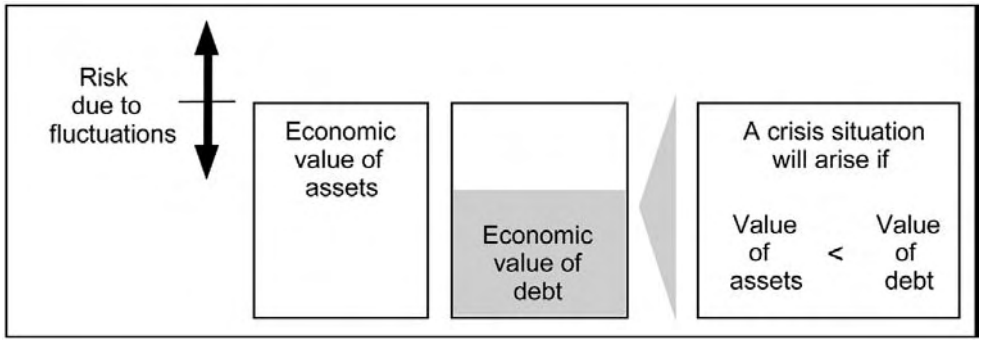

Chart 21: General Premise of Option Pricing Models[34]

In the option pricing model, the loan taken out by the company is associated with the purchase of an option which would allow the equity investors to satisfy the claims of the debt lenders by handing over the company instead of repaying the debt in the case of default.[35] The price the company pays for this option corresponds to the risk premium included in the interest on the loan. The price of the option can be calculated using option pricing models commonly used in the market. This calculation also yields the probability that the option will be exercised, that is, the probability of default.

[32] *Cf. FÜSER, K., Mittelstandsrating mit Hilfe neuronaler Netzwerke, p. 372; cf. HEITMANN, C. , Neuro-Fuzzy, p. 20.*

[33] *Cf. SCHIERENBECK, H., Ertragsorientiertes Bankmanagement Vol. 1.*

[34] *Adapted from GERDSMEIER, S./KROB, B., Bepreisung des Ausfallrisikos mit dem Optionspreismodell, p. 469ff.*

[35] *Cf. KIRMSSE, S., Optionspreistheoretischer Ansatz zur Bepreisung.*

The parameters required to calculate the option price (= risk premium) are the duration of the observation period as well as the following:[36]

– Economic value of the debt[37]
– Economic value of the equity
– Volatility of the assets.

Due to the required data input, the option pricing model cannot even be considered for applications in retail business.[38] However, generating the data required to use the option pricing model in the corporate segment is also not without its problems, for example because the economic value of the company cannot be estimated realistically on the basis of publicly available information. For this purpose, in-house planning data are usually required from the company itself. The company's value can also be calculated using the discounted cash flow method. For exchange-listed companies, volatility is frequently estimated on the basis of the stock price's volatility, while reference values specific to the industry or region are used in the case of unlisted companies.

In practice, the option pricing model has only been implemented to a limited extent in German-speaking countries, mainly as an instrument of credit assessment for exchange-listed companies.[39] However, this model is also being used for unlisted companies as well.

As a credit default on the part of the company is possible at any time during the observation period (not just at the end), the risk premium and default rates calculated with a European-style[40] option pricing model are conservatively interpreted as the lower limits of the risk premium and default rate in practice.

Qualitative company valuation criteria are only included in the option pricing model to the extent that the market prices used should take this information (if available to the market participants) into account. Beyond that, the option pricing model does not cover qualitative criteria. For this reason, the application of option pricing models should be restricted to larger companies for which one can assume that the market price reflects qualitative factors sufficiently (cf. section 4.2.3).

3.3.2 Cash Flow (Simulation) Models

Cash flow (simulation) models are especially well suited to credit assessment for specialized lending transactions, as creditworthiness in this context depends primarily on the future cash flows arising from the assets financed. In this case, the transaction itself (and not a specific borrower) is assessed explicitly, and the result is therefore referred to as a *transaction rating*.

Cash flow-based models can also be presented as a variation on option pricing models in which the economic value of the company is calculated on the basis of cash flow.

[36] *The observation period is generally one year, but longer periods are also possible.*

[37] *For more information on the fundamental circular logic of the option pricing model due to the mutual dependence of the market value of debt and the risk premium as well as the resolution of this problem using an iterative method, see JANSEN, S, Ertrags- und volatilitätsgestützte Kreditwürdigkeitsprüfung, p. 75 (footnote) and VARNHOLT B., Modernes Kreditrisikomanagement, p. 107 ff. as well as the literature cited there.*

[38] *Cf. SCHIERENBECK, H., Ertragsorientiertes Bankmanagement Vol. 1.*

[39] *Cf. SCHIERENBECK, H., Ertragsorientiertes Bankmanagement Vol. 1.*

[40] *In contrast to an American option, which can be exercised at any time during the option period.*

In principal, it is possible to define cash flow from various perspectives which are considered equally suitable. Accordingly, a suitable valuation model can be selected according to the individual requirements of the organization performing the valuation and in line with the purpose of the valuation. In this context, the total free cash flow available is particularly relevant to valuation.[41]

For the purpose of company valuation on capital markets, free cash flow is calculated as EBITDA[42] minus investments.[43] The average free cash flow over the last five years generally serves as the point of departure for calculating a company's value. Dividing the average free cash flow by the weighted capital costs for equity and debt financing[44] yields a company value which can be used as an input parameter in the option pricing model. The volatility of this value can be calculated in an analogous way based on the time series used to determine the average free cash flow.

Two types of methods can be used to extrapolate future cash flow on the basis of past cash flow data:

Analytical methods are based on time series analysis methods, which come in two forms:

– *Regression models* create a functional model of the time series and optimize the model parameters by minimizing the deviations observed.
– *Stochastic time series models* depict the time series as the realization of a stochastic process and calculate optimum estimates for the process parameters.

Simulation methods generate and weight possible future realizations of cash flow on the basis of historical data or – in the approach more commonly used in practice – by developing macroeconomic models which depict the input values for cash flow in relation to certain scenarios (e.g. the overall development of the economy).

3.4 Hybrid Forms

In practice, the models described in the previous sections are only rarely used in their pure forms. Rather, heuristic models are generally combined with one of the two other model types (statistical models or causal models). This approach can generally be seen as favorable, as the various approaches complement each other well. For example, the advantages of statistical and causal models lie in their objectivity and generally higher classification performance in comparison to heuristic models. However, statistical and causal models can only process a limited number of creditworthiness factors. Without the inclusion of credit experts' knowledge in the form of heuristic modules, important information on the borrower's creditworthiness would be lost in individual cases. In addition, not all statistical models are capable of processing qualitative information directly (as is the case with discriminant analysis, for example), or they require a large amount of data in order to function properly (e.g. logistic regression); these data are frequently unavailable in banks. In order to obtain a complete pic-

[41] *Cf. JANSEN, S., Ertrags- und volatilitätsgestützte Kreditwürdigkeitsprüfung.*

[42] *EBITDA: earnings before interest, tax, depreciation and amortization.*

[43] *Cf. KEMPF, M., Dem wahren Aktienkurs auf der Spur. More detailed explanations on the general conditions of the cash flow method can be found in JANSEN, S., Ertrags- und volatilitätsgestützte Kreditwürdigkeitsprüfung.*

[44] *This interest rate is required in order to discount future earnings to their present value.*

ture of the borrower's creditworthiness in such cases, it thus makes sense to assess qualitative data using a supplementary heuristic model.

This heuristic component also involves credit experts more heavily in the rating process than in the case of automated credit assessment using a statistical or causal model, meaning that combining models will also serve to increase user acceptance.

In the sections below, three different architectures for the combination of these model types are presented.

3.4.1 Horizontal Linking of Model Types

As statistical and causal models demonstrate particular strength in the assessment of quantitative data, and at the same time most of these models cannot process qualitative data without significant additional effort, this combination of model types can be encountered frequently in practice. A statistical model or causal model is used to analyze annual financial statements or (in broader terms) to evaluate a borrower's financial situation. Qualitative data (e.g. management quality) is evaluated using a heuristic module included in the model. IT is then possible to merge the output produced by these two modules to generate an overall credit assessment.

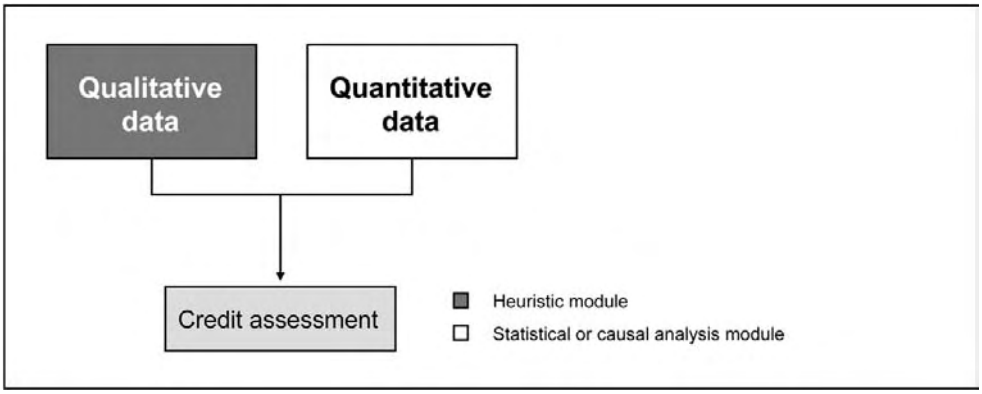

Chart 22: Horizontal Linking of Rating Models

Applied example:

One practical example of this combination of rating models can be found in the Deutsche Bundesbank's credit assessment procedure described in section 3.1.4. Annual financial statements are analyzed by means of statistical discriminant analysis. This quantitative creditworthiness analysis is supplemented with additional qualitative criteria which are assessed using a fuzzy logic system. The overall system is shown in chart 23.

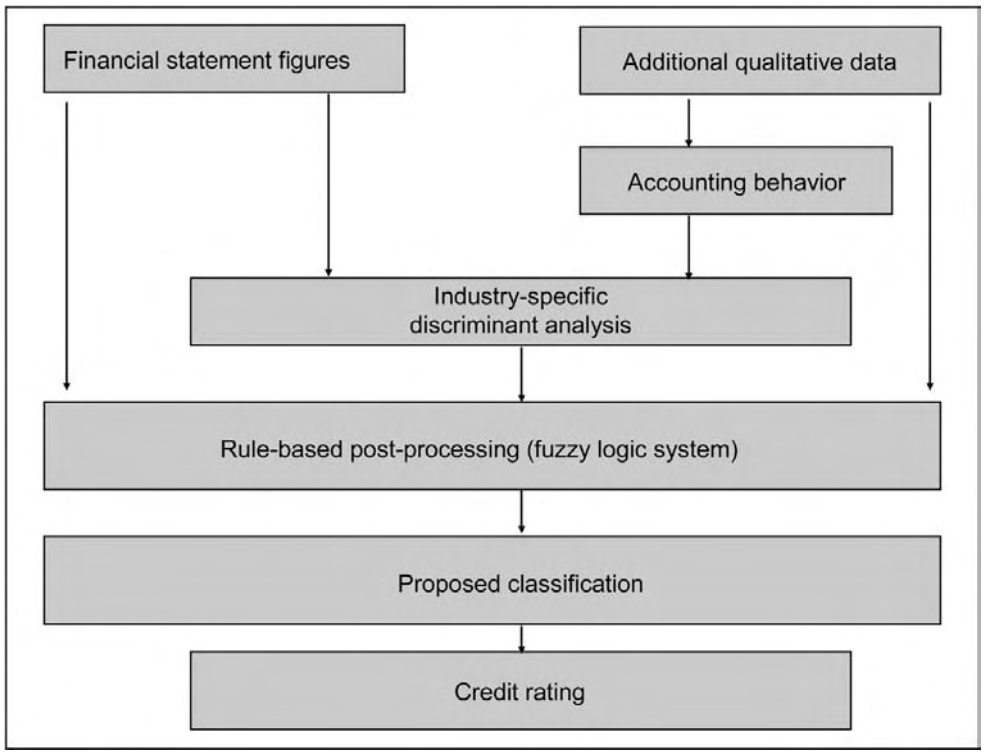

Chart 23: Architecture of Deutsche Bundesbank's Credit Assessment Procedure[45]

3.4.2 Vertical Linking of Model Types Using Overrides

This approach links partial ratings to generate a proposed classification, which can then be modified by a credit expert to yield the final credit rating. This downstream modification component based on expert knowledge constitutes a separate type of hybrid model. This combination first assesses quantitative as well as qualitative creditworthiness characteristics using a statistical or causal model. The result of this assessment is a proposed classification which can then be modified (within certain limits) by credit analysts on the basis of their expert knowledge.

In these combined models, it is important to define precisely the cases and the range in which overrides can be used.

In particular, facts which have already been used in a statistical or causal analysis module should not serve as the basis for later modifications by credit analysts. Instead, the heuristic component is important in order to include factors relevant to creditworthiness which are only known to the credit analyst and which could not be covered by the upstream module.

If the upstream module is modeled properly, however, overrides should only be necessary in some cases. The excessive use of overrides may indicate a lack of user acceptance or a lack of understanding of the rating model and should therefore be reviewed carefully in the course of validation.

[45] *Adapted from. BLOCHWITZ, STEFAN/EIGERMANN, JUDITH, Bonitätsbeurteilungsverfahren der Deutschen Bundesbank.*

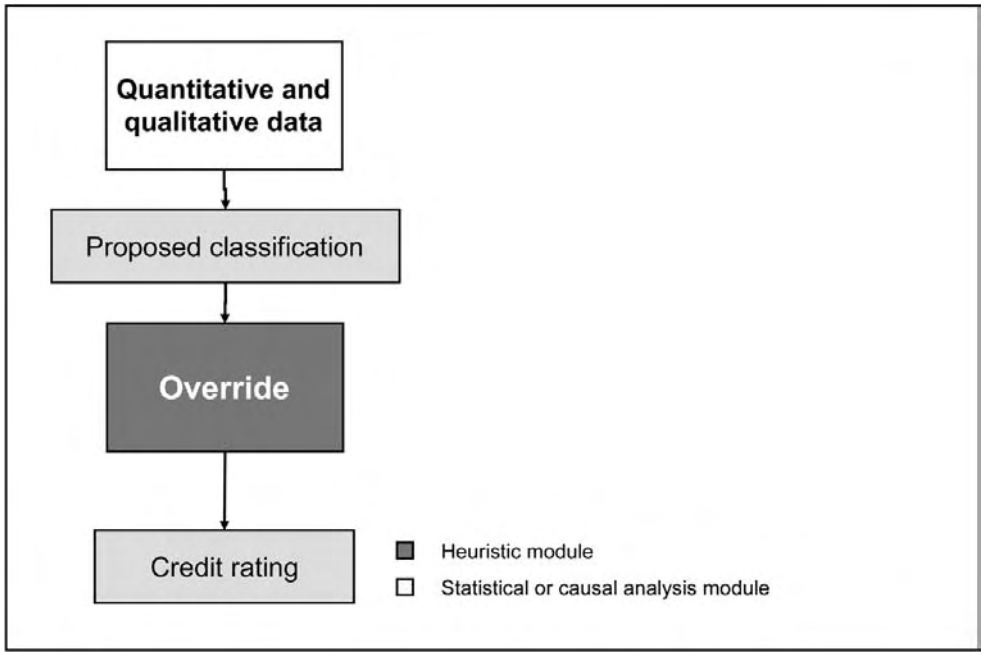

Chart 24: Vertical Linking of Rating Models Using Overrides

3.4.3 Upstream Inclusion of Heuristic Knock-Out Criteria

The core element of this combination of model types is the statistical module. However, this module is preceded by knock-out criteria defined on the basis of the practical experience of credit experts and the bank's individual strategy. If a potential borrower fulfills a knock-out criterion, the credit assessment process does not continue downstream to the statistical module.

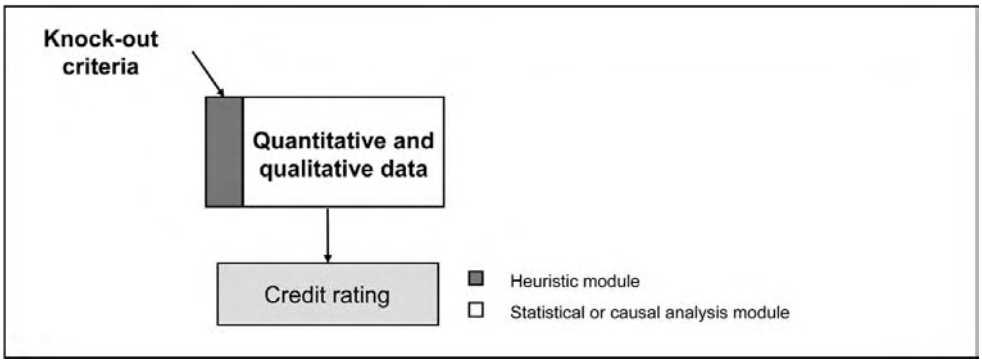

Chart 25: Hybrid Forms – Upstream Inclusion of Heuristic Knock-Out Criteria

Knock-out criteria form an integral part of credit risk strategies and approval practices in credit institutions. One example of a knock-out criterion used in practice is a negative report in the consumer loans register. If such a report is found, the bank will reject the credit application even before determining a differentiated credit rating using its in-house procedures.

4 Assessing the Models' Suitability for Various Rating Segments

In general, credit assessment procedures have to fulfill a number of requirements regardless of the rating segments in which they are used. These requirements are the result of business considerations applied to credit assessment as well as documents published on the IRB approaches under Basel II. The fundamental requirements are listed in chart 26 and explained in detail further below.

Target value PD
It must be possible to represent the rating result as a default probability.

Completeness
The rating result should take into account all available information relevant to creditworthiness.

Objectivity
The rating result should be reproducible by various subjects, i.e. with the same information, different analysts should come to as similar a rating result as possible.

Acceptance
In the eyes of the user, the rating model should assess the borrower's creditworthiness accurately.

Consistency
The rating must not contradict accepted theories and methods.

Chart 26: Fundamental Requirements of Rating Models

4.1 Fulfillment of Essential Requirements

4.1.1 PD as Target Value

The probability of default reflected in the rating forms the basis for risk management applications such as risk-based loan pricing. Calculating PD as the target value is therefore a basic prerequisite for a rating model to make sense in the business context.

The data set used to calculate PD is often missing in *heuristic models* and might have to be accumulated by using the rating model in practice. Once this requirement is fulfilled, it is possible to calibrate results to default probabilities even in the case of heuristic models (see section 5.3).

Statistical models are developed on the basis of an empirical data set, which makes it possible to determine the target value PD for individual rating classes by calibrating results with the empirical development data. Likewise, it is possible to calibrate the rating model (ex post) in the course of validation using the data gained from practical deployment.

One essential benefit of *logistic regression* is the fact that it enables the direct calculation of default probabilities. However, calibration or rescaling may also

be necessary in this case if the default rate in the sample deviates from the average default rate of the rating segment depicted.

In the case of *causal models*, the target value PD can be calculated for individual rating classes using data gained from practical deployment. In this case, the model directly outputs the default parameter to be validated.

4.1.2 Completeness

In order to ensure the completeness of credit rating procedures, Basel II requires banks to take all available information into account when assigning ratings to borrowers or transactions.[46] This should also be used as a guideline for best practices.

As a rule, *classic rating questionnaires* only use a small number of characteristics relevant to creditworthiness. For this reason, it is important to review the completeness of factors relevant to creditworthiness in a critical light. If the model has processed a sufficient quantity of expert knowledge, it can cover all information categories relevant to credit assessment.

Likewise, *qualitative systems* often use only a small number of creditworthiness characteristics. As expert knowledge is also processed directly when the model is applied, however, these systems can cover most information categories relevant to creditworthiness.

The computer-based processing of information enables *expert systems* and *fuzzy logic systems* to take a large number of creditworthiness characteristics into consideration, meaning that such a system can cover all of those characteristics if it is modeled properly.

As a large number of creditworthiness characteristics can be tested in the development of *statistical models*, it is possible to ensure the completeness of the relevant risk factors if the model is designed properly. However, in many cases these models only process a few characteristics of high discriminatory power when these models are applied, and therefore it is necessary to review completeness critically in the course of validation.

Causal models derive credit ratings using a theoretical business-based model and use only a few – exclusively quantitative – input parameters without explicitly taking qualitative data into account. The information relevant to creditworthiness is therefore only complete in certain segments (e.g. specialized lending, large corporate customers).

4.1.3 Objectivity

In order to ensure the objectivity of the model, a *classic rating questionnaire* should contain questions on creditworthiness factors which can be answered clearly and without room for interpretation.

Achieving high discriminatory power in *qualitative systems* requires that the rating grades generated by qualitative assessments using a predefined scale are as objective as possible. This can only be ensured by a precise, understandable and plausible user's manual and the appropriate training measures.

As *expert systems* and *fuzzy logic systems* determine the creditworthiness result using defined algorithms and rules, different credit analysts using the

[46] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 8.*

same information and the "correct" model inputs will receive the same rating results. In this respect, the model can be considered objective.

In *statistical models*, creditworthiness characteristics are selected and weighted using an empirical data set and objective methods. Therefore, we can regard these models as objective rating procedures. When the model is supplied properly with the same information, different credit analysts will be able to generate the same results.

If *causal models* are supplied with the "correct" input parameters, these models can also be regarded as objective.

4.1.4 Acceptance

As *heuristic models* are designed on the basis of expert opinions and the experience of practitioners in the lending business, we can assume that these models will meet with high acceptance.

The explanatory component in an *expert system* makes the calculation of the credit assessment result transparent to the user, thus enhancing the acceptance of such models.

In *fuzzy logic systems*, acceptance may be lower than in the case of expert systems, as the former require a greater degree of expert knowledge due to their modeling of "fuzziness" with linguistic variables. However, the core of fuzzy logic systems models the experience of credit experts, which means it is possible for this model type to attain the necessary acceptance despite its increased complexity.

Statistical rating models generally demonstrate higher discriminatory power than heuristic models. However, it can be more difficult to gain methodical acceptance for statistical models than for heuristic models. One essential reason for this is the large amount of expert knowledge required for statistical models. In order to increase acceptance and to ensure that the model is applied in an objectively proper manner, user training seminars are indispensable.

One severe disadvantage for the acceptance of *artificial neural networks* is their "black box" nature.[47] The increase in discriminatory power achieved by such methods can mainly be attributed to the complex network's ability to learn and the parallel processing of information within the network. However, it is precisely this complexity of network architecture and the distribution of information across the networks which make it difficult to comprehend the rating results. This problem can only be countered by the appropriate training measures (e.g. sensitivity analyses can make the processing of information in artificial neural network appear more plausible and comprehensible to the user).

Causal models generally meet with acceptance when users understand the fundamentals of the underlying theory and when the input parameters are defined in an understandable way which is also appropriate to the rating segment.

Acceptance in individual cases will depend on the accompanying measures taken in the course of introducing the rating model, and in particular on the transparency of the development process and adaptations as well as the quality of training seminars.

[47] *Cf. FÜSER, K., Mittelstandsrating mit Hilfe neuronaler Netzwerke, p. 374.*

4.1.5 Consistency

Heuristic models do not contradict recognized scientific theories and methods, as these models are based on the experience and observations of credit experts.

In the data set used to develop *empirical statistical rating models*, relationships between indicators may arise which contradict actual business considerations. Such contradictory indicators have to be consistently excluded from further analyses. Filtering out these problematic indicators will serve to ensure consistency.

Causal models depict business interrelationships directly and are therefore consistent with the underlying theory.

4.2 Suitability of Individual Model Types

The suitability of each model type is closely related to the data requirements for the respective rating segments (see chapter 2). The most prominent question in model evaluation is whether the quantitative and qualitative data used for credit assessment in individual segments can be processed properly. While quantitative data generally fulfills this condition in all models, differences arise with regard to qualitative data in statistical models.

In terms of discriminatory power and calibration, statistical models demonstrate clearly superior performance in practice compared to heuristic models. Therefore, banks are increasingly replacing or supplementing heuristic models with statistical models in practice. This is especially true in those segments for which it is possible to compile a sufficient data set for statistical model development (in particular corporate customers and mass-market banking). For these customer segments, statistical models are the standard.

However, the quality and suitability of the rating model used cannot be assessed on the basis of the model type alone. Rather, validation should involve regular reviews of a rating model's quality on the basis of ongoing operations. Therefore, we only describe the essential, observable strengths and weaknesses of the rating models for each rating segment below, without attempting to recommend, prescribe or rule out rating models in individual segments.

4.2.1 Heuristic Models

In principle, heuristic models can be used in all rating segments. However, in terms of discriminatory power, statistical models are clearly superior to heuristic models in the corporate customer segment and in mass-market banking. Therefore, the use of statistical models is preferable in those particular segments if a sufficient data set is available.

When heuristic models are used in practice, it is important in any case to review their discriminatory power and forecasting accuracy in the course of validation.

Classic Rating Questionnaires

The decisive success component in a classic rating questionnaire is the use of creditworthiness criteria for which the user can give clear and understandable answers. This will increase user acceptance as well as the objectivity of the model. Another criterion is the plausible and comprehensible assignment of points to specific answers. Answers which experience has shown to indicate high

creditworthiness have to be assigned a larger number of points than answers which point to lower creditworthiness. This ensures consistency and is a fundamental prerequisite for acceptance among users and external interest groups.

Qualitative Systems
The business-based user's manual is crucial to the successful deployment of a qualitative system. This manual has to define in a clear and understandable manner the circumstances under which users are to assign certain ratings for each creditworthiness characteristic. Only in this way is it possible to prevent credit ratings from becoming too dependent on the user's subjective perceptions and individual levels of knowledge. Compared to statistical models, however, qualitative systems remain severely limited in terms of objectivity and performance capabilities.

Expert Systems
Suitable rating results can only be attained using an expert system if they model expert experience in a comprehensible and plausible way, and if the inference engine developed is capable of making reasonable conclusions.

Additional success factors for expert systems include the knowledge acquisition component and the explanatory component. The advantages of expert systems over classic rating questionnaires and qualitative systems are their more rigorous structuring and their greater openness to further development. However, it is important to weigh these advantages against the increased development effort involved in expert systems.

Fuzzy Logic Systems
The comments above regarding expert systems also apply to these systems. However, fuzzy logic systems are substantially more complex than expert systems due to their additional modeling of "fuzziness," therefore they involve even greater development effort. For this reason, the application of a fuzzy logic system does not appear to be appropriate for mass-market banking or for small businesses (cf. section 2.3) compared to conventional expert systems.

4.2.2 Statistical Models

In the development stage, statistical models always require a sufficient data set, especially with regard to defaulted borrowers. Therefore, is often impossible to apply these statistical models to all rating segments in practice. For example, default data on governments and the public sector, financial service providers, exchange-listed/international companies, as well as specialized lending operations are rarely available in a quantity sufficient to develop statistically valid models.

The requirements related to sample sizes sufficient for developing a statistical model are discussed in section 5.1.3. In that section, we present one possible method of obtaining valid model results using a smaller sample (bootstrap/resampling).

In addition to the required data quantity, the representativity of data also has to be taken into account (cf. section 5.1.2).

Compared to heuristic models, statistical models generally demonstrate higher discriminatory power, meaning that heuristic models can be complemented by statistical model components if a sufficient data set is available. In practice, automated credit decision-making is often only possible with statistical models due to the high goodness-of-fit requirements involved.

Multivariate Discriminant Analysis

As a method, discriminant analysis can generally be applied to all rating segments. However, limitations do arise in the case of qualitative data, which cannot be processed directly in this form of analysis. Therefore, this type of rating model is especially suitable for analyzing *quantitative data*, for example annual financial statements for corporate customers, bank account activity data in various rating segments, as well as financial information provided by retail customers.

When assessing the applicability of a discriminant analysis model, it is at least necessary to check whether it fulfills the formal mathematical requirements, especially the normal distribution of creditworthiness characteristics. In practice, however, these requirements are often disregarded for quantitative indicators.

If the assumption of normal distribution is not fulfilled, the resulting model could be less than optimal, that is, the rating model will not necessarily attain its maximum discriminatory power. Therefore, banks should review the effect on the model output at the latest during validation.

Regression Models

As methods, regression models can generally be employed in all rating segments. No particular requirements are imposed on the statistical characteristics of the creditworthiness factors used, which means that all types of *quantitative and qualitative creditworthiness characteristics* can be processed without problems.

However, when ordinal data are processed, it is necessary to supply a sufficient quantity of data for each category in order to enable statistically significant statements; this applies especially to defaulted borrowers.[48]

Another advantage of regression models is that their results can be interpreted directly as default probabilities. This characteristic facilitates the calibration of the rating model (see section 5.3.1).

Artificial Neural Networks

In terms of method, artificial neural networks can generally be employed in all rating segments. Artificial neural networks do not impose formal mathematical requirements on the input data, which means that these models can process both quantitative and qualitative data without problems.

In order to "learn" connections properly, however, artificial neural networks require a substantially larger quantity of data in the development stage than other statistical models. Methods which can be applied to regression models and discriminant analysis with a small sample size (e.g. bootstrap/resampling)

[48] *Cf. EIGERMANN, J.; Quantitatives Credit-Rating mit qualitativen Merkmalen, p. 356.*

cannot be employed in artificial neural networks. For this reason, artificial neural networks can only be used for segments in which a sufficiently large quantity of data can be supplied for rating model development.

4.2.3 Causal Models

Option Pricing Models
In general, it is only possible to determine the input parameters required for these models (market value of equity, volatility of assets, etc.) reliably for *exchange-listed companies* and *financial service providers*, as in these cases the market value of equity and the volatility of assets can be derived from stock prices with relative ease. Using cash flow (simulation) models and additional modeling assumptions, the option pricing model can also be suitable for large companies which prepare balance sheets if a sufficiently long time series of the necessary balance sheet data is available and cash flows can be calculated reliably on the basis of planning data. In the case of smaller borrowers, the effort necessary for (company) valuation is too high and the calculation of parameters is too uncertain. However, should a bank decide to develop option pricing models for such rating segments nonetheless, it is necessary to review the calculated input parameters critically in terms of adequacy.

Cash Flow (Simulation) Models
Cash flow (simulation) models are especially well suited to *specialized lending*, as the primary source of funds for repaying the exposure is the income produced by the assets financed. This means that creditworthiness essentially depends on the future cash flows arising from the assets. Likewise, cash flow (simulation) models can be used as a preliminary processing module for option pricing models. In principle, cash flow (simulation) models can also be used for exchange-listed companies and in some cases for large companies which prepare balance sheets.

The decisive factor in the success of a cash flow (simulation) model is the suitable calculation of future cash flows and discounting factors. If cash flows are calculated directly on the basis of historical values, it is important to ensure that the data set used is representative of the credit institution and to review the forecasting power of the historical data.

5 Developing a Rating Model

In the previous sections, we discussed rating models commonly used in the market as well as their strengths and weaknesses when applied to specific rating segments. A model's suitability for a rating segment primarily depends on the data and information categories required for credit assessment, which were defined in terms of best business practices in chapter 3.

The *fundamental decision* to use a specific rating model for a certain rating segment is followed by the actual development of the rating procedure. This chapter gives a detailed description of the essential steps in the development of a rating procedure under the best-practice approach.

The procedure described in this document is based on the development of a *statistical* rating model, as such systems involve special requirements regarding

the data set and statistical testing. The success of statistical rating procedures in practice depends heavily on the development stage. In many cases, it is no longer possible to remedy critical development errors once the development stage has been completed (or they can only be remedied with considerable effort).

In contrast, *heuristic* rating models are developed on the basis of expert experience which is not verified until later with statistical tests and an empirical data set. This gives rise to considerable degrees of freedom in developing heuristic rating procedures. For this reason, it is not possible to present a generally applicable development procedure for these models. As expert experience is not verified by statistical tests in the development stage, validation is especially important in the ongoing use of heuristic models.

Roughly the same applies to *causal models*. In these models, the parameters for a financial theory-based model are derived from external data sources (e.g. volatility of the market value of assets from stock prices in the option pricing model) without checking the selected input parameters against an empirical data set in the development stage. Instead, the input parameters are determined on the basis of theoretical considerations. Suitable modeling of the input parameters is thus decisive in these rating models, and in many cases it can only be verified in the ongoing operation of the model.

In order to ensure the *acceptance* of a rating model, it is crucial to include the *expert experience* of practitioners in credit assessment throughout the development process. This is especially important in cases where the rating model is to be deployed in multiple credit institutions, as is the case in data pooling solutions.

The first step in rating development is *generating the data set*. Prior to this process, it is necessary to define the precise requirements of the data to be used, to identify the data sources, and to develop a stringent data cleansing process.

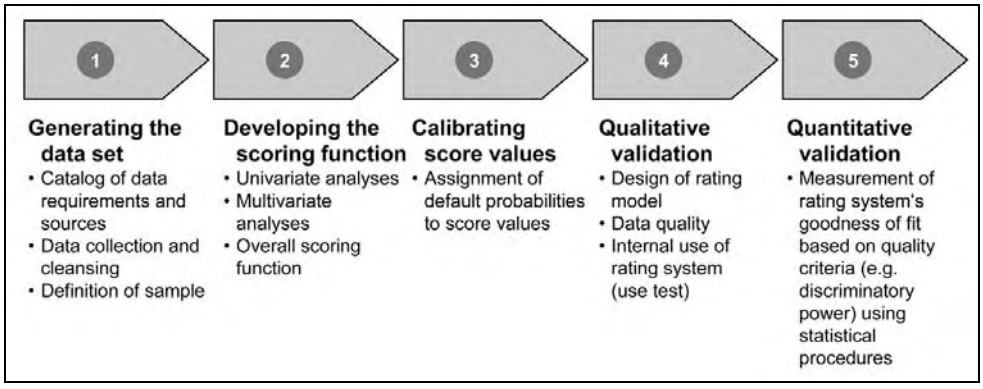

Chart 27: Procedure for Developing a Rating Model

Important requirements for the empirical data set include the following:
- Representativity of the data for the rating segment
- Data quantity (in order to enable statistically significant statements)
- Data quality (in order to avoid distortions due to implausible data).

For the purpose of *developing a scoring function*, it is first necessary to define a catalog of criteria to be examined. These criteria should be plausible from the business perspective and should be examined individually for their discriminatory power (*univariate analysis*). This is an important preliminary stage before

examining the interaction of individual criteria in a scoring function (*multivariate analysis*), as it enables a significant reduction in the number of relevant creditworthiness criteria. The multivariate analyses yield partial scoring functions which are combined in an overall scoring function in the model's architecture.

The scoring function determines a score which reflects the borrower's creditworthiness. The score alone does not represent a default probability, meaning that it is necessary to assign default probabilities to score values by means of *calibration*, which concludes the actual rating development process.

The ensuing steps (4 and 5) are not part of the actual process of rating development; they involve the *validation* of rating systems during the actual operation of the model. These steps are especially important in reviewing the performance of rating procedures in ongoing operations. The aspects relevant in this context are described in chapter 6 (Validation).

5.1 Generating the Data Set

A rating model developed on an empirical basis can only be as good as the underlying data. The quality of the data set thus has a decisive influence on the goodness of fit and the discriminatory power of a rating procedure. The data collection process requires a great deal of time and effort and must undergo regular quality assurance reviews. The main steps in the process of generating a suitable data set are described below.

Chart 28: Procedure for Generating the Data Set

5.1.1 Data Requirements and Sources

Before actual data collection begins, it is necessary to define the data and information to be gathered according to the respective rating segment. In this process, all data categories relevant to creditworthiness should be included. We discussed the data categories relevant to individual rating segments in chapter 2.

First, it is necessary to specify the data to be collected more precisely on the basis of the defined data categories. This involves defining various quality assurance requirements for quantitative, qualitative and external data.

Quantitative data such as annual financial statements are subject to various legal regulations, including those stipulated under commercial law. This means that they are largely standardized, thus making it is possible to evaluate a company's economic success reliably using accounting ratios calculated from annual financial statements. However, in some cases special problems arise where various accounting standards apply. This also has to be taken into account when indicators are defined (see "Special Considerations in International Rating Models" below).

In many cases, other quantitative data (income and expense accounts, information from borrowers on assets/liabilities, etc.) are frequently not available in a standardized form, making it necessary to define a clear and comprehensible

data collection procedure. This procedure can also be designed for use in different jurisdictions.

Qualitative questions are always characterized by subjective leeway in assessment. In order to avoid unnecessarily high quality losses in the rating system, it is important to consider the following requirements:

- The questions have to be worded as simply, precisely and unmistakably as possible. The terms and categorizations used should be explained in the instructions.
- It must be possible to answer the questions unambiguously.
- When answering questions, users should have no leeway in assessment. It is possible to eliminate this leeway by (a) defining only a few clearly worded possible answers, (b) asking yes/no questions, or (c) requesting "hard" numbers.
- In order to avoid unanswered questions due to a lack of knowledge on the users' part, users must be able to provide answers on the basis of their existing knowledge.
- Different analysts must be able to generate the same results.
- The questions must appear sensible to the analyst. The questions should also give the analyst the impression that the qualitative questions will create a valid basis for assessment.
- The analyst must not be influenced by the way in which questions are asked.

If credit assessment also relies on *external data* (external ratings, credit reporting information, capital market information, etc.), it is necessary to monitor the quality, objectivity and credibility of the data source at all times.

When defining data requirements, the bank should also consider the availability of data. For this reason, it is important to take the possible sources of each data type into account during this stage. Possible data collection approaches are discussed in section 5.1.2.

Special Considerations in International Rating Models

In the case of international rating models, the selection and processing of the data used for the rating model should account for each country's specific legal framework and special characteristics. In this context, only two aspects are emphasized:

- Definition of the default event
- Use of deviating accounting standards.

The *default event* is the main target value in rating models used to determine PD. For this reason, it is necessary to note that the compatibility of the "default event" criterion among various countries may be limited due to country-specific options in defining a Basel II-compliant default event as well as differences in bankruptcy law or risk management practice. When a rating model is developed for or applied in multiple countries, it is necessary to ensure the uniform use of the term "default event."

Discrepancies in the use of *financial indicators* may arise between individual countries due to different accounting standards. These discrepancies may stem from different names for individual data fields in the annual financial statements or from different valuation options and reporting requirements for individual items in the statements. It is therefore necessary to ensure the uniform meaning

of specific data fields when using data from countries or regions where different accounting standards apply (e.g. Austria, EU, CEE). This applies analogously to rating models designed to process annual financial statements drawn up using varying accounting standards (e.g. balance sheets compliant with the Austrian Commercial Code or IAS).

One possible way of handling these different accounting standards is to develop translation schemes which map different accounting standards to each other. However, this approach requires in-depth knowledge of the accounting systems to be aligned and also creates a certain degree of imprecision in the data. It is also possible to use mainly (or only) those indicators which can be applied in a uniform manner regardless of specific details in accounting standards. In the same way, leeway in valuation within a single legal framework can be harmonized by using specific indicators. However, this may limit freedom in developing models to the extent that the international model cannot exhaust the available potential.

In addition, it is also possible to use *qualitative rating criteria* such as the country or the utilization of existing leeway in accounting valuation in order to improve the model.

5.1.2 Data Collection and Cleansing

In addition to ensuring data quality by defining data requirements at a very early stage in rating development, it is also important to keep data quantities in mind. Collecting a sufficient number of borrowers for rating procedure development ensures the statistical significance of statements on the suitability of specific creditworthiness criteria and of the scoring function(s) developed. Developing a rating model on the basis of empirical data requires both good and bad cases.

In line with the definition of default used in Basel II, bad borrowers are defined as follows in the draft EU directive on regulatory capital requirements:[49]

- *The credit institution considers that the obligor is unlikely to pay its credit obligations to the credit institution, the parent undertaking or any of its subsidiaries in full, without recourse by the credit institution to actions such as realising security (if held).*
- *The obligor is past due more than 90 days on any material credit obligation to the credit institution, the parent undertaking or any of its subsidiaries.. Overdrafts shall be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than current outstandings.*

When default probabilities in the rating model are used as the basic parameter PD for the IRB approach, the definition of those probabilities must conform with the reference definition in all cases. When the default definition is determined in the course of developing a rating model, the observability and availability of the default characteristic are crucial in order to identify bad borrowers consistently and without doubt.

Generating a *data set for model development* usually involves sampling, a full survey within the credit institution, or data pooling. In this context, a full survey is always preferable to sampling. In practice, however, the effort involved in

[49] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 1, No. 46. Annex D-5, No. 43 of the draft directive lists concrete indications of imminent insolvency.*

full surveys is often too high, especially in cases where certain data (such as qualitative data) are not stored in the institution's IT systems but have to be collected from paper-based files. For this reason, sampling is common in banks participating in data pooling solutions. This reduces the data collection effort required in banks without falling short of the data quantity necessary to develop a statistically significant rating system.

Full Surveys

Full data surveys in credit institutions involve collecting the required data on all borrowers assigned to a rating segment and stored in a bank's operational systems. However, full surveys (in the development of an individual rating system) only make sense if the credit institution has a sufficient data set for each segment considered.

The main advantage of full surveys is that the empirical data set is representative of the credit institution's portfolio.

Data Pooling – Opportunities and Implementation Barriers

As in general only a small number of borrowers default and the data histories in IT systems are not sufficiently long, gathering a sufficient number of bad cases is usually the main challenge in the data collection process. Usually, this problem can only be solved by means of *data pooling* among several credit institutions.[50] Data pooling involves collecting the required data from multiple credit institutions. The rating model for the participating credit institutions is developed on the basis of these data. Each credit institution contributes part of the data to the empirical data set, which is then used jointly by these institutions. This makes it possible to enlarge the data set and at the same time spread the effort required for data collection across multiple institutions. In particular, this approach also enables credit institutions to gather a sufficient amount of qualitative data. Moreover, the larger data set increases the statistical reliability of the analyses used to develop scoring functions.

Unlike full surveys within individual banks, a data pool implies that the empirical data set contains cases from multiple banks. For this reason, it is necessary to ensure that all banks contributing to the data pool fulfill the relevant data quality requirements. The basic prerequisite for high data quality is that the participating banks have smoothly functioning IT systems. In addition, it is crucial that the banks contributing to the pool use a uniform default definition.

Another important aspect of data pooling is adherence to any applicable legal regulations, for example in order to comply with data protection and banking secrecy requirements. For this reason, data should at least be made anonymous by the participating banks before being transferred to the central data pool. If the transfer of personal data from one bank to another cannot be avoided in data pooling, then it is necessary in any case to ensure that the other bank cannot determine the identity of the customers. In any case, the transfer of personal data requires a solid legal basis (legal authorization, consent of the respective customer).

[50] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 52/53.*

If no full survey is conducted in the course of data pooling, the data collection process has to be designed in such a way that the data pool is representative of the business of all participating credit institutions.

The *representativity* of the data pool means that it reflects the relevant structural characteristics – and their proportions to one another – in the basic population represented by the sample. In this context, for example, the basic population refers to all transactions in a given rating segment for the credit institutions contributing to the data pool. One decisive consequence of this definition is that we cannot speak of the data pool's representativity for the basic population in general, rather its representativity with reference to certain characteristics in the basic population which can be inferred from the data pool (and vice versa).

For this reason, it is first necessary to define the relevant representative characteristics of data pools. The characteristics themselves will depend on the respective rating segment. In corporate ratings, for example, the distribution across regions, industries, size classes, and legal forms of business organization might be of interest. In the retail segment, on the other hand, the distribution across regions, professional groups, and age might be used as characteristics. For these characteristics, it is necessary to determine the frequency distribution in the basic population.

If the data pool compiled only comprises a sample drawn from the basic population, it will naturally be impossible to match these characteristics exactly in the data pool. Therefore, an acceptable deviation bandwidth should be defined for a characteristic at a certain confidence level (see chart 29).

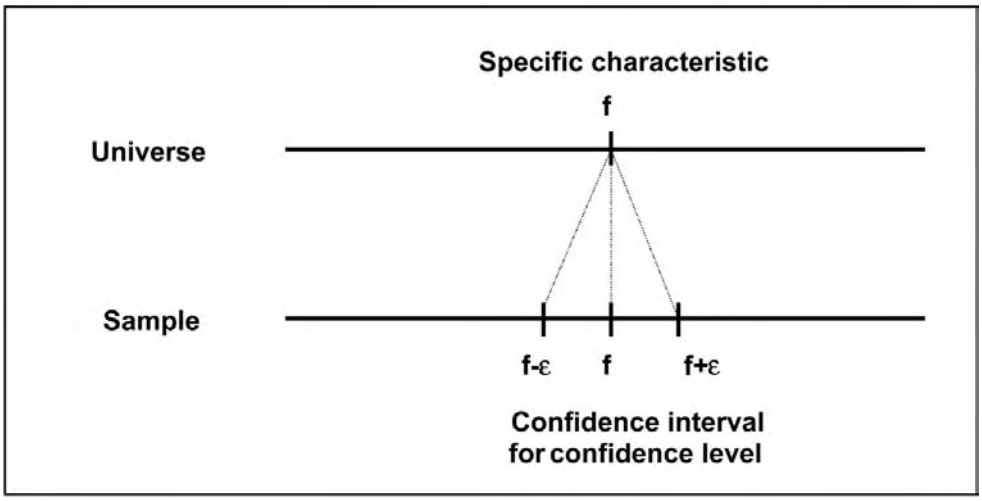

Chart 29: Mathematical Depiction of Representativity

The example below is intended to illustrate this representativity concept:

Assume that 200 banks contribute to a data pool. In order to minimize the time and effort required for data collection, the individual banks do not conduct full surveys. Instead, each bank is required to contribute 30 cases in the respective rating segment, meaning that the sample will comprise a total of 6,000 cases.

The population consists of all transactions in the relevant rating segment for the credit institutions contributing to the data pool. As the segment in question is the corporate segment, the distribution across industries is defined as an

essential structural characteristic. The sample only consists of part of the basic population, which means that the distribution of industries in the sample will not match that of the population exactly. However, it is necessary to ensure that the sample is representative of the population. Therefore, it is necessary to verify that deviations in industry distribution can only occur within narrow bandwidths when data are collected.

Data Pooling – Data Collection Process

This method of ensuring representativity has immediate consequences in the selection of credit institutions to contribute to a data pool. If, for example, regional distribution plays an important role, then the banks selected for the data collection stage should also be well distributed across the relevant regions in order to avoid an unbalanced regional distribution of borrowers.

In addition to the selection of suitable banks to contribute to the data pool, the following preliminary tasks are necessary for data collection:

- Definition of the catalog of data to be collected (see section 5.1.1) as well as the number of good and bad cases to be collected per bank.
- In order to ensure uniform understanding, it is crucial to provide a comprehensive business-based guide describing each data field.
- For the sake of standardized data collection, the banks should develop a uniform data entry tool. This tool will facilitate the central collection and evaluation of data.
- A user's manual is to be provided for data collection procedures and for the operation of the data entry tool.

In order to ensure that the sample captures multiple stages of the business cycle, it should include *data histories* spanning several years. In this process, it is advisable to define *cutoff dates* up to which banks can evaluate and enter the available information in the respective data history.

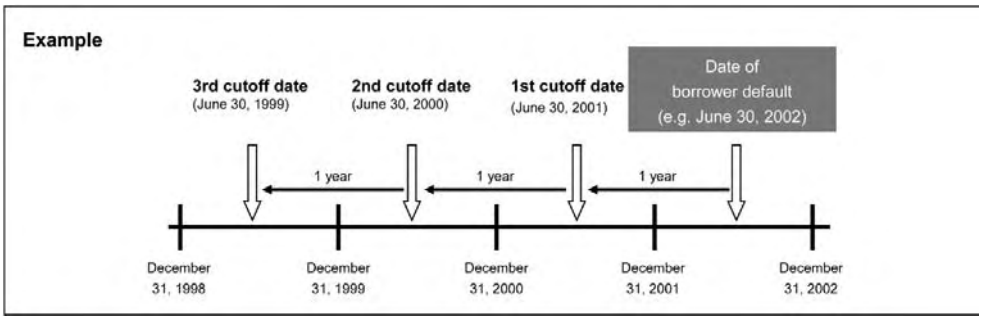

Chart 30: Creating a Data History of Bad Cases

For *bad cases*, the time of default provides a "natural" point of reference on the basis of which the cutoff dates in the data history can be defined. The interval selected for the data histories will depend on the desired forecasting horizon. In general, a one-year default probability is to be estimated, meaning that the cutoff dates should be set at 12-month intervals. However, the rating model can also be calibrated for longer forecasting periods. This procedure is illustrated in chart 30 on the basis of 12-month intervals.

Good cases refer to borrowers for whom no default has been observed up to the time of data collection. This means that no "natural" reference time is avail-

able as a basis for defining the starting point of the data history. However, it is necessary to ensure that these cases are indeed good cases, that is, that they do not default within the forecasting horizon after the information is entered. For good cases, therefore, it is only possible to use information which was available within the bank at least 12 months before data collection began (= 1st cutoff date for good cases). Only in this way is it possible to ensure that no credit default has occurred (or will occur) over the forecasting period. Analogous principles apply to forecasting periods of more than 12 months. If the interval between the time when the information becomes available and the start of data collection is shorter than the defined forecasting horizon, the most recently available information cannot be used in the analysis.

If *dynamic indicators* (i.e. indicators which measure changes over time) are to be defined in rating model development, quantitative information on at least two successive and viable cutoff dates have to be available with the corresponding time interval.

When data are compiled from various information categories in the data record for a specific cutoff date, the timeliness of the information may vary. For example, bank account activity data are generally more up to date than qualitative information or annual financial statement data. In particular, annual financial statements are often not available within the bank until 6 to 8 months after the balance sheet date.

In the organization of decentralized data collection, it is important to define in advance how many good and bad cases each bank is to supply for each cutoff date. For this purpose, it is necessary to develop a stringent data collection process with due attention to possible time constraints. It is not possible to make generally valid statements as to the time required for decentralized data collection because the collection period depends on the type of data to be collected and the number of banks participating in the pool. The workload placed on employees responsible for data collection also plays an essential role in this context. From practical experience, however, we estimate that the collection of qualitative and quantitative data from 150 banks on 15 cases each (with a data history of at least 2 years) takes approximately four months.

In this context, the actual process of collecting data should be divided into several blocks, which are in turn subdivided into individual stages. An example of a data collection process is shown in chart 31.

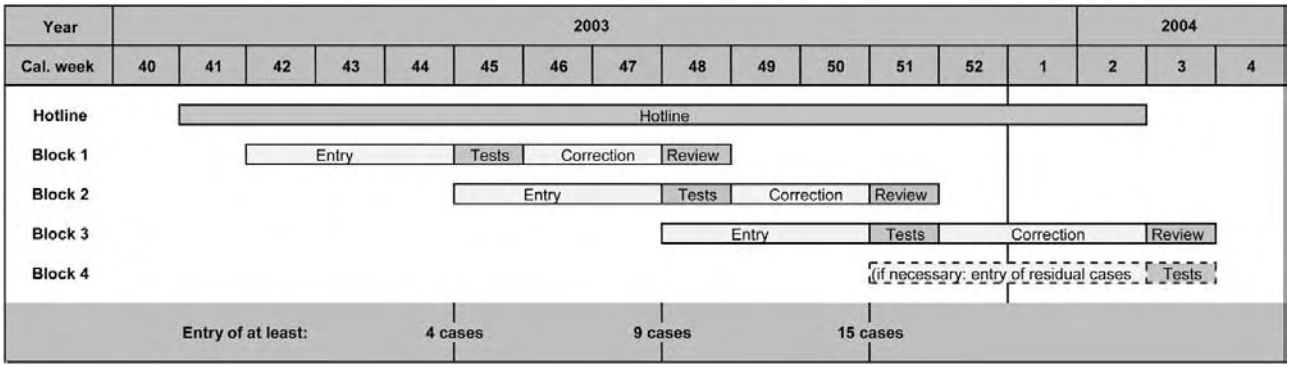

Chart 31: Example of a Data Collection Process

For each block, interim objectives are defined with regard to the cases entered, and the central project office monitors adherence to these objectives. This block-based approach makes it possible to detect and remedy errors (e.g. failure to adhere to required proportions of good and bad cases) at an early stage. At the end of the data collection process, it is important to allow a sufficient period of time for the (ex post) collection of additional cases required.

Another success factor in decentralized data collection is the constant availability of a hotline during the data collection process. Intensive support for participants in the process will make it possible to handle data collection problems at individual banks in a timely manner. This measure can also serve to accelerate the data collection process. Moreover, feedback from the banks to the central hotline can also enable the central project office to identify frequently encountered problems quickly. The office can then pass the corresponding solutions on to all participating banks in order to ensure a uniform understanding and the consistently high quality of the data collected.

Data Pooling – Decentralized Data Collection Process

It is not possible to prevent errors in data collection in decentralized collection processes, even if a comprehensive business-based user's manual and a hotline are provided. For this reason, it is necessary to include stages for testing and checking data in each data collection block (see chart 21). The quality assurance cycle is illustrated in detail in chart 32.

The objectives of the quality assurance cycle are as follows:

– to avoid systematic and random entry errors
– to ensure that data histories are depicted properly
– to ensure that data are entered in a uniform manner
– to monitor the required proportions of good and bad cases.

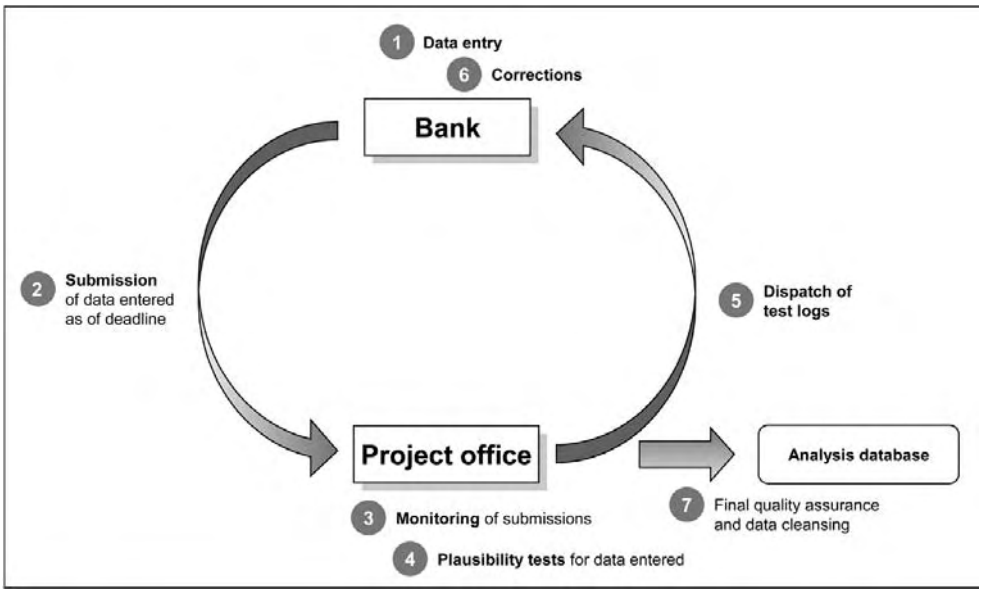

Chart 32: Data Collection and Cleansing Process in Data Pooling Solutions

Within the *quality assurance model* presented here, the banks contributing to the data pool each enter the data defined in the "Data Requirements and Data Sources" stage independently. The banks then submit their data to a central project office, which is responsible for monitoring the timely delivery of data as well as performing data plausibility checks. Plausibility checks help to ensure uniform data quality as early as the data entry stage. For this purpose, it is necessary to develop plausibility tests for the purpose of checking the data collected.

In these *plausibility tests*, the items to be reviewed include the following:

- Does the borrower entered belong to the relevant rating segment?
- Were the structural characteristics for verifying representativity entered?
- Was the data history created according to its original conception?
- Have all required data fields been entered?
- Are the data entered correctly? (including a review of defined relationships between positions in annual financial statements, e.g. assets = liabilities)

Depending on the rating segments involved, additional plausibility tests should be developed for as many information categories as possible.

Plausibility tests should be automated, and the results should be recorded in a *test log* (see chart 33).

Sequ.. Cust. No.	Customer No.	Sub-segment	Industry	Postal code	Provision created on	Processing-time	T1	Remarks	T2	Remarks	T3	Remarks
107	900.3966.1	1	3	96047	31.08.1998	150min	OK		OK		OK	
108	101.535.100	1	11	10409	01.12.1997	150min	OK		OK		OK	
109	900.2445.1	2	8	39114	30.10.1995	240min	OK		OK		OK	
110	900.5455.4	1	1	70173	31.08.1999	90min	OK		OK		OK	
111	900.8903.6	2	10	91489	30.09.2001	120min	Warning	Entry missing	OK		OK	
112	901.0044.2	1	3	99867	28.02.2001	240min	OK		OK		OK	
113	900.4201.9	2	12	36199	30.11.1998	180min	OK		OK		OK	
114	900.3363.0	2	7	03055	31.12.1997	150min	OK		Error!	linterval. < 12	OK	
115	900.8909.5	2	11	96199	30.09.2001	90min	OK		OK		OK	
116	901.0097.9	2	8	99885	30.03.2001	120min	OK		OK		OK	
117	900.6047.9	2	5	45127	31.12.2000	150min	OK		OK		OK	
118	900.7216.2	2	9	08358	31.12.1996	120min	OK		OK		OK	
119	900.7247.0	2	12	09212	31.12.2000	75min	OK		OK		OK	
120	900.6522.9	2	7	08371	30.11.2000	120min	OK		OK		OK	

Chart 33: Sample Test Log

The central project office should return the test log promptly so that the bank can search for and correct any errors. Once the data have been corrected, they can be returned to the project office and undergo the data cleansing cycle once again.

Data Pooling – Centralized Data Collection Process

Once the decentralized data collection and cleansing stages are completed, it is necessary to create a central *analysis database* and to perform a final quality check. The steps in this process are shown in chart 34.

In the first step, it is necessary to extract the data from the banks' data entry tools and merge them in an overall database.

The second step involves the final data cleansing process for this database.

The first substep of this process serves to ensure data integrity. This is done using a data model which defines the relationships between the data elements. In this context, individual banks may have violated integrity conditions in the course of decentralized data collection. The data records for which these conditions cannot be met are to be deleted from the database. Examples of possible integrity checks include the following:

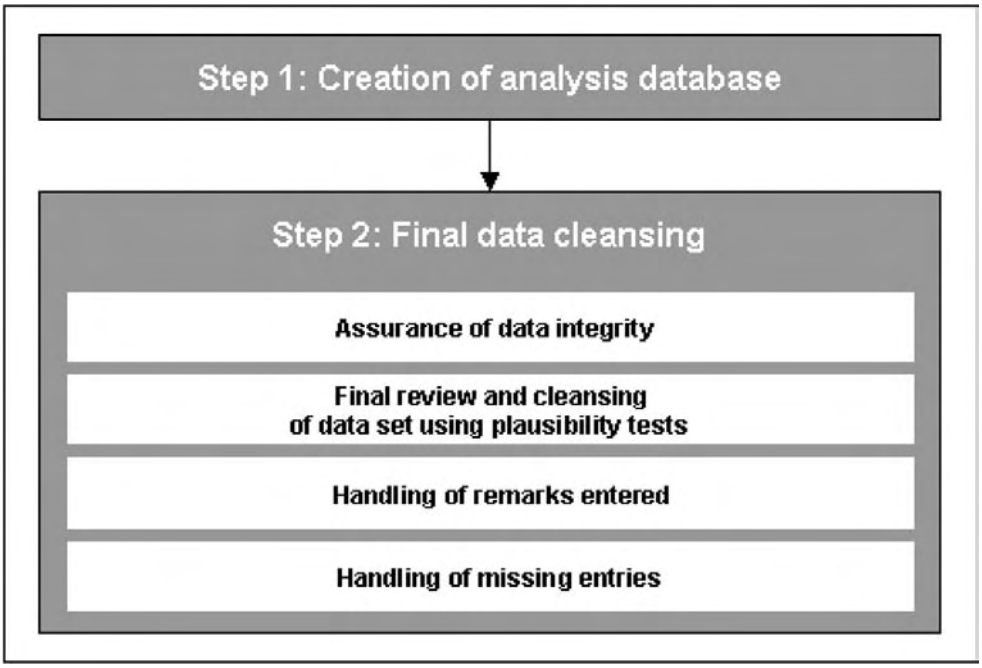

Chart 34: Creating the Analysis Database and Final Data Cleansing

– Were balance sheets entered which cannot be assigned to an existing borrower?
– Is there qualitative information which cannot be assigned to a cutoff date?
– Have different borrowers been entered with the same borrower number?
– Have different banks been entered under the same bank identifier code?
– Have any banks been entered for which no borrowers exist?

Once data integrity has been ensured, all of the remaining cases should be checked once again using the plausibility tests developed for decentralized data collection. This serves to ensure that all of the data found in the analysis database are semantically correct. In order to avoid reducing the size of the data set more than necessary, it is advisable to correct rather than delete data records wherever possible.

In order to enable banks to enter important notes on the use of a data record, the data should also include an additional field for remarks. These remarks are to be viewed and evaluated in the data collection process. In individual cases, it will then be necessary to decide whether the remarks have an effect on the use of a case or not.

For information on dealing with missing values, please refer to section 5.2.1.

The combination of decentralized and centralized data cleansing ensures a high level of *data quality* in data pooling solutions. This is a fundamental prerequisite for developing meaningful statistical models.

Data Pooling in the Validation and Ongoing Development of Rating Models

If a data pool was used to develop a rating model, it is necessary to ensure that decentralized data collection (in compliance with the high requirements of the rating development process) also continues into the rating validation and ongo-

ing development stages. However, data pooling in the validation and continued development of rating models is usually less comprehensive, as in this case it is only necessary to retrieve the rating criteria which are actually used. However, additional data requirements may arise for any necessary further developments in the pool-based rating model.

5.1.3 Definition of the Sample

The data gathered in the data collection and cleansing stages represent the overall sample, which has to be divided into an analysis sample and a validation sample. The analysis sample supports the actual development of the scoring functions, while the validation sample serves exclusively as a hold-out sample to test the scoring functions after development. In general, one can expect sound discriminatory power from the data records used for development. Testing the model's applicability to new (i.e. generally unknown) data is thus the basic prerequisite for the recognition of any classification procedure. In this context, it is possible to divide the overall sample into the analysis and validation samples in two different ways:

– Actual division of the database into the analysis and validation samples
– Application of a bootstrap procedure

In cases where sufficient data (especially regarding bad cases) are available to enable actual division into two sufficiently large subsamples, the first option should be preferred. This ensures the strict separation of the data records in the analysis and validation samples. In this way, it is possible to check the quality of the scoring functions (developed using the analysis sample) using the unknown data records in the validation sample.

In order to avoid bias due to subjective division, the sample should be split up by random selection (see chart 35). In this process, however, it is necessary to ensure that the data are representative in terms of their defined structural characteristics (see section 5.1.2).

Only those cases which fulfill certain minimum data quality requirements can be used in the analysis sample. In general, this is already ensured during the data collection and cleansing stage. In cases where quality varies within a database, the higher-quality data should be used in the analysis sample. In such cases, however, the results obtained using the validation sample will be considerably less reliable.

Borrowers included in the analysis sample must not be used in the validation sample, even if different cutoff dates are used. The analysis and validation samples thus have to be disjunct with regard to borrowers.

With regard to weighting good and bad cases in the analysis sample, two different procedures are conceivable:

– The analysis sample can be created in such a way that the proportion of bad cases is representative of the rating segment to be analyzed. In this case, calibrating the scoring function becomes easier (cf. section 5.3). For example, the result of logistic regression can be used directly as a probability of default (PD) without further processing or rescaling. This approach is advisable whenever the number of cases is not subject to restrictions in the data collection stage, and especially when a sufficient number of bad cases can be collected.

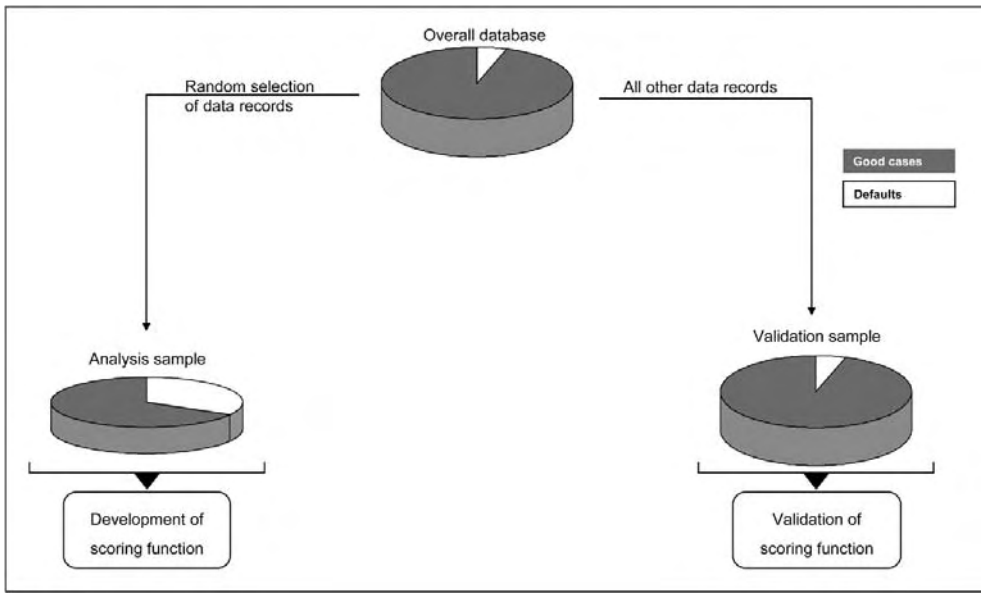

Chart 35: Creating the Analysis and Validation Samples

– If restrictions apply to the number of cases which banks can collect in the data collection stage, a higher proportion of bad cases should be collected. In practice, approximately one fourth to one third of the analysis sample comprises bad cases. The actual definition of these proportions depends on the availability of data in rating development. This has the advantage of maximizing the reliability with which the statistical procedure can identify the differences between good and bad borrowers, even for small quantities of data. However, this approach also requires the calibration and rescaling of calculated default probabilities (cf. section 5.3).

As an alternative or a supplement to splitting the overall sample, the *bootstrap method (resampling)* can also be applied. This method provides a way of using the entire database for development and at the same time ensuring the reliable validation of scoring functions.

In the bootstrap method, the overall scoring function is developed using the entire sample without subdividing it. For the purpose of validating this scoring function, the overall sample is divided several times into pairs of analysis and validation samples. The allocation of cases to these subsamples is random.

The coefficients of the factors in the scoring function are each calculated again using the analysis sample in a manner analogous to that used for the overall scoring function. Measuring the fluctuation margins of the coefficients resulting from the test scoring functions in comparison to the overall scoring function makes it possible to check the stability of the scoring function.

The resulting discriminatory power of the test scoring functions is determined using the validation samples. The mean and fluctuation margin of the resulting discriminatory power values are likewise taken into account and serve as indicators of the overall scoring function's discriminatory power for unknown data, which cannot be determined directly.

In cases where data availability is low, the bootstrap method provides an alternative to actually dividing the sample. Although this method does not

include out-of-sample validation, it is a statistically valid instrument which enables the optimal use of the information contained in the data without neglecting the need to validate the model with unknown data.

In this context, however, it is necessary to note that excessively large fluctuations – especially changes in the sign of coefficients and inversions of high and low coefficients in the test scoring functions – indicate that the sample used is too small for statistical rating development. In such cases, highly inhomogeneous data quantities are generated in the repeated random division of the overall sample into analysis and validation samples, which means that a valid model cannot be developed on the basis of the given sample.

5.2 Developing the Scoring Function

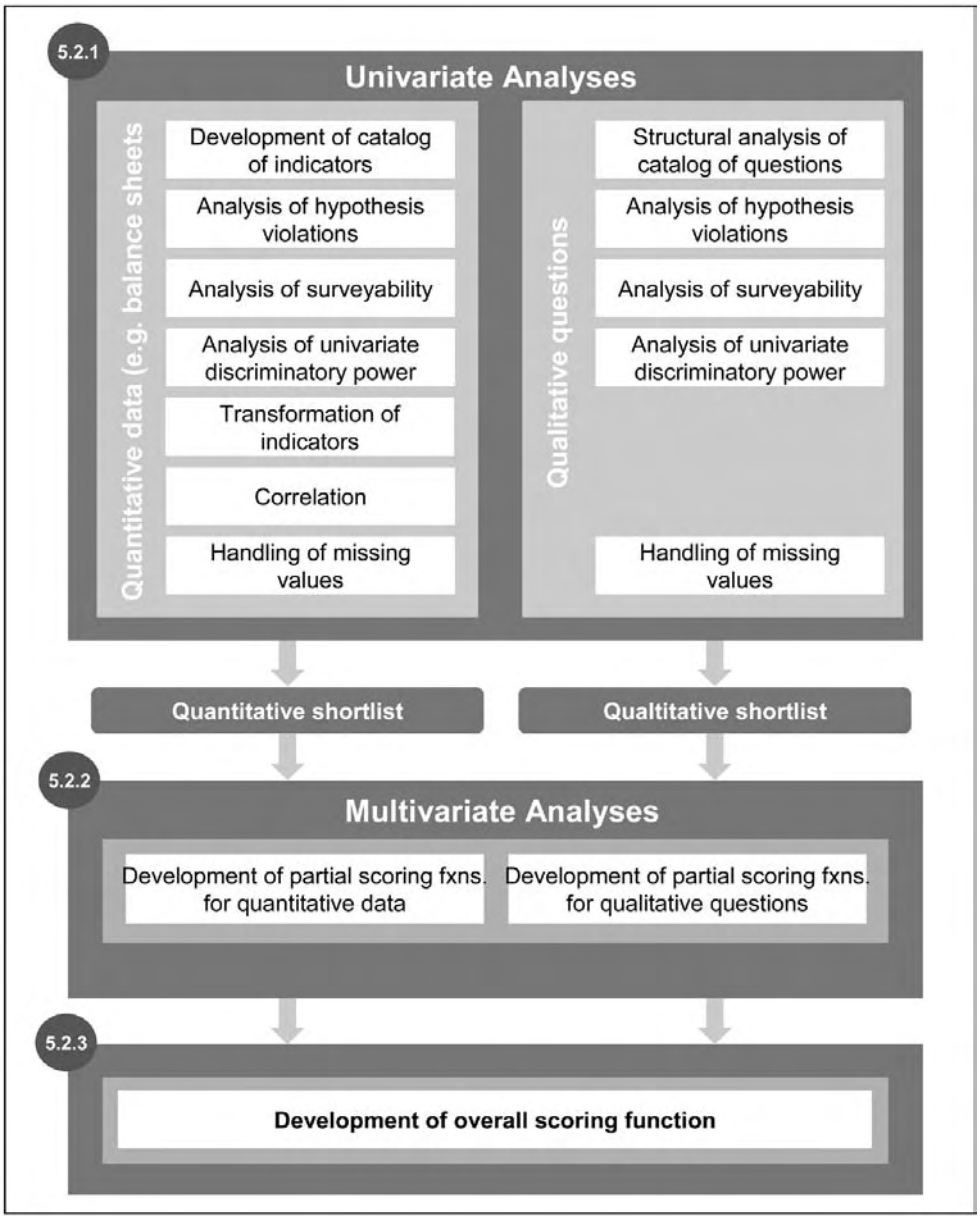

Chart 36: Scoring Function Development Procedure

Once a quality-assured data set has been generated and the analysis and validation samples have been defined, the actual development of the scoring function can begin. In this document, a "scoring function" refers to the core calculation component in the rating model. No distinction is drawn in this context between rating and scoring models for credit assessment (cf. introduction to chapter 3).

The development of the scoring function is generally divided into three stages, which in turn are divided into individual steps. In this context, we explain the fundamental procedure on the basis of an architecture which includes partial scoring functions for quantitative as well as qualitative data (see chart 36).

The individual steps leading to the overall scoring function are very similar for quantitative and qualitative data. For this reason, detailed descriptions of the individual steps are based only on quantitative data in order to avoid repetition. The procedure is described for qualitative data only in the case of special characteristics.

5.2.1 Univariate Analyses

The purpose of univariate analyses is to identify creditworthiness characteristics which make sense in the business context, can be surveyed with some ease, and show high discriminatory power for the purpose of developing the scoring function. The result of these analyses is a shortlist of fundamentally suitable creditworthiness characteristics. Preselecting creditworthiness characteristics reduces the complexity of the ensuing multivariate analyses, thus facilitating the process substantially.

The steps shown in chart 36 are described in detail below for quantitative data (e.g. from balance sheet analysis).

Developing a Catalog of Indicators

The first step is to develop a comprehensive catalog of indicators on the basis of the quantitative data from the data collection process. This catalog should include indicators from all business-related information categories which can support the assessment of a borrower's situation in terms of assets, finances, and income. These information categories determine the structure of the catalog of indicators. The indicators defined for each information category should ensure that a comprehensive assessment of the area is possible and that different aspects of each area are covered. For the purpose of assessing individual aspects, indicators can be included in different variants which may prove to be more or less suitable in the univariate and multivariate analyses. For this reason, a very large number of indicators – and in some cases very similar indicators – are defined in order to enable the best variants to be selected later in the process.

Another condition which indicators have to fulfill is that it must be possible to calculate them for all of the cases included in a segment. This deserves special attention in cases where a rating segment contains companies to which simplified accounting standards apply and for which not all balance sheet items are available.

In practice, indicators are usually defined in two steps:

1. In the first step, the quantitative data items are combined to form *indicator components*. These indicator components combine the numerous items into sums which are meaningful in business terms and thus enable the information to be structured in a manner appropriate for economic analysis. However, these absolute indicators are not meaningful on their own.
2. In order to enable comparisons of borrowers, the second step calls for the definition of *relative indicators* due to varying sizes. Depending on the type of definition, a distinction is made between constructional figures, relative figures, and index figures.[51]

For each indicator, it is necessary to postulate a *working hypothesis* which describes the significance of the indicator in business terms. For example, the working hypothesis "G > B" means that the indicator will show a higher average value for *g*ood companies than for *b*ad companies. Only those indicators for which a clear and unmistakable working hypothesis can be given are useful in developing a rating. The univariate analyses serve to verify whether the presumed hypothesis agrees with the empirical values.

In this context, it is necessary to note that the existence of a *monotonic working hypothesis* is a crucial prerequisite for all of the statistical rating models presented in chapter 3 (with the exception of artificial neural networks). In order to use indicators which conform to non-monotonic working hypotheses such as revenue growth (average growth is more favorable than low or excessively high growth), it is necessary to transform these hypotheses in such a way that they describe a monotonic connection between the transformed indicator's value and creditworthiness or the probability of default. The transformation to default probabilities described below is one possible means of achieving this end.

Analyzing Indicators for Hypothesis Violations

The process of analyzing indicators for hypothesis violations involves examining whether the empirically determined relationship confirms the working hypothesis. Only in cases where an indicator's working hypothesis can be confirmed empirically is it possible to use the indicator in further analyses. If this is not the case, the indicator cannot be interpreted in a meaningful way and is thus unsuitable for the development of a rating system which is comprehensible and plausible from a business perspective. The working hypotheses formulated for indicators can be analyzed in two different ways.

The first approach uses a *measure of discriminatory power* (e.g. the Powerstat value[52]) which is already calculated for each indicator in the course of univariate analysis. In this context, the algorithm used for calculation generally assumes "G > B" as the working hypothesis for indicators.[53] If the resulting discriminatory power value is positive, the indicator also supports the empirical hypothesis G > B. In the case of negative discriminatory power values, the empirical hypothesis is G < B. If the sign before the calculated discriminatory power value does not agree with that of the working hypothesis, this is considered a violation of the hypothesis and the indicator is excluded from further analyses. Therefore,

[51] *For more information on defining indicators, see BAETGE, J./HEITMANN, C., Kennzahlen.*

[52] *Powerstat (Gini coefficient, accuracy ratio) and alternative measures of discriminatory power are discussed in section 6.2.1.*

[53] *In cases where the working hypothesis is "G < B," the statements made further below are inverted.*

an indicator should only be used in cases where the empirical value of the indicator in question agrees with the working hypothesis at least in the analysis sample. In practice, working hypotheses are frequently tested in the analysis and validation samples as well as the overall sample.

One alternative is to calculate the *medians* of each indicator separately for the good and bad borrower groups. Given a sufficient quantity of data, it is also possible to perform this calculation separately for all time periods observed. This process involves reviewing whether the indicator's median values differ significantly for the good and bad groups of cases and correspond to the working hypothesis (e.g. for G > B: the group median for good cases is greater than the group median for bad cases). If this is not the case, the indicator is excluded from further analyses.

Analyzing the Indicators' Availability and Dealing with Missing Values

The analysis of an indicator's availability involves examining how often an indicator cannot be calculated in relation to the overall sample of cases. We can distinguish between two cases in which indicators cannot be calculated:

– The information necessary to calculate the indicator is not available in the bank because it cannot be determined using the bank's operational processes or IT applications. In such cases, it is necessary to check whether the use of this indicator is relevant to credit ratings and whether it will be possible to collect the necessary information in the future. If this is not the case, the rating model cannot include the indicator in a meaningful way.
– The indicator cannot be calculated because the denominator is zero in a division calculation. This does not occur very frequently in practice, as indicators are preferably defined in such a way that this does not happen in meaningful financial base values.

In multivariate analyses, however, a value must be available for each indicator in each case to be processed, otherwise it is not possible to determine a rating for the case. For this reason, it is necessary to handle missing values accordingly. It is generally necessary to deal with missing values before an indicator is transformed.

In the process of *handling missing values*, we can distinguish between four possible approaches:

3. Cases in which an indicator cannot be calculated are excluded from the development sample.
4. Indicators which do not attain a minimum level of availability are excluded from further analyses.
5. Missing indicator values are included as a separate category in the analyses.
6. Missing values are replaced with estimated values specific to each group.

Procedure (1) is often impracticable because it excludes so many data records from analysis that the data set may be rendered empirically invalid.

Procedure (2) is a proven method of dealing with indicators which are difficult to calculate. The lower the fraction of valid values for an indicator in a sample, the less suitable the indicator is for the development of a rating because its value has to be estimated for a large number of cases. For this reason, it is necessary to define a limit up to which an indicator is considered suitable for rating development in terms of availability. If an indicator can be calculated in less than approximately 80% of cases, it is not possible to ensure that missing values can

be handled in a statistically valid manner.[54] In such cases, the indicator has to be excluded from the analysis.

Procedure (3) is very difficult to apply in the development of scoring functions for quantitative data, as a missing value does not constitute independent information. However, for qualitative data it is entirely possible to use missing values as a separate category in the development of scoring functions. Due to the ordinal nature of this type of data, it is indeed possible to determine a connection between a value which cannot be determined and its effects on creditworthiness. For this reason, Procedure (3) can only be used successfully in the case of qualitative data.

For quantitative analyses, *Procedure (4)* is a suitable and statistically valid procedure for handling missing indicator values which reach the minimum surveyability level of 80% but are not valid in all cases. Suitable group-specific estimates include the medians for the groups of good and bad cases. The use of group-specific averages is not as suitable because averages can be dominated heavily by outliers within the groups. Group-specific estimates are essential in the analysis sample because the indicator's univariate discriminatory power cannot be analyzed optimally in the overall scoring function without such groupings. In the validation sample, the median of the indicator value for all cases is applied as the general estimate.

– *Estimate for analysis sample:* Separate medians of the indicator for good and bad cases.
– *Estimate for validation sample:* Median of the indicator for all cases.

As the validation sample is intended to simulate the data to be assessed with the rating model in the future, the corresponding estimate does not differentiate between good and bad cases. If a group-specific estimate were also used here, the discriminatory power of the resulting rating model could easily be overestimated using unknown data.

The result of the process of handling missing values is a database in which a valid value can be found for each shortlisted indicator in each case. This database forms the basis for multivariate analyses in the development of partial scoring functions for quantitative data.

Analysis of Univariate Discriminatory Power

In order to be used in a statistically valid manner, an indicator has to exhibit a certain level of discriminatory power in the univariate context. However, univariate discriminatory power only serves as an indication that the indicator is suitable for use within a rating model. Indicators which do not attain a discriminatory value which differs significantly from zero do not support their working hypotheses and should thus be excluded from the final rating model wherever possible.

In any case, it is necessary to perform the analysis of the indicators' univariate discriminatory power before handling missing values. Only those cases which return valid indicator values – not those with missing or invalid values – should be used in the univariate discriminatory power analyses.

[54] *Cf. also HEITMANN, C., Neuro-Fuzzy, p. 139 f.; JERSCHENSKY, A., Messung des Bonitätsrisikos von Unternehmen, p. 137; THUN, C., Entwicklung von Bilanzbonitätsklassifikatoren, p. 135.*

Transformation of Indicators

In order to make it easier to compare and process the various indicators in multivariate analyses, it is advisable to transform the indicators using a uniform scale. Due to the wide variety of indicator definitions used, the indicators will be characterized by differing value ranges. While the values for a constructional figure generally fall within the range $[0, 1]$, as in the case of expense rates, relative figures are seldom restricted to predefined value intervals and can also be negative. The best example is return on equity, which can take on very low (even negative) as well as very high values. Transformation standardizes the value ranges for the various indicators using a uniform scale.

One transformation commonly used in practice is the *transformation of indicators into probabilities of default (PD)*. In this process, the average default rates for disjunct intervals in the indicators' value ranges are determined empirically for the given sample. For each of these intervals, the default rate in the sample is calculated. The time horizon chosen for the default rate is generally the intended forecasting horizon of the rating model. The nodes calculated in this way (average indicator value and default probability per interval) are connected by nonlinear interpolation. The following logistic function might be used for interpolation:

$$TI = l + \frac{u - l}{1 + \exp(-aI + b)}.$$

In this equation, K and TK represent the values of the untransformed and transformed indicator, and o and u represent the upper and lower limits of the transformation. The parameters a and b determine the steepness of the curve and the location of the inflection point. The parameters a, b, u, and o have to be determined by nonlinear interpolation.

The result of the transformation described above is the assignment of a sample-based default probability to each possible indicator value. As the resulting default probabilities lie within the range $[u; o]$, outliers for very high or very low indicator values are effectively offset by the S-shaped curve of the interpolation function.

It is important to investigate hypothesis violations using the untransformed indicator because every indicator will meet the conditions of the working hypothesis G < B after transformation into a default probability. This is plausible because transformation into PD values indicates the probability of default on the basis of an indicator value. The lower the probability of default is, the "better" the borrower is.

Analyzing Indicator Correlations

The analysis of each indicator for hypothesis violations, discriminatory power and availability can serve to reduce the size of the indicator catalog substantially. However, the remaining indicators will show more or less strong similarities or correlations. In general, similar indicators depict the same information. For this reason, it is advantageous to use uncorrelated indicators wherever possible when developing a rating model, as this will ensure that the rating reflects various information categories. In addition, high correlations can lead to stability problems in the estimation of coefficients for the scoring functions. In such cases, the estimation algorithm used will not be able to uniquely identify the

coefficients of the linear combinations of indicators. The relevant literature does not indicate any binding guidelines for the maximum size of correlations between indicators. As a rule of thumb, however, pairs of indicators which show correlation coefficients greater than 0.3 should only be included in scoring functions with great caution.

One tool which can be used to examine indicator correlations is *hierarchical cluster analysis*.[55] Hierarchical cluster analysis involves creating groups (i.e. clusters) of indicators which show high levels of correlation within the clusters but only low levels of correlation between the various clusters.[56]

Those indicators which return high Powerstat values are selected from the clusters created. Indicators which have very similar definitions (e.g. various types of return) but for which it is not possible to decide at the time of cluster analysis which variant will be most suitable can be used in parallel in multivariate analysis. For the sake of the model's stability, however, it is advisable to avoid including highly correlated variables in the final rating model.

5.2.2 Multivariate Analysis

Indicator preselection yields a shortlist of indicators, and the objective of multivariate analysis is to develop a scoring function for these indicators. In this section, we present a scoring function for quantitative data. The procedure for developing scoring functions for qualitative information categories is analogous. The catalogs of indicators/questions reduced in univariate analyses form the basis for this development process.

In practice, banks generally develop multiple scoring functions in parallel and then select the function which is most suitable for the overall rating model. The following general requirements should be imposed on the development of the scoring functions:

- Objective indicator selection based on the empirical procedure
- Attainment of high discriminatory power
  For this purpose, as few indicators as possible should be used in order to increase the stability of the scoring function and to ensure an efficient rating procedure.
- Inclusion of as many different information categories as possible (e.g. assets situation, financial situation, income situation)
- Explicit selection or explicit exclusion of certain indicators in order to enhance or allow statements which are meaningful in business terms

On the basis of the shortlist of indicators, various scoring functions are then determined with attention to the requirements listed above. Banks which use discriminant analyses or regression models can estimate the indicators' coefficients using optimization algorithms from statistics software programs. The scoring function's coefficients are always optimized using the data in the analysis sample. The validation sample serves the exclusive purpose of testing

[55] *Cf. BACKHAUS ET AL., Multivariate Analysemethoden, chapter 6.*

[56] *Rank order correlation (Spearman correlation) is an especially well suited measure of correlation in untransformed indicators. In comparison to the more commonly used Pearson correlation, this method offers the advantage of performing calculations using only the ranks of indicator values, not the indicator values themselves. Rank order correlation also delivers suitable results for indicators which are not normally distributed and for small samples. For this reason, this method can be applied in particular to indicators which are not uniformly scaled. Cf. SACHS, L., Angewandte Statistik, sections 5.2/5.3.*

the scoring functions developed using the analysis sample (see also section 5.1.3).

The final scoring function can be selected from the body of available functions according to the following criteria, which are explained in greater detail further below:

– Checking the signs of coefficients
– Discriminatory power of the scoring function
– Stability of discriminatory power
– Significance of individual coefficients
– Coverage of relevant information categories

Checking the Signs of Coefficients

The coefficients determined in the process of developing the model have to be in line with the working business hypotheses postulated for the indicators. Therefore, indicators for which the working hypothesis is G > B should be input with positive signs, while indicators whose hypotheses are G < B should be entered with negative signs if larger function values are to indicate higher levels of creditworthiness.[57] In cases where this sign rule is violated, it is necessary to eliminate the scoring function because it cannot be interpreted in meaningful business terms. This situation arises frequently in the case of highly correlated indicators with unstable coefficients. For this reason, it is often possible to remedy this error by changing the indicators selected.

If all of the indicators to be included in the scoring function have been transformed into a uniform working hypotheses (e.g. by transforming them into default probabilities, cf. section 5.2.1), all coefficients have to bear the same sign.

Discriminatory Power of the Scoring Function

In cases where multiple scoring functions with plausible coefficients are available, the discriminatory power of the scoring functions for the forecasting horizon serves as the decisive criterion. In practice, discriminatory power is frequently measured using Powerstat values.[58]

Stability of Discriminatory Power

In addition to the level of discriminatory power, its stability is also a significant factor. In this context, it is necessary to differentiate between the stability of the scoring function when applied to unknown data (out-of-sample validation) and its stability when applied to longer forecasting horizons.

Scoring functions for which discriminatory power turns out to be substantially lower for the validation sample (see section 5.1.3) than for the analysis sample are less suitable for use in rating models because they fail when applied to unknown data. When selecting scoring functions, therefore, it is important to favor those functions which only show a slight decrease in discriminatory power in out-of-sample validation or in the calculation of average discriminatory power using the bootstrap method. In general, further attempts to

[57] *If higher function values imply lower creditworthiness, for example in the logistic regression results (which represent default probabilities), the signs are reversed.*

[58] *Powerstat (Gini coefficient, accuracy ratio) and alternative measures of discriminatory power are discussed in section 6.2.1.*

optimize the model should be made in cases where the difference in discriminatory power between the analysis and validation samples exceeds 10% as measured in Powerstat values.

Moreover, the stability of discriminatory power in the analysis and validation samples also has to be determined for time periods other than the forecasting horizon used to develop the model. Suitable scoring functions should show sound discriminatory power for forecasting horizons of 12 months as well as longer periods.

Significance of Individual Coefficients

In the optimization of indicator coefficients, a statistical hypothesis in the form of "coefficient ? 0" is postulated. This hypothesis can be tested using the significance measures (e.g. F-Tests) produced by most optimization programs.[59] On the basis of this information, it is also possible to realize algorithms for automatic indicator selection. In this context, all indicators whose optimized coefficients are not equal to zero at a predefined level of significance are selected from the sample. These algorithms are generally included in software packages for multivariate analysis.[60]

Coverage of Relevant Information Categories

An important additional requirement condition in the development of scoring functions is the coverage of all information categories (where possible). This ensures that the rating represents a holistic assessment of the borrower's economic situation.

Should multivariate analysis yield multiple scoring functions which are equivalent in terms of the criteria described, the scoring function which contains the most easily understandable indicators should be chosen. This will also serve to increase user acceptance.

Once the scoring function has been selected, it is possible to scale the score values (e.g. to a range of 0 to 100). This enables partial scores from various information categories to be presented in a simpler and more understandable manner.

5.2.3 Overall Scoring Function

If separate partial scoring functions are developed for quantitative and qualitative data, these functions have to be linked in the model's architecture to form an overall scoring function. The objective in this context is to determine the optimum weighting of the two data types.

In general, the personal traits of the business owner or manager influence the credit quality of enterprises in smaller-scale rating segments more heavily than in larger companies. For this reason, we can observe in practice that the influence of qualitative information categories on each overall scoring function increases as the size of the enterprises in the segment decreases. However, the weighting shown in chart 37 is only to be seen as a rough guideline and not as a binding requirement of all rating models suitable for use in practice.

[59] *Cf. (for example) SACHS, L., Angewandte Statistik, section 3.5.*

[60] *e.g. SPSS.*

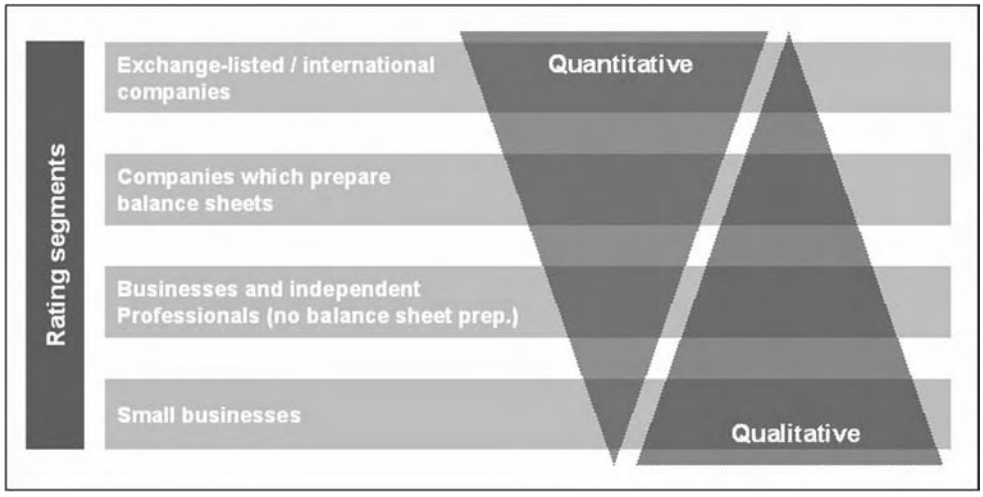

Chart 37: Significance of Quantitative and Qualitative Data in Different Rating Segments

In individual cases, banks can choose various approaches to attaining the optimum weighting of partial scoring functions in terms of discriminatory power. These approaches include the following:

– Optimization using multivariate discriminant analysis
– Optimization using a regression model
– Purely heuristic weighting of partial scoring functions
– Combined form: Heuristic definition of weights based on statistical results

Using statistical models offers the advantage of allowing the bank to determine the optimum weighting of partial scores objectively with a view to improving discriminatory power.

As an alternative, it is possible to assign relative weights to partial scoring functions exclusively on the basis of expert judgment. This would bring about a higher level of user acceptance, but it also has the disadvantage of potentially high losses in discriminatory power compared to the optimum level which can be attained using statistical methods.

Therefore, hybrid forms which use linear combination for partial scoring functions are common in practice. For this purpose, the Powerstat value of the overall scoring function for both the analysis and validation samples is calculated for various linear weighting possibilities (see chart 38). This makes it possible to define a range for overall scoring functions with very high discriminatory power.

Finally, the opinions of credit business practitioners are used to determine the weighting within the range identified. In summary, this approach offers the following advantages:

– It makes the influence of quantitative and qualitative data on the credit rating transparent due to the use of linear weighting.
– It ensures high user acceptance due to the inclusion of expert opinions.
– It also ensures high discriminatory power due to the inclusion of statistical methods.

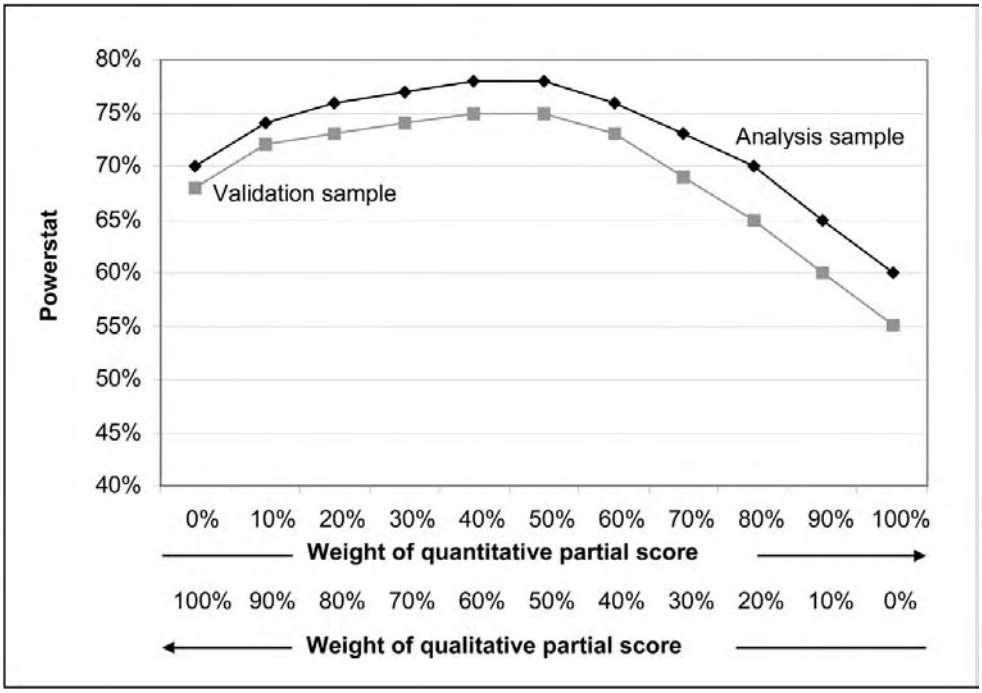

Chart 38: Example of Weighting Optimization for Partial Scoring Functions

5.3 Calibrating the Rating Model

The objective of calibration is to assign a *default probability* to each possible overall score, which may be a grade or other score value. The default probabilities themselves can be classified into as many as about 20 *rating classes*, mainly in order to facilitate reporting. Assigning default probabilities to rating results is crucial in order to meet the minimum requirements of the IRB approach under Basel II and the proposed EU directive.[61]

In order to fulfill these minimum requirements, the *rating scale* used has to include at least seven rating classes (i.e. grades) for non-defaulted borrowers and one class for defaulted borrowers, except in the retail segment.[62]

In practice, banks frequently use what is referred to as a *master scale*, that is, a uniform rating scale which is used throughout the bank and into which all rating results for segment-specific rating procedures are mapped. The advantage of this approach is the resulting comparability of rating results across all rating segments.

As each segment is characterized by specific features, especially with regard to average default probability rates, separate segment-specific calibration is necessary for each rating model. In the calibration process, the bandwidth of rating results (e.g. score value ranges) to be assigned to each rating class on the master scale is determined (see chart 39). With regard to model types, the following factors are to be differentiated in calibration:

– *Logistic regression* (see section 3.2.2) already yields rating results in the form of sample-dependent default probabilities, which may have to be rescaled to each segment's average default probability (see section 5.3.1).

[61] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-1, No. 1.*

[62] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 8.*

– For all other *statistical* and *heuristic rating models*, it is necessary to assign default probabilities in the calibration process. In such cases, it may also be necessary to rescale results in order to offset sample effects (see section 5.3.2).
– The *option pricing model* already yields sample-independent default probabilities.

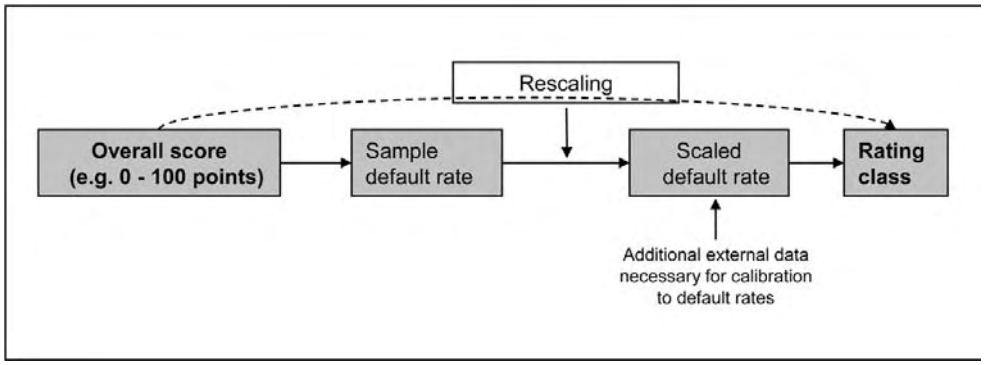

Chart 39: Calibration Scheme

5.3.1 Calibration for Logistic Regression

The results output by logistic regression are already in the form of default probabilities. The average of these default probabilities for all cases in the sample corresponds to the proportion of bad cases included a priori in the analysis sample. If logistic regression is only one of several modules in the overall rating model (e.g. in hybrid systems) and the rating result cannot be interpreted directly as a default probability, the procedure described under 5.3.2 is to be applied.

Rescaling default probabilities is therefore necessary whenever the proportion of good and bad cases in the sample does not match the actual composition of the portfolio in which the rating model is meant to be used. This is generally the case when the bank chooses not to conduct a full data survey. The average default probability in the sample is usually substantially higher than the portfolio's average default probability. This is especially true in cases where predominantly bad cases are collected for rating system development.

In such cases, the sample default probabilities determined by logistic regression have to be scaled to the average market or portfolio default probability. The scaling process is performed in such a way that the segment's "correct" average default probability is attained using a sample which is representative of the segment (see chart 39). For example, it is possible to use all good cases from the data collected as a representative sample, as these represent the bank's actual portfolio to be captured by the rating model.

In order to perform calibration, it is necessary to know the segment's average default rate. This rate can be estimated using credit reporting information, for example. In this process, it is necessary to ensure that external sources are in a position to delineate each segment with sufficient precision and in line with the bank's in-house definitions.

In addition, it is necessary to pay attention to the default criterion used by the external source. If this criterion does not match the one used in the process of developing the rating model, it will be necessary to adjust estimates of the segment's average default rate. If, for example, the external information source deviates from Basel II guidelines and uses the declaration of bankruptcy as the default criterion, and if the "loan loss provision" criterion is used in developing the model, the segment's estimated average default probability according to the external source will have to be adjusted upward. This is due to the fact that not every loan loss provision leads to bankruptcy and therefore more loan loss provision defaults than bankruptcy defaults occur.

Sample default rates are not scaled directly by comparing the default probabilities in the sample and the portfolio, but indirectly using relative default frequencies (RDFs), which represent the ratio of bad cases to good cases in the sample.

RDF is directly proportional to the general probability of default (PD):

$$RDF = \frac{PD}{1 - PD} \quad \text{or} \quad PD = \frac{RDF}{1 + RDF}$$

The process of rescaling the results of logistic regression involves six steps:

1. Calculation of the average default rate resulting from logistic regression using a sample which is representative of the non-defaulted portfolio
2. Conversion of this average sample default rate into RDF_{sample}
3. Calculation of the average portfolio default rate and conversion into $RDF_{portfolio}$
4. Representation of each default probability resulting from logistic regression as $RDF_{unscaled}$
5. Multiplication of $RDF_{unscaled}$ by the scaling factor specific to the rating model
$$RDF_{scaled} = RDF_{unscaled} \cdot \frac{RDF_{portfolio}}{RDF_{sample}}$$
6. Conversion of the resulting scaled RDF into a scaled default probability.

This makes it possible to calculate a scaled default probability for each possible value resulting from logistic regression. Once these default probabilities have been assigned to grades in the rating scale, the calibration is complete.

5.3.2 Calibration in Standard Cases

If the results generated by the rating model are not already sample-dependent default probabilities but (for example) score values, it is first necessary to assign default probabilities to the rating results. One possible way of doing so is outlined below. This approach includes rescaling as discussed in 5.3.1).

7. The rating model's value range is divided into several intervals according to the granularity of the value scale and the quantity of data available. The intervals should be defined in such a way that the differences between the corresponding average default probabilities are sufficiently large, and at the same time the corresponding classes contain a sufficiently large number of cases (both good and bad). As a rule, at least 100 cases per interval are necessary to enable a fairly reliable estimate of the default rate. A minimum

of approximately 10 intervals should be defined.[63] The interval widths do not necessarily have to be identical.

8. An $RDF_{unscaled}$ is calculated for each interval. This corresponds to the ratio of bad cases to good cases in each score value interval of the overall sample used in rating development.
9. Multiplication of $RDF_{unscaled}$ by the rating model's specific scaling factor, which is calculated as described in section 5.3.1:

$$RDF_{scaled} = RDF_{unscaled} \cdot \frac{RDF_{portfolio}}{RDF_{sample}}$$

10. Conversion of RDFscaled into scaled probabilities of default (PD) for each interval.

This procedure assigns rating model's score value intervals to scaled default probabilities. In the next step, it is necessary to apply this assignment to all of the possible score values the rating model can generate, which is done by means of interpolation.

If rescaling is not necessary, which means that the sample already reflects the correct average default probability, the default probabilities for each interval are calculated directly in step 2 and then used as input parameters for interpolation; steps 3 and 4 can thus be omitted.

For the purpose of interpolation, the scaled default probabilities are plotted against the average score values for the intervals defined. As each individual score value (and not just the interval averages) is to be assigned a probability of default, it is necessary to smooth and interpolate the scaled default probabilities by adjusting them to an approximation function (e.g. an exponential function).

Reversing the order of the rescaling and interpolation steps would lead to a miscalibration of the rating model. Therefore, if rescaling is necessary, it should always be carried out first.

Finally, the score value bandwidths for the individual rating classes are defined by inverting the interpolation function. The rating model's score values to be assigned to individual classes are determined on the basis of the defined PD limits on the master scale.

As "only" the data from the collection stage can be used to calibrate the overall scoring function and the estimation of the segments' average default probabilities frequently involves a certain level of uncertainty, it is essential to validate the calibration regularly using a data sample gained from ongoing operation of the rating model in order to ensure the functionality of a rating procedure (cf. section 6.2.2). Validating the calibration in quantitative terms is therefore one of the main elements of rating model validation, which is discussed in detail in chapter 6.

[63] *In the case of databases which do not fulfill these requirements, the results of calibration are to be regarded as statistically uncertain. Validation (as described in section 6.2.2) should therefore be carried out as soon as possible.*

5.4 Transition Matrices

The rating result generated for a specific customer[64] can change over time. This is due to the fact that a customer has to be re-rated regularly both before and after the conclusion of a credit agreement due to regulatory requirements and the need to ensure the regular and current monitoring of credit risk from a business perspective. In line with best business practices, the requirements arising from Basel II call for ratings to be renewed regularly (at least on an annual basis); this is to be carried at even shorter intervals in the case of noticeably higher risk.[65] This information can be used to improve risk classification and to validate rating models.

In addition to the exact assignment of default probabilities to the individual rating classes (a process which is first performed only for a defined time horizon of 12 months), it is also possible to determine how the rating will change in the future for longer-term credit facilities. The transition matrices specific to each rating model indicate the probability of transition for current ratings (listed in columns) to the various rating classes (listed in rows) during a specified time period. In practice, time periods of one or more years are generally used for this purpose.

This section only presents the methodical fundamentals involved in determining transition matrices. Their application, for example in risk-based pricing, is not covered in this document. For information on back-testing transition matrices, please refer to section 6.2.3.

5.4.1 The One-Year Transition Matrix

In order to calculate the transition matrix for a time horizon of one year, it is necessary to identify the rating results for all customers rated in the existing data set and to list these results over a 12-month period. Using this data, all observed changes between rating classes are counted and compiled in a table. Chart 40 gives an example of such a matrix.

Rating classes after 12 months

Original rating classes	1a	1b	1c	1d	1e	2a	2b	2c	2d	2e	3a	3b	3c	3d	3e	Default	Total
1a	1															X	1
1b		18	18	12	6	3	1	0	0	0	0	0	0	0	0	X	58
1c		37	185	84	32	26	10	5	2	1	0	1	0	0	0	X	383
1d		19	125	202	96	60	22	7	4	9	4	4	1	0	1	X	554
1e		7	48	93	92	93	35	12	15	4	5	2	0	1	1	X	408
2a		3	30	56	79	151	57	27	19	12	9	3	3	1	2	X	452
2b		4	12	16	23	84	103	40	20	9	8	7	5	2	8	X	341
2c		0	9	8	12	19	45	66	40	17	14	10	12	7	8	X	267
2d		1	11	20	13	16	24	47	60	45	26	17	7	4	12	X	303
2e		1	15	13	8	20	11	22	62	54	50	23	20	9	18	X	326
3a		2	8	22	8	22	19	15	29	42	71	55	18	14	20	X	345
3b		1	5	4	10	21	6	20	15	19	28	52	30	26	35	X	272
3c		0	0	1	2	7	6	4	8	8	18	29	52	26	39	X	200
3d		0	1	0	2	7	5	2	5	4	14	15	22	39	52	X	168
3e		0	0	0	1	2	5	5	4	7	16	19	29	51	254	X	393

Chart 40: Matrix of Absolute Transition Frequencies (Example)

[64] *The explanations below also apply analogously to transaction-specific ratings.*

[65] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, Nos. 27 and 29.*

With regard to the time interval between consecutive customer ratings, it is necessary define a margin of tolerance for the actual time interval between rating results for, as the actual intervals will only rarely be exactly one year. In this context, it is necessary to ensure that the average time interval for the rating pairs determined matches the time horizon for which the transition matrix is defined. At the same time, the range of time intervals around this average should not be so large that a valid transition matrix cannot be calculated. The range of time intervals considered valid for calculating a transition matrix should also be consistent with the bank's in-house guidelines for assessing whether customer re-ratings are up to date and performed regularly.

Actual credit defaults are frequently listed as a separate class (i.e. in their own column). This makes sense insofar as a default describes the transition of a rated borrower to the "defaulted loans" class.

Frequently cases will accumulate along the main diagonal of the matrix. These cases represent borrowers which did not migrate from their original rating class over the time horizon observed. The other borrowers form a band around the main diagonal, which becomes less dense with increasing distance from the diagonal. This concentration around the main diagonal correlates with the number of existing rating classes as well as the stability of the rating procedure. The more rating classes a model uses, the more frequently rating classes will change and the lower the concentration along the main diagonal will be. The same applies in the case of decreasing stability in the rating procedure.

In order to calculate transition probabilities, it is necessary to convert the absolute numbers into percentages (row probabilities). The resulting probabilities indicate the fraction of cases in a given class which actually remained in their original class. The transition probabilities of each row – including the default probability of each class in the last column – should add up to 100%.

Rating classes after 12 months

Original rating classes	1a	1b	1c	1d	1e	2a	2b	2c	2d	2e	3a	3b	3c	3d	3e	Default	Total
1a	100%															0.07%	100.00%
1b		31%	31%	21%	10%	5%	2%									0.10%	100.00%
1c		10%	48%	22%	8%	7%	3%	1%	1%	0%		0%				0.15%	100.00%
1d		3%	23%	36%	17%	11%	4%	1%	1%	2%	1%	1%	0%		0%	0.23%	100.00%
1e		2%	12%	23%	22%	23%	9%	3%	4%	1%	1%	0%		0%	0%	0.35%	100.00%
2a		1%	7%	12%	17%	33%	13%	6%	4%	3%	2%	1%	1%	0%	0%	0.50%	100.00%
2b		1%	3%	5%	7%	24%	30%	12%	6%	3%	2%	2%	1%	1%	2%	0.75%	100.00%
2c			3%	3%	4%	7%	17%	24%	15%	6%	5%	4%	4%	3%	3%	1.10%	100.00%
2d		0%	4%	6%	4%	5%	8%	15%	19%	15%	8%	6%	2%	1%	4%	1.70%	100.00%
2e		0%	4%	4%	2%	6%	3%	7%	19%	16%	15%	7%	6%	3%	5%	2.60%	100.00%
3a		1%	2%	6%	2%	6%	5%	4%	8%	12%	20%	15%	5%	4%	6%	4.00%	100.00%
3b		0%	2%	1%	3%	7%	2%	7%	5%	7%	10%	18%	10%	9%	12%	6.00%	100.00%
3c				0%	1%	3%	3%	2%	4%	4%	8%	13%	24%	12%	18%	8.00%	100.00%
3d			1%		1%	4%	3%	1%	3%	2%	7%	8%	11%	20%	27%	13.50%	100.00%
3e					0%	0%	1%	1%	1%	1%	3%	4%	6%	10%	52%	20.00%	100.00%
Default																100.0%	100.00%

Chart 41: Empirical One-Year Transition Matrix

Especially with a small number of observations per matrix field, the empirical transition matrix derived in this manner will show inconsistencies. Inconsistencies refer to situations where large steps in ratings are more probable than

smaller steps in the same direction for a given rating class, or where the probability of ending up in a certain rating class is more probable for more remote rating classes than for adjacent classes. In the transition matrix, inconsistencies manifest themselves as probabilities which do not decrease monotonically as they move away from the main diagonal of the matrix. Under the assumption that a valid rating model is used, this is not plausible.

Inconsistencies can be removed by smoothing the transition matrix. Smoothing refers to optimizing the probabilities of individual cells without violating the constraint that the probabilities in a row must add up to 100%. As a rule, smoothing should only affect cell values at the edges of the transition matrix, which are not statistically significant due to their low absolute transition frequencies. In the process of smoothing the matrix, it is necessary to ensure that the resulting default probabilities in the individual classes match the default probabilities from the calibration.

Chart 42 shows the smoothed matrix for the example given above. In this case, it is worth noting that the default probabilities are sometimes higher than the probabilities of transition to lower rating classes. These apparent inconsistencies can be explained by the fact that in individual cases the default event occurs earlier than rating deterioration. In fact, it is entirely conceivable that a customer with a very good current rating will default, because ratings only describe the average behavior of a group of similar customers over a fairly long time horizon, not each individual case.

Rating classes after 12 months

Original rating classes	1a	1b	1c	1d	1e	2a	2b	2c	2d	2e	3a	3b	3c	3d	3e	Default	Total
1a	71%	21%	7%													0.07%	100.00%
1b	24%	24%	24%	16%	8%	4%	1%									0.10%	100.00%
1c	6%	9%	46%	21%	8%	6%	2%	1%								0.15%	100.00%
1d		3%	23%	31%	18%	11%	4%	1%	1%	1%	1%					0.23%	100.00%
1e		2%	12%	23%	23%	23%	9%	3%	4%	1%	1%					0.35%	100.00%
2a		1%	7%	12%	18%	33%	13%	6%	4%	3%	2%	1%	1%			0.50%	100.00%
2b		1%	4%	5%	7%	25%	31%	12%	6%	3%	2%	2%	1%	1%	1%	0.75%	100.00%
2c			3%	3%	5%	8%	18%	26%	16%	7%	6%	4%	3%	1%	1%	1.10%	100.00%
2d			3%	3%	4%	6%	8%	17%	21%	16%	9%	6%	2%	1%	1%	1.70%	100.00%
2e			3%	3%	3%	4%	4%	7%	20%	18%	17%	8%	7%	3%	3%	2.60%	100.00%
3a			3%	3%	3%	3%	3%	5%	9%	13%	22%	17%	6%	4%	4%	4.00%	100.00%
3b			2%	2%	2%	2%	2%	6%	6%	8%	11%	21%	12%	10%	10%	6.00%	100.00%
3c				1%	1%	2%	2%	2%	4%	4%	9%	15%	26%	13%	13%	9.00%	100.00%
3d					1%	1%	1%	1%	2%	2%	8%	9%	13%	23%	23%	13.50%	100.00%
3e							0%	0%	1%	1%	3%	3%	5%	11%	55%	20.00%	100.00%
Default																100.0%	100.00%

Chart 42: Smoothed Transition Matrix (Example)

Due to the large number of parameters to be determined, the data requirements for calculating a valid transition matrix are very high. For a rating model with 15 classes plus one default class (as in the example above), it is necessary to compute a total of 225 transition probabilities plus 15 default probabilities. As statistically valid estimates of transition frequencies are only possible given a sufficient number of observations per matrix field, these requirements amount to several thousand observed rating transitions – assuming an even distribution of transitions across all matrix fields. Due to the generally observed concentration

around the main diagonal, however, the actual requirement for valid estimates are substantially higher at the edges of the matrix.

5.4.2 Multi-Year Transition Matrices

If a sufficiently large database is available, multi-year transition matrices can be calculated in a manner analogous to the procedure described above using the corresponding rating pairs and a longer time interval.

If it is not possible to calculate multi-year transition matrices empirically, they can also be determined on the basis of the one-year transition matrix. One procedure which is common in practice is to assume stationarity or the Markov property in the one-year transition matrix, and to calculate the n-year transition matrix by raising the one-year matrix to the nth power. In this way, for example, the two-year transition matrix is calculated by multiplying the one-year matrix by itself.

In order to obtain useful results with this procedure, it is important to note that the stationarity assumption is not necessarily fulfilled for a transition matrix. In particular, economic fluctuations have a strong influence on the tendency of rating results to deteriorate or improve, meaning that the transition matrix is not stable over longer periods of time. Empirically calculated multi-year transition matrices are therefore preferable to calculated transition matrices. In particular, the multi-year (cumulative) default rates in the last column of the multi-year transition matrix can often be calculated directly in the process of calibrating and back-testing rating models.

In the last column of the n-year matrix (default), we see the cumulative default rate (cumDR). For each rating class, this default rate indicates the probability of transition to the default class within n years. Chart 43 shows the cumulative default rates of the rating classes in the example used here.

The cumulative default rates should exhibit the following two properties:

1. In each rating class, the cumulative default rates increase along with the length of the term and approach 100% over infinitely long terms.
2. If the cumulative default rates are plotted over various time horizons, the curves of the individual rating classes do not intersect, that is, the cumulative default rate in a good rating class will be lower than in the inferior classes for all time horizons.

The first property of cumulative default probabilities can be verified easily:

- Over a 12-month period, we assume that the rating class-dependent probability of observing a default in a randomly selected loan equals 1%, for example.
- If this loan is observed over a period twice as long, the probability of default would have to be greater than the default probability for the first 12 months (1%), as the probability of default for the second year cannot become zero even if the case migrates to a different rating class at the end of the first year.
- Accordingly, the cumulative default probabilities for 3, 4, 5, and more years form a strictly monotonic ascending sequence. This sequence has an upper limit (maximum probability of default for each individual case = 100 %) and is therefore convergent.

– The limit of this sequence is 100%, as over an infinitely long term every loan will default due to the fact that the default probability of each rating class cannot equal zero.

The property of non-intersection in the cumulative default probabilities of the individual rating classes results from the requirement that the rating model should be able to yield adequate creditworthiness forecasts not only for a short time period but also over longer periods. As the cumulative default probability correlates with the total risk of a loan over its multi-year term, consistently lower cumulative default probabilities indicate that (ceteris paribus) the total risk of a loan in the rating class observed is lower than that of a loan in an inferior rating class.

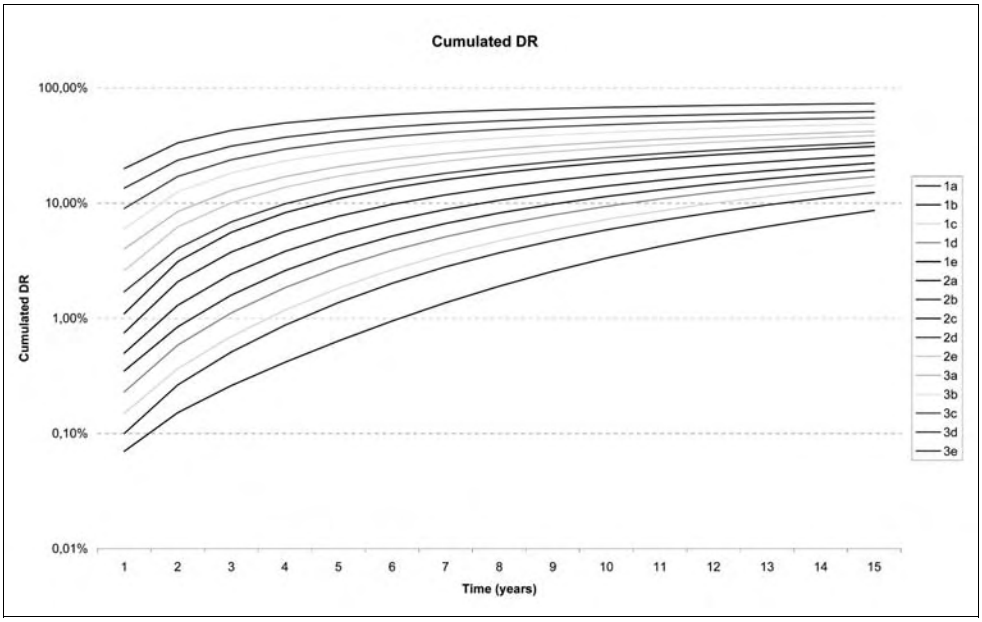

Chart 43: Cumulative Default Rates (Example: Default Rates on a Logarithmic Scale)

The cumulative default rates are used to calculate the marginal default rates ($marg_{DR}$) for each rating class. These rates indicate the change in cumulative default rates from year to year, that is, the following applies:

$$marg_{DR,n} = \begin{cases} cum_{DR,n} & \text{for } n = 1 \text{ year} \\ cum_{DR,n} - cum_{DR,n-1} & \text{for } n = 2, 3, 4 \ldots \text{years} \end{cases}$$

Due to the fact that cumulative default rates ascend monotonically over time, the marginal default rates are always positive. However, the curves of the marginal default rates for the very good rating classes will increase monotonically, whereas monotonically decreasing curves can be observed for the very low rating classes (cf. chart 44). This is due to the fact that the good rating classes show a substantially larger potential for deterioration compared to the very bad rating classes. The rating of a loan in the best rating class cannot improve, but it can indeed deteriorate; a loan in the second-best rating class can only improve by one class, but it can deteriorate by more than one class, etc. The situation is analogous for rating classes at the lower end of the scale. For this reason, even in a symmetrical transition matrix we can observe an ini-

tial increase in marginal default rates in the very good rating classes and an initial decrease in marginal default probabilities in the very poor rating classes. From the business perspective, we know that low-rated borrowers who "survive" several years pose less risk than loans which are assigned the same rating at the beginning but default in the meantime. For this reason, in practice we can also observe a tendency toward lower growth in cumulative default probabilities in the lower rating classes.

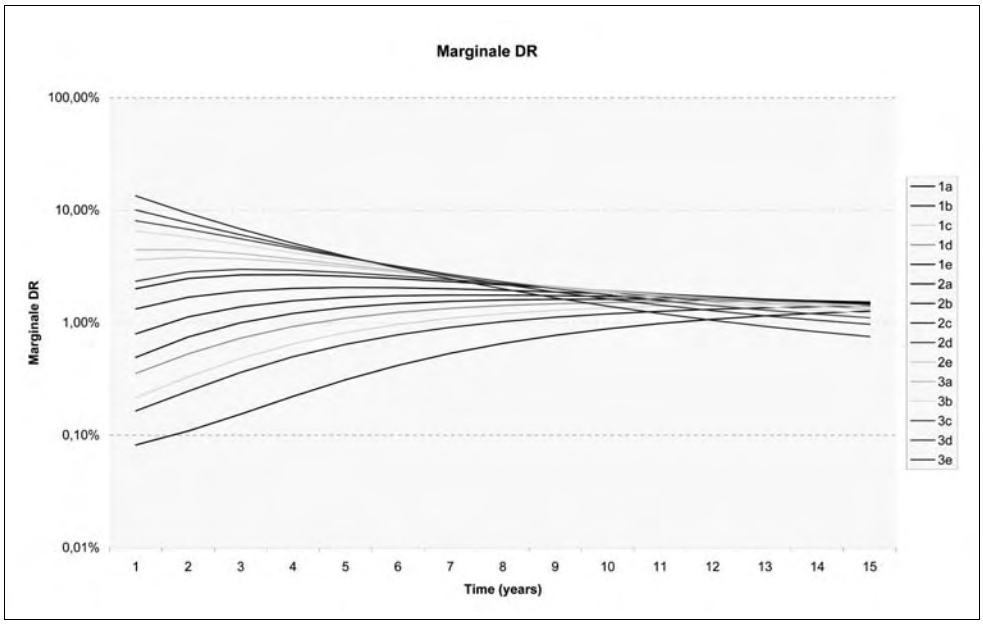

Chart 44: Marginal Default Rates (Example: Default Rates on a Logarithmic Scale)

Conditional default rates ($cond_{DR}$) indicate the probability that a borrower will default in the nth year assuming that the borrower has survived the first (n-1) years. These conditional default rates can be calculated using cumulative and marginal default rates as follows:

$$cond_{DR,n} = \begin{cases} marg_{DR,n} & \text{for } n = 1 \text{ year} \\ \dfrac{marg_{DR,n}}{1 - cum_{DR,n-1}} & \text{for } n = 2, 3, 4 \ldots \text{years} \end{cases}$$

Chart 45 shows the curve of conditional default rates for each rating class. In this context, it is necessary to ensure that the curves for the lower rating classes remain above those of the good rating classes and that none of the curves intersect. This reflects the requirement that a rating model should be able to discriminate creditworthiness and classify borrowers correctly over several years. Therefore, the conditional probabilities also have to show higher values in later years for a borrower who is initially rated lower than for a borrower who is initially rated higher. In this context, conditional default probabilities account for the fact that defaults cannot have happened in previous years when borrowers are compared in later years.

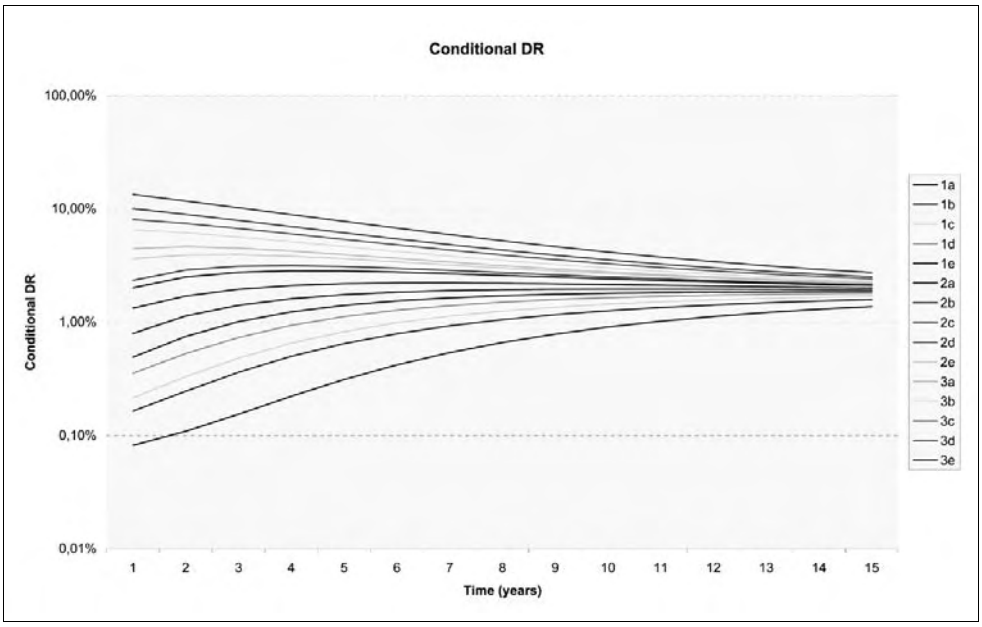

Chart 45: Conditional Default Rates (Example: Default Rates on a Logarithmic Scale)

When the Markov property is applied to the transition matrix, the conditional default rates converge toward the portfolio's average default probability for all rating classes as the time horizon becomes longer. In this process, the portfolio attains a state of balance in which the frequency distribution of the individual rating classes no longer shifts noticeably due to transitions. In practice, however, such a stable portfolio state can only be observed in cases where the rating class distribution remains constant in new business and general circumstances remain unchanged over several years (i.e. seldom).

6 Validating Rating Models

The term "validation" is defined in the minimum requirements of the IRB approach as follows:

The institution shall have a regular cycle of model validation that includes monitoring of model performance and stability; review of model relationships; and testing of model outputs against outcomes.[66]

Chart 46 below gives an overview of the essential aspects of validation.

The area of *quantitative validation* comprises all validation procedures in which statistical indicators for the rating procedure are calculated and interpreted on the basis of an empirical data set. Suitable indicators include the model's α and β errors, the differences between the forecast and realized default rates of a rating class, or the Gini coefficient and AUC as measures of discriminatory power.

In contrast, the area of *qualitative validation* fulfills the primary task of ensuring the applicability and proper application of the quantitative methods in practice. Without a careful review of these aspects, the rating's intended purpose cannot be achieved (or may even be reversed) by unsuitable rating procedures due to excessive faith in the model.[67]

[66] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 18.*

[67] *Cf. EUROPEAN COMMISSION, Annex D-5, No. 41 ff.*

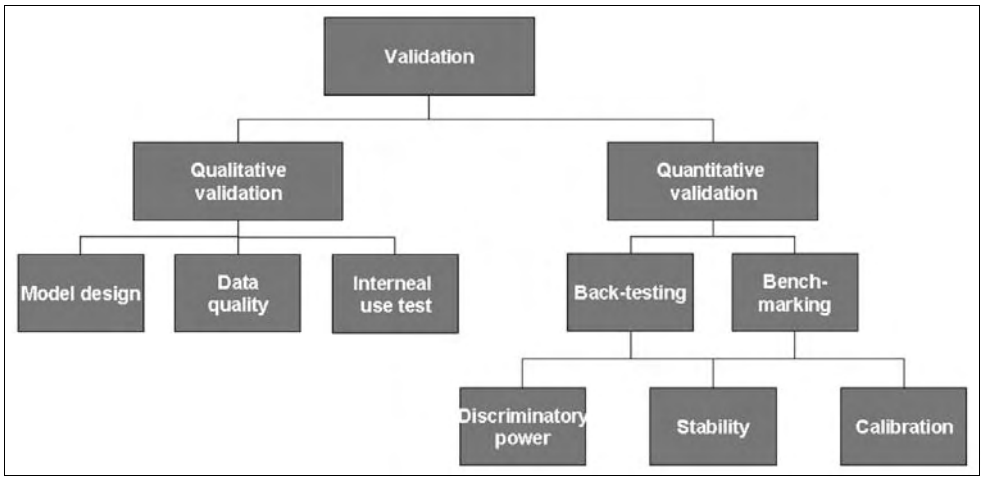

Chart 46: Aspects of Rating Model Validation[68]

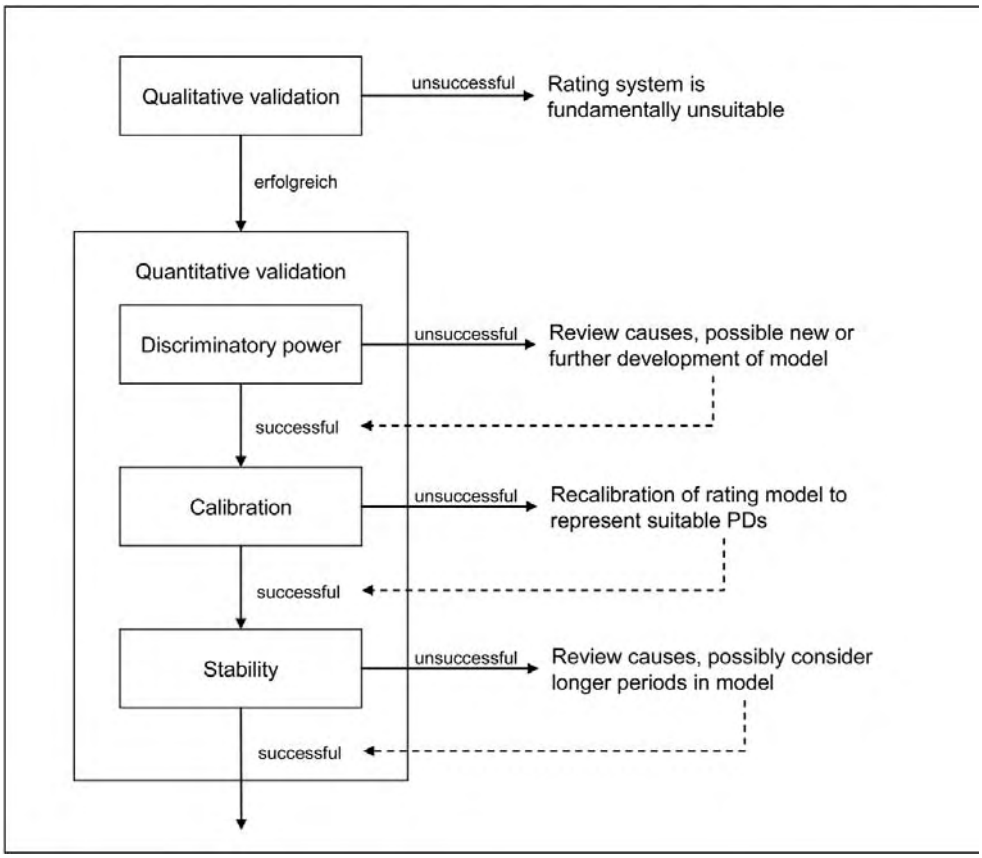

Chart 47: Validation Procedure for Rating Models

These two aspects of validation complement each other. A rating procedure should only be applied in practice if it receives a positive assessment in the qualitative area. A positive assessment only in quantitative validation is not sufficient. This also applies to rating procedures used within an IRB approach.

[68] *Adapted from DEUTSCHE BUNDESBANK, Monthly Report for September 2003, Approaches to the validation of internal rating systems.*

Conversely, a negative quantitative assessment should not be considered decisive to the general rejection of a rating procedure. This is especially true because the statistical estimates themselves are subject to random fluctuations, and the definition of a suitable tolerance range allows a certain degree of freedom in the interpretation of analysis results. It is therefore necessary to place greater emphasis on qualitative validation.

With regard to validation, Basel II imposes the following additional requirements on banks using the IRB approach:

The validation process must be described in the rating model's *documentation*. This is an explicit[69] requirement for statistical models, for which validation is already an essential factor in model development. In this family of models, validation must also include out-of-sample and out-of-time performance tests which review the behavior of the model's results using unknown data (i.e. data not used in developing the model).

The credit risk control unit should be responsible for *carrying out the validation process*; in this context, it is especially important to separate validation activities from the front office.[70] However, the organizational aspects of validation will not be discussed in greater detail here.

Validation methodologies must not be influenced by changes in general economic conditions. However, the interpretation of deviations identified in the validation process between model predictions and reality should take external influences such as economic cycles into account.[71] This is discussed further in the presentation of *stress tests* (see section 6.4).

If significant deviations arise between the parameters estimated using the models and the values actually realized, the models have to be adapted.[72]

6.1 Qualitative Validation

The qualitative validation of rating models can be divided into three core areas:[73]

- Model design
- Data quality
- Internal use ("use test")

Model Design

The model's design is validated on the basis of the rating model's *documentation*. In this context, the scope, transparency and completeness of documentation are already essential validation criteria. The documentation of statistical models should at least cover the following areas:

- Delineation criteria for the rating segment
- Description of the rating method/model type/model architecture used
- Reason for selecting a specific model type
- Completeness of the (best practice) criteria used in the model
- Data set used in statistical rating development
- Quality assurance for the data set

[69] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 21.*

[70] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 39.*

[71] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 98.*

[72] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 98.*

[73] *DEUTSCHE BUNDESBANK, Monthly Report for September 2003, Approaches to the validation of internal rating systems.*

– Model development procedure
  Model architecture and fundamental business assumptions
  Selection and assessment of model parameters
  Analyses for model development
– Quality assurance/validation during model development
– Documentation of all model functions
– Calibration of model output to default probabilities
– Procedure for validation/regular review
– Description of the rating process
– Duties and responsibilities with regard to the rating model

For heuristic and causal models, it is possible to omit the description of the data set and parts of the analysis for model development. However, in these model families it is necessary to ensure the transparency of the assumptions and/or evaluations which form the basis of the rating model's design.

The *rating method* should be selected with attention to the portfolio segment to be analyzed and the data available. The various model types and their general suitability for individual rating segments are described in chapter 4.

The influence of individual *factors* on the rating result should be comprehensible and in line with the current state of business research and practice. For example, it is necessary to ensure that the factors in a statistical balance sheet analysis system are plausible and comprehensible according to the fundamentals of financial statement analysis.

In statistical models, special emphasis is to be placed on documenting the model's *statistical foundations*, which have to be in line with the standards of quantitative validation.

Data Quality

In statistical models, data quality stands out as a goodness-of-fit criterion even during model development. Moreover, a comprehensive *data set* is an essential prerequisite for quantitative validation. In this context, a number of aspects have to be considered:

– Completeness of data in order to ensure that the rating determined is comprehensible
– Volume of available data, especially data histories
– Representativity of the samples used for model development and validation
– Data sources
– Measures taken to ensure quality and cleanse raw data.

The minimum data requirements under the draft EU directive only provide a basis for the validation of data quality.[74] Beyond that, the best practices described for generating data in rating model development (section 5.1) can be used as guidelines for validation.

Internal Use ("Use Test")

Validating the internal use of the rating models ("use test") refers to the actual integration of rating procedures and results into the bank's in-house risk management and reporting systems. With regard to internal use, the essential

[74] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 18, 21, 31–33.*

aspects of the requirements imposed on banks using the IRB approach under Basel II include:[75]

- Design of the bank's internal processes which interface with the rating procedure as well as their inclusion in organizational guidelines
- Use of the rating in risk management (in credit decision-making, risk-based pricing, rating-based competence systems, rating-based limit systems, etc.)
- Conformity of the rating procedures with the bank's credit risk strategy
- Functional separation of responsibility for ratings from the front office (except in retail business)
- Employee qualifications
- User acceptance of the procedure
- The user's ability to exercise freedom of interpretation in the rating procedure (for this purpose, it is necessary to define suitable procedures and process indicators such as the number of overrides)

Banks which intend to use an IRB approach will be required to document these criteria completely and in a verifiable way. Regardless of this requirement, however, complete and verifiable documentation of how rating models are used should form an integral component of in-house use tests for any bank using a rating system.

6.2 Quantitative Validation

In statistical models, quantitative validation represents a substantial part of model development (cf. section 5.2). For heuristic and causal models, on the other hand, an empirical data set is not yet available during rating development. Therefore, the quantitative validation step is omitted during model development in this family of models.

However, quantitative validation is required for all rating models. For this purpose, validation should primarily use the data gained during practical operation of the model. Comparison or benchmark data can also be included as a supplement. This is particularly advisable when the performance of multiple rating models is to be compared using a common sample.

The criteria to be reviewed in quantitative validation are as follows:[76]

- Discriminatory power
- Calibration
- Stability

A sufficient data set for quantitative validation is available once all loans have been rated for the first time (or re-rated) and observed over the forecasting horizon of the rating model; this is usually the case approximately two years after a new rating model is introduced.

6.2.1 Discriminatory Power

The term "discriminatory power" refers to the fundamental ability of a rating model to differentiate between good and bad cases.[77] The term is often used as a synonym for "classification accuracy." In this context, the categories good

[75] *DEUTSCHE BUNDESBANK, Monthly Report for September 2003, Approaches to the validation of internal rating systems.*

[76] *DEUTSCHE BUNDESBANK, Monthly Report for September 2003, Approaches to the validation of internal rating systems.*

[77] *Instead of being restricted to borrower ratings, the descriptions below also apply to rating exposures in pools. For this reason, the terms used are not differentiated.*

and bad refer to whether a credit default occurs (bad) or does not occur (good) over the forecasting horizon after the rating system has classified the case.

The forecasting horizon for PD estimates in IRB approaches is 12 months. This time horizon is a direct result of the minimum requirements in the draft EU directive.[78] However, the directive also explicitly requires institutions to use longer time horizons in rating assessments.[79] Therefore, it is also possible to use other forecasting horizons in order to optimize and calibrate a rating model as long as the required 12-month default probabilities are still calculated. In this section, we only discuss the procedure applied for a forecasting horizon of 12 months.

In practice, the discriminatory power of an application scoring function for installment loans, for example, is often optimized for the entire period of the credit transaction. However, forecasting horizons of less than 12 months only make sense where it is also possible to update rating data at sufficiently short intervals, which is the case in account data analysis systems, for example.

The discriminatory power of a model can only be reviewed ex post using data on defaulted and non-defaulted cases. In order to generate a suitable data set, it is first necessary to create a sample of cases for which the initial rating as well as the status (good/bad) 12 months after assignment of the rating are known.

In order to *generate the data set* for quantitative validation, we first define two cutoff dates with an interval of 12 months. End-of-year data are often used for this purpose. The cutoff dates determine the rating's application period to be used in validation. It is also possible to include data from several previous years in validation. This is especially necessary in cases where average default rates have to be estimated over several years.

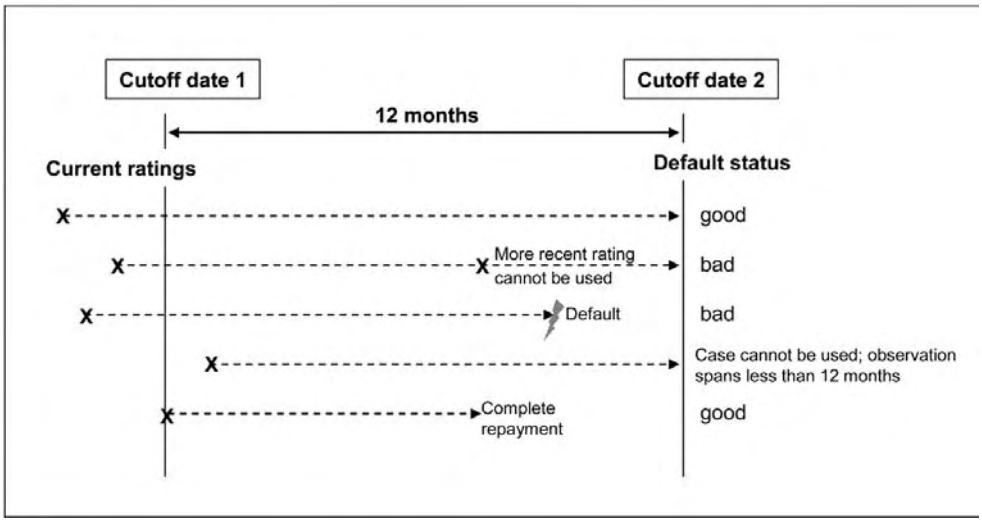

Chart 48: Creating a Rating Validation Sample for a Forecasting Horizon of 12 Months

[78] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-3, Nos. 1 and 16.*

[79] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 15.*

The rating information available on all cases as of the earlier cutoff date (1; see chart 48) is used. The second step involves adding status information as of the later cutoff date (2) for all cases. In this process, all cases which were assigned to a default class at any point between the cutoff dates are classified as bad; all others are considered good. Cases which no longer appear in the sample as of cutoff date (2) but did not default are also classified as good. In these cases, the borrower generally repaid the loan properly and the account was deleted. Cases for which no rating information as of cutoff date (1) is available (e.g. new business) cannot be included in the sample as their status could not be observed over the entire forecasting horizon.

Rating class	Number of good cases	Number of bad cases	Total	Proportion per rating class	Default rate per rating class
1	50	0	50	0,5%	0,0%
2	1785	3	1788	17,5%	0,2%
3	1870	6	1876	18,4%	0,3%
4	3330	15	3345	32,7%	0,4%
5	1208	15	1223	12,0%	1,2%
6	839	17	856	8,4%	2,0%
7	330	12	342	3,3%	3,5%
8	205	9	214	2,1%	4,2%
9	243	22	265	2,6%	8,3%
10	224	33	257	2,5%	12,8%
All classes	**10084**	**132**	**10216**	**100%**	**1,3%**

Chart 49: Example of Rating Validation Data

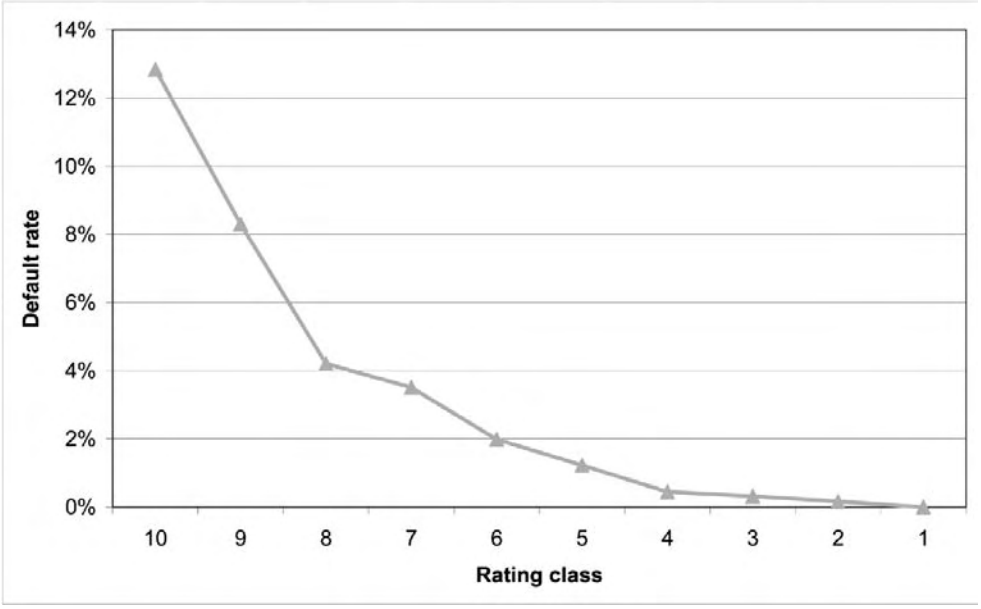

Chart 50: Curve of the Default Rate for each Rating Class in the Data Example

On the basis of the resulting sample, various analyses of the rating procedure's discriminatory power are possible. The example shown in chart 49 forms the basis of the explanations below. The example refers to a rating model with 10 classes. However, the procedures presented can also be applied to a far finer observation scale, even to individual score values. At the same time, it is necessary to note that statistical fluctuations predominate in the case of small

numbers of cases per class observed, thus it may not be possible to generate meaningful results.

In the example below, the default rate (i.e. the proportion of bad cases) for each rating class increases steadily from class 1 to class 10. Therefore, the underlying rating system is obviously able to classify cases by default probability. The sections below describe the methods and indicators used to quantify the discriminatory power of rating models and ultimately to enable statements such as "Rating system A discriminates better/worse/just as well as rating system B."

Frequency Distribution of Good and Bad Cases

The frequency density distributions and the cumulative frequencies of good and bad cases shown in the table and charts below serve as the point of departure for calculating discriminatory power. In this context, cumulative frequencies are calculated starting from the worst class, as is generally the case in practice.

Rating class	Density function for good cases	Density function for bad cases	Cumulative frequency (good)	Cumulative frequency (bad)	Cumulative frequency (all)
1	0,5%	0,0%	100,0%	100,0%	100,0%
2	17,7%	2,3%	99,5%	100,0%	99,5%
3	18,5%	4,5%	81,8%	97,7%	82,0%
4	33,0%	11,4%	63,3%	93,2%	63,6%
5	12,0%	11,4%	30,2%	81,8%	30,9%
6	8,3%	12,9%	18,3%	70,5%	18,9%
7	3,3%	9,1%	9,9%	57,6%	10,6%
8	2,0%	6,8%	6,7%	48,5%	7,2%
9	2,4%	16,7%	4,6%	41,7%	5,1%
10	2,2%	25,0%	2,2%	25,0%	2,5%
All classes	**100%**	**100%**	0	0	0

Chart 51: Density Functions and Cumulative Frequencies for the Data Example

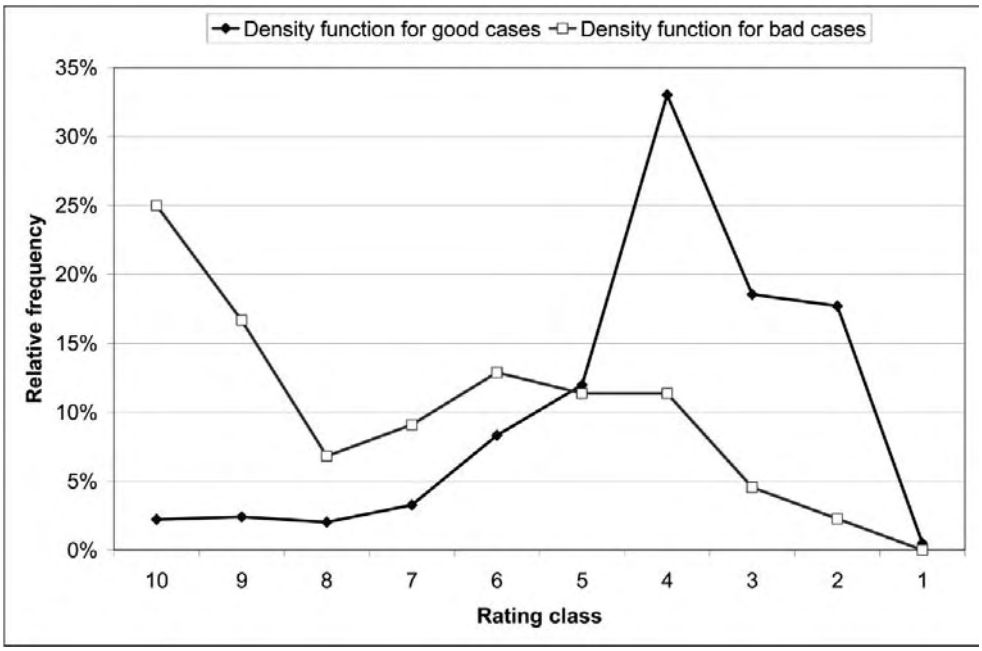

Chart 52: Curve of the Density Functions of Good/Bad Cases in the Data Example

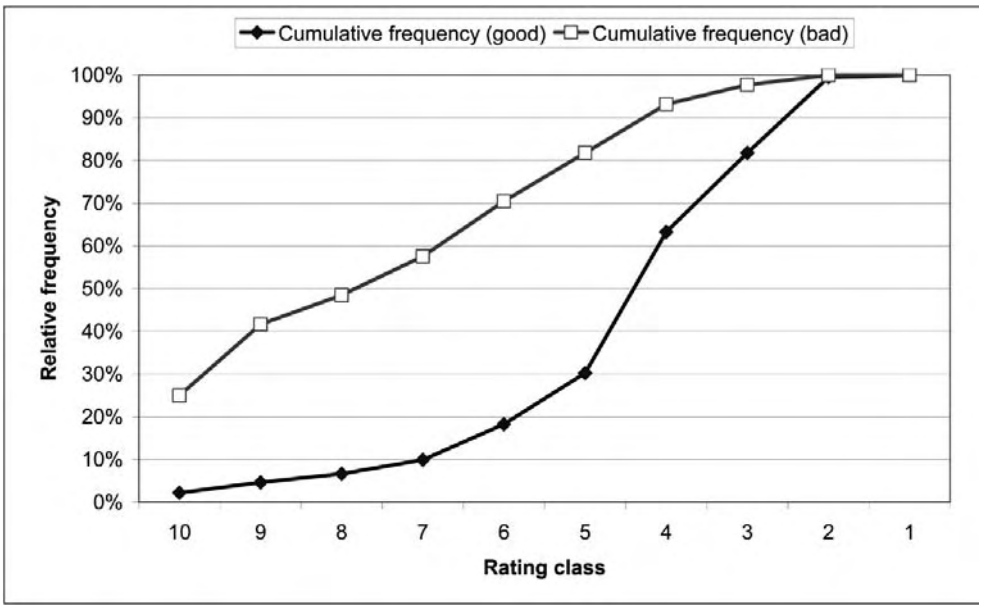

Chart 53: Curve of the Cumulative Probabilities Functions of Good/Bad Cases in the Data Example

The density functions show a substantial difference between good and bad cases. The cumulative frequencies show that approximately 70% of the bad cases – but only 20% of the good cases – belong to classes 6 to 10. On the other hand, 20% of the good cases – but only 2.3% of the bad cases – can be found in classes 1 to 3. Here it is clear that the cumulative probability of bad cases is greater than that of the good cases for almost all rating classes when the classes are arranged in order from bad to good. If the rating classes are arranged from good to bad, we can simply invert this statement accordingly.

α and β errors

α and β errors can be explained on the basis of our presentation of density functions for good and bad cases. In this context, a case's rating class is used as the decision criterion for credit approval. If the rating class is lower than a predefined cutoff value, the credit application is rejected; if the rating class is higher than that value, the credit application is approved. In this context, two types of error can arise:

– α error (type 1 error): A case which is actually bad is not rejected.
– β error (type 2 error): A case which is actually good is rejected.

In practice, α errors cause damage due to credit defaults, while β errors cause comparatively less damage in the form of lost business. When the rating class is used as the criterion for the credit decision, therefore, it is important to define the cutoff value with due attention to the costs of each type of error.

Usually, α and β errors are not indicated as absolute numbers but as percentages. α error refers to the proportion of good cases below the cutoff value, that is, the cumulative frequency F_{cum}^{good} of good cases starting from the worst rating class. α error, on the other hand, corresponds to the proportion of bad cases above the cutoff value, that is, the complement of the cumulative frequency of bad cases F_{cum}^{bad}.

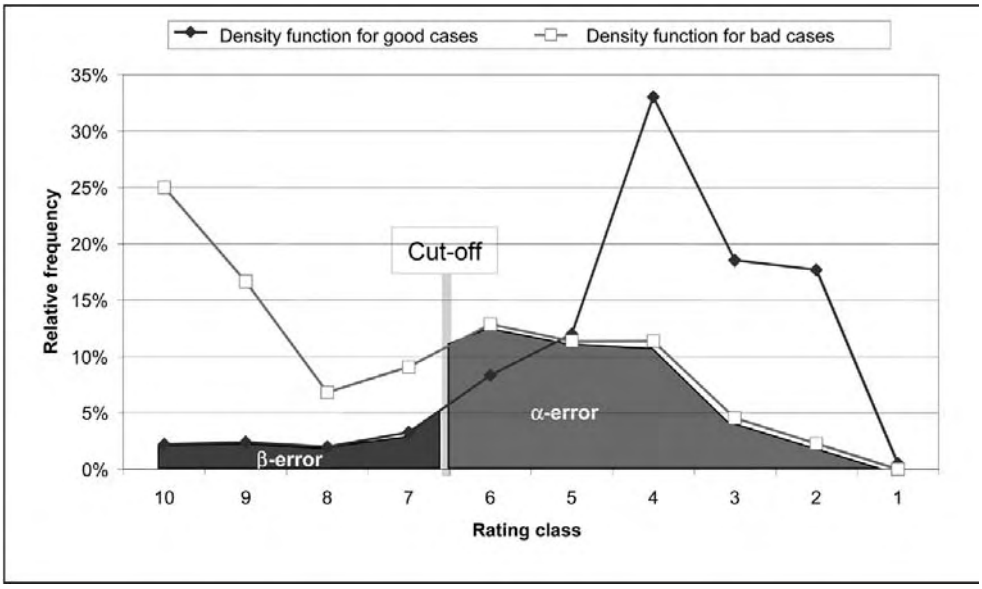

Chart 54: Depiction of α and β errors with Cutoff between Rating Classes 6 and 7

ROC Curve

One common way of depicting the discriminatory power of rating procedures is the ROC Curve,[80] which is constructed by plotting the cumulative frequencies of bad cases as points on the y axis and the cumulative frequencies of good cases along the x axis. Each section of the ROC curve corresponds to a rating class, beginning at the left with the worst class.

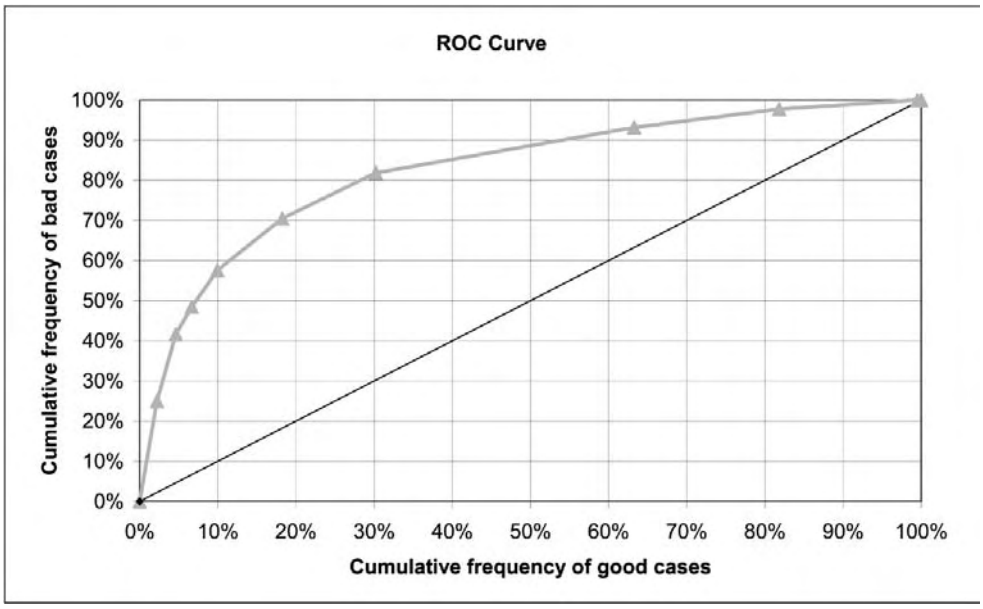

Chart 55: Shape of the ROC Curve for the Data Example

[80] *Receiver Operating Characteristic.*

An ideal rating procedure would classify all actual defaults in the worst rating class. Accordingly, the ROC curve of the ideal procedure would run vertically from the lower left point (0%, 0%) upwards to point (0%, 100%) and from there to the right to point (100%, 100%). The x and y values of the ROC curve are always equal if the frequency distributions of good and bad cases are identical. The ROC curve for a rating procedure which cannot distinguish between good and bad cases will run along the diagonal.

If the objective is to review rating classes (as in the given example), the ROC curve always consists of linear sections. The slope of the ROC curve in each section reflects the ratio of bad cases to good cases in the respective rating class. On this basis, we can conclude that the ROC curve for rating procedures should be concave (i.e. curved to the right) over the entire range. A violation of this condition will occur when the expected default probabilities do not differ sufficiently, meaning that (due to statistical fluctuations) an inferior class will show a lower default probability than a rating class which is actually superior. This may point to a problem with classification accuracy in the rating procedure and should be examined with regard to its significance and possible causes. In this case, one possible cause could be an excessively fine differentiation of rating classes.

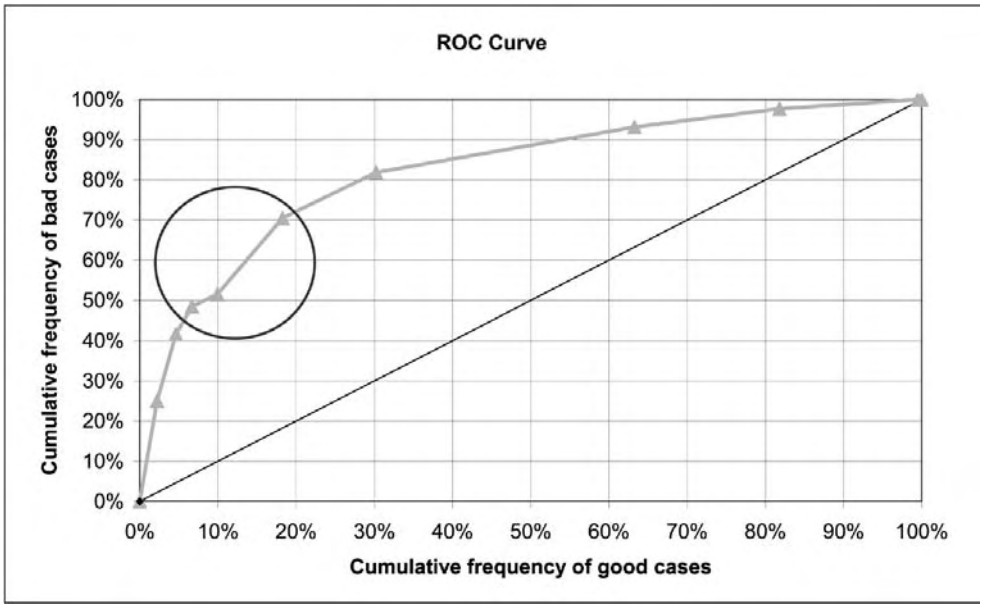

Chart 56: Non-Concave ROC Curve

α–β error curve

The depiction of the α–β error curve is equivalent to that of the ROC curve. This curve is generated by plotting the α error against the β error. The α–β error curve is equivalent to the ROC curve with the axes exchanged and then tilted around the horizontal axis. Due to this property, the discriminatory power measures derived from the α–β error curve are equivalent to those derived from the ROC curve; both representations contain exactly the same information on the rating procedure examined.

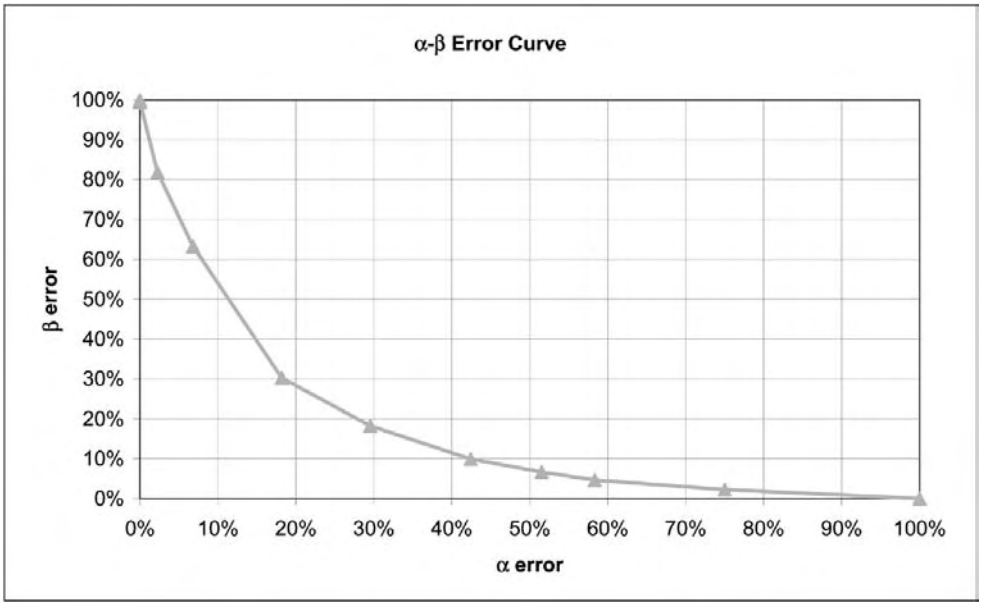

Chart 57: Shape of the α–β Error Curve for the Data Example

Area under Curve (AUC) as a Measure of Discriminatory Power

AUC (area under curve) is a graphic measure of a rating procedure's discriminatory power derived from the ROC curve and refers to the area under the ROC curve (expressed in units where 100% = 1 for both axes).[81] In ideal rating models, AUC = 1, for models which cannot differentiate between good and bad cases, AUC = 1/2. Values where AUC < 1/2 are possible but indicate that the rating system in question classifies the cases at least partly in the wrong order. The higher the AUC value is, the higher the discriminatory power of the rating model is.

However, AUC is a one-dimensional measure of discriminatory power. Essential information on the shape of the ROC curve and the properties of the rating model examined is lost in the calculation of this value. Therefore, when two different rating models using the same sample are compared on the basis of AUC alone, it is not immediately clear which of the procedures is better in terms of performance. This is especially true when the ROC curves intersect. Chart 58 shows an example of two ROC curves with the same AUC. In practice, the procedure which shows a steeper curve in the lower range of scores on the left (dark line/squares) would be preferable because fewer α errors occur at the same level of β error in this range, even though this ROC curve is not consistently concave in the good score ranges.

Besides its geometric interpretation as the area under the ROC curve, AUC can also be interpreted as the probability that a bad case randomly selected from the given sample will actually have a lower rating than a randomly drawn good case.

[81] *Cf. LEE, Global Performances of Diagnostic Tests und LEE/HSIAO, Alternative Summary Indices as well as the references cited in those works.*

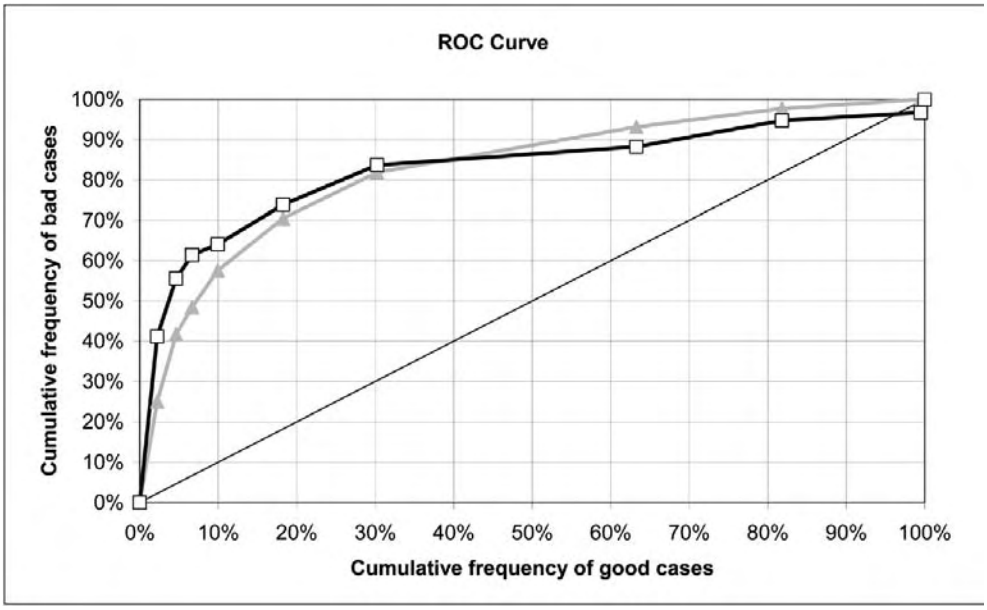

Chart 58: Comparison of Two ROC Curves with the Same AUC Value

There is also an area-based measure of discriminatory power for the α–β error curve: the error area, which represents the area underneath the α–β error curve. Based on the definition of the α–β error curve, the following is true:

Error area = 1 – AUC.

Therefore, AUC and the error area are equivalent measures of discriminatory power. The smaller the error area is, the higher the discriminatory power of the rating model is. The same limitations with regard to one-dimensionality as in the case of AUC also apply in this context.

The Pietra Index as a Measure of Discriminatory Power

Another one-dimensional measure of discriminatory power which can be derived from the ROC curve is the Pietra Index. In geometric terms, the Pietra Index is defined as twice the area of the largest triangle which can be drawn between the diagonal and the ROC curve. In the case of a concave ROC curve, the area of this triangle can also be calculated as the product of the diagonal's length and the largest distance between the ROC curve and the diagonal.[82]

The Pietra Index can take on values between 0 (non-differentiating procedure) and 1 (ideal procedure). It can also be interpreted as the maximum difference between the cumulative frequency distributions of good and bad cases.[83]

[82] *This can also be interpreted as the largest distance between the ROC curve and the diagonal, divided by the maximum possible distance* $(1/\sqrt{2})$ *in an ideal procedure.*

[83] *The relevant literature contains various definitions of the Pietra Index based on different standardizations; the form used here with the value range [0, 1] is the one defined in LEE: Global Performances of Diagnostic Tests.*

Interpretation of the Pietra Index using Probability Theory

In the explanation below, $\triangle P$ represents the average absolute difference in default probabilities in the sample with and without knowledge of the classification of the cases in the c rating classes:

$$\triangle P = \sum_c p_c |P(D|c) - P(D)|.$$

With summation over all rating classes c, the variable p_c refers to the relative frequency of the assignment of cases to individual rating classes, $P(D|c)$ denotes the default rate in class c, and $P(D)$ stands for the default rate in the sample examined. Thus the following relation is true:[84]

$$\triangle P = 2 \cdot P(D) \cdot (1 - P(D)) \cdot [\text{Pietra Index}].$$

In an ideal procedure which places all truly bad cases (and only those cases) in the worst rating class, the following applies:

$$\triangle P_{max} = 2 \cdot P(D) \cdot (1 - P(D)).$$

Therefore, we can simplify the relation above as follows:

$$[\text{Pietra Index}] = \frac{\triangle P}{\triangle P_{max}}.$$

Kolmogorov-Smirnov Test for the Pietra Index

As can be proven, it is also possible to interpret the Pietra Index as the maximum difference between the cumulative frequency distributions of good and bad cases.

$$\text{Pietra Index} = \max \left[F_{cum}^{good} - F_{cum}^{bad} \right].$$

Interpreting the Pietra Index as the maximum difference between the cumulative frequency distributions for the score values of good and bad cases makes it possible to perform a statistical test for the differences between these distributions. This is the Kolmogorov-Smirnov Test (KS Test) for two independent samples. However, this test only yields very rough results, as all kinds of differences between the distribution shapes are captured and evaluated, not only the differences in average rating scores for good and bad case (which are especially relevant to discriminatory power) but also differences in variance and the higher moments of the distribution, that is, the shape of the distributions in general.

The null hypothesis tested is: "The score distributions of good and bad cases are identical." This hypothesis is rejected at level q if the Pietra Index equals or exceeds the following value:

$$D = D_q \Big/ \sqrt{N \cdot p(1-p)}$$

N denotes the number of cases in the sample examined and p refers to the observed default rate. If the Pietra Index $>= D$, therefore, significant differences exist between the score values of good and bad cases.

The values D_q for the individual significance levels (q) are listed in chart 61.

[84] *Cf. LEE, Global Performances of Diagnostic Tests.*

CAP Curve (Powercurve)

Another form of representation which is similar to the ROC curve is the CAP curve,[85] in which the cumulative frequencies of all cases are placed on the x axis instead of the cumulative frequencies of the good cases alone. The ROC and CAP curves are identical in terms of information content, a fact which manifests itself in the associated discriminatory power measures (AUC for the ROC curve, Gini Coefficient for the CAP curve).

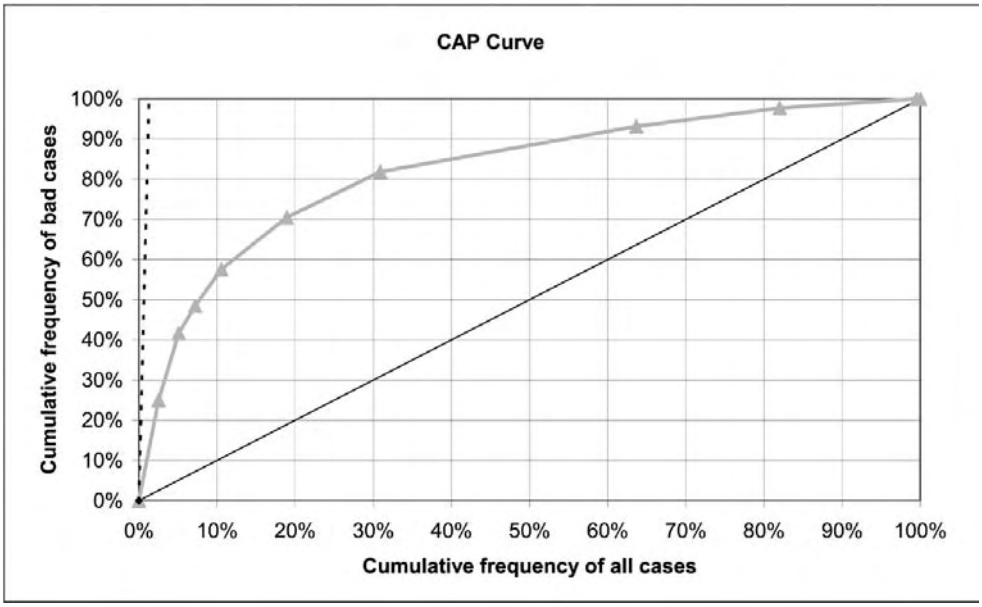

Chart 59: Shape of the CAP Curve for the Data Example

The CAP curve can be interpreted as follows: y% of the cases which actually defaulted over a 12-month horizon can be found among the worst-rated x% of cases in the portfolio. In our example, this means that approximately 80% of later defaults can be found among the worst-rated 30% of cases in the portfolio (i.e. the rating classes 5 to 10); approximately 60% of the later defaults can be found among the worst 10% (classes 7 to 10); etc.

An ideal rating procedure would classify all bad cases (and only those cases) in the worst rating class. This rating class would then contain the precise share p of all cases, with p equaling the observed default rate in the sample examined. For an ideal rating procedure, the CAP curve would thus run from point $(0, 0)$ to point $(p, 1)$[86] and from there to point $(1, 1)$. Therefore, a triangular area in the upper left corner of the graph cannot be reached by the CAP curve.

Gini Coefficient (Accuracy Ratio, AR, Powerstat)

A geometrically defined measure of discriminatory power also exists for the CAP curve: the Gini Coefficient.[87] The Gini Coefficient is calculated as the quotient of the area which the CAP curve and diagonal enclose and the corresponding area in an ideal rating procedure.

[85] *Cumulative Accuracy Profile.*

[86] *i.e. the broken line in the diagram.*

[87] *This is also frequently referred to as the Accuracy Ratio (AR) or Powerstat. Cf. LEE, Global Performances of Diagnostic Tests, and KEENAN/SOBEHART, Performance Measures, as well as the references cited in those works.*

The following relation applies to the ROC and CAP curves' measures of discriminatory power: Gini Coefficient = 2 * AUC – 1.

Therefore, the information contained in the summary measures of discriminatory power derived from the CAP and ROC curves is equivalent.

The table below (chart 60) lists Gini Coefficient values which can be attained in practice for different types of rating models.

Model	Gini Coefficient
Univariate models (individual balance sheet / P&L indicators)	In general, good individual indicators can reach 30-40%, special indicators may reach approx. 55% in selected samples.
Classic rating questionnaire / qualitative systems	Frequently below 50%
Option pricing model	>65% for exchange-listed companies
Multivariate models (discriminant analysis and logit regression)	Practical models with quantitative indicators reach approximately 60-70%.
Multivariate models with quantitative and qualitative factors	Practical models reach approximately 70-80%.
Neural networks	Up to 80% in heavily cleansed samples; however, in practice this value is hardly attainable.

Chart 60: Typical Values Obtained in Practice for the Gini Coefficient as a Measure of Discriminatory Power

Interpretation of the Gini Coefficient using Probability Theory

$\triangle If$ $\triangle P^*$ denotes the average absolute difference in default probabilities for two cases randomly selected from the sample, defined as:

$$\triangle P^* = \sum_c \sum_l p_c \cdot p_l |(P|c) - P(D|l)|,$$

where for summation over all rating classes c and l, the variables p_c and p_l refer to the relative frequency of the assignment of cases to individual rating classes, $P(D|c)$ and $P(D|l)$ denote the default rates in classes c and l, the following relation is true:[88]

$$\triangle P^* = 2 \cdot P(D) \cdot (1 - P(D)) \cdot [\text{Gini coeffizient}].$$

As we can reproduce in several steps, the following applies to an ideal procedure which classifies all bad cases in the worst rating class:

$$\triangle {P_{max}}^* = 2 \cdot P(D) \cdot (1 - P(D)).$$

This means that we can simplify the relation above as follows:

$$[\text{Gini coeffizient}] = \frac{\triangle P^*}{\triangle {P_{max}}^*}.$$

Confidence Levels for the Gini Coefficient and AUC

As one-dimensional measures of discriminatory power, the Gini Coefficient and AUC are statistical values subject to random fluctuations. In general, two procedures can be used to calculate confidence levels for these values:

– Analytical estimation of confidence levels by constructing confidence bands around the CAP or ROC curve[89]
– Heuristic estimation of confidence levels by means of resampling.[90]

[88] *Cf. LEE, Global Performances of Diagnostic Tests.*

[89] *Cf. FAHRMEIR/HENKING/HÜLS, Vergleich von Scoreverfahren and the references cited there.*

[90] *Cf. SOBEHART/KEENAN/STEIN, Validation methodologies.*

In the analytical estimation of confidence levels, confidence bands are placed around the CAP or ROC curve. These bands indicate the area of the diagram in which the overall curve is located at a predefined probability (i.e. the confidence level). As the Gini Coefficient and AUC area measures are summary properties of the overall curve, simultaneous confidence bands are preferable to point-based confidence bands.

We will now demonstrate the process of constructing confidence bands around the ROC curve. For each point in the ROC curve, we use Kolmogorov distribution to define the upper and lower limits of the x and y values for the desired confidence level (1-q), thus creating a rectangle around each point. The probability that the overall ROC curve will be located within these rectangles is $(1\text{-}q)^2$. Linking the outer corners of all confidence rectangles forms the ROC curve envelope for the confidence level $(1\text{-}q)^2$. In turn, the upper and lower limits of the parameter AUC can be calculated on the basis of this envelope. These values form the confidence interval for the AUC value.

The confidence rectangles are constructed by adding the value

$$\pm D_q \Big/ \sqrt{N^+}$$

to the x values and the value

$$\pm D_q \Big/ \sqrt{N^-}$$

to the y values of the points on the ROC curve. N^+ and N^- refer to the number of good and bad cases in the sample examined. The values for D_q can be found in the known Kolmogorov distribution table.

q	20%	15%	10%	5%	1%	0,1%
1-q	80%	85%	90%	95%	99%	99,9%
(1-q)^2	64%	72%	81%	90%	98%	99,8%
Dq	1,07	1,14	1,22	1,36	1,63	1,95

Chart 61: Kolmogorov Distribution Table for Selected Confidence Levels

Chart 62 shows simultaneous confidence bands around the ROC curve at the confidence level $(1\text{-}q)^2 = 90\%$ for the example used here. The confidence interval for the value AUC = 82.8% (calculated using the sample) is between 69.9% and 91.5%.

The table below (chart 63) shows the confidence intervals of the parameter AUC calculated analytically using simultaneous confidence bands for various confidence levels.

Heuristic estimation of confidence intervals for the parameter AUC is based on resampling methods. In this process, a large number of subsamples are drawn from the existing sample. These subsamples can be drawn without replacement (each case in the sample occurs exactly once or not at all in a subsample) or with replacement (each case from the sample can occur multiple times in a subsample). The resulting subsamples should each contain the same number of cases in order to ensure sound comparability. The ROC curve is drawn and the parameter AUC is calculated for each of the subsamples.

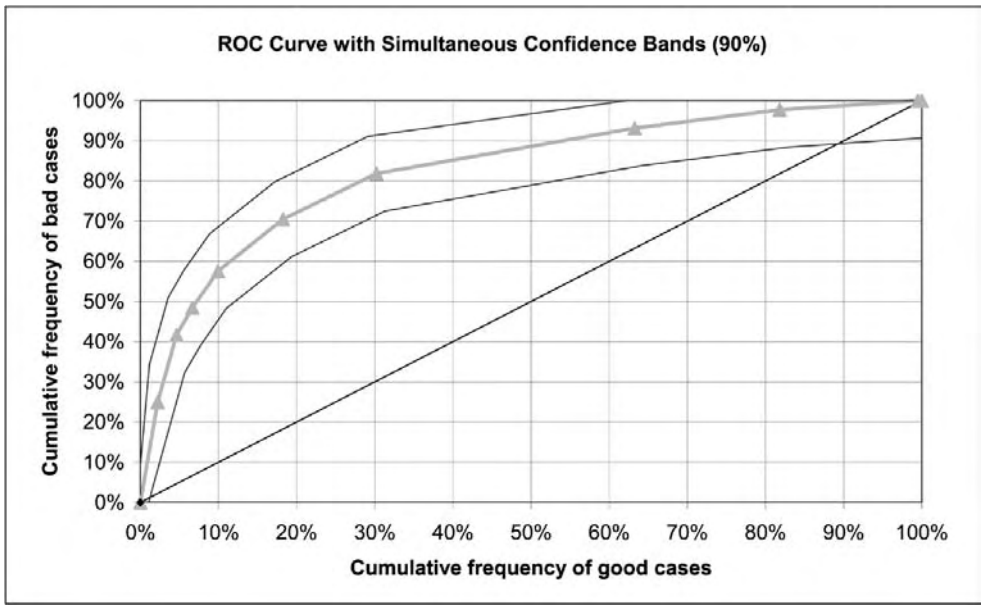

Chart 62: ROC Curve with Simultaneous Confidence Bands at the 90% Level for the Data Example

Confidence level	Lower limit for AUC	Upper limit for AUC
99,8%	64,4%	94,2%
98%	67,4%	92,8%
90%	69,9%	91,5%
81%	71,2%	90,8%
72%	71,9%	90,5%
64%	72,6%	90,1%
Calculated value for AUC = 82.8%		

Chart 63: Table of Upper and Lower AUC Limits for Selected Confidence Levels in the Data Example

Given a sufficient number of subsamples, this procedure yields estimates of the average and variance of the AUC value. However, this procedure often only shows an apparently exact AUC parameter calculated from the same sample, as especially in homogenous samples the AUC fluctuations in subsamples are rather low. However, the procedure is useful for small samples where the confidence bands are very wide due to the small number of cases.

Bayesian Error Rate

As a measure of discriminatory power, the Bayesian error rate is defined as the minimum error rate occurring in the sample examined (α error plus β error). In this technique, the minimum is search for among all cutoff values (the score value beyond which the debtor is classified as bad):

$$ER = \min_{C} [\, p \cdot \alpha(C) + (1 - p) \cdot \beta(C)].$$

The α and β errors for each cutoff value C are weighted with the sample's default rate p or its complement (1-p). For the example with 10 rating classes used here, the table below (chart 64) shows the α and β errors for all cutoff values as well as the corresponding Bayesian error rates for various values of p. The Bayesian error rate means the following: "In the optimum use of the rating model, a proportion *ER* of all cases will still be misclassified."

Rating classes better than cutoff value	◊ error	◊ error	Bayesian error rate with a sample default rate of p =			
			50%	10%	3%	1%
none	0,0%	100,0%	50,0%	90,0%	97,0%	99,0%
1 (best class)	0,0%	99,5%	49,8%	89,6%	96,5%	98,5%
1 - 2	2,3%	81,8%	42,0%	73,8%	79,4%	81,0%
1 - 3	6,8%	63,3%	35,0%	57,6%	61,6%	62,7%
1 - 4	18,2%	30,2%	24,2%	29,0%	29,9%	30,1%
1 - 5	29,5%	18,3%	**23,9%**	19,4%	18,6%	18,4%
1 - 6	42,4%	9,9%	26,2%	13,2%	10,9%	10,3%
1 - 7	51,5%	6,7%	29,1%	11,1%	8,0%	7,1%
1 - 8	58,3%	4,6%	31,5%	10,0%	6,2%	5,2%
1 - 9	75,0%	2,2%	38,6%	**9,5%**	4,4%	2,9%
1 - 10 (all classes)	100,0%	0,0%	50,0%	10,0%	**3,0%**	**1,0%**

Chart 64: Table of Bayesian Error Rates for Selected Sample Default Rates in the Data Example

As the table shows, one severe drawback of using the Bayesian error rate to calculate a rating model's discriminatory power is its heavy dependence on the default rate in the sample examined. Therefore, direct comparisons of Bayesian error rate values derived from different samples are not possible. In contrast, the measures of discriminatory power mentioned thus far (AUC, the Gini Coefficient and the Pietra Index) are independent of the default probability in the sample examined.

The Bayesian error rate is linked to an optimum cutoff value which minimizes the total number of misclassifications (α error plus β error) in the rating system. However, as the optimum always occurs when no cases are rejected ($\alpha = 100\%$, $\beta = 0\%$), it becomes clear that the Bayesian error rate hardly allows the differentiated selection of an optimum cutoff value for the low default rates p occurring in the validation of rating models.

Due to the various costs of α and β errors, the cutoff value determined by the Bayesian error rate is not optimal in business terms, that is, it does not minimize the overall costs arising from misclassification, which are usually substantially higher for α errors than for β errors.

For the default rate $p = 50\%$, the Bayesian error rate equals exactly half the sum of α and β for the point on the α–β error curve which is closest to point $(0, 0)$ with regard to the total α and β errors (cf. chart 65). In this case, the following is also true (α and β errors are denoted as α and β):[91]

$$ER = \tfrac{1}{2} \min[\alpha + \beta] = -\tfrac{1}{2} \max[1 - \alpha - \beta - 1] = -\tfrac{1}{2} \max[F_{cum}^{bad} - F_{cum}^{good}] + \tfrac{1}{2} =$$
$$= \tfrac{1}{2} (1 - \text{Pietra Index})$$

However, this equivalence of the Bayesian error rate to the Pietra Index only applies where $p = 50\%$.[92]

[91] *The Bayesian error rate ER is defined at the beginning of the equation for the case where p = 50%. The error values α and β are related to the frequency distributions of good and bad cases, which are applied in the second to last expression. The last expression follows from the representation of the Pietra Index as the maximum difference between these frequency distributions.*

[92] *Cf. LEE, Global Performances of Diagnostic Tests.*

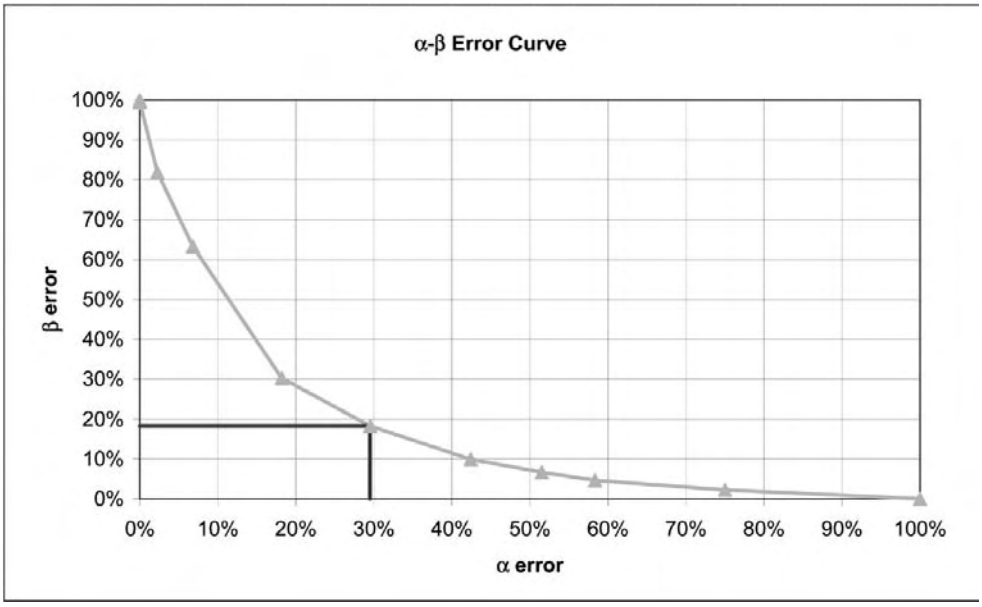

Chart 65: Interpretation of the Bayesian Error Rate as the Lowest Overall Error where p = 50%

Entropy-Based Measures of Discriminatory Power

These measures of discriminatory power assess the information gained by using the rating model. In this context, information is defined as a value which is measurable in absolute terms and which equals the level of knowledge about a future event.

Let us first assume that the average default probability of all cases in the segment in question is unknown. If we look at an individual case in this scenario without being able to estimate its credit quality using rating models or other assumptions, we do not possess any information about the (good or bad) future default status of the case. The maximum in this scenario is the information an observer gains by waiting for the future status.

If, however, the average probability of a credit default is known, the information gained by actually observing the future status of the case is lower in this scenario due to the previously available information.

These considerations lead to the definition of the "information entropy" value, which is represented as follows for dichotomous events with a probability of occurrence p for the "1" event (in this case the credit default):

$$H_0 = -\{p \log_2(p) + (1-p) \log_2(1-p)\}$$

H_0 refers to the absolute information value which is required in order to determine the future default status, or conversely the information value which is gained by observing the "credit default/no credit default" event. Thus entropy can also be interpreted as a measure of uncertainty as to the outcome of an event.

H_0 reaches its maximum value of 1 when p = 50%, that is, when default and non-default are equally probable. H_0 equals zero when p takes the value 0 or 1, that is, the future default status is already known with certainty in advance.

Conditional entropy is defined with conditional probabilities $p(\cdot|c)$ instead of absolute probabilities p; the conditional probabilities are based on condition c.

For the purpose of validating rating models, the condition in the definition of conditional entropy is the classification in rating class c and default event to be depicted (D). For each rating class c, the conditional entropy is h_c:

$$h_c = -\{p(D|c)\log_2(p(D|c)) + (1-p(D|c))\log_2(1-p(D|c))\}.$$

The conditional entropy h_c of a rating class thus corresponds to the uncertainty remaining with regard to the future default status after a case is assigned to a rating class. Across all rating classes in a model, the conditional entropy H_1 (averaged using the observed frequencies of the individual rating classes' p_c) is defined as:

$$H_1 = -\sum_c p_c \cdot h_c.$$

The average conditional entropy H1 corresponds to the uncertainty remaining with regard to the future default status after application of the rating model. Using the entropy H_0, which is available without applying the rating model if the average default probability of the sample is known, it is possible to define a relative measure of the information gained due to the rating model. The conditional information entropy ratio (CIER) is defined as:[93]

$$CIER = \frac{H_0 - H_1}{H_0} = 1 - \frac{H_1}{H_0}.$$

The value CIER can be interpreted as follows:

– If no additional information is gained by applying the rating model, $H_1 = H_0$ and $CIER = 0$.
– If the rating model is ideal and no uncertainty remains regarding the default status after the model is applied, $H_1 = 0$ and $CIER = 1$.

The higher the CIER value is, the more information regarding the future default status is gained from the rating system.

However, it should be noted that information on the properties of the rating model is lost in the calculation of CIER, as is the case with AUC and the other one-dimensional measures of discriminatory power. As an individual indicator, therefore, CIER has only limited meaning in the assessment of a rating model.

Rating class	Proportion per rating class	Default rate per rating class	Conditional entropy per rating class
1	0,5%	0,00%	0,00%
2	17,5%	0,17%	1,79%
3	18,4%	0,32%	3,11%
4	32,7%	0,45%	4,14%
5	12,0%	1,23%	9,55%
6	8,4%	1,99%	14,07%
7	3,3%	3,51%	21,93%
8	2,1%	4,21%	25,16%
9	2,6%	8,30%	41,27%
10	2,5%	12,84%	55,30%
Alle Klassen	**100%**	**1,3%**	**8,29%**
		Unconditional entropy without rating classes	**Conditional entropy (overall model)**
		9,96%	**8,29%**
Conditional Information Entropy Ratio (CIER)		**16,80%**	

Chart 66: Entropy-Based Measures of Discriminatory Power for the Data Example

[93] *The difference $(H_0 - H_1)$ is also referred to as the Kullback-Leibler distance. Therefore, CIER is a standardized Kullback-Leibler distance.*

The table below (chart 67) shows a comparison of Gini Coefficient values and CIER discriminatory power indicators from a study of rating models for American corporates.

Model	Gini Coefficient	CIER
Univariate model (Indicator: return on assets)	53%	0,06
Multivariate discriminant analysis model (Z-Score)	56%	0,09
Option pricing model	67%	0,14
Non-linear multivariate model (logit regression)	73%	0,19

Chart 67: Gini Coefficient and CIER Values from a Study of American Corporates[94]

6.2.2 Back-Testing the Calibration

The assignment of default probabilities to a rating model's output is referred to as calibration. The quality of calibration depends on the degree to which the default probabilities predicted by the rating model match the default rates actually realized. Therefore, reviewing the calibration of a rating model is frequently referred to as back-testing.

The basic data used for back-testing are: the default probabilities forecast over a rating class for a specific time horizon (usually 12 months), the number of cases assigned to the respective rating class by the model, and the default status of those cases once the forecasting period has elapsed, starting from the time of rating (i.e. usually 12 months after the rating was assigned). Calibration involves assigning forecast default probabilities to the individual rating classes (cf. section 5.3). In this process, it is also possible to use longer forecasting horizons than the 12-month horizon required of IRB banks; these other time horizons also have to undergo back-testing.

The results of various segment-specific rating procedures are frequently depicted on a uniform master scale of default probabilities.

In the course of quantitative validation, significant differences may be identified between the default rates on the master scale and the default rates actually realized for individual rating classes in a segment-specific rating procedure. In order to correct these deviations, two different approaches are possible:

– In a fixed master scale, the predefined default probabilities are not changed; instead, only the assignment of results from the rating procedure under review to rating classes on the master scale is adjusted.
– In a variable master scale, the predefined default probabilities are changed, but the assignment of rating results from the rating procedure under review to rating classes on the master scale is not adjusted.

As any changes to the master scale will affect all of the rating procedures used in a bank – including those for which no (or only minor) errors in calibration have been identified – fixed master scales are generally preferable. This is especially true in cases where the default probability of rating classes serves as the basis for risk-based pricing, which would be subject to frequent changes if a variable master scale were used.

[94] *Cf. KEENAN/SOBEHART, Performance Measures.*

However, changes to the master scale are still permissible. Such changes might be necessary in cases where new rating procedures are to be integrated or finer/rougher classifications are required for default probabilities in certain value ranges. However, it is also necessary to ensure that data history records always include a rating class's default probability, the rating result itself, and the accompanying default probability in addition to the rating class.

Chart 68 shows the forecast and realized default rates in our example with ten rating classes. The average realized default rate for the overall sample is 1.3%, whereas the forecast – based on the frequency with which the individual rating classes were assigned as well as their respective default probabilities – was 1.0%. Therefore, this rating model underestimated the overall default risk.

Rating class	Default rate on master scale	Realized default rate
1	0,05%	0,0%
2	0,10%	0,2%
3	0,25%	0,3%
4	0,50%	0,4%
5	0,75%	1,2%
6	1,30%	2,0%
7	2,00%	3,5%
8	3,40%	4,2%
9	6,00%	8,3%
10	10,00%	12,8%
All classes	**1,0%**	**1,3%**

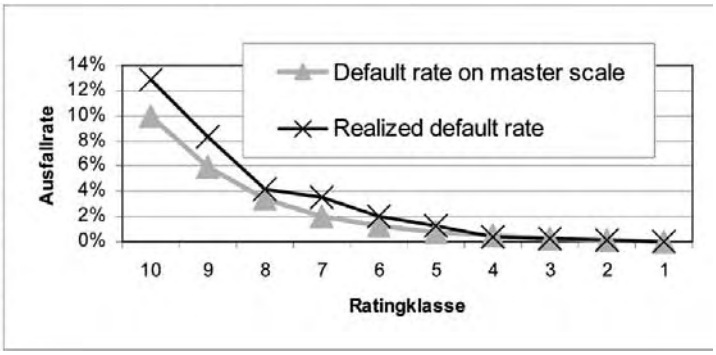

Chart 68: Comparison of Forecast and Realized Default Rates in the Data Example

Brier Score

The average quadratic deviation of the default rate forecast for each case of the sample examined from the rate realized in that case (1 for default, 0 for no default) is known as the Brier Score:[95]

$$BS = \frac{1}{N}\sum_{n=1}^{N}(p_n^{forecast} - y_n)^2 \text{ where } y_n \begin{cases} 1 \text{ for default in n} \\ 0 \text{ for no default in n} \end{cases}$$

In the case examined here, which is divided into C rating classes, the Brier Score can also be represented as the total for all rating classes c:

$$BS = \frac{1}{N}\sum_{c=1}^{K} N_k\left[p_c^{observed}(1 - p_c^{forecast})^2 + (1 - p_c^{observed})(p_c^{forecast})^2\right]$$

In the equation above, N_c denotes the number of cases rated in rating class c, while $p_{observed}$ and $p_{forecast}$ refer to the realized default rate and the forecast default rate (both for rating class c). The first term in the sum reflects the defaults in class c, and the second term shows the non-defaulted cases. This Brier Score equation can be rewritten as follows:

$$BS = \frac{1}{N}\sum_{c=1}^{K} N_c\left[p_c^{observed}(1 - p_c^{observed}) + (p_c^{forecast} - p_c^{observed})^2\right]$$

[95] *Cf. BRIER, G. W., Brier-Score.*

The lower the Brier Score is, the better the calibration of the rating model is. However, it is also possible to divide the Brier Score into three components, only one of which is directly linked to the deviations of real default rates from the corresponding forecasts.[96] The advantage of this division is that the essential properties of the Brier Score can be separated.

$$BS = \underbrace{p^{observed}(1-p^{observed})}_{\text{Uncertainty/variation}} + \underbrace{\frac{1}{N}\sum_{c=1}^{K} N_c\left[(p_c^{forecast} - p_c^{observed})^2\right]}_{\text{Calibration/reliability}} - \underbrace{\frac{1}{N}\sum_{c=1}^{K} N_c\left[\left(p^{observed} - p_c^{observed}\right)^2\right]}_{\text{Resolution}}$$

The first term $BS_{ref} = p^{observed}(1-p^{observed})$ describes the variance of the default rate observed over the entire sample ($p^{observed}$). This value is independent of the rating procedure's calibration and depends only on the observed sample itself. It represents the minimum Brier Score attainable for this sample with a perfectly calibrated but also *trivial rating model* which forecasts the observed default rate precisely for each case but only comprises one rating class. However, the trivial rating model does not differentiate between more or less good cases and is therefore unsuitable as a rating model under the requirements of the IRB approach.

The second term

$$\frac{1}{N}\sum_{c=1}^{K} N_c\left[\left(p_c^{forecast} - p_c^{observed}\right)^2\right]$$

represents the average quadratic deviation of forecast and realized default rates in the C rating classes. A well-calibrated rating model will show lower values for this term than a poorly calibrated rating model. The value itself is thus also referred to as the "calibration."

The third term

$$\frac{1}{N}\sum_{ck=1}^{K} N_c\left[\left(p^{forecast} - p_c^{observed}\right)^2\right]$$

describes the average quadratic deviation of observed default rates in individual rating classes from the default rate observed in the overall sample. This value is referred to as "resolution." While the resolution of the trivial rating model is zero, it is not equal to zero in discriminating rating systems. In general, the resolution of a rating model rises as rating classes with clearly differentiated observed default probabilities are added. Resolution is thus linked to the discriminatory power of a rating model.

The different signs preceding the calibration and resolution terms make it more difficult to interpret the Brier Score as an individual value for the purpose of assessing the classification accuracy of a rating model's calibration. In addition, the numerical values of the calibration and resolution terms are generally far lower than the variance. The table below (chart 69) shows the values of the Brier Score and its components for the example used here.

[96] *Cf. MURPHY, A. H., Journal of Applied Meteorology.*

Rating class	Number of cases	Default rates		Squared difference: observed -	Squared difference: observed -
		Default rate on master scale	Realized default rate		
1	50	0,05%	0,00%	0,0000%	0,0167%
2	1788	0,10%	0,17%	0,0000%	0,0126%
3	1876	0,25%	0,32%	0,0000%	0,0095%
4	3345	0,50%	0,45%	0,0000%	0,0071%
5	1223	0,75%	1,23%	0,0023%	0,0000%
6	856	1,30%	1,99%	0,0047%	0,0048%
7	342	2,00%	3,51%	0,0228%	0,0491%
8	214	3,40%	4,21%	0,0065%	0,0849%
9	265	6,00%	8,30%	0,0530%	0,4914%
10	257	10,00%	12,84%	0,0807%	1,3336%
All classes	**10216**	**0,97%**	**1,29%**	**0,0010%**	**0,0000%**
			1,28%	**0,0050%**	**0,0565%**
			Variance	**Calibration**	**Resolution**
			1,2239%		
			Brier Score		

Chart 69: Calculation of Brier Score for the Data Example

In practice, a standardized measure known as the Brier Skill Score (BSS) is often used instead of the Brier Score. This measure scales the Brier Score to the variance term:

$$BSS = 1 - \frac{BS}{p^{observed}(1 - p^{observed})}.$$

In the trivial, ideal model, the Brier Skill Score takes the value zero. For the example above, the resulting value is BSS = 4.04 %.

Reliability Diagrams

Additional information on the quality of calibration for rating models can be derived from the reliability diagram, in which the observed default rates are plotted against the forecast rate in each rating class. The resulting curve is often referred to as the calibration curve.

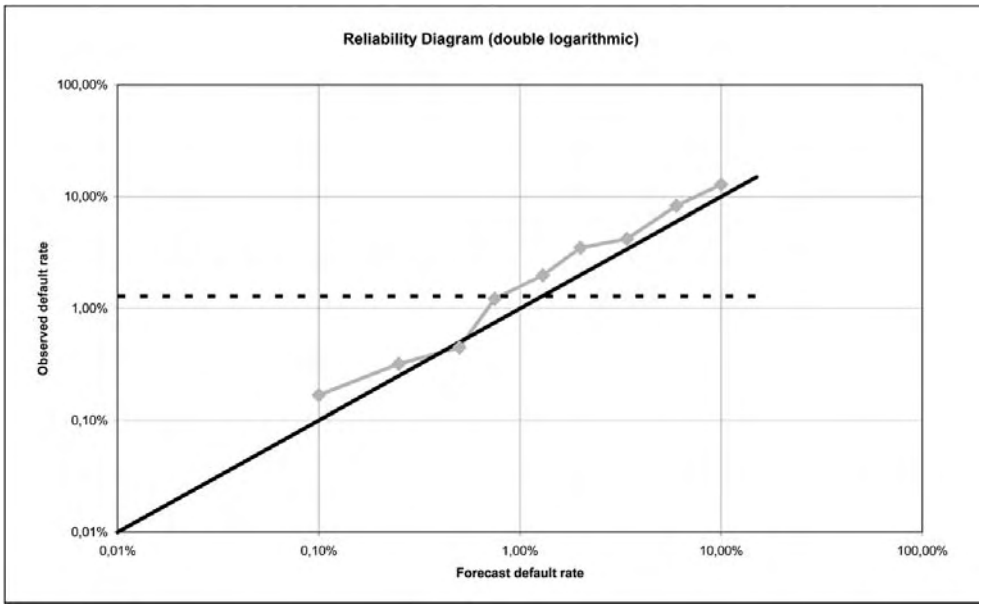

Chart 70: Reliability Diagram for the Data Example

Chart 70 shows the reliability diagram for the example used here. In the illustration, a double logarithmic representation was selected because the default probabilities are very close together in the good rating classes in particular. Note that the point for rating class 1 is missing in this graph. This is because no defaults were observed in that class.

The points of a well-calibrated system will fall close to the diagonal in the reliability diagram. In an ideal system, all of the points would lie directly on the diagonal. The "calibration" term of the Brier Score represents the average (weighted with the numbers of cases in each rating class) squared deviation of points on the calibration curve from the diagonal. This value should be as low as possible.

The resolution of a rating model is indicated by the average (weighted with the numbers of cases in the individual rating classes) squared deviation of points in the reliability diagram from the broken line, which represents the default rate observed in the sample. This value should be as high as possible, which means that the calibration curve should be as steep as possible. However, the steepness of the calibration curve is primarily determined by the rating model's discriminatory power and is independent of the accuracy of default rate estimates.

An ideal trivial rating system with only one rating class would be represented in the reliability diagram as an isolated point located at the intersection of the diagonal and the default probability of the sample.

Like discriminatory power measures, one-dimensional indicators for calibration and resolution can also be defined as standardized measures of the area between the calibration curve and the diagonal or the sample default rate.[97]

Checking the Significance of Deviations in the Default Rate

In light of the fact that realized default rates are subject to statistical fluctuations, it is necessary to develop indicators to show how well the rating model estimates the parameter PD. In general, two approaches can be taken:

– Assumption of uncorrelated default events
– Consideration of default correlation

Empirical studies show that default events are generally not uncorrelated. Typical values of default correlations range between 0.5% and 3%. Default correlations which are not equal to zero have the effect of strengthening fluctuations in default probabilities. The tolerance ranges for the deviation of realized default rates from estimated values may therefore be substantially larger when default correlations are taken into account. In order to ensure conservative estimates, therefore, it is necessary to review the calibration under the initial assumption of uncorrelated default events.

The statistical test used here checks the null hypothesis "The forecast default probability in a rating class is correct" against the alternative hypothesis "The forecast default probability is incorrect" using the data available for back-testing. This test can be one-sided (checking only for significant overruns of the forecast default rate) or two-sided (checking for significant overruns and underruns of the forecast default probability). From a management standpoint, both significant underestimates and overestimates of risk are relevant. A one-sided test can

[97] *See also HASTIE/TIBSHIRANI/FRIEDMAN, Elements of statistical learning.*

also be used to check for risk underestimates. One-sided and two-sided tests can also be converted into one another.[98]

Calibration Test using Standard Normal Distribution

One simple test for the calibration of default rates under the assumption of uncorrelated default events uses standard normal distribution.[99] In the formulas below, Φ^{-1} denotes the inverse cumulative distribution function for the standard normal distribution, N_c stands for the number of cases in rating class c, and p_c refers to the default rate:

– (one-sided test): If

$$\left(p_c^{observed} - p_c^{forecast}\right) > \Phi^{-1}(q) \cdot \sqrt{\frac{p_c^{forecast}(1 - p_c^{forecast})}{N_c}},$$

the default rate in class c is significantly underestimated at the confidence level q.

– (one-sided test): If

$$\left(p_c^{prognose} - p_c^{observed}\right) > \Phi^{-1}(q) \cdot \sqrt{\frac{p_c^{forecast}(1 - p_c^{forecast})}{N_c}},$$

the default rate in class c is significantly overestimated at the confidence level q.

– (two-sided test): If

$$\left|p_c^{observed} - p_c^{forecast}\right| > \Phi^{-1}\left(\frac{q+1}{2}\right) \cdot \sqrt{\frac{p_c^{forecast}(1 - p_c^{forecast})}{N_c}},$$

the default rate in class c is significantly misestimated at confidence level q.

The table below (chart 71) shows the results of the one-sided test for risk underestimates. Significant overruns of the forecast default rates are shaded in gray for individual significance levels. The default rates in classes 5 to 7 were significantly underestimated at a level of 95%, while the default rates in classes 9 and 10 were only underestimated at the 90% level. The average default probability of the overall sample was underestimated even at the level of 99.9%.[100]

This shows that a highly significant misestimate of the average default rate for the entire sample can arise even if the default rates in the individual classes remain within their tolerance ranges. This is due to the inclusion of the number of cases N in the denominator of the test statistic; this number reduces the test statistic when all rating classes are combined, thus making the test more sensitive.

[98] *The limit of a two-sided test at level q is equal to the limit of a one-sided test at the level ½(q + 1). If overruns/underruns are also indicated in the two-sided test by means of the plus/minus signs for differences in default rates, the significance levels will be identical.*

[99] *Cf. CANTOR, R./FALKENSTEIN, E., Testing for rating consistencies in annual default rates.*

[100] *The higher the significance level q is, the more statistically certain the statement is; in this case, the statement is the rejection of the null hypothesis asserting that PD is estimated correctly. The value (1-q) indicates the probability that this rejection of the null hypothesis is incorrect, that is, the probability that an underestimate of the default probability is identified incorrectly.*

Rating class	Default rates			Tolerance ranges for significance level:			
	Default rate on master scale	Realized default rate	Difference	90%	95%	99%	99,9%
1	0,05%	0,0%	-0,05%	0,41%	0,52%	0,74%	0,98%
2	0,10%	0,2%	0,07%	0,10%	0,12%	0,17%	0,23%
3	0,25%	0,3%	0,07%	0,15%	0,19%	0,27%	0,36%
4	0,50%	0,4%	-0,05%	0,16%	0,20%	0,28%	0,38%
5	0,75%	1,2%	0,48%	0,32%	0,41%	0,57%	0,76%
6	1,30%	2,0%	0,69%	0,50%	0,64%	0,90%	1,20%
7	2,00%	3,5%	1,51%	0,97%	1,25%	1,76%	2,34%
8	3,40%	4,2%	0,81%	1,59%	2,04%	2,88%	3,83%
9	6,00%	8,3%	2,30%	1,87%	2,40%	3,39%	4,51%
10	10,00%	12,8%	2,84%	2,40%	3,08%	4,35%	5,78%
All classes	**1,0%**	**1,3%**	**0,32%**	**0,12%**	**0,16%**	**0,23%**	**0,30%**

Chart 71: Identification of Significant Deviations in Calibration in the Data Example
(Test with Standard Normal Distribution)

For the purpose of interpreting confidence levels, a "traffic lights approach" has been proposed for practice in Germany.[101] In this approach, deviations of realized and forecast default rates below a confidence level of 95% should not be regarded as significant ("green" range). Deviations at a confidence level of at least 99.9% are then considered significant and should definitely be corrected ("red" range). Deviations which are significant at confidence levels between 95% and 99.9% may need to be corrected ("yellow" range).

For the example above, this means that the overall default rate – which was underestimated by the rating model – should be corrected upward, preferably by making the appropriate adjustments in rating classes 5 to 7.

Binomial Calibration Test

The test using normal distribution (described above) is a generalization of the binomial test for frequencies of uncorrelated binary events. The binomial test is described in detail below.

For low default probabilities and low numbers of cases in the individual rating classes, the prerequisites for using normal distribution are not always met. The table below (chart 72) lists the minimum number of cases required for a sound approximation of test values with standard normal distribution when testing various default probabilities.[102]

If individual classes contain fewer cases than the minimum number indicated, the binomial test should be carried out. In the formula below, N_c^- denotes the number of defaults observed in class c, and N_c refers to the number of cases in class c. Summation is performed for all defaults in class c.

– (one-sided test): If

$$\sum_{n=0}^{N_c^-} \binom{N_c}{n} (p_c^{forecast})^n (1 - p^{forecast})^{N_c - n} > q,$$

the default rate in class c is significantly underestimated at confidence level q.[103]

[101] *Cf. TASCHE, D., A traffic lights approach to PD validation.*

[102] *The strict condition for the application of standard normal distribution as an approximation of binomial distribution is $N_p(1-p) > 9$, where N is the number of cases in the rating class examined and p is the forecast default probability (cf. SACHS, L., Angewandte Statistik, p. 283).*

[103] *This is equivalent to the statement that the probability of occurrence $P[n < N_k^- | p_k^{prognose}; N_k]$ (assumed to be binomially distributed) of a maximum of N_c defaults among the N_c cases in class c must not be greater than q.*

Default rate to be tested	Minimum number of cases for normal
0,10%	9010
0,25%	3610
0,50%	1810
1%	910
2%	460
3%	310
5%	190
10%	101
20%	57
50%	37

Chart 72: Theoretical Minimum Number of Cases for the Normal Test based on Various Default Rates

The table below (chart 73) shows the results of the binomial test for significant underestimates of the default rate. It is indeed conspicuous that the results largely match those of the test using normal distribution despite the fact that several classes did not contain the required minimum number of cases.

Rating class	Number of bad cases	Total	Cum. binomial distribution	Test result at significance level:			
				90%	95%	99%	99,9%
1	0	50	97,5%	Underestimate	Underestimate		
2	3	1788	89,3%				
3	6	1876	80,6%				
4	15	3345	39,6%				
5	15	1223	97,5%	Underestimate	Underestimate		
6	17	856	96,6%	Underestimate	Underestimate		
7	12	342	97,8%	Underestimate	Underestimate		
8	9	214	80,5%				
9	22	265	95,0%	Underestimate	Underestimate		
10	33	257	94,3%	Underestimate			
All classes	**132**	**10216**	**99,9%**	Underestimate	Underestimate	Underestimate	Underestimate

Chart 73: Identification of Significant Deviations in Calibration for the Data Example (Binomial Test)

One interesting deviation appears in rating class 1, where the binomial test yields a significant underestimate at the 95% level for an estimated default probability of 0.05% and no observed defaults. In this case, the binomial test does not yield reliable results, as the following already applies when no defaults are observed:

$$P\left[n \leq 0 | p_1^{prognose}; N_1\right] = (1 - p_1^{prognose})^{N_1} > 90\%.$$

In general, the test using normal distribution is faster and easier to perform, and it yields useful results even for small samples and low default rates. This test is thus preferable to the binomial test even if the mathematical prerequisites for its application are not always met. However, it is important to bear in mind that the binomial test and its generalization using normal distribution are based on the assumption of uncorrelated defaults. A test procedure which takes default correlations into account is presented below.

Calibration Test Procedure Based on Default Correlation

The assumption of uncorrelated defaults generally yields an overestimate of the significance of deviations in the realized default rate from the forecast rate. This is especially true of risk underestimates, that is, cases in which the realized default rate is higher than the forecast rate. From a conservative risk assessment standpoint, overestimating significance is not critical in the case of risk under-

estimates, which means that it is entirely possible to operate under the assumption of uncorrelated defaults. In any case, however, persistent overestimates of significance will lead to more frequent recalibration of the rating model, which can have negative effects on the model's stability over time. It is therefore necessary to determine at least the approximate extent to which default correlations influence PD estimates.

Default correlations can be modeled on the basis of the dependence of default events on common and individual random factors.[104] For correlated defaults, this model also makes it possible to derive limits for assessing deviations in the realized default rate from its forecast as significant at certain confidence levels.

In the approximation formula below, q denotes the confidence level (e.g. 95% or 99.9%), N_c the number of cases observed per rating class, p_k the default rate per rating class, Φ the cumulative standard normal distribution, ϕ the probability density function of the standard normal distribution, and ρ the default correlation:

(one-sided test): If

$$p_c^{observed} > Q + \frac{1}{2N_c}\left(2Q - 1 - \frac{Q\cdot(1-Q)}{\phi\left(\sqrt{\frac{\rho}{1-\rho}}\Phi^{-1}(1-q) - \frac{t}{\sqrt{1-\rho}}\right)}\left(\frac{(1-2\rho)\cdot\Phi^{-1}(1-q) - t\sqrt{\rho}}{\sqrt{\rho(1-\rho)}}\right)\right)$$

the default rate in class c is significantly underestimated at the confidence level q.

In this context, $t = \Phi^{-1}(p_c^{forecast}); Q = \Phi^{-1}\left(\frac{\sqrt{\rho}\,\Phi^{-1}(q) + t}{\sqrt{1-\rho}}\right)$.

The test above can be used to check for overestimates of the default rate in individual rating classes as well as the overall sample.[105]

The table below (chart 74) compares the results of the calibration test under the assumption of uncorrelated defaults with the results based on a default correlation of 0.01. A deviation in the realized default rate from the forecast rate is considered significant whenever the realized default rate exceeds the upper limit indicated for the respective confidence level.

Rating class	Default rates			Upper limit of DR for confidence level: Uncorrelated		Correlation: 0,01	
	Default rate on master scale	Realized default rates	Number of cases per rating class	95%	99,9%	95%	99,9%
1	0,05%	0,00%	50	0,57%	1,03%	4,79%	9,41%
2	0,10%	0,17%	1788	0,22%	0,33%	0,30%	0,53%
3	0,25%	0,32%	1876	0,44%	0,61%	0,54%	0,89%
4	0,50%	0,45%	3345	0,70%	0,88%	0,86%	1,31%
5	0,75%	1,23%	1223	1,16%	1,51%	1,38%	2,13%
6	1,30%	1,99%	856	1,94%	2,50%	2,29%	3,45%
7	2,00%	3,51%	342	3,25%	4,34%	3,88%	5,96%
8	3,40%	4,21%	214	5,44%	7,23%	6,47%	9,78%
9	6,00%	8,30%	265	8,40%	10,51%	9,62%	13,50%
10	10,00%	12,84%	257	13,08%	15,78%	14,76%	19,71%
All classes	**1,0%**	**1,29%**	**10216**	**1,13%**	**1,27%**	**1,48%**	**2,14%**

Chart 74: Identification of Significant Deviations in Calibration in the Data Example
(Comparison of Tests for Uncorrelated Cases and Results for a Default Correlation of 0.01)

[104] *Vasicek One-Factor Model, cf. e.g. TASCHE, D, A traffic lights approach to PD validation.*

[105] *In this context, it is necessary to note that crossing the boundary to uncorrelated defaults ($\rho \to 0$) is not possible in the approximation shown. For this reason, the procedure will yield excessively high values in the upper default probability limit for very low default correlations (< 0.005).*

The assumption of weak default correlation is already sufficient in this example to refute the assumption of miscalibration for the model at both the 95% as well as the 99.9% level. Under the assumption of uncorrelated defaults, on the other hand, deviations in the yellow range appear in rating classes 5 to 7, and a deviation in the red range (as defined in the aforementioned traffic lights approach) is shown for the overall sample.

The table below (chart 75) indicates the upper limits in the one-sided test at a 95% significance level for various default correlations. In this context, it is important to note that with higher default correlations the limits yielded by this test become lower for rating class 1 due to the very low default rate and the small number of cases in that class.

A number of additional test procedures exist for the validation of rating model calibrations. The approach presented here is distinguished by its relative simplicity. Besides analytical models,[106] simulation models can also be implemented, but for default correlations between 0.01 and 0.05 they should yield similar results to those presented here. The absolute amount of the default correlations themselves is also subject to estimation; however, this will not be discussed in further detail at this point.[107] However, the assumed correlation's significant influence on the validation result should serve to illustrate that it is necessary to determine this parameter as accurately as possible in order to enable meaningful statements regarding the quality of calibration.

Rating class	Default rates			Upper limit of DR for confidence level 95%				
				Correlation				
	Default rate on master scale	Realized default rate	Number of cases per rating class	0,5%	1,0%	2,0%	5,0%	10,0%
1	0,05%	0,00%	50	6,71%	4,79%	3,44%	2,23%	1,62%
2	0,10%	0,17%	1788	0,34%	0,30%	0,29%	0,32%	0,38%
3	0,25%	0,32%	1876	0,55%	0,54%	0,57%	0,68%	0,84%
4	0,50%	0,45%	3345	0,81%	0,86%	0,96%	1,21%	1,54%
5	0,75%	1,23%	1223	1,36%	1,38%	1,49%	1,82%	2,27%
6	1,30%	1,99%	856	2,25%	2,29%	2,47%	3,00%	3,72%
7	2,00%	3,51%	342	4,01%	3,88%	4,00%	4,63%	5,61%
8	3,40%	4,21%	214	6,75%	6,47%	6,61%	7,52%	8,99%
9	6,00%	8,30%	265	9,58%	9,62%	10,15%	11,84%	14,25%
10	10,00%	12,84%	257	14,50%	14,76%	15,66%	18,18%	21,64%
All classes	**1,0%**	**1,29%**	**10216**	**1,34%**	**1,48%**	**1,70%**	**2,18%**	**2,78%**

Chart 75: Upper Default Rate Limits at the 95% Confidence Level for Various Default Correlations in the Example

6.2.3 Back-Testing Transition Matrices

In general, the methods applied for default probabilities can also be used to back-test the transition matrix. However, there are two essential differences in this procedure:

- Back-testing transition matrices involves simultaneous testing of a far larger number of probabilities. This imposes substantially higher data requirements on the sample used for back-testing.
- Transition probability changes which become necessary during back-testing must not bring about inconsistencies in the transition matrix.

The number of data fields in the transition matrix increases sharply as the number of rating classes increases (cf. section 5.4.1). For each column of the

[106] *Cf. TASCHE, D., A traffic lights approach to PD validation.*

[107] *Cf. DUFFIE, D./SINGLETON, K. J., Simulating correlated defaults and ZHOU, C., Default correlation: an analytical result, as well as DUFFIE, D./SINGLETON, K. J., Credit Risk, Ch. 10.*

transition matrix, the data requirements for back-testing the matrix are equivalent to the requirements for back-testing default probabilities for all rating classes.

Specifically, back-testing the default column in the transition matrix is equivalent to back-testing the default probabilities over the time horizon which the transition matrix describes. The entries in the default column of the transition matrix should therefore no longer be changed once the rating model has been calibrated or recalibrated. This applies to one-year transition matrices as well as multi-year transition matrices if the model is calibrated for longer time periods.

When back-testing individual transition probabilities, it is possible to take a simplified approach in which the transition matrix is divided into individual elementary events. In this approach, the dichotomous case in which "The transition of rating class x leads to rating class y (Event A)" or "The transition of rating class x does *not* lead to rating class y (Event B)" is tested for all pairs of rating classes. The forecast probability of transition from x to y ($p^{forecast}$) then has to be adjusted to Event A's realized frequency of occurrence.

The procedure described is highly simplistic, as the probabilities of occurrence of all possible transition events are correlated with one another and should therefore not be considered separately. However, lack of knowledge about the size of the correlation parameters as well as the analytical complexity of the resulting equations present obstacles to the realization of mathematically correct solutions. As in the assumption of uncorrelated default events, the individual transition probabilities can be subjected to isolated back-testing for the purpose of initial approximation.

For each individual transition event, it is necessary to check whether the transition matrix has significantly underestimated or overestimated its probability of occurrence. This can be done using the binomial test. In the formulas below, N^A denotes the observed frequency of the transition event (Event A), while N^B refers to the number of cases in which Event B occurred (i.e. in which ratings in the same original class migrated to any other rating class).

– (one-sided test): If

$$\sum_{n=0}^{N^A} \binom{N^A + B^B}{n} (p^{forecast})^n (1 - p^{forecast})^{N^A + N^B - n} > q,$$

the transition probability was significantly underestimated at confidence level q.[108]

– (one-sided test): If

$$\sum_{n=0}^{N^A} \binom{N^A + B^B}{n} (p^{forecast})^n (1 - p^{forecast})^{N^A + N^B - n} < 1 - q,$$

the transition probability was significantly overestimated at confidence level q.

The binomial test is especially suitable for application even when individual matrix rows contain only small numbers of cases. If large numbers of cases are

[108] *This is equivalent to the statement that the (binomially distributed) probability* $P[n < N^A | p^{forecast}; N^A + N^B]$ *of occurrence of a maximum of* N^A *defaults for* $N^A + N^B$ *cases in the original rating class must not be greater than* q.

available, the binomial test can be replaced with a test using standard normal distribution.[109]

Chart 76 below shows an example of how the binomial test would be performed for one row in a transition matrix. For each matrix field, we first calculate the test value on the left side of the inequality. We then compare this test value to the right side of the inequality for the 90% and 95% significance levels. Significant deviations between the forecast and realized transition rates are identified at the 90% level for transitions to classes 1b, 3a, 3b, and 3e. At the 95% significance level, only the underruns of forecast rates of transition to classes 1b and 3b are significant.

In this context, class 1a is a special case because the forecast transition rate equals zero in this class. In this case, the test value used in the binomial test always equals 1, even if no transitions are observed. However, this is not considered a misestimate of the transition rate. If transitions are observed, the transition rate is obviously not equal to zero and would have to be adjusted accordingly.

Forecast Migration Matrix

Original score class	Score class after 12 months: 1a	1b	1c	1d	...	3a	3b	3c	3d	3e	Ausfall	Alle	
...													
2a	0,0%	0,7%	6,6%	12,3%	...	2,0%	0,7%	0,7%	0,4%	0,2%	**0,5%**	**100%**	<- Values to be tested for migration matrix
...													

Migrations Observed

2a	0	0	31	56	...	13	0	1	2	2	**2**	**449**	<- Number of observations
	0,0%	0,0%	6,9%	12,5%	...	2,9%	0,0%	0,2%	0,4%	0,4%	**0,4%**	**100%**	<- Realized frequencies

Binomial Test

Test value	100,0%	4,3%	64,8%	57,9%	...	92,9%	4,3%	17,8%	73,2%	93,8%	61,1%	
Significance	!				...	!				!		<- Significant overrun
90%		!			...		!					<- Significant underrun
Significance	!				...							<- Significant overrun
95%		!			...		!					<- Significant underrun of forecast value

Chart 76: Data Example for Back-testing a Transition Matrix with the Binomial Test

If significant deviations are identified between the forecast and realized transition rates, it is necessary to adjust the transition matrix. In the simplest of cases, the inaccurate values in the matrix are replaced with new values which are determined empirically. However, this can also bring about inconsistencies in the transition matrix which then make it necessary to smooth the matrix.

There is no simple algorithmic method which enables the parallel adjustment of all transition frequencies to the required significance levels with due attention to the consistency rules. Therefore, practitioners often use pragmatic solutions in which parts of individual transition probabilities are shifted to neighboring classes out of heuristic considerations in order to adhere to the consistency rules.[110]

If multi-year transition matrices can be calculated directly from the data set, it is not absolutely necessary to adhere to the consistency rules. However, what is essential to the validity and practical viability of a rating model is the consis-

[109] *Cf. the comments in section 6.2.2 on reviewing the significance of default rates.*

[110] *In mathematical terms, it is not even possible to ensure the existence of a solution in all cases.*

tency of the cumulative and conditional default probabilities calculated from the one-year and multi-year transition matrices (cf. section 5.4.2).

Once a transition matrix has been adjusted due to significant deviations, it is necessary to ensure that the matrix continues to serve as the basis for credit risk management. For example, the new transition matrix should also be used for risk-based pricing wherever applicable.

6.2.4 Stability

When reviewing the stability of rating models, it is necessary to examine two aspects separately:

- Changes in the discriminatory power of a rating model given forecasting horizons of varying length and changes in discriminatory power as loans become older
- Changes in the general conditions underlying the use of the model and their effects on individual model parameters and on the results the model generates.

In general, rating models should be robust against the aging of the loans rated and against changes in general conditions. In addition, another essential characteristic of these models is a sufficiently high level of discriminatory power for periods longer than 12 months.

Changes in Discriminatory Power over Various Forecast Horizons

In section 6.2.1, we described the process of testing the discriminatory power of rating models over a time horizon of 12 months. However, it is also possible to measure discriminatory power over longer periods of time if a sufficient data set is available. In this context, any measure of discriminatory power, such as the Gini Coefficient (Powerstat), can be used.

When rating models are optimized for a period of 12 months, their discriminatory power decreases for longer time horizons. Here it is necessary to ensure that the discriminatory power of a rating model only deteriorates steadily, that is, without dropping abruptly to excessively low values. Sound rating models should also demonstrate sufficient discriminatory power over forecasting horizons of three or more years.

Another aspect of the time stability of rating models is the decrease in discriminatory power as loans become older. This is especially relevant in the case of application scores where the discriminatory power for an observed quantity of new business cases decreases noticeably over a period of 6 to 36 months after an application is submitted. This is due to the fact that the data used in application scoring become less significant over time. Therefore, practitioners frequently complement application scoring models with behavior scoring models. The latter models evaluate more recent information from the development of the credit transaction and therefore provide a better indicator of creditworthiness than application scoring models alone. However, behavior scoring is not possible until a credit facility has reached a certain level of maturity, that is, once behavior-related data are actually available.

Changes in the General Conditions for Model Use

The assessment of changes in the general conditions under which a model is used has strong qualitative elements. On the one hand, it is necessary to review whether developments in the economic, political, or legal environment will have an influence on the rating model or individual model parameters and criteria. On the other hand, internal factors at the bank such as changes in business strategies, the expansion of activities in certain market segments, or changes in organizational structures may also affect the performance of a rating model substantially.

Changes in the economic environment include the business cycle in particular, which can cause major fluctuations in the parameter PD during periods of recovery and decline. However, factors such as technical progress or political and legal developments can also influence the effectiveness of rating models.

In particular, country and regional ratings depend heavily on changes in the political environment. However, such scenarios should already be integrated into a rating model, otherwise one would have to question the suitability of the model itself.

Examples in the legal environment include changes in commercial law or accounting standards which may have a positive or negative influence on the effectiveness and significance of certain financial indicators. Changes in legislation which are relevant in this context also include revisions of the minimum subsistence level (i.e. the salary amount which cannot be attached) or changes in bankruptcy procedures. In particular, these changes can cause shifts in the risk parameter LGD.

Quantifying the effects of changes in general conditions on the functionality of rating models requires an in-depth analysis of the model parameters and should therefore accompany the ongoing development of the model. Rating models have to undergo further development whenever their performance decreases due to changes in general conditions.

On the other hand, the bank may also decide to develop a new rating model if experts believe that a potential or planned change in general conditions would lead to a substantial loss in the performance of the current model.

6.3 Benchmarking

In the quantitative validation of rating models, it is necessary to distinguish between back-testing and benchmarking.[111] Back-testing refers to validation on the basis of a bank's in-house data. In particular, this term describes the comparison of forecast and realized default rates in the bank's credit portfolio.

In contrast, benchmarking refers to the application of a rating model to a reference data set (benchmark data set). Benchmarking specifically allows quantitative statistical rating models to be compared using a uniform data set. It can be proven that the indicators used in quantitative validation – in particular discriminatory power measures, but also calibration measures – depend at least in part on the sample examined.[112] Therefore, benchmarking results are preferable

[111] *DEUTSCHE BUNDESBANK, Monthly Report for September 2003, Approaches to the validation of internal rating systems.*
[112] *HAMERLE/RAUHMEIER/RÖSCH, Uses and misuses of measures for credit rating accuracy.*

to internal back-testing results in the comparative evaluation of different rating models.

In terms of method and content, back-testing and benchmarking involve the same procedures. However, the two procedures' results are interpreted differently in terms of scope and orientation:

Back-Testing for Discriminatory Power
Poor results from discriminatory power tests in the bank's own data set primarily indicate weaknesses in the rating model and should prompt the development of a new or revised model. However, sound discriminatory power results from back-testing alone do not provide a reliable indication of a rating model's goodness of fit compared to other models.

Testing Discriminatory Power by Benchmarking
Poor results from discriminatory power tests using external reference data may indicate structural differences between the data sets used for rating development and benchmarking. This is especially true when the discriminatory power of the model is comparatively high in back-testing. However, good results in back-testing discriminatory power compared to other rating models should not be considered the only criterion for a rating model's goodness of fit. The results should also be compared to the discriminatory power derived from the bank's internal back-tests.

Testing Calibration by Back-Testing
In general, the calibration of a rating system should always be reviewed using the bank's internal data and adjusted according to the results. Only in this way is it possible to ensure that the rating model's risk estimates accurately reflect the structure of the portfolio analyzed.

Testing Calibration by Benchmarking
The calibration of a rating model should not be tested using benchmark data alone, as the structure of reference data would have to precisely match the segment in which the rating model is used. Only then could one expect reliable results from calibrations to benchmark data when applying the model to the bank's in-house data. Therefore, it is only advisable to calibrate a rating model to benchmark or reference data in cases where the quality or quantity of the bank's internal data is insufficient to enable calibration with an acceptable level of statistical precision. In particular, this may be the case in segments with low numbers of defaults or in highly specialized segments.

Quality Requirements for Benchmark Data
A number of requirements apply to the data set used in benchmarking:

– *Data quality:*
  The data quality of the benchmark sample has to at least fulfill the requirements which apply to the development and validation of rating models within the bank.
– *Consistency of input data fields:*
  It is important to ensure that the content of data fields in the benchmark

sample matches that of the input data fields required in the rating model. Therefore, it is absolutely necessary that financial data such as annual financial statements are based on uniform legal regulations. In the case of qualitative data, individual data field definitions also have to match; however, highly specific wordings are often chosen in the development of qualitative modules. This makes benchmark comparisons of various rating models exceptionally difficult. It is possible to map data which are defined or categorized differently into a common data model, but this process represents another potential source of errors. This may deserve special attention in the interpretation of benchmarking results.

– *Consistency of target values:*
  In addition to the consistency of input data fields, the definition of target values in the models examined must be consistent with the data in the benchmark sample. In benchmarking for rating models, this means that all models examined as well as the underlying sample have to use the same definition of a default. If the benchmark sample uses a narrower (broader) definition of a credit default than the rating model, this will increase (decrease) the model's discriminatory power and simultaneously lead to overestimates (underestimates) of default rates.
– *Structural consistency:*
  The structure of the data set used for benchmarking has to depict the respective rating models' area of application with sufficient accuracy. For example, when testing corporate customer ratings it is necessary to ensure that the company size classes in the sample match the rating models' area of application. The sample may have to be cleansed of unsuitable cases or optimized for the target area of application by adding suitable cases. Other aspects which may deserve attention in the assessment of a benchmark sample's representativity include the regional distribution of cases, the structure of the industry, or the legal form of business organizations. In this respect, the requirements imposed on the benchmark sample are largely the same as the representativity requirements for the data set used in rating model development.

6.4 Stress Tests

6.4.1 Definition and Necessity of Stress Tests

In general, stress tests can be described as instruments for estimating the potential effects an extraordinary – but plausible – event may have on an institution.

The term "extraordinary" in this definition implies that stress tests evaluate the consequences of events which have a low probability of occurrence. However, crisis events must not be so remote from practice that they become implausible. Otherwise, the stress test would yield unrealistic results from which no meaningful measures could be derived.

The specific need for stress tests in lending operations can be illustrated by the following experiences with historical crisis events:

Changes in correlations
One of the primary objectives of credit portfolio management is to diversify the portfolio, thereby making it possible minimize risks under normal economic conditions. However, past experience has shown that generally applicable correlations are no longer valid under crisis conditions, meaning that even a well diversified portfolio may suddenly exhibit high concentration risks. The credit portfolio's usual risk measurement mechanisms are therefore not always sufficient and have to be complemented with stress tests.

Rapid Propagation of Crisis Situations
In recent years, the efficiency of markets has increased substantially due to the introduction of modern communication technologies and the globalization of financial markets. However, these developments have also served to accelerate the propagation of crisis situations on the financial markets, which means that banks may no longer be capable of responding to such situations in a timely manner. Stress tests draw attention to risks arising under extraordinary conditions and can be used to define countermeasures well in advance. These measures can then be taken quickly if a crisis situation should actually arise.

Therefore, stress tests should always be regarded as a necessary complement to a bank's other risk management tools (e.g. rating systems, credit portfolio models). Whereas stress tests can help a bank estimate its risk in certain crisis situations, the other risk management tools support risk-based credit portfolio management under "normal" business conditions. In addition, Basel II requires IRB banks to perform stress tests for the purpose of assessing capital adequacy.[113] The procedure described here, however, goes well beyond the objectives and requirements of Basel II.

Section 6.4.2 describes the main characteristics a stress test should have, after which we present a general procedure for developing stress tests in section 6.4.3.

6.4.2 Essential Factors in Stress Tests

The essential factors in the development and application of stress tests are as follows:
- Consideration of portfolio composition and general conditions
- Completeness of risk factors included in the model
- Extraordinary changes in risk factors
- Acceptance
- Reporting
- Definition of countermeasures
- Regular updating
- Documentation and approval

Consideration of Portfolio Composition and General Conditions
As stress tests serve to reveal portfolio-specific weaknesses, it is important to keep the composition of the institution's individual credit portfolio in mind when developing stress tests.

[113] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 34.*

In order to ensure plausibility, as many internal and external experts as possible from various professional areas should participate in the development of stress tests.

Completeness of Risk Factors Included in the Model
Past experience has shown that when crisis situations arise, multiple risk factors tend to show clearly unfavorable changes at the same time. Comprehensive and realistic crisis scenarios will thus include simultaneous changes in all essential risk factors wherever possible. Stress tests which consider the effects of a change in only one risk factor (one-factor stress tests) should only be performed as a complement for the analysis of individual aspects. (For an example of how risk factors can be categorized, please see chart 78)

Extraordinary Changes in Risk Factors
Stress tests should only measure the effects of large-scale and/or extraordinary changes in risk factors. The bank's everyday risk management tools can (and should) capture the effects of "normal" changes.

Acceptance
In order to encourage management to acknowledge as a sensible tool for improving the bank's risk situation as well, stress tests primarily have to be plausible and comprehensible. Therefore, management should be informed as early as possible about stress tests and – if possible – be actively involved in developing these tests in order to ensure the necessary acceptance.

Reporting
Once the stress tests have been carried out, their most relevant results should be reported to the management. This will provide them with an overview of the special risks involved in credit transactions. This information should be submitted as part of regular reporting procedures.

Definition of Countermeasures
Merely analyzing a bank's risk profile in crisis situations is not sufficient. In addition to stress-testing, it is also important to develop potential countermeasures (e.g. the reversal or restructuring of positions) for crisis scenarios. For this purpose, it is necessary to design sufficiently differentiated stress tests in order to enable targeted causal analyses for potential losses in crisis situations.

Regular Updating
As the portfolio's composition as well as political and economic conditions can change at any time, stress tests have to be adapted to the current situation on an ongoing basis in order to identify and evaluate changes in the bank's risk profile in a timely manner.

Documentation and Approval
The objectives, procedures, responsibilities, and all other aspects associated with stress tests have to be documented and submitted for management approval.

6.4.3 Developing Stress Tests

This section presents one possible procedure for developing and performing stress tests. The procedure presented here can be divided into six stages:

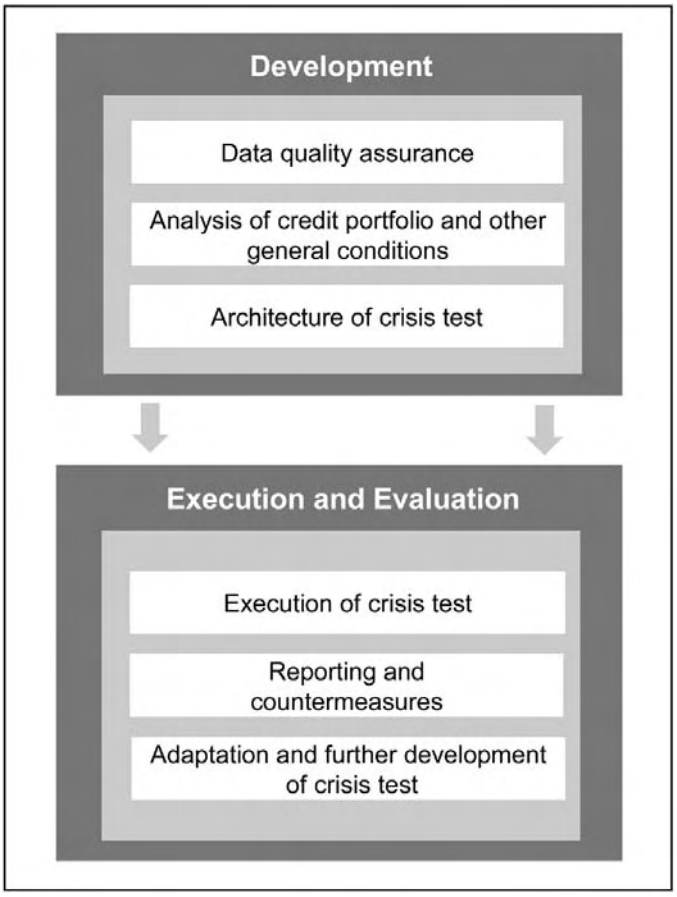

Chart 77: Developing and Performing Stress Tests

Step 1: Ensuring Data Quality

One basic prerequisite for successful stress-testing is a high-quality data set. Only when the data used are accurate and up to date can stress tests yield suitable results from which effective countermeasures can be derived.

For example, it is crucial to ensure that ratings are always up to date and valid. If this is not the case, the borrowers' creditworthiness (and thus also the corresponding PD) may change substantially without the knowledge of the credit institution. The stress test would then be based on an outdated risk situation, and would thus be unable to generate meaningful forecasts under crisis conditions.

Other important credit portfolio data include the outstanding volume of each credit facility, the interest rate, as well as any available collateral.

Using inaccurate or dated collateral values, for example, can also distort the risk situation. This is precisely the case when excessively high collateral values (which cannot be attained in the case of realization) are entered.

Important market data which may simultaneously represent risk factors include interest rates, exchange rates and stock indices. This information is

required in particular to valuate trading book positions which involve credit risk.

If the bank uses credit portfolio models, the data quality underlying the default rate volatility and correlations between individual credit facilities or borrowers is especially important.

Step 2: Analyzing the Credit Portfolio and Other General Conditions
One essential characteristic of a reliable stress test is the inclusion of an institution's individual credit portfolio composition as well as the prevailing political and economic conditions (see section 6.4.2).

For this reason, it is first necessary to compile a list of the credit products currently in use and to supplement the list with potential new credit products as well. The decisive risk factors should be identified for each individual credit product.

It is then necessary to sort the factors by relevance, and to group those risk factors which influence each other strongly under normal conditions or in crisis situations.

These groups make it possible to "stress" not only individual risk factors but all relevant factors simultaneously in the development of stress tests.

In the next step, it is necessary to analyze the prevailing social, economic, and political conditions and to filter as many potential crisis situations as possible out of this analysis. For this purpose, it is important to use in-house as well as external expertise. In particular, it is crucial to include the bank's own experts from various areas and hierarchical levels in order to ensure that the stress tests attain the necessary level acceptance. This will facilitate any later implementation of potentially drastic countermeasures resulting from stress tests.

Possible risk factor types which may arise from the analyses mentioned above are presented in chart 78. This presentation is meant to serve as a guide for a bank's individual design of stress tests and can be expanded as necessary.

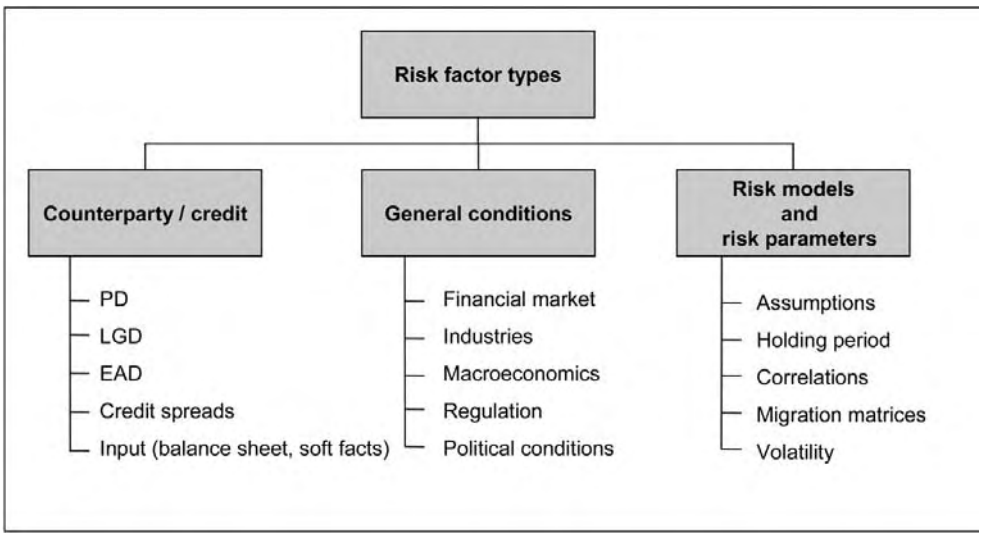

Chart 78: Risk Factor Types

Counterparty-based and credit facility-based risk factors: These scenarios can be realized with relative ease by estimating credit losses after modeling a change in PD and/or LGD/EAD. The methods of modeling stress tests include the following examples:

– Downgrading all borrowers by one rating class
– Increasing default probabilities by a certain percentage
– Increasing LGD by a certain percentage
– Increasing EAD by a certain percentage for variable credit products (justification: customers are likely to utilize credit lines more heavily in crisis situations, for example)
– Assumption of negative credit spread developments (e.g. parallel shifts in term structures of interest rates) for bonds
– Modeling of input factors (e.g. balance sheet indicators)

The approaches listed above can also be combined with one another as desired in order to generate stress tests of varying severity.

With regard to *general conditions*, examples might include stress tests for specific industries or regions. Such tests might involve the following:

– Downgrading all borrowers in one or more crisis-affected industries
– Downgrading all borrowers in one or more crisis-affected regions

Macroeconomic risk factors include interest rates, exchange rates, etc. These factors should undergo stress-testing especially when the bank uses them as the basis for credit risk models which estimate PD or credit losses. If the bank uses models, these stress tests are to be performed by adjusting the parameters and then recalculating credit losses.

Examples include:

– Unfavorable changes (increases/decreases, depending on portfolio composition) in the underlying interest rate by a certain number of basis points
– Unfavorable changes (increases/decreases, depending on portfolio composition) in crucial exchange rates by a certain percentage

It is particularly important to examine political risk factors when significant parts of the credit portfolio consist of borrowers from politically unstable countries. Due to the complex interrelationships involved, however, developing plausible stress tests for political risk factors involves far more effort than designing tests for macroeconomic risk factors, for example. It is therefore advisable to call in specialists to develop stress tests for political risk factors in order to assess the relevant effects on financial and macroeconomic conditions.

If the bank uses *risk models* (such as credit portfolio models or credit pricing models), it is necessary to perform stress tests which show whether the assumptions underlying the risk models will also be fulfilled in crisis situations. Only then will the models be able to provide the appropriate guidance in crisis situations as well.

Other risk model-related stress tests might focus on risk parameters such as correlations, transition matrices, and default rate volatilities. In particular, it appear sensible to use different correlation parameters in stress tests because

past experience has shown that the historical average correlations assumed under general conditions no longer apply in crisis situations.

Research has also proven that macroeconomic conditions (economic growth or recession) have a significant impact on transition probabilities.[114] For this reason, it makes sense to develop transition matrices under the assumption of certain crisis situations and to re-evaluate the credit portfolio using these transition matrices.

Examples of such crisis scenarios include the following:

– Increasing the correlations between individual borrowers by a certain percentage
– Increasing the correlations between individual borrowers in crisis-affected industries by a certain percentage
– Increasing the probability of transition to lower rating classes and simultaneously decreasing the probability of transition to higher rating classes
– Increasing the volatility of default rates by a certain percentage

The *size of changes* in risk factors for stress tests can either be defined by subjective expert judgment or derived from past experience in crisis situations.

When past experience is used, the observation period should cover at least one business cycle and as many crisis events as possible. Once the time interval has been defined, it is possible to define the amount of the change in the risk factor as the difference between starting and ending values or as the maximum change within the observation period, for example.

Step 3: Architecture of Stress Tests

On the basis of the previous analyses, it is possible to use one of the structures shown in chart 79 for the stress test.

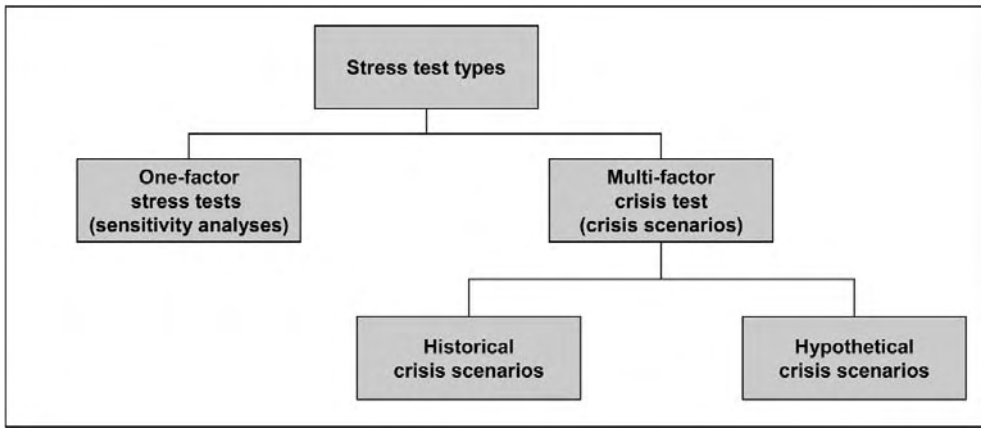

Chart 79: Systematic Overview of Stress Tests

One-factor stress tests measure the effect of drastic changes in individual risk factors on certain credit positions and the credit portfolio. When actual crisis events occur, however, multiple risk factors are always affected at the same time. Therefore, due to their lack of plausibility one-factor stress tests are only suitable to a limited extent. However, these stress tests can help the bank iden-

[114] *See BANGIA, ANIL/DIEBOLD, FRANCIS X./ SCHUERMANN, TIL, Ratings Migration and the Business Cycle.*

tify decisive factors influencing individual positions and to elucidate interrelationships more effectively.

Multi-factor stress tests attempt to simulate reality more closely and examine the effects of simultaneous changes in multiple risk factors. This type of stress test is also referred to as a scenario stress test.

Scenarios can be designed either top-down or bottom-up. In the top-down approach, a crisis event is assumed in order to identify its influence on risk factors. The bottom-up approach involves direct changes in the risk factors without assuming a specific crisis event.

However, what is more decisive in the development of a multi-factor stress test is whether the risk factors and the accompanying assumed changes are developed on the basis of past experience (historical crisis scenarios) or hypothetical events (hypothetical crisis scenarios).

Historical crisis scenarios offer the advantage of enabling the use of historical changes in risk factors, thus ensuring that all relevant risk factors are taken into account and that the assumed changes are plausible on the basis of past experience.

In this context, the primary challenge is to select scenarios which are suited to the credit portfolio and also applicable to the potential changes in general conditions. Ultimately, no one crisis will ever be identical to another, which means that extreme caution is required in the development of multi-factor stress tests based on past experience.

Using *hypothetical crisis situations* is especially appropriate when the available historical scenarios do not fit the characteristics of the credit portfolio, or when it is desirable to examine the effects of new combinations of risk factors and their changes.

In the construction of hypothetical crisis scenarios, it is especially important to ensure that no relevant risk factors are omitted and that the simultaneous changes in risk factors are sensible, comprehensible and plausible in economic terms.

The main challenge in constructing these crisis scenarios is the fact that the number of risk factors to be considered can be extremely high in a well-diversified portfolio. Even if it is possible to include all relevant risk factors in the crisis scenario, a subjective assessment of the interrelationships (correlations) between the changes in individual risk factors is hardly possible.

For this reason, hypothetical crisis scenarios can also be developed systematically with various mathematical tools and methods.[115]

6.4.4 Performing and Evaluating Stress Tests

Once the crisis scenarios, the accompanying risk factors, and the size of the changes in those factors have been defined, it is necessary to re-evaluate the credit portfolio using these scenarios. If the bank uses quantitative models, it can perform the stress tests by adjusting the relevant input factors (risk factors).

If no quantitative models exist, it is still possible to perform stress tests. However, in such cases they will require greater effort because the effects on the credit portfolio can only be estimated roughly in qualitative terms.

[115] *See MONETARY AUTHORITY OF SINGAPORE, Credit Stress-Testing, p. 42–44.*

Reporting and Countermeasures
Once the stress tests have been carried out, it is necessary to report the results to the relevant levels of management. In this context, it is crucial to present only those results which are truly decisive. Such reports should cover the results of routine stress tests as well as those of new tests based specifically on the prevailing economic situation. The targeted selection of decisive results will also facilitate the process of developing countermeasures. In order to derive countermeasures from stress tests, the tests have to be designed in such a way that they enable causal analysis.

Adaptation and Ongoing Development of Stress Tests
As the portfolio composition as well as economic and political conditions change constantly, it is also necessary to adapt stress tests on an ongoing basis. This point is decisive in ensuring plausible results from which suitable countermeasures can be derived.

III ESTIMATING AND VALIDATING LGD/EAD AS RISK COMPONENTS

7 Estimating Loss Given Default (LGD)

While it is common practice in many banks to calculate PDs without using the results to calculate Basel II regulatory capital requirements, these institutions are now also focusing on estimating LGD and EAD due to the requirements of Basel II. It is not possible to discuss the estimation of these two parameters without referring to Basel II, as no independent concepts have been developed in this field to date. Therefore, institutions which plan to implement their own LGD estimation procedures in compliance with the draft EU directive face a number of special challenges. First, in contrast to PD estimates, LGD estimation procedures cannot rely on years of practical experience or established industry standards. Second, many institutions do not have comprehensive loss databases at their disposal.

This chapter presents potential solutions for banks' in-house estimation of LGD as well as the current state of development in this area. We cannot claim that this chapter presents a conclusive discussion of LGD estimation or that the procedure presented is suitable for all conceivable portfolios. Instead, the objective of this chapter is to encourage banks to pursue their own approaches to improving LGD estimates.

The LGD estimation procedure is illustrated in chart 80 below.

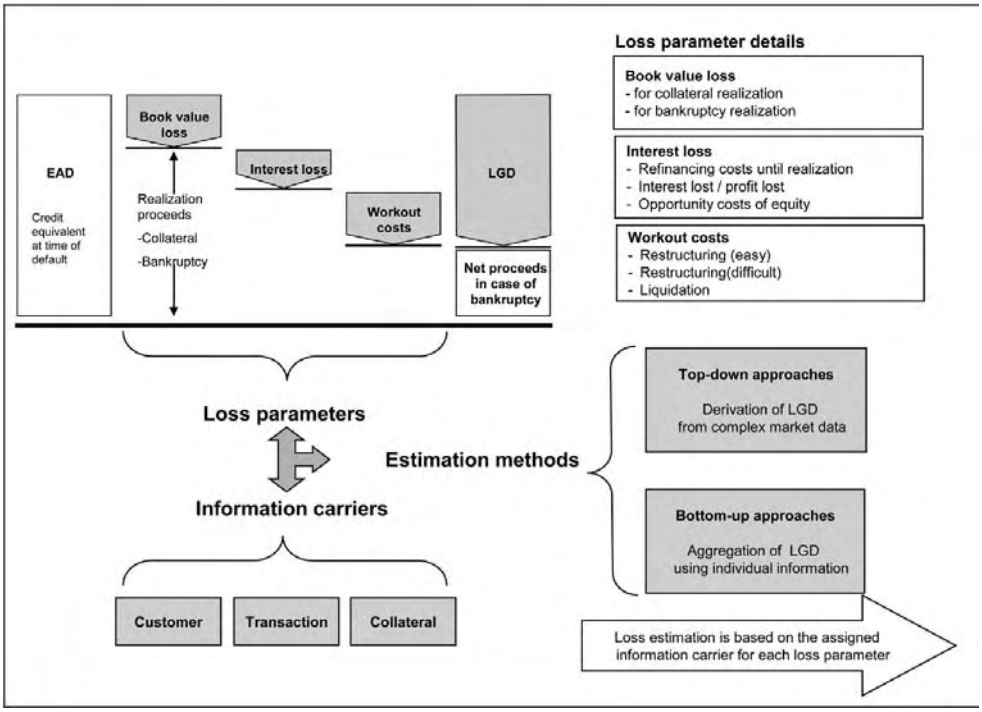

Chart 80: LGD Estimation Procedure

In this chapter, we derive the loss parameters on the basis of the definitions of default and loss presented in the draft EU directive and then link them to the main segmentation variables identified: customers, transactions, and collateral. We then discuss procedures which are suitable for LGD estimation. Finally, we

present a number of approaches to implementing LGD estimation methodologies on this basis.

7.1 Definition of Loss

A clear definition of default is a basic prerequisite for estimating loss given default. The Basel II-compliant definition of default used for calculating PD (see section 5.1.2) also applies in this context.[116]

The second major prerequisite is a definition of the term "loss." Loss given default has been defined in various ways in practice and in the literature to date. The draft EU directive creates a uniform basis with its definition of loss in LGD estimation:

For the purpose of LGD estimation, the term "loss" shall mean economic loss. The measurement of economic loss should take all relevant factors into account, including material discount effects, and material direct and indirect costs associated with collecting on the instrument.[117]

For LGD estimation, the use of this definition means that it is also necessary to take losses arising from restructured credit facilities into account (in addition to liquidated credit facilities). These facilities generally involve a lower level of loss than liquidated facilities. Out of business considerations, banks only opt for restructuring (possibly with a partial write-off) if the probable loss in the case of successful restructuring is lower than in the case of liquidation. Accordingly, taking only liquidated facilities into account in LGD estimation would lead to a substantial exaggeration of loss. For this reason, it is crucial to consider all defaulted credit facilities (including facilities recovered from default), especially in the data collection process.

7.2 Parameters for LGD Calculation

Based on the definition of loss given above, this section identifies the relevant loss components which may be incurred in a credit default. These loss components form the basis for LGD estimation. Depending on the type of liquidation or restructuring, not all loss components will be relevant.

As the draft EU directive allows pooling in the retail segment, the loss parameters are discussed separately for retail and non-retail segments.

7.2.1 LGD-Specific Loss Components in Non-Retail Transactions

The essential components of loss are the amount receivable to be written off after realization, interest loss, and liquidation costs. The relationship between EAD, LGD and the individual loss components is presented in chart 81.

For further consideration of the loss components, we recommend a cash flow-based perspective. In such a perspective, any further payments are regarded as costs and payments received (recoveries) on the basis of EAD. Additional payments received essentially refer to the part of the amount receivable which has not yet been written off and for which a corresponding return flow of funds is expected. The underlying time periods for the lost payments either result from the originally agreed interest and principal repayment dates or have

[116] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 1, No. 46, and Annex D-5, No. 43.*

[117] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 1, No. 47.*

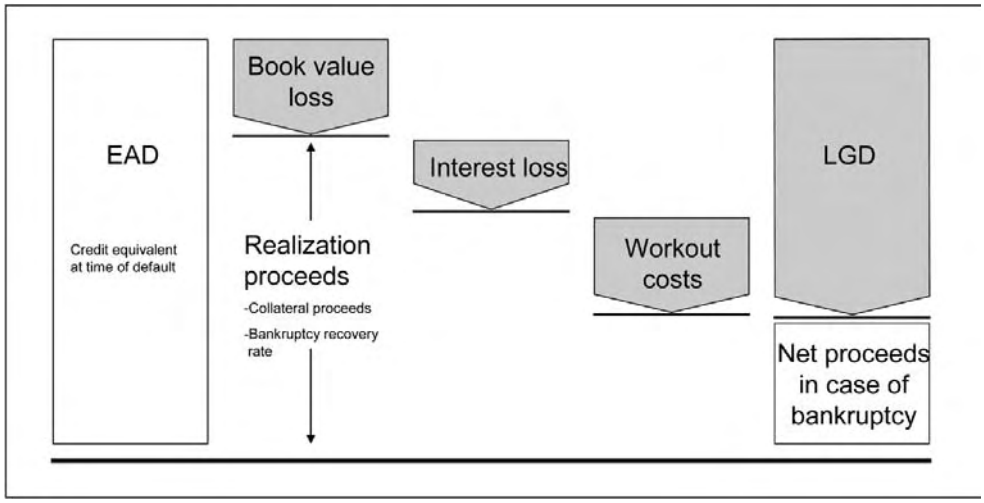

Chart 81: Loss Components in LGD

to be estimated explicitly, as in the case of recoveries. The selection of the discounting factor or factors depends on the desired level of precision. Chart 82 illustrates this cash flow-based perspective.

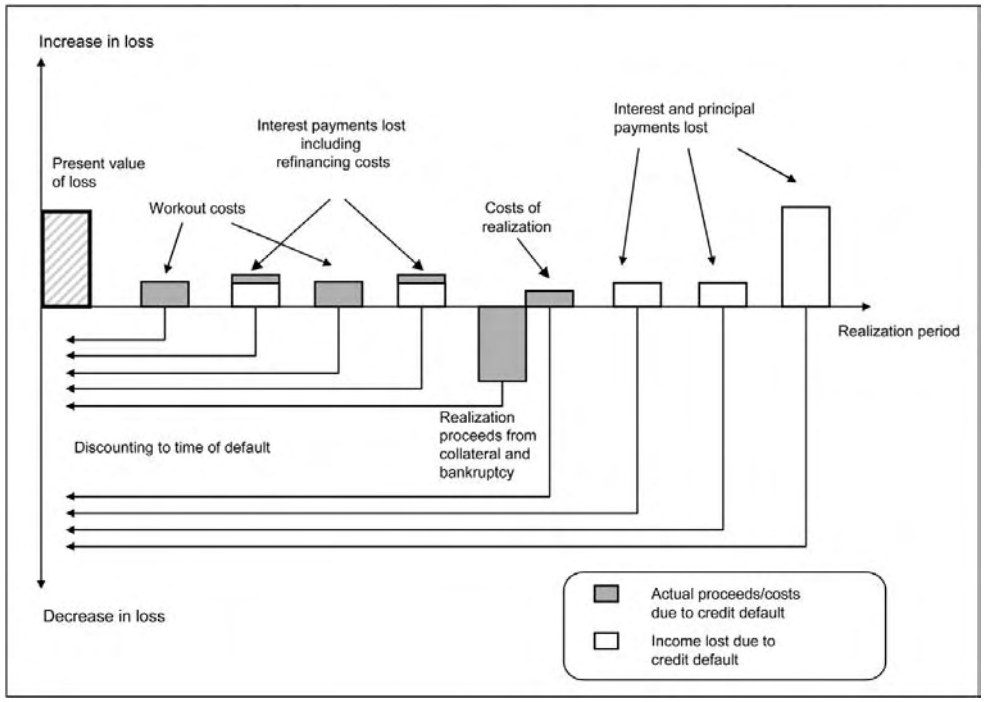

Chart 82: LGD in the Cash Flow-Based Perspective

Book Value Loss/Recoveries

When calculating book value loss (i.e. the amount receivable), it is necessary to differentiate between restructuring and liquidation. In the case of restructuring, the book value loss results from a partial write-off, and in the case of liquidation this loss is equal to EAD less recoveries. As assets are not realized in the course of loan restructuring, the amount of a partial write-off can vary widely and can-

not be calculated on the basis of expected recoveries. From the business perspective, restructuring generally only makes sense in cases where the loss due to the partial write-off is lower than in the case of liquidation. The other loss components notwithstanding, a partial write-off does not make sense for cases of complete collateralization, as the bank would receive the entire amount receivable if the collateral were realized. Therefore, the expected book value loss arising from liquidation is generally the upper limit for estimates of the partial write-off.

As the book value loss in the case of liquidation is merely the difference between EAD and recoveries, the actual challenge in estimating book value loss is the calculation of recoveries. In this context, we must make a fundamental distinction between realization in bankruptcy proceedings and the realization of collateral. As a rule, the bank has a claim to a bankruptcy dividend unless the bankruptcy petition is dismissed for lack of assets. In the case of a collateral agreement, however, the bank has additional claims which can isolate the collateral from the bankruptcy estate if the borrower provided the collateral from its own assets. Therefore, collateral reduces the value of the bankruptcy estate, and the reduced bankruptcy assets lower the recovery rate. In the sovereigns/central governments, banks/institutions and large corporates segments, unsecured loans are sometimes granted due to the borrower's market standing or the specific type of transaction. In the medium-sized to small corporate customer segment, bank loans generally involve collateral agreements. In this customer segment, the secured debt capital portion constitutes a considerable part of the liquidation value, meaning that the recovery rate will tend to take on a secondary status in further analysis.

A large number of factors determine the amount of the respective recoveries. For this reason, it is necessary to break down recoveries into their essential determining factors:

Recoveries from Collateral		**Recoveries from Bankruptcy**	
Value of collateral at time of default	Market or nominal value of collateral	Value of enterprise at time of default	Liquidation value of the enterprise
Collateral recovery rate	Liquidation period	Bankruptcy recovery rate	Duration of bankruptcy proceedings
	Market value of collateral at time of realization		Liquidation value of enterprise at time of realization
	Realization costs		Bankruptcy costs
	Markdown on market price due to forced sale		Markdown on market price due to forced sale

Chart 83: Factors Determining Recoveries

The point of departure in estimating recoveries is the identification of the assessment base. This is the collateral value in the case of collateral realization and the liquidation value of the enterprise in the case of bankruptcy proceedings. In the first step, it is necessary to estimate these values at the time of default, as at that time the bank or bankruptcy administrator receives the power of disposal over the assets to be realized. In this context, it is necessary to mark down the collateral value at the time of default, especially for tangible fixed

assets, as the danger exists that measures to maintain the value of assets may have been neglected before the default due to liquidity constraints.

As the value of assets may fluctuate during the realization process, the primary assessment base used here is the collateral value or liquidation value at the time of realization. As regards guarantees or suretyships, it is necessary to check the time until such guarantees can be exercised as well as the credit standing (probability of default) of the party providing the guarantee or surety.

Another important component of recoveries is the cost of realization or bankruptcy. This item consists of the direct costs of collateral realization or bankruptcy which may be incurred due to auctioneers' commissions or the compensation of the bankruptcy administrator.

It may also be necessary to discount the market price due to the limited liquidation horizon, especially if it is necessary to realize assets in illiquid markets or to follow specific realization procedures.

Interest Loss

Interest loss essentially consists of the interest payments lost from the time of default onward. In line with the analysis above, the present value of these losses can be included in the loss profile. In cases where a more precise loss profile is required, it is possible to examine interest loss more closely on the basis of the following components:

– Refinancing costs until realization
– Interest payments lost in case of provisions/write-offs
– Opportunity costs of equity

Workout Costs

In the case of workout costs, we can again distinguish between the processing costs involved in restructuring and those involved in liquidation. Based on the definition of default used here, the restructuring of a credit facility can take on various levels of intensity. Restructuring measures range from rapid renegotiation of the commitment to long-term, intensive servicing.

In the case of liquidation, the measures taken by a bank can also vary widely in terms of their intensity. Depending on the degree of collateralization and the assets to be realized, the scope of these measures can range from direct write-offs to the complete liquidation of multiple collateral assets. For unsecured loans and larger enterprises, the main emphasis tends to be on bankruptcy proceedings managed by the bankruptcy administrator, which reduces the bank's internal processing costs.

7.2.2 LGD-Specific Loss Components in Retail Transactions

Depending on the type and scope of a bank's retail business, it may be necessary to make a distinction between mass-market banking and private banking in this context. One essential characteristic of mass-market banking is the fact that it is possible to combine large numbers of relatively small exposures in homogenous groups and to treat them as pools. This perspective does not have to cover all of the specific characteristics of retail customers and transactions, which means that in the case of private banking customers it may be appropriate to view loss using approaches applied to non-retail customers.

With regard to pooling in mass-market banking, Basel II requires the following minimum segmentation:[118]

– Exposures secured by real estate
– Qualifying revolving retail exposures
– All other retail exposures

Beyond this basic segmentation, banks can also segment exposures according to additional criteria. In this context, it may also be sensible from a business prespective to subdivide exposures further by product type and degree of collateralization. One essential requirement is that each group consists of a large number of homogenous exposures.[119]

Due to the pooling of exposures, specific transactions – and thus also their potential recoveries – can no longer be regarded individually. Accordingly, pooling makes it possible to apply the pool's historical book value loss percentage, which applies equally to all transactions in a pool. This is also the case for the two other loss parameters (interest loss and processing costs). The table below gives an overview of the loss components relevant to mass-market banking.

Loss Parameter for Retail Pool	Components
Book value loss	Amount receivable written off
Interest loss	Refinancing costs of defaulted loans until realization
	Interest payments lost due to provisions/write-offs
	Opportunity costs of equity
Workout costs	Costs of direct write-offs/easy renegotiation
	Restructuring/liquidation costs

Chart 84: Loss Components in Mass-Market Banking

7.3 Identifying Information Carriers for Loss Parameters

For the purpose of selecting and assessing individual methods of estimating LGD, it is necessary to identify the main information carriers for each loss parameter. Breaking down loss parameters into their separate components enables direct assignment and at the same time reduces complexity in the application of estimation methods.

7.3.1 Information Carriers for Specific Loss Parameters

Non-Retail Information Carriers

The following information carriers are relevant to the loss parameters in non-retail business:

– *Customers:* Creditworthiness information, assigned collateral and transactions, customer master data (customer type, industry, region, etc.)

[118] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 47, No. 7.*
[119] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Article 47, No. 5.*

– *Collateral:* Collateral value, collateral master data (collateral type, collateral provider, etc.),
– *Transactions:* Book value, assigned collateral, transaction master data (product type, interest rate, repayment structure, etc.).

The table below shows how information carriers are assigned to loss parameters.

Loss Parameter	Components	Information Carrier
Value of enterprise at default	Liquidation value of the enterprise	Customer
Bankruptcy recoveries	Liquidation value of enterprise at time of realization	Customer
	Duration of bankruptcy proceedings	Customer
	Bankruptcy costs	Customer
	Markdown on market price due to forced sale	Customer
Collateral value at default	Market or nominal value of collateral	Collateral
Collateral recoveries	Market value of collateral at time of realization	Collateral
	Liquidation period	Collateral
	Realization costs	Collateral
	Markdown on market price due to forced sale	Collateral
Interest loss	Refinancing costs of defaulted loan until realization	Transaction
	Interest payments lost for provisions/write-offs	Transaction
	Opportunity costs of equity	Transaction
Workout costs	Administrative costs	Transaction/customer
	Restructuring costs	Transaction/customer
	Liquidation costs	Transaction/customer

Chart 85: Assignment of Information Carriers to Loss Parameters (Non-Retail)

For the purpose of estimating the recoveries from bankruptcy, it is necessary to use customer-specific data. Based on the customer type and country of domicile, the bank can estimate whether bankruptcy proceedings will bring in any relevant payments, for example. The customer's industry and creditworthiness may also enable conclusions as to the type of assets, their ability to be realized, and the realization period. Collateral can also provide information on whether realization in bankruptcy proceedings is of material relevance to the bank.

For the purpose of estimating the proceeds from collateral realization, collateral data can provide information as to the relevant means of realization, the realization period in the corresponding markets, as well as the volatility and liquidity of those markets.

Calculating interest loss requires transaction-specific information, such as the agreed interest rate. If more precise calculations are required, it is also possible to use the target costs calculated in contribution margin analysis. Moreover, market interest rates and bank-specific interest claims can also be included in these calculations.

It is possible to calculate processing costs and general expenses for transaction types as well as customer types according to the type and scope of cost accounting. Depending on the level of detail in cost center and/or cost unit

accounting, it is possible to calculate costs down to the level of individual transactions. In this way, for example, the costs of collateral realization can be assigned to individual transactions by way of the transaction-specific dedication of collateral. The costs of realizing customer-specific collateral can be allocated to the customer's transactions arithmetically or using a specific allocation key.

Retail Information Carriers

Unless the loss components in retail transactions are depicted as they are in the non-retail segment, the collective banking book account for the pool to which the transactions are assigned can serve as the main source of information. The segmentation of the pool can go beyond the minimum segmentation required by the draft EU directive and provide for a more detailed classification according to various criteria, including additional product types, main collateral types, degrees of collateralization or other criteria. For each of the loss parameters listed below, it is advisable to define a separate collective account for each pool.

Loss Parameter (Retail Pool)	Components	Information Carrier
Book Value Loss	Amount receivable written off per pool	Collective pool account
Interest loss	Refinancing costs of defaulted loan until realization	Collective pool account
	Interest payments lost for provisions/write-offs	Collective pool account
	Opportunity costs of equity	Collective pool account
Processing costs	Administrative costs	Collective pool account
	Restructuring costs	Collective pool account
	Liquidation costs	Collective pool account

Chart 86: Assignment of Information Carriers to Loss Parameters (Retail)

7.3.2 Customer Types

As customer types are required in order to calculate losses due to bankruptcy or composition proceedings, it is advisable to define further subdivisions for these types because the nature and scope of bankruptcy proceedings depend heavily on the customer type. The segmentation shown below is essentially based on the customer segmentation described for PD estimation earlier.

On the basis of this segmentation, it is already possible to derive the most essential information for further estimates of bankruptcy proceeds. In the case of corporate customers, the industry in which they operate provides important additional information with which the bank can estimate the value content and liquidity of assets, for example. Beyond the information mentioned above, lenders can only gain additional insight through a more detailed analysis of defaulted customers similar to the analysis performed in order to evaluate a company when determining its liquidation value.

Main Group	Subgroup	Individual Level
Governments and the public sector	Governments	
	Other public agencies	Regional governments
		Local authorities
		Public sector entities (PSEs)
Financial service providers	Credit institutions	
	Insurance companies	
	Financial institutions	
Corporate customers	Enterprises/business owners	Capital market-oriented / international companies
		Other companies which prepare balance sheets
		Businesses and independent professionals (not preparing balance sheets)
		Small businesses
		Start-ups
		NPOs (non-profit organizations)
	Specialized lending	Project finance
		Object finance
		Commodities finance
		Financing of Income-producing real estate
Retail	Mass-market banking	Current accounts
		Consumer loans
		Residential construction loans
	Private banking	

Chart 87: Overview of Customer Types

7.3.3 Types of Collateral

In order to determine collateral recoveries, it is advisable to categorize collateral using the following types:

Collateral Class	Collateral Category	Collateral Type	Collateral Value
Uncollateralized	Unsecured		-
Financial collateral	Cash	Cash balances	Nominal value
		Netting	Nominal value
Financial collateral	Securities	Equities	Market value
		Bonds	Market value
		Investment funds	Market value
		Precious metals and certificates	Market value
	Claims against financial service providers	Endowment policies	Nominal value
		Credit balances with other banks, including building and loan associations	Nominal value
Personal collateral	Guarantees	Deficiency suretyship	Nominal value
		Ordinary suretyship	Nominal value
		Guarantees	Nominal value
	Credit derivatives	Total return swaps	Nominal value
		Credit default swap	Nominal value
		Other credit derivatives	Nominal value
Physical collateral	Real estate	Residential buildings	Nominal value
		Office and commercial buildings	Nominal value
		Undeveloped properties	Nominal value
		Special-purpose buildings	Nominal value
	Other assets	Registered assets (ships, aircraft, etc.)	Nominal value
		Vehicles	Nominal value
		Machines	Nominal value
		Raw materials	Nominal value
		Finished products	Market value
Miscellaneous collateral	Trade accounts receivable		Nominal value
	Receivables from continuous obligations	Rentals/leasing	Nominal value – present value of payment stream
		Assignment of salary	Nominal value – present value of payment stream
	Rights	Patents	Market value
		Licenses	Market value
		Royalties	Nominal value or present value of payment stream
		Registered rights (life annuities, easements on real estate, etc.)	Market value
		Miscellaneous rights (e.g. film rights)	Market value
		Private equity	Market value

Chart 88: Overview of Collateral Types

The value of collateral forms the basis for calculating collateral recoveries and can either be available as a nominal value or a market value. Nominal values are characterized by the fact that they do not change over the realization period. If the collateral is denominated in a currency other than that of the secured

transaction, however, it is also necessary to account for changes in value due to exchange rate fluctuations.

As regards guarantees and credit derivatives, it is necessary to take the collateral provider's default probability into account explicitly. However, with regard to regulatory capital requirements under Basel II, it is possible to include the credit risk mitigation effect of guarantees and credit derivatives in PD instead of LGD estimates. In such cases, return flows of funds from such guarantees can no longer be included in LGD estimates.

In the case of market values, we can draw a distinction between financial collateral and physical collateral. Financial collateral is characterized by the fact that its value is generally independent of the borrower's creditworthiness and not subject to depreciation. Moreover, the markets for financial collateral are usually far more liquid than the realization markets for physical collateral, and extensive databases of historical market prices are available for this type of collateral.

Physical collateral can be differentiated according to various criteria. Two of the most important criteria are the liquidity of the relevant markets and the existence of a market price index. In addition, aspects such as susceptibility to poor maintenance, useful life, and smooth legal liquidation are particularly significant.

With regard to miscellaneous collateral, we can distinguish between saleable and non-saleable assets. Non-saleable collateral such as salary assignments or life annuities can only be realized by means of the underlying payment stream or by selling the goods received regularly over time. In such cases, the collateral's present value is subject to the same risks as the other forms of collateral, depending on whether the payment is based on the creditworthiness of a third party or on the future development of the collateral's value. As in the case of physical collateral, the recoveries from saleable assets will result from the market value of the rights.

7.3.4 Types of Transaction

Differentiating by transaction type allows a more detailed classification of losses according to the components interest loss and workout costs. In this context, banks generally distinguish the following types of credit facilities:

- Lines of credit
- Loans
- Consumer loans
- Leasing transactions
- Purchase of receivables
- Bonds in the banking book
- Guarantee credit

Transactions are further categorized by:

- Purpose (e.g. real estate loan, rental payment guarantee)
- Type of transaction (e.g. syndicated loans)
- Type of underlying transaction (e.g. acceptance credit)
- Degree of standardization (e.g. private banking/mass-market banking)
- Customer type (e.g. current account, start-up loan)
- Organizational units (e.g. project finance, ship finance)

This categorization will depend heavily on the organizational structure of each bank. Product catalogs frequently list various combinations of the categories mentioned above.

7.3.5 Linking of Collateral Types and Customer Types

In the table below, typical collateral types are mapped to individual customer groups.

	Customers			
Collateral	Governments Financial Service Providers Large Joint Stock Companies	SMEs Independent Professionals Start-Ups Private Banking	Specialized Lending	Mass-Market Banking
Unsecured	X	X		X
Cash balances		X	X	
Netting	X	X	X	
Equities	X	X	X	
Bonds	X	X	X	X
Investment funds		X		X
Gold	X	X	X	
Deficiency guarantees / suretyships	X	X	X	X
Other guarantees / suretyships	X	X	X	
Total return swaps	X		X	
Credit default swaps	X		X	
Other credit derivatives	X		X	
Endowment policies		X		X
Balances with building and loan associations		X		X
Miscellaneous credit balances with other banks		X		X
Residential buildings		X	X	X
Office buildings and commercial buildings	X	X	X	
Undeveloped properties	X	X	X	
Special-purpose buildings		X	X	
Registered assets (ships, etc.)		X	X	
Vehicles		X	X	X
Machines		X	X	
Specialized machines		X	X	
Raw materials		X	X	
Finished products		X		
Trade accounts receivable		X		
Rentals/leasing		X	X	X
Assignment of salary				X
Patents	X	X	X	
Licenses	X	X	X	
Royalties		X		
Registered rights		X		X
Misc. rights	X	X	X	
Private equity		X	X	

Chart 89: Typical Collateral for Various Customer Types

Implementing an in-house LGD estimation procedure should begin with an analysis of the collateral/customer combinations which are most significant for the individual bank. First of all, it is necessary to ensure that the basic prerequisite of a sufficient quantitative and qualitative data set is fulfilled. Depending on the type and extent of data requirements, collateral can be further subdivided according to its value content, the materiality of individual collateral types, as well as the complexity of collateral agreements. It may also be useful to define categories based on transaction types. On the basis of these preliminary tasks, it is then possible to select the appropriate LGD estimation method specifically for each loss parameter. These methods are presented in the next section.

7.4 Methods of Estimating LGD Parameters

In this section, we give a general presentation of the procedures currently being discussed and/or used in LGD estimation. Section 7.5 then gives specific applied examples of how the loss components are estimated in practice.

In general, we can differentiate top-down and bottom-up approaches to LGD estimation.

7.4.1 Top-Down Approaches

Top-down approaches use freely available market data to derive LGD by cleansing or breaking down the complex external information available, such as recovery rates or expected loss. This can be done in two ways:

– Using explicit loss data
– Using implicit loss data

For the sake of completeness, it is worth noting here that the draft EU directive mentions another method in addition to these two methods of calculating LGD from external data. For purchased corporate receivables and in the retail segment, LGD can also be estimated on the basis of internally available loss data (expected loss) if suitable estimates of default probability are possible.

Explicit Loss Data

There are two possible ways to use explicit loss data. The first possibility is to use historical loss information (such as recovery rates for bonds) provided by specialized agencies. The recovery rate corresponds to the insolvency payout and indicates the amount reimbursed in bankruptcy proceedings as a percentage of the nominal amount receivable. LGD can then be calculated using the following formula:

$$LGD = 1 - Recovery\ Rate$$

Historical recovery rates are currently available in large quantities, predominantly for US bonds and corporate loans. In addition, loss data are also available on banks and governments. Even if the definitions of loss and default are consistent, these data probably only apply to borrowers from Austrian banks to a limited extent. For example, aspects such as collateralization, bankruptcy procedures, balance sheet structures, etc. are not always comparable. Therefore, historical recovery rates from the capital market are generally only suitable for unsecured transactions with governments, international financial service providers and large international companies.

Moreover, due to the definition of loss used here, the relation "LGD = 1 – recovery rate" is not necessarily ensured. Of all loss parameters, only the book value loss is completely covered by this relation. With regard to interest loss, the remarks above only cover the interest owed after the default, as this interest increases the creditor's claim. The other components of interest loss are specific to individual banks and are therefore not included. As regards workout costs, only the costs related to bankruptcy administration are included; accordingly, additional costs to the bank are not covered in this area. These components have to be supplemented in order to obtain a complete LGD estimate.

The second possibility involves the direct use of secondary market prices. The established standard is the market value 30 days after the bond's default. In this context, the underlying hypothesis is that the market can already estimate actual recovery rates at that point in time and that this manifests itself accordingly in the price. Uncertainty as to the actual bankruptcy recovery rate is therefore reflected in the market value.

With regard to market prices, the same requirements and limitations apply to transferability as in the case of recovery rates. Like recovery rates, secondary market prices do not contain all of the components of economic loss. Secondary market prices include an implicit premium for the uncertainty of the actual recovery rate. The market price is more conservative than the recovery rate, thus the former may be preferable.

Implicit Loss Data

When implicit loss data are used, LGD estimates are derived from complex market information on the basis of a verified or conjectured relationship between the underlying data and LGD. In this context, it is possible to utilize not only information on defaulted loans but also data on transactions which are not overdue. The two best-known data elements are credit risk spreads and ratings. When credit risk spreads are used, the assumption is that the spread determined between the yield of a traded bond and the corresponding risk-free interest rate is equal to the expected loss (EL) of the issue. If PD is known, it is then possible to calculate LGD using the equation EL (%) = PD * LGD. This requires that PD can be calculated without ambiguity. If an external or internal rating is available, it can be assigned to a PD or PD interval. As external ratings are sometimes equal to EL, it is necessary to ensure that the rating used has a clear relationship to PD, not to EL.

In the derivation of LGD from credit risk spreads, the same general requirements and limitations apply as in the case of explicit data. Capital market data are only available on certain customer groups and transaction types, thus the transferability of data has to be reviewed in light of general economic conditions. In addition, it must be possible to extract the implicit information contained in these data. If the credit risk spread corresponds to EL, it is to be quantified as EL. The varying liquidity of markets in particular makes this more difficult. Moreover, an unambiguous PD value has to be available, that is, any ratings used to derive PD must not contain implicit LGD aspects as well. This means that the ratings have to be pure borrower ratings, not transaction ratings. Naturally, credit risk spreads do not contain bank-specific loss compo-

nents; this is analogous to the use of explicit loss data from secondary market prices (see previous section).

In top-down approaches, one of the highest priorities is to check the transferability of the market data used. This also applies to the consistency of default and loss definitions used, in addition to transaction types, customer types, and collateral types. As market data are used, the information does not contain bank-specific loss data, thus the resulting LGDs are more or less incomplete and have to be adapted according to bank-specific characteristics.

In light of the limitations explained above, using top-down approaches for LGD estimation is best suited for unsecured transactions with governments, international financial service providers and large capital market companies.

7.4.2 Bottom-Up Approaches

Bottom-up approaches involve compressing specific information on the three loss parameters into an LGD value. This analysis is based on the assumption of various scenarios describing how the exposure will develop after the default. Possible scenarios include:

- Complete servicing of the outstanding debt, possible renegotiation of terms and returning the loan's status from "defaulted" to the "normal" range
- Restructuring of the loan with creation of partial provisions
- Liquidation of the loan with realization of collateral
- Liquidation of the loan without collateral

The following methods are currently being implemented or discussed for the individual loss parameters in LGD estimation:

Loss Parameter	Information Carrier	Method of direct/ simplified estimation	Components of Loss Parameters	Method of indirect/detailed estimation
Book value – Realized recovery rate	Customer	Estimation using loss data	Liquidation value of the enterprise at default	Calculation of liquidation value by company valuation based on net value of tangible assets
			Bankruptcy costs	Expert estimation
			Markdown on market price due to forced sale	Expert estimation
Book value – Recovery rate for collateral realization	Collateral	Estimation using loss data	Value of collateral at time of default	Estimation using loss data, cash flow model, realization-based expert valuation of collateral
			Liquidation period	Expert estimation, estimation using loss data
			Realization costs	Expert estimation, estimation using loss data
			Markdown on market price due to forced sale	Expert estimation, estimation using loss data
Book value – Partial write-off (restructuring)	Customer	Estimation using loss data		
Interest loss	Transaction	Calculation of lost interest	Refinancing costs until realization	Contribution margin analysis, financial calculations
			Lost interest	Contribution margin analysis, financial calculations
			Opportunity costs	Contribution margin analysis, financial calculation
Workout costs	Customer/ transaction	Expert estimation, cost and activity accounting	Direct write-off/ easy restructuring	Expert estimation, cost and activity accounting
			Medium restructu-ring/liquidation	Expert estimation, cost and activity accounting
			Difficult restructu-ring/liquidation	Expert estimation, cost and activity accounting

Chart 90: Loss Parameters and Estimation Methods

LGD calculation is based on estimates of individual loss parameters. In this process, it is necessary to differentiate at least by individual customer type, collateral type, and transaction type according to the given level of materiality. If the existing loss history does not allow the direct estimation of parameters, it is possible to estimate individual loss parameters on the basis of their individual components (see table above). For each loss parameter, we explain the application of these methods in greater detail below.

Book Value Loss

Book value losses can arise in the course of restructuring due to a partial write-off or in the course of liquidation. The liquidation itself can be based on bankruptcy realization or the realization of collateral.

In the case of realization in *bankruptcy/composition proceedings*, historical default data are used for the purpose of direct estimation. In order to reduce the margin of fluctuation around the average recovery rate, it is advisable to perform segmentation by individual customer type. In the case of corporate customers, additional segmentation by industry may be helpful in order to account for their typical assets structure and thus also their probable recovery rates. As

those assets which serve as collateral for loans from third parties are isolated from the bankruptcy estate, the recovery rate will be reduced accordingly for the unsecured portion of the loan. For this reason, it is also advisable to further differentiate customers by their degree of collateralization from balance sheet assets.[120]

If possible, segments should be defined on the basis of statistical analyses which evaluate the discriminatory power of each segmentation criterion on the basis of value distribution. If a meaningful statistical analysis is not feasible, the segmentation criteria can be selected on the basis of expert decisions. These selections are to be justified accordingly.

If historical default data do not allow direct estimates on the basis of segmentation, the recovery rate should be excluded from use at least for those cases in which collateral realization accounts for a major portion of the recovery rate. If the top-down approach is also not suitable for large unsecured exposures, it may be worth considering calculating the recovery rate using an alternative business valuation method based on the net value of tangible assets. Appropriately conservative estimates of asset values as well as the costs of bankruptcy proceedings and discounts for the sale of assets should be based on suitable documentation.

When estimating LGD for a *partial write-off* in connection with loan restructuring, it is advisable to filter out those cases in which such a procedure occurs and is materially relevant. If the available historical data do not allow reliable estimates, the same book value loss as in the bankruptcy proceedings should be applied to the unsecured portion of the exposure.

In the case of *collateral realization,* historical default data are used for the purpose of direct estimation. Again, it is advisable to perform segmentation by collateral type in order to reduce the margin of fluctuation around the average recovery rate (see section 7.3.3).

In the case of personal collateral, the payment of the secured amount depends on the creditworthiness of the collateral provider at the time of realization. This information is implicit in historical recovery rates. In order to differentiate more precisely in this context, it is possible to perform segmentation based on the ratings of collateral providers. In the case of guarantees and credit derivatives, the realization period is theoretically short, as most contracts call for payment at first request. In practice, however, realization on the basis of guarantees sometimes takes longer because guarantors do not always meet payment obligations immediately upon request. For this reason, it may be appropriate to differentiate between institutional and other guarantors on the basis of the bank's individual experience.

In the case of securities and positions in foreign currencies, potential value fluctuations due to market developments or the liquidity of the respective market are implicitly contained in the recovery rates. In order to differentiate more

[120] *For example, this can be done by classifying customers and products in predominantly secured and predominantly unsecured product/customer combinations. In non-retail segments, unsecured transactions tend to be more common in the case of governments, financial service providers and large capital market-oriented companies. The company's revenues, for example, might also serve as an alternative differentiating criterion. In the retail segment, for example, unsecured transactions are prevalent in standardized business, in particular products such as credit cards and current account overdraft facilities. In such cases, it is possible to estimate book value loss using the retail pooling approach.*

precisely in this context, it may be advisable to consider segmentation based on the historical volatility of securities as well as market liquidity.

The recovery rates for physical collateral implicitly contain the individual components (collateral value at time of default, realization period, realization costs and markdown on market price for illiquid markets). In order to improve discriminatory power with regard to the recovery rate for each segment, it is advisable to perform further segmentation based on these components. The definition of segments should be analogous to the selection of segmentation criteria for bankruptcy recovery rates based on statistical analyses wherever possible. If a meaningful statistical analysis is not feasible, the segmentation criteria can be selected on the basis of justified expert decisions.

As an alternative, it is possible to estimate the value of components individually, especially in the case of physical collateral. This is especially common practice in the case of large objects (real estate, ships, aircraft, etc.). Capital equipment is generally valuated using business criteria in such a way that the collateral's value depends on the income it is expected to generate (present value of cash flow). In such cases, suitable methods include cash flow models, which can be coupled with econometric models for the purpose of estimating rent developments and occupancy rates, for example. Instead of cash flow simulation, the present value and appropriate markdowns can form the basis for estimates of the collateral value at default, which is calculated by means of expert valuation for real estate and large movable property (e.g. ships). Private consumer goods such as passenger vehicles can be valuated using the secondary market prices of goods with comparable characteristics. In contrast, saleability is uncertain in the case of physical collateral for which liquid and established secondary markets do not exist; this should be taken into account accordingly.

It is then necessary to adjust the resulting present value conservatively using any applicable markdowns (e.g. due to neglected maintenance activities) and miscellaneous market developments up to the time of default. In addition to the realization period, the specific realization costs (expert opinions, auctioneers' commissions) and any markdowns on the market price due to the realization market's liquidity also deserve special attention. As these costs generally remain within known ranges, it is advisable to use expert estimates for these components. In this process, the aspects covered and the valuation should be comprehensible and clearly defined.

Interest Loss

The basis for calculating interest loss is the interest payment streams lost due to the default. As a rule, the agreed interest rate implicitly includes refinancing costs, process and overhead costs, premiums for expected and unexpected loss, as well as the calculated profit. The present value calculated by discounting the interest payment stream with the risk-free term structure of interest rates represents the realized interest loss. For a more detailed analysis, it is possible to use contribution margin analyses to deduct the cost components which are no longer incurred due to the default from the agreed interest rate.

In addition, it is possible to include an increased equity portion for the amount for which no loan loss provisions were created or which was written off. This higher equity portion results from uncertainty about the recoveries

during the realization period. The bank's individual cost of equity can be used for this purpose.

If an institution decides to calculate opportunity costs due to lost equity, it can include these costs in the amount of profit lost (after calculating the risk-adjusted return on equity).

Workout Costs

In order to estimate workout costs, it is possible to base calculations on internal cost and activity accounting. Depending on how workout costs are recorded in cost unit accounting, individual transactions may be used for estimates. The costs of collateral realization can be assigned to individual transactions based on the transaction-specific dedication of collateral.

When cost allocation methods are used, it is important to ensure that these methods are not applied too broadly. It is not necessary to assign costs to specific process steps. The allocation of costs incurred by a liquidation unit in the retail segment to the defaulted loans is a reasonable approach to relatively homogenous cases. However, if the legal department only provides partial support for liquidation activities, for example, it is preferable to use internal transfer pricing.

In cases where the bank's individual cost and activity accounting procedures cannot depict the workout costs in a suitable manner, expert estimates can be used to calculate workout costs. In this context, it is important to use the basic information available from cost and activity accounting (e.g. costs per employee and the like) wherever possible.

When estimating workout costs, it is advisable to differentiate on the basis of the intensity of liquidation. In this context, it is sufficient to differentiate cases using two to three categories. For each of those categories, a probability of occurrence can be determined on the basis of historical defaults. If this is not possible, the rate can be based on conservative expert estimates. The bank might also be able to assume a standard restructuring/liquidation intensity for certain customer and/or product types.

7.5 Developing an LGD Estimation Model

The procedural model for the development of an LGD estimation model consists of the following steps:

1. Analysis of data availability and quality of information carriers
2. Data preparation
3. Selection of suitable estimation methods for individual loss parameters
4. Combination of individual estimation methods to create an overall model
5. Validation

Data availability and quality are the main limiting factors in the selection of suitable methods for LGD estimation. As a result, it is necessary to analyze the available data set before making decisions as to the type and scope of the estimation methods to be implemented. In the course of data preparation, it may also be possible to fill gaps in the data set. The quality requirements for the data set are the same as those which apply to PD estimates.

Loss data analyses are frequently complemented by expert validations due to statistically insufficient data sets. A small data set is generally associated with a high degree of variance in results. Accordingly, this loss of precision in the inter-

pretation of results deserves special attention. In the course of development, the bank can use a data pool in order to provide a broader data set (cf. section 5.1.2). In the short term, estimated values can be adjusted conservatively to compensate for a high degree of variance. In the medium and long term, however, it is advisable to generate a comprehensive and quality-assured historical data set. These data provide an important basis for future validation and back-testing activities, as well as enabling future changes in estimation methodology. Moreover, Basel II and the draft EU directive require the creation of loss histories, even for the IRB Foundation Approach.[121]

When selecting methods, the bank can take the materiality of each loss component into account with regard to the effort and precision involved in each method. Based on a bank's individual requirements, it may be appropriate to implement specific LGD estimation tools for certain customer and transaction segments. For this purpose, individual combinations of loss parameters and information carriers can be aggregated to create a business segment-specific LGD tool using various estimation methods. This tool should reflect the significance of individual loss components. Throughout the development stage, it is also important to bear validation requirements in mind as an ancillary condition.

In the sections that follow, we present an example of how to implement estimation methods for each of the loss parameters: book value loss, interest loss, and workout costs.

Estimating Book Value Loss
(Example: Recovery Rates for Physical Collateral)

In the course of *initial practical implementations* at various institutions, segmentation has emerged as the best-practice approach with regard to implementability, especially for the recovery rates of physical collateral. In this section, we briefly present a segmentation approach based on Chart 91 below.

It is first necessary to gather recovery rates for all realized collateral over as long a time series as possible. These percentages are placed on one axis ranging from 0% to the highest observed recovery rate. In order to differentiate recovery rates more precisely, it is then possible to segment them according to various criteria. These criteria can be selected either by statistical means using discriminatory power tests or on the basis of expert estimates and conjectured relationships.

[121] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-5, No. 33.*

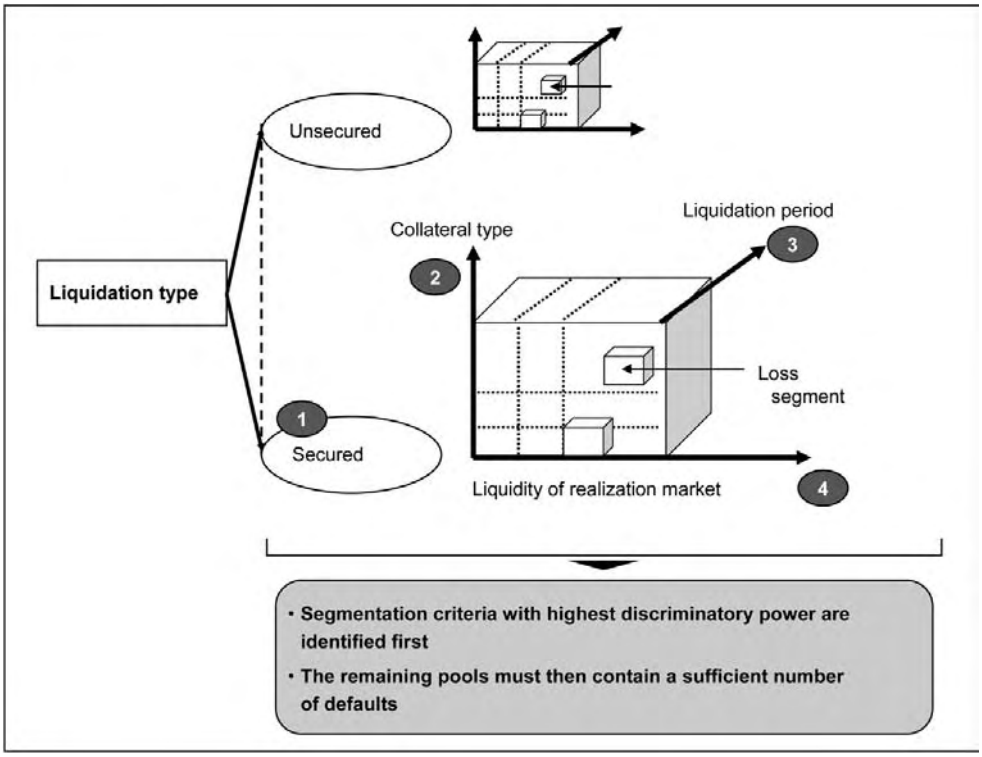

Chart 91: Example of Segmentation for Estimating LGD

The diagram below shows an example of the distribution of historical recovery rates from the realization of real estate collateral based on the type of real estate:

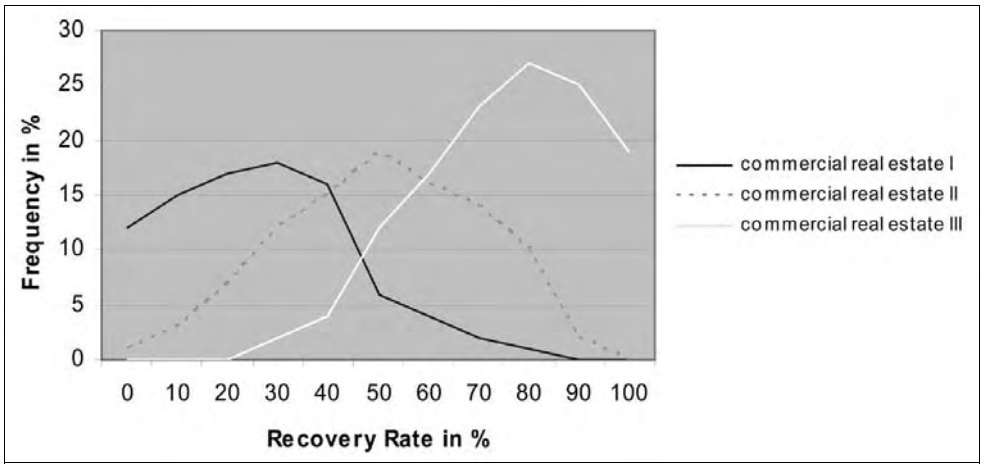

Chart 92: Example of Recovery Rates for Default by Customer Group

Even at first glance, the distribution in the example above clearly reveals that the segmentation criterion is suitable due to its discriminatory power. Statistical tests (e.g. Kolmogorov-Smirnov Test, U-Test) can be applied in order to analyze the discriminatory power of possible segmentation criteria even in cases where their suitability is not immediately visible.

It is possible to specify segments even further using additional criteria (e.g. liquidity of realization markets and liquidation period). Such specification does not necessarily make sense for every segment. When selecting criteria, it is important to ensure that a sufficiently large group can be assigned to each segment. At the same time, the criteria should not overlap excessively in terms of information content. For example, the property type in real estate constitutes a complex data element which contains implicit information on the relevant realization market, its liquidity, etc. Additional subdivisions can serve to enhance the information value, although the absolute information gain tends to drop as the fineness of the categorization increases.

In the calculation of book value loss, the collateral of an active loan is assigned to a segment according to its specific characteristics. The assigned recovery rate is equal to the arithmetic mean of historical recovery rates for all realized collateral assigned to the segment. The book value loss for the secured portion of the loan is thus equal to the secured book value minus the recovery rate.[122] In the course of quantitative validation (cf. chapter 6), it is particularly necessary to review the standard deviations of realized recovery rates critically. In cases where deviations from the arithmetic mean are very large, the mean should be adjusted conservatively.

One highly relevant practical example is the LGD-Grading procedure used by the Verband deutscher Hypothekenbanken (VDH, the Association of German Mortgage Banks),[123] which consists of approximately 20 institutions. The basis for this model was a sample of some 2,500 defaulted loans (including 1,900 residential and 600 commercial construction loans) which the participating institutions had contributed to a pool in anonymous form. For each data record, the experts preselected and surveyed 30 characteristics. Due to the market presence of the participating institutions, the sample can be assumed to contain representative loss data. On the basis of the 30 characteristics selected, the developers carried out suitable statistical analyses in order to identify 22 discriminating segments with regard to recovery rates. Segmentation is based on the property type, which is currently divided into 9 specific types; efforts are underway to subdivide this category further into 19 types. Additional segmentation criteria include the location and characteristics of the realization market, for example. This historical recovery rate is then applied to the market value in the case of liquidation. For this purpose, the current market value (expert valuation) is extrapolated for the time of liquidation using a conservative market value forecast and any applicable markdowns.

In another practical implementation for object financing transactions, segmentation is based on a far smaller sample due to the relative infrequency of defaults. In this case, object categories (aircraft, etc.) were subdivided into individual object types (in the case of aircraft: long-haul freight, long-haul passenger, etc.). Due to the relatively small data set, experts were called in to validate the segment assignments. Additional segmentation criteria included the liquid-

[122] *For the unsecured portion, the bankruptcy recovery rate can be estimated using a specific segmentation approach (based on individual criteria such as the legal form of business organization, industry, total assets, and the like) analogous to the one described for collateral recovery rates.*

[123] *Various documents on the implementation of this model are available at http://www.hypverband.de/hypverband/attachments/aktivlgd_gdw.pdf (in German), or at http://www.pfandbrief.org (menu path: lending/mortgages/LGD-Grading).*

ity of the realization market and the marketability of the object. A finer differentiation would not have been justifiable due to the size of the data set used. Due to the high variance of recovery rates within segments, the results have to be interpreted conservatively. The recovery rates are applied to the value at realization. In this process, the value at default, which is calculated using a cash flow model, is adjusted conservatively according to the expected average liquidation period for the segment.

Estimating Interest Loss

In practice, interest loss is estimated on the basis of the interest payments lost due to the default. In this process, the agreed interest for the residual term is discounted, for example using the current risk-free term structure of interest rates. Therefore, the resulting present value implicitly contains potential refinancing costs as well as the spread components process and overhead costs, risk premiums, and profit. There are also practical approaches which account for the increased equity portion required to refinance the amount not written off over the liquidation period. The bank's individual cost of equity can be used for this purpose. The practical examples implemented to date have not taken the opportunity costs of lost equity into account.

Estimating Workout Costs

The estimation method selected for calculating workout costs depends on the organizational structure and the level of detail used in cost unit and cost center accounting. The type and scope of the estimation method should reflect the significance of workout costs for the specific customer or transaction type using the available accounting information. If a bank has a separate restructuring and liquidation unit for a specific business segment, for example, it is relatively easy to allocate the costs incurred by that department.

In one practical implementation of a model for object financing transactions, experts estimated the time occupied by typical easy and difficult restructuring/liquidation cases for an employee with the appropriate qualifications. Based on accounting data, costs per employee were allocated to the time occupied, making it possible to determine the cost rates for easy and difficult liquidation cases. Historical rates for easy and difficult liquidation cases were used to weight these cost rates with their respective probabilities of occurrence.

Combining Book Value Loss, Interest Loss and Workout Costs to Yield LGD

In order to calculate LGD, the individual loss component estimates have to be merged. In this context, it is important to note that the collateral recoveries are expressed as a percentage of the secured portion and bankruptcy proceeds as a percentage of the unsecured portion of the loan, and that workout costs are more specific to cases than volumes. In order to calculate LGD, the individual components have to be merged accordingly for the credit facility in question. In this context, estimated probabilities of occurrence first have to be assigned to the post-default development scenarios preselected for the specific facility type (cf. section 7.4.2). Then it is necessary to add up the three estimated loss components (book value loss, interest loss, and workout costs). It is not neces-

sary – but may, of course, be useful – to implement a cohesive LGD estimation tool for this purpose.

8 Estimating Exposure at Default (EAD)

EAD is the only parameter which the bank can influence in advance by predefiníng limits on credit approvals for certain PD/LGD combinations. In active agreements, the bank can also impose limits by agreeing on additional covenants. The level of EAD itself is determined by the transaction type and customer type.

8.1 Transaction Types

Concerning transaction types, we can make a general distinction between balance sheet items and off-balance-sheet transactions. In the case of balance sheet items, EAD is equal to the current book value of the loan. In the case of off-balance-sheet transactions, an estimated credit conversion factor (CCF) is used to convert granted and undrawn credit lines into EAD values. In the case of a default, EAD is always equal to the current book value. In general, off-balance-sheet transactions can no longer be utilized by the borrower due to the termination of the credit line in the case of default. Therefore, EAD estimates using CCFs attempt to estimate the expected utilization of the off-balance-sheet transaction granted at the time of estimation. The following product types are among the relevant off-balance-sheet transactions:

- Lines of credit (revolving credit for corporate customers, current account overdraft facilities for retail customers)
- Loan commitments (not or only partly drawn)
- Letters of credit
- Guarantee credit (guarantees for warranty obligations, default guarantees, rental payment guarantees)

Under the draft EU directive, foreign exchange, interest rate, credit and commodity derivatives are exempt from banks' internal CCF estimation.[124] In these cases, the replacement costs plus a premium for potential future exposure are entered according to the individual products and maturity bands.

It is not necessary to estimate EAD in the case of undrawn credit commitments which can be cancelled immediately if the borrower's credit standing deteriorates. In such cases, the bank has to ensure that it can detect deterioration in the borrower's credit standing in time and reduce the line of credit accordingly.

[124] *Cf. EUROPEAN COMMISSION, draft directive on regulatory capital requirements, Annex D-4, No. 3.*

The level of utilization for off-balance-sheet transactions can range between 0 and 100% at the time of default. The chart below illustrates this point:

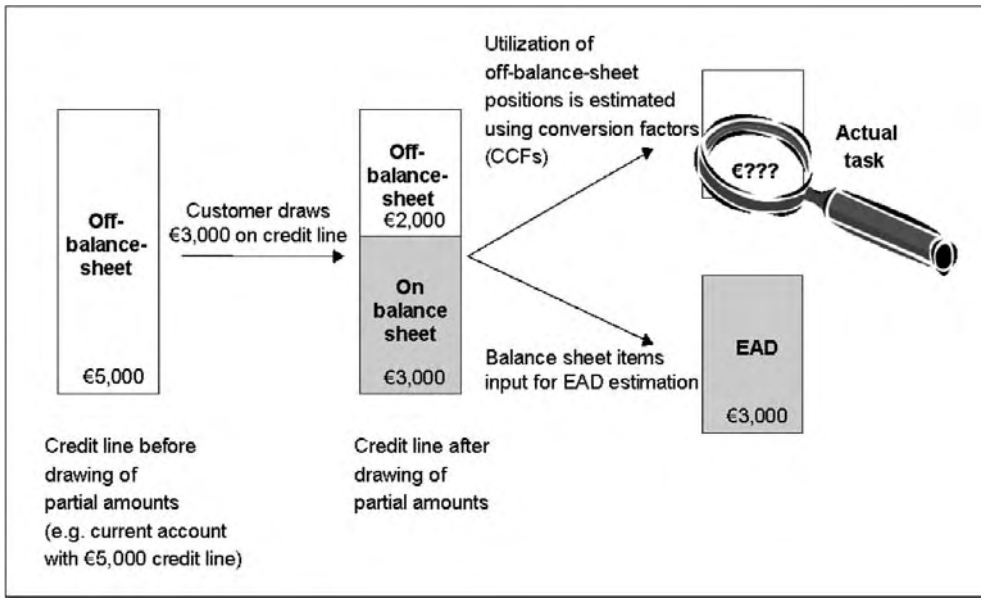

Chart 93: Objective in the Calculation of EAD for Partial Utilization of Credit Lines

In the case of guarantees for warranty obligations, the guarantee can only be utilized by the third party to which the warranty is granted. In such a case, the bank has a claim against the borrower. If the borrower defaults during the period for which the bank granted the guarantee, the utilization of this guarantee would increase EAD. The utilization itself does not depend on the borrower's creditworthiness.

In the bank's internal treatment of expected loss, the repayment structure of off-balance-sheet transactions is especially interesting over a longer observation horizon, as the borrower's probability of survival decreases for longer credit terms and the loss exposure involved in bullet loans increases.

8.2 Customer Types

The differentiation of customer types is relevant with regard to varying behavior in credit line utilization. Studies on the EAD of borrowers on the capital market and other large-scale borrowers have shown that lines of credit are often not completely utilized at the time of default. Moreover, it has been observed that the EAD for borrowers with whom the bank has agreed on covenants tends to decrease as the borrower's creditworthiness deteriorates, and that a large number of possible ways to raise debt capital also tends to lower EAD. In contrast, retail customers as well as small and medium-sized enterprises are more likely as borrowers to overdraw approved lines of credit. It is rather unusual to agree on covenants in these customer segments, and the possible ways of raising debt capital are also more limited than in the case of large companies. The table below can serve as a basis for differentiating individual customer groups. In some cases, it may also be advisable to aggregate individual customer types.

Main Group	Subgroup	Individual Level
Governments and the public sector	Governments	
	Other public agencies	Regional governments
		Local authorities
		Public sector entities (PSEs)
Financial service providers	Credit institutions	
	Insurance companies	
	Financial institutions	
Corporate customers	Enterprises/business owners	Capital market-oriented / international companies
		Other companies which prepare balance sheets
		Business owners and independent professionals
		Small businesses
		Start-ups
		NPOs (non-profit organizations)
	Specialized lending	Project finance
		Object finance
		Commodities finance
		Financing of income-producing real estate (IPRE)
Retail	Mass-market banking and private banking	Current accounts
		Rental payment guarantees
		Residential construction loans (commitments)
		Credit cards

Chart 94: Overview of Customer Types

8.3 EAD Estimation Methods

As in the case of LGD estimation, the initial implementations of EAD estimation models have primarily used the segmentation approach. CCFs are estimated on the basis of historical loss data for certain combinations of transactions and customers (and possibly other segmentation criteria such as the credit term survived, etc.). The chart below illustrates this point:

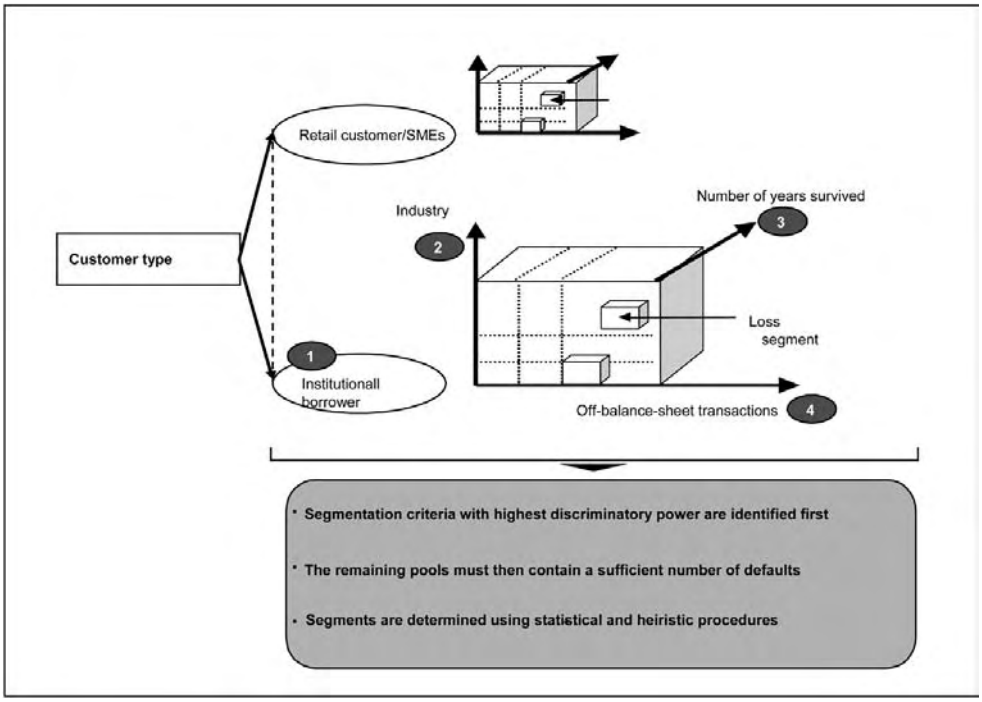

Chart 95: Example of Segmentation in CCF Estimation

It is first necessary to collect data on defaulted lines of credit over as long a time series as possible. In this process, it is important to ensure that loans which later recovered from default are also included. The percentage drawn at the time of default is determined for each of these credit facilities. These percentages are placed on one axis ranging from 0% to the highest observed utilization. In order differentiate CCFs more precisely, it is possible to segment them according to various criteria. These criteria can be selected on the basis of either statistical analyses or theoretical considerations. As an example, the diagram below shows the distribution of historical utilization rates at default using the customer type as the segmentation criterion:

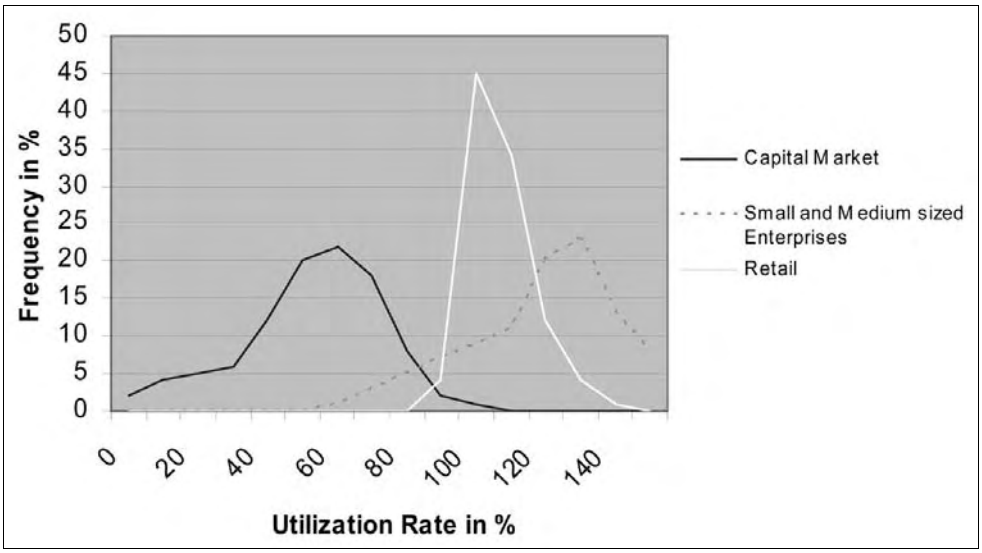

Chart 96: Example of Utilization Rates at Default by Customer Group

Even at first glance, the sample distribution above clearly shows that the criterion is suitable for segmentation (cf. chart 92 in connection with recovery rates for LGD estimates). Statistical tests can be used to perform more precise checks of the segmentation criteria's discriminatory power with regard to the level of utilization at default. It is possible to specify segments even further using additional criteria (e.g. off-balance-sheet transactions). However, this specification does not necessarily make sense for every segment. It is also necessary to ensure that the number of defaulted loans assigned to each segment is sufficiently large. Data pools can also serve to enrich the bank's in-house default data (cf. section 5.1.2).

In the calculation of CCFs, each active credit facility is assigned to a segment according to its specific characteristics. The assigned CCF value is equal to the arithmetic mean of the credit line utilization percentages for all defaulted credit facilities assigned to the segment. The draft EU directive also calls for the use of CCFs which take the effects of the business cycle into account.

In the course of quantitative validation (cf. chapter 6), it is necessary to check the standard deviations of realized utilization rates. In cases where deviations from the arithmetic mean are very large, the mean (as the segment CCF) should be adjusted conservatively. In cases where PD and the CCF value exhibit strong positive dependence on each other, conservative adjustments should also be made.

IV REFERENCES

Backhaus, Klaus/Erichson, Bernd/Plinke, Wulff/Weiber, Rolf, Multivariate Analysemethoden: Eine anwendungsorientierte Einführung, 9th ed., Berlin 1996 (Multivariate Analysemethoden)

Baetge, Jörg, Bilanzanalyse, Düsseldorf 1998 (Bilanzanalyse)

Baetge, Jörg, Möglichkeiten der Objektivierung des Jahreserfolges, Düsseldorf 1970 (Objektivierung des Jahreserfolgs)

Baetge, Jörg/Heitmann, Christian, Kennzahlen, in: Lexikon der internen Revision, Lück, Wolfgang (ed.), Munich 2001, 170–172 (Kennzahlen)

Basler Ausschuss für Bankenaufsicht, Consultation Paper – The New Basel Capital Accord, 2003 (Consultation Paper 2003)

Black, F./Scholes, M., The Pricing of Options and Corporate Liabilities, in: The Journal of Political Economy 1973, Vol. 81, 63–654 (Pricing of Options)

Blochwitz, Stefan/Eigermann, Judith, Effiziente Kreditrisikobeurteilung durch Diskriminanzanalyse mit qualitativen Merkmale, in: Handbuch Kreditrisikomodelle und Kreditderivate, Eller, R./Gruber, W./Reif, M. (eds.), Stuttgart 2000, 3-22 (Effiziente Kreditrisikobeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen)

Blochwitz, Stefan/Eigermann, Judith, Unternehmensbeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen, in: Zfbf Feb/2000, 58–73 (Unternehmensbeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen)

Blochwitz, Stefan/Eigermann, Judith, Das modulare Bonitätsbeurteilungsverfahren der Deutschen Bundesbank, in: Deutsche Bundesbank, Tagungsdokumentation – Neuere Verfahren zur kreditgeschäftlichen Bonitätsbeurteilung von Nichtbanken, Eltville 2000 (Bonitätsbeurteilungsverfahren der Deutschen Bundesbank)

Brier, G. W., Monthly Weather Review, 75 (1952), 1–3 (Brier Score)

Bruckner, Bernulf (2001), Modellierung von Expertensystemen zum Rating, in: Rating – Chance für den Mittelstand nach Basel II, Everling, Oliver (ed.), Wiesbaden 2001, 387–400 (Expertensysteme)

Cantor, R./Falkenstein, E., Testing for rating consistencies in annual default rates, Journal of fixed income, September 2001, 36ff (Testing for rating consistencies in annual default rates)

Deutsche Bundesbank, Monthly Report for Sept. 2003, Approaches to the validation of internal rating systems

Deutsche Bundesbank, Tagungsband zur Veranstaltung „Neuere Verfahren zur kreditgeschäftlichen Bonitätsbeurteilung von Nichtbanken", Eltville 2000 (Neuere Verfahren zur kreditgeschäftlichen Bonitätsbeurteilung)

Duffie, D./Singleton, K. J., Simulating correlated defaults, Stanford, preprint 1999 (Simulating correlated defaults)

Duffie, D./Singleton, K. J., Credit Risk: Pricing, Measurement and Management, Princeton University Press, 2003 (Credit Risk)

Eigermann, Judith, Quantitatives Credit-Rating mit qualitativen Merkmalen, in: Rating – Chance für den Mittelstand nach Basel II, Everling, Oliver (ed.), Wiesbaden 2001, 343–362 (Quantitatives Credit-Rating mit qualitativen Merkmalen)

Eigermann, Judith, Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale, Kaiserslautern 2001 (Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale)

European Commission, Review of Capital Requirements for Banks and Investment Firms – Commission Services Third Consultative Document – Working Paper, July 2003, (draft EU directive on regulatory capital requirements)

Fahrmeir/Henking/Hüls, Vergleich von Scoreverfahren, risknews 11/2002, http://www.risknews.de (Vergleich von Scoreverfahren)

Financial Services Authority (FSA), Report and first consultation on the implementation of the new Basel and EU Capital Adequacy Standards, Consultation Paper 189, July 2003 (Report and first consultation)

Füser, Karsten, Mittelstandsrating mit Hilfe neuronaler Netzwerke, in: Rating – Chance für den Mittelstand nach Basel II, Everling, Oliver (ed.), Wiesbaden 2001, 363–386 (Mittelstandsrating mit Hilfe neuronaler Netze)

Gerdsmeier, S./Krob, Bernhard, Kundenindividuelle Bepreisung des Ausfallrisikos mit dem Optionspreismodell, Die Bank 1994, No. 8; 469–475 (Bepreisung des Ausfallrisikos mit dem Optionspreismodell)

Hamerle/Rauhmeier/Rösch, Uses and misuses of measures for credit rating accuracy, Universität Regensburg, preprint 2003 (Uses and misuses of measures for credit rating accuracy)

Hartung, Joachim/Elpelt, Bärbel, Multivariate Statistik, 5th ed., Munich 1995 (Multivariate Statistik)

Hastie/Tibshirani/Friedman, The elements of statistical learning, Springer 2001 (Elements of statistical learning)

Heitmann, Christian, Beurteilung der Bestandfestigkeit von Unternehmen mit Neuro-Fuzzy, Frankfurt am Main 2002 (Neuro-Fuzzy)

Jansen, Sven, Ertrags- und volatilitätsgestützte Kreditwürdigkeitsprüfung im mittelständischen Firmenkundengeschäft der Banken; Vol. 31 of the "Schriftenreihe des Zentrums für Ertragsorientiertes Bankmanagement in Münster," Rolfes, B./Schierenbeck, H. (eds.), Frankfurt am Main (Ertrags- und volatilitätsgestützte Kreditwürdigkeitsprüfung)

Jerschensky, Andreas, Messung des Bonitätsrisikos von Unternehmen: Krisendiagnose mit Künstlichen Neuronalen Netzen, Düsseldorf 1998 (Messung des Bonitätsrisikos von Unternehmen)

Kaltofen, Daniel/Möllenbeck, Markus/Stein, Stefan, Große Inspektion: Risikofrüherkennung im Kreditgeschäft mit kleinen und mittleren Unternehmen, Wissen und Handeln 02, http://www.ruhr-uni-bochum.de/ikf/wissenundhandeln.htm (Risikofrüherkennung im Kreditgeschäft mit kleinen und mittleren Unternehmen)

Keenan/Sobehart, Performance Measures for Credit Risk Models, Moody's Research Report #1-10-10-99 (Performance Measures)

Kempf, Markus, Geldanlage: Dem wahren Aktienkurs auf der Spur, in Financial Times Deutschland, May 9, 2002 (Dem wahren Aktienkurs auf der Spur)

Kirmße, Stefan, Die Bepreisung und Steuerung von Ausfallrisiken im Firmenkundengeschäft der Kreditinstitute – Ein optionspreistheoretischer Ansatz, Vol. 10 of the "Schriftenreihe des Zentrums für Ertragsorientiertes Bankmanagement in Münster," Rolfes, B./Schierenbeck, H. (eds.), Frankfurt am Main 1996 (Optionspreistheoretischer Ansatz zur Bepreisung)

Kirmße, Stefan/Jansen, Sven, BVR-II-Rating: Das verbundeinheitliche Ratingsystem für das mittelständische Firmenkundengeschäft, in: Bankinformation 2001, No. 2, 67–71 (BVR-II-Rating)

Lee, Wen-Chung, Probabilistic Analysis of Global Performances of Diagnostic Tests: Interpreting the Lorenz Curve-based Summary Measures, Stat. Med. 18 (1999) 455 (Global Performances of Diagnostic Tests)

Lee, Wen-Chung/Hsiao, Chuhsing Kate, Alternative Summary Indices for the Receiver Operating Characteristic Curve, Epidemiology 7 (1996) 605 (Alternative Summary Indices)

Murphy, A. H., Journal of Applied Meteorology, 11 (1972), 273–282 (Journal of Applied Meteorology)

Sachs, L., Angewandte Statistik, 9th ed., Springer 1999 (Angewandte Statistik)

Schierenbeck, H., Ertragsorientiertes Bankmanagement, Vol. 1: Grundlagen, Marktzinsmethode und Rentabilitäts-Controlling, 6th ed., Wiesbaden 1999 (Ertragsorientiertes Bankmanagement Vol. 1)

Sobehart/Keenan/Stein, Validation methodologies for default risk models, Moody's, preprint 05/2000 (Validation Methodologies)

Stuhlinger, Matthias, Rolle von Ratings in der Firmenkundenbeziehung von Kreditgenossenschaften, in: Rating – Chance für den Mittelstand nach Basel II, Everling, Oliver (ed.), Wiesbaden 2001, 63–78 (Rolle von Ratings in der Firmenkundenbeziehung von Kreditgenossenschaften)

Tasche, D., A traffic lights approach to PD validation, Deutsche Bundesbank, preprint (A traffic lights approach to PD validation)

Thun, Christian, Entwicklung von Bilanzbonitätsklassifikatoren auf der Basis schweizerischer Jahresabschlüsse, Hamburg 2000 (Entwicklung von Bilanzbonitätsklassifikatoren)

Varnholt, B., Modernes Kreditrisikomanagement, Zurich 1997 (Modernes Kreditrisikomanagement)

Zhou, C., Default correlation: an analytical result, Federal Reserve Board, preprint 1997 (Default correlation: an analytical result)

V FURTHER READING

Further Reading on Ratings/PD

Albrecht, Jörg/Baetge, Jörg/Jerschensky, Andreas/Roeder, Klaus-Hendrick, Risikomanagement auf der Basis von Insolvenzwahrscheinlichkeiten, in: Die Bank 1999, No. 7, 494–499

Berens, Wolfgang, Beurteilung von Heuristiken: Neuorientierung und Vertiefung am Beispiel logistischer Probleme, Wiesbaden 1991

Corsten, Hans/May, Constantin, Anwendungsfelder Neuronaler Netze und ihre Umsetzung, in: Neuronale Netze in der Betriebswirtschaft: Anwendung in Prognose, Klassifikation und Optimierung – Ein Reader, Corsten, Hans/May, Constantin (eds.), Wiesbaden 1996, 1–11

Crosbie/Bohn, Modelling Default Risk, KMV LLC 2001, http://www.kmv.com/insight/index.html (Modelling Default Risk)

Erxleben, K. et al., Klassifikation von Unternehmen, Ein Vergleich von Neuronalen Netzen und Diskriminanzanalyse, in: ZfB 1992, No. 11, 1237–1262

Fahrmeir, L./Frank, M./Hornsteiner, U., Bonitätsprüfung mit alternativen Methoden der Diskriminanzanalyse, in: Die Bank 1994, No. 6, 368–373.

Fahrmeier, L./Hamerle, A./Tutz, G. (eds.), Multivariate statistische Verfahren, Berlin/New York 1996

Feulner, Waldemar, Moderne Verfahren bei der Kreditwürdigkeitsprüfung im Konsumentenkreditgeschäft, Frankfurt a. M. 1980

Fischer, Jürgen H., Computergestützte Analyse der Kreditwürdigkeitsprüfung auf Basis der Mustererkennung, in: Betriebswirtschaftliche Schriften zur Unternehmensführung, Vol. 23: Kreditwesen, Düsseldorf 1981

Gabriel, Roland, Wissensbasierte Systeme in der betrieblichen Praxis, Hamburg/New York 1990

Gabriel, Roland/Frick, Detlev, Expertensysteme zur Lösung betriebswirtschaftlicher Problemstellungen, in: ZfbF 1991, No. 6, 544–565

Gaida, S., Kreditrisikokosten-Kalkulation mit Optionspreisansätzen, Die empirische Anwendung eines Modells von Longstaff und Schwartz auf risikobehaftete Finanztitel, Münster 1997

Gaida, S., Bewertung von Krediten mit Optionspreisansätzen, in: Die Bank 1998, No. 3, 180–184

Hauschildt, Jürgen/Leker, Jens, Kreditwürdigkeitsprüfung, inkl. automatisierte, in: Handwörterbuch des Bank- und Finanzwesen, Gerke, Wolfgang/ Steiner, Manfred (eds.), 2nd ed., Stuttgart 1995, 251–262

Hüls, Dagmar, Früherkennung insolvenzgefährdeter Unternehmen, in: Schriften des Instituts für Revisionswesen der Westfälischen Wilhelms-Universität Münster, Baetge, Jörg (ed.), Düsseldorf 1995

Jacobs, Otto H., Bilanzanalyse: EDV-gestützte Jahresabschlussanalyse als Planungs- und Entscheidungsrechnung, 2nd ed., Munich 1994

Jacobs, Otto H./Oestreicher, Andreas/Piotrowski-Allert, Susanne, Die Einstufung des Fehlerrisikos im handelsrechtlichen Jahresabschluss anhand von Regressionen aus empirisch bedeutsamen Erfolgsfaktoren, in: ZfbF 1999, No. 6, 523–549

Krakl, Johann/Nolte-Hellwig, K. Ulf, Computergestützte Bonitätsbeurteilung mit dem Expertensystem „CODEX", in: Die Bank 1990, No. 11, 625–634

Krause, Clemens, Kreditwürdigkeitsprüfung mit Neuronalen Netzen, in: Schriften des Instituts für Revisionswesen der Westfälischen Wilhelms-Universität Münster, Baetge, Jörg (ed.), Düsseldorf 1993

Kurbel, Karl, Entwicklung und Einsatz von Expertensystemen: Eine anwendungsorientierte Einführung in wissensbasierte Systeme, 2nd ed., Berlin 1992

Mechler, Bernhard, Intelligente Informationssysteme, Bonn 1995

Nauck, Detlef/Klawonn, Frank/Kruse, Rudolf, Neuronale Netze und Fuzzy-Systeme: Grundlagen des Konnektionismus, Neuronale Fuzzy-Systeme und der Kopplung mit wissensbasierten Methoden, 2nd ed., Braunschweig 1996

Pytlik, Martin, Diskriminanzanalyse und Künstliche Neuronale Netze zur Klassifizierung von Jahresabschlüssen: Ein empirischer Vergleich, in: Europäische Hochschulschriften, No. 5, Volks- und Betriebswirtschaft, Vol. 1688, Frankfurt a. M. 1995

Sachs, L., Angewandte Statistik, 9th ed., Springer 1999 (Angewandte Statistik)

Schnurr, Christoph, Kreditwürdigkeitsprüfung mit Künstlichen Neuronalen Netzen: Anwendung im Konsumentenkreditgeschäft, Wiesbaden 1997

Schultz, Jens/Mertens, Peter, Expertensystem im Finanzdienstleistungssektor: Zahlen aus der Datensammlung, in: KI 1996, No. 4, 45–48

Weber, Martin/Krahnen, Jan/Weber, Adelheid, Scoring-Verfahren – häufige Anwendungsfehler und ihre Vermeidung, in: DB 1995; No. 33, 1621–1626

Further Reading on LGD/EAD

Altmann, Edward I./Resti, Andrea/Sironi, Andrea, Analyzing and Explaining Default Recovery Rates, The International Swaps & Dervatives Association, 2001

Altmann, Edward I./Resti, Andrea/Sironi, Andrea, The Link between Default and Recovery Rates: Effects on the Procyclicality of Regulatory Capital Ratios, BIS Working Papers, No. 113, 2002

Altmann, Eward I./Brady, Brooks/Resti, Andrea/Sironi, Andrea, The Link between Default and Recovery Rates: Implications for Credit Risk Models and Procyclicality, 2003

Araten, M. und Jacobs M. Jr., Loan Equivalents for Revolving Credits and Advised Lines, The RMA Journal, May 2001, 34–39

Asarnow, E. und J. Marker, Historical Performance of the U.S. Corporate Loan Market: 1988–1993, The Journal of Commercial Lending (1995), Vol. 10 (2), 13–32.

Bakshi, G./Dilip Madan, Frank Zhang, Understanding the Role of Recovery in Default Risk Models: Empirical Comparisons and Implied Recovery Rates, in: Finance and Economics Discussion Series, 2001-37, Federal Reserve Board of Governors, Washington D.C.

Basel Committee on Banking Supervision, Credit Risk Modeling: Current Practices and Applications, Bank for International Settlements, 1999

Bürgisser, Peter/Kurth, Alexander/Wagner, Armin, Incorporating Severity Variations into Credit Risk, in: Journal of Risk 2001, Volume 3, Number 4

Frye, Jon, Depressing Recoveries, Federal Reserve Bank of Chicago, Working Papers, 2000

Gordy, Michael B., A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules, Board of Governors of the Federal Reserve System, 2002

Gupton, Greg M./Gates, Daniel/Carty, Lea V., Bank Loan Loss Given Default, Moody's Investors Service – Global Credit Research, 2000

Gupton, Greg M./Stein, Roger M., Loss CalcTM: Moody's Model for Predicting Loss Given Default (LGD), Moody's Investors Service-Global Credit Research, 2002

Holter, Rebecca/Marburger, Christian, Basel II – Description of LGD-Grading project for the Verein deutscher Hypothekenbanken (Association of German Mortgage Banks), http://www.hypverband.de/hypverband/attachments/aktiv|gd_gdw.pdf (in German) or http://www.pfandbrief.org (menu path: lending/mortgages/LGD-Grading).

Jokivuolle, Esa/Peura, Samu, A Model for Estimating Recovery Rates and Collateral Haircuts for Bank Loans, in: Bank of Finland-Discussion Papers 2/2000

Jokivuolle, Esa/Peura, Samu, Incorporating Collateral Value Uncertainty in Loss Given Default: Estimates and Loan-to-value Ratios, 2003

Katzengruber, Bruno, Loss Given Default: Ratingagenturen und Basel 2, in: Österreichisches Bank-Archiv 2003, No. 10, 747–752

Van de Castle, Karen/Keisman, David, Recovering your Money: Insights into Losses from Defaults, in: Standard & Poor's CreditWeek 1999, 28–34

Van de Castle, Karen/Keisman, David, Suddenly Structure Mattered: Insights into Recoveries of Defaulted Loans, in: Standard & Poor's Corporate Ratings 2000

Verband Deutscher Hypothekenbanken, Professionelles Immobilien-Banking: Fakten und Daten, Berlin 2002

Wehrspohn, Uwe, CreditSmartRiskTM Methodenbeschreibung, CSC Ploenzke AG, 2001