European
Credit
Research
Institute

Retail Loans & Basel II

Using Portfolio Segmentation to Reduce Capital Requirements

Daniel Kaltofen

Stephan Paul

Stefan Stein

ECRI Research Report No. 8

RETAIL LOANS & BASEL II

USING PORTFOLIO SEGMENTATION TO REDUCE CAPITAL REQUIREMENTS

DANIEL KALTOFEN
STEPHAN PAUL
STEFAN STEIN

ECRI RESEARCH REPORT NO. 8
AUGUST 2006

The European Credit Research Institute (ECRI) is an independent research institution established in March 1999 in partnership with the Centre for European Policy Studies (CEPS) in Brussels. Its principal goal is to promote the study of the retail financial services sector at the EU level. ECRI's activities include statistics and econometrics, analysis and advice for a better understanding of credit markets.

This report was prepared by Dr. Daniel Kaltofen, Analyst Credit and Risk, Citibank Privatkunden AG & Co. KGaA, 40213 Düsseldorf, Germany, Dr. Stephan Paul, Professor of Banking and Finance, Ruhr-University Bochum, 44780 Bochum, Germany, and Dr. Stefan Stein, Managing Director, ikf institut für kredit- und finanzwirtschaft [institute for banking and finance], Ruhr-University (corresponding author, phone +49 234 3225344; fax +49 234 3205344, stefan.stein@rub.de).

ISBN 92-9079-660-X

European Credit Research Institute
Place du Congrès 1, B-1000 Brussels
Tel: 32(0)2 229.39.11 Fax: 32(0)2 219.41.51
E-mail: info@ecri.be
Website: http://www.ecri.be

Contents

RETAIL LOANS & BASEL II:
USING PORTFOLIO SEGMENTATION TO REDUCE CAPITAL REQUIREMENTS

ECRI RESEARCH REPORT NO. 8

Daniel Kaltofen, Stephan Paul & Stefan Stein

EXECUTIVE SUMMARY

In 2004, the Basel Committee for Banking Supervision adopted recommendations for a revised framework for risk measurement and corresponding equity capital standards to further strengthen the soundness and stability of the international banking system ('Basel II'). The implementation of the new framework in EU member jurisdictions is scheduled for year-end 2006 although some advanced approaches to risk measurement will only become available at year-end 2007.

The key concept of the revised framework is that the existing regulations pertaining to credit risk will be differentiated to a greater extent through the integration of external ratings or individualised through reference to the internal ratings of financial institutions. Certain methods that aim to determine the necessary capital requirements, particularly those developed in-house by banks must only be applied by a bank once they have been comprehensively audited by the supervisory authority (the so-called Supervisory Review Process). This can only happen once the bank has informed the financial markets of the structure of its systems within the framework of its disclosure obligations.

Under the revised framework, banks will for the first time be permitted to group their loans to private individuals (e.g. credit cards, overdrafts, residential mortgages, home equity loans and other personal loans such as education or car loans) and small corporate clients into a 'retail portfolio'. In this respect, Basel II will permit banks to choose between the internal ratings based (IRB) approach (dealt with in this paper) and the standardised approach for calculating their capital requirements for the credit risk of their retail portfolio.

While significant progress has been made in understanding the risk of commercial credits, we find that far less research has been undertaken on measuring (and managing) credit risk in retail portfolios, on either a theoretical or practical basis. This is surprising given the narrow range of options that the revised framework will grant to banks in calculating the capital requirements on their retail portfolios: there will be no differentiation between a foundation and an advanced IRB approach that is found in other asset classes (e.g. corporate portfolios). The retail IRB approach is an advanced approach and is therefore far more demanding: banks must themselves estimate all of the loss variables, which are entered into the Basel formula, in order to determine the capital requirement of the credit risk. When estimating these variables, banks should not

value their retail claims individually but group them together into ‘homogenous pools’. Basel II will leave banks with considerable margin for manoeuvre as to how to group similar loans into pools with homogenous loss characteristics.

This paper takes the retail IRB requirements as its central feature. In accordance with the regulatory guidelines, we develop an ‘optimised segmentation approach’ with regard to the credit default event and measure the implications for regulatory capital requirements. The approach presented will enable banks to improve their measurement of credit risk and the corresponding management of equity capital resources.

We identify risk drivers and operating figures, describing them in order to efficiently separate ‘good’ from ‘bad’ loans. As regards methodology, we present an innovative technique and test it on a data set of approximately 413,000 motor vehicle loans. We use a ‘recursive partitioning procedure’, which overcomes the disadvantages of the more familiar parametric methods and delivers more robust results.

By classifying loans according to selective predictors of default, we find that banks can achieve significant savings in terms of ensuring a lower regulatory capital requirement. As segmentation quality increases, the capital requirement for performing loans falls significantly. In principle, there is a direct correlation between segmentation quality and the level of the regulatory capital requirement. This provides banks with the opportunity to increase lending capacity.

Moreover, we calculate a comparatively low capital requirement (3.02% for the unexpected loss portion of the credit risk) for our reference portfolio. This is due to the fact that our portfolio is made up of motor vehicle loans only. In the event of a loan default, there is the potential for a bank to recover up to approximately 75% of the exposure by selling the vehicles accepted as collateral in comparatively efficient secondary markets. At the same time, borrowers show particular sensitivity when motor vehicles are used as collateral. The anxiety of borrowers to avoid repossession of their car by the bank means that they are more willing to keep up payments on the loan versus payments on other instalment loans for goods/services that are regarded by the borrower as less important.

With the help of the method introduced here, a meaningful differentiation of the credit risk becomes possible and the default probability and the amount of possible credit losses can be consistently and accurately assessed. For bank profitability, it is of the utmost importance that banks include their risk assessment for different groups of borrowers in their credit pricing and indeed earn the calculated risk premiums in the market. In this respect, risk-based pricing signals also promote the goal of consumer protection by supporting responsible borrowing decisions, which, in view of the growing debt burden of private households, has become increasingly important.

Finally, we call for a proportionate, pragmatic treatment of different *modi operandi* used by the banks assessed under the new Supervisory Review Process, which includes in-house approaches to modelling credit risk. Although the Committee of European Banking Supervisors (CEBS) has committed itself in this respect to a flexible approach that grants banks considerable scope for action to promote innovation within this sector, we highlight an important conflict inherent in the new concept of qualitative

supervision. If the provisions of the revised framework remain in a general rather than a specific form, they will be left open to interpretation by the supervised institutions and supervising institutions will have considerable margin for manoeuvre in how they evaluate those institutions. On the one hand, this entails a risk that actions will be hidden from view by banks, while on the other, there is a danger that the banks will receive unequal treatment. As such, the objective of competitive equality could be jeopardised. The alternative, i.e. fixing rigid compliance to highly detailed regulations, would mean that supervisory guidelines essentially pre-define risk management criteria. If this were to occur, there is a danger that one of the original aims of Basel II, i.e. the promotion of further development and improvement of risk management through competition among financial institutions, would be lost.

JEL classification: C14; C25; C53; G21; G28

Keywords: Basel II, retail portfolio, capital requirement for credit risks, portfolio segmentation, classification algorithms

RETAIL LOANS & BASEL II

USING PORTFOLIO SEGMENTATION TO REDUCE CAPITAL REQUIREMENTS

ECRI RESEARCH REPORT NO. 8

Daniel Kaltofen, Stephan Paul & Stefan Stein

1. Basel II: A new framework for risk measurement and corresponding equity capital standards

1.1 Overview of Basel II

Since the early 1980s, there have been calls – at least among the major economic nations – for the urgent harmonisation of the numerous and disparate international supervisory standards that exist. A driving force behind these calls has been the need to eradicate regulatory arbitrage, where transactions are undertaken in those countries subject to the weakest regulatory supervision. Following a joint campaign by the national banking supervisory bodies of the United States and of Great Britain, in 1988 the recommendations of the Basel Committee for Banking Supervision were passed by the central bank governors of the 'Group of Ten' (the so-called 'Basel Capital Accord', now termed 'Basel I').

The publication, entitled "International Convergence of Capital Measurement and Capital Standards – A Revised Framework", issued in June 2004[1] (Basel Committee, 2004), followed several consultation papers and was an attempt by the Basel Committee to almost completely revise the 1988 Accord ('New Basel Capital Accord' or 'Basel II'). Through this revision, the committee's aim is to "further strengthen the soundness and stability of the international banking system". While maintaining the current overall regulatory capital requirement, the new regulation aims to manage bank risks on a more comprehensive, risk-sensitive and individual basis. At the same time, the committee intends to improve competitive equality among banks and to encourage banks to further develop risk measurement and management systems.

Under this new framework, a three-pillar approach will be used to achieve these objectives (for an overview see Dierick et al., 2005, Allen et al., 2004; Allen & Saunders, 2004). For decades, banks have been exposed to quantitative capital requirements (Pillar 1) with regard to credit risks, and since 1988, even to market risks. Under the new framework, however, the existing regulations pertaining to credit risk will be differentiated to a greater extent through the integration of external ratings, or

[1] In July 2006 the Basel Committee issued a comprehensive version, which is a "compilation of the June 2004 Basel II Framework, the elements of the 1988 Accord that were not revised during the Basel II process, the 1996 Amendment to the Capital Accord to Incorporate Market Risks, and the 2005 paper on the Application of Basel II to Trading Activities and the Treatment of Double Default Effects. No new elements have been introduced in this compilation".

individualised through reference to the internal ratings of financial institutions. At the same time, so-called operational risks will be limited by quantitative regulations for the first time ever. With regard to these two areas of risk, the committee is pursuing a ‘forward-looking approach’ (as before for market risks) to ensure that the framework keeps pace with market developments and advances in risk management practices. Banks will be allowed to choose between either the standardised approaches or more sophisticated in-house approaches. Although the latter call for more development input, the regulators prefer these methods given their greater precision and risk-sensitivity, and indeed encourage banks to apply them.

In the United States, the resources and internal flows of a bank are traditionally subject to a regular audit. This then forms the basis for the levy of any possible additional charges relating to equity capital requirements necessary under the quantitative standards. This practice, which has up until now only been used in very few national European bank supervisory regimes, will be the focus of the 2nd pillar of the Basel framework. A Supervisory Review Process (SRP) will determine the individual risk profile of a bank and as part of this, the bank will be subject to regular on-site audits, examining its key capabilities and operating processes, including its in-house approaches to modelling credit risk, as outlined above.

Under the 3rd pillar, the Basel Committee aims to increase the transparency of the banks’ exposure, so that financial marketers can discipline banks via their return requirements.

It is important, though, that the three pillars are treated as a single framework rather than in isolation. Certain methods that aim to determine the necessary capital requirement (Pillar 1), particularly those developed in-house by banks, must only be applied by a bank once they have been comprehensively audited by the supervisory authority (Pillar 2). This can only happen once the bank has informed the financial markets of the structure of its systems within the framework of its disclosure obligations (Pillar 3).[2]

Implementation of the new framework in EU member jurisdictions is scheduled for year-end 2006 although some advanced approaches to risk measurement will only become available at year-end 2007 (Capital Adequacy Directive 3 (CAD 3)). To a large extent, CAD 3 follows the rules proposed by the Basel Committee. However, it is not an exact copy. The main differences relate to the scope of application and the range of approaches available to financial institutions in calculating their regulatory capital requirements (for further information on the key differences, see Cluse & Cremer, 2006). By contrast with the revised framework, which is advisory, CAD 3 is legislative. It is the responsibility of EU member states to transpose the directive into their national jurisdictions, thus making it binding on all financial institutions that are domiciled in the EU (Dierick et al., 2005).

---

[2] For more details about the general outline of the new framework, see Paul, 2006 and Dierick et al., 2005.

1.2 Grouping exposures into homogenous pools as the central feature of Basel's retail IRB approach

Under the new capital adequacy framework of Basel II, banks will for the first time be permitted to group their loans to private individuals[3] and small corporate clients into a 'retail portfolio'. As a result, they will be able to calculate the capital requirements for the credit risk of these retail portfolios rather than for the individual accounts (Basel Committee, 2006; Paul, 2006).

While significant progress has been made in understanding the risk of commercial credits, we find far less research has been undertaken on measuring (and managing) credit risk in retail portfolios, on either a theoretical or practical basis (see Allen, DeLong & Saunders, 2004 and Claessens, Krahnen & Lang, 2005 for an overview of the current state of play in the field of retail credit research and practice). This is surprising given the narrow range of options that Basel II will grant to banks in calculating the capital requirements on their retail portfolios. There will be no differentiation between a foundation and an advanced internal ratings based (IRB) approach that is found in other asset classes (e.g. corporate portfolios). The retail IRB approach is an advanced approach and is therefore far more demanding.[4] First, the new guidelines state that banks must themselves estimate all of the loss variables, which are entered into the Basel formula, in order to determine the capital requirement for the credit risk. To do this, they must collect data on the probability of default (PD), the loss given default (LGD) and the exposure at default (EAD).

Second, Basel II envisions that, when estimating these variables, banks should not value their retail claims individually but group them together into 'homogenous pools'. According to the 2006 Basel Committee on Banking Supervision, Revised Framework, section 401, banks "must assign each exposure that falls within the definition of retail for IRB purposes into a particular pool. Banks must demonstrate that this process provides for a meaningful differentiation of risk, provides for a grouping of sufficiently homogenous exposures and allows for accurate and consistent estimation of loss characteristics at pool level". According to this, banks are required to estimate the loss variables PD, EAD and LGD at the pool level (section 402) in order to derive the capital requirement for the grouped loans.

Basel II contains no binding instructions on how banks should group similar loans into pools with homogenous loss characteristics. The Basel Committee recommends only that they take at least three risk drivers into account (section 402): borrower risk characteristics, transaction risk characteristics (including product and/or collateral types) and delinquency of exposure. Each risk driver can be described using different

---

[3] Including, e.g. credit cards, overdrafts, residential mortgages, home equity loans and other personal loans such as education, instalment or motor vehicle loans.

[4] Basel II allows banks to choose between the IRB approach (dealt with in this paper) and the standardised approach for calculating their capital requirements for credit risk. The standardised approach differs only marginally from the present standard with a uniform capital requirement (that is not related to credit quality) for loans to private customers and corporate clients. In the retail segment, the key modifications relate to the 25% reduction of the risk weighting across-the-board, compared with the status quo, and the multi-faceted risk weights for defaulted loans.

operating figures (although Basel provides no further information on these and does not set any standards), and for each risk driver, at least one of the operating figures should be included in the segmentation process.

This means that the choice of suitable risk drivers and the respective operating figures is left to the banks' discretion, giving them considerable freedom. However, banks must document their procedures in order for the regulatory authorities to examine them (as outlined in Pillar 2).

At this point, Basel II again offers a high degree of flexibility in the design and implementation of this pool formation process, meaning that there are no binding supervisory specifications on how such a 'pool landscape' is meant to look. Furthermore, the Basel framework restricts neither the choice of criteria and threshold values nor the number of sub-segments. In addition to choosing suitable operating figures for the above-mentioned three risk drivers, the banks are also free to identify other relevant risk drivers and operating figures describing them and to include these in the pool formation process. As a qualitative minimum requirement, Basel stipulates only that there must be a sufficient number of loans in a pool to allow estimates to be made for the risk parameters that can be statistically validated and are stable over time. After a transitional period of three years, a data history spanning at least five years must be kept. At the same time, it is necessary to avoid undue concentrations of borrowers with regard to a particular level of credit risk (Basel Committee on Banking Supervision (2006): section 406). Basel also expects banks to conduct checks at least once a year in order to verify the validity of the segmentation process.

In this paper, we take the requirement to form homogenous risk pools as the central feature of the retail IRB approach. In accordance with the regulatory guidelines, we develop an 'optimised segmentation approach' with regard to the credit default event, which will enable banks to improve their measurement of credit risk and the corresponding management of equity capital resources.

When devising our method, we drew on research by Lang & Santomero (2002), Gross & Souleles (2002) and Hand (2001), who each refer to the advantages of statistical methods in optimising credit scoring models and of classification procedures in identifying patterns in borrower data that might be relevant to a study of default. Lang & Santomero (2002, p. 19) also find, that "overall required capital will be lower if PDs are estimated at a finer level of segmentation". However, they provide no empirical demonstration of these methods. We have used the above research as a starting point and developed the ideas with supporting empirical evidence.

Thus, the analysis presented here has both a methodological and a descriptive purpose. For the latter, risk drivers and operating figures describing them are identified in order to efficiently separate 'good' from 'bad' loans. Furthermore, we measure the implications for regulatory capital requirements. As regards methodology, we focus on a non-parametric 'recursive partitioning procedure' which, we believe, is suitable for grouping together the individual retail claims into homogenous pools – according to PD – and overcomes some disadvantages of the more familiar parametric methods.

2. Approaches for identifying homogenous risk pools

2.1 Overview of mathematical-statistical methods for the classification of retail loans

From a formal point of view, the key task is to derive a sensible and selective function of risk indicators from a set of k predictors $X_1,\ldots,X_k$ in order to quantify credit risk in terms of default. In case default is (as here) modelled as a dichotomous variable Y with the classes $\{c_1, c_2\}$ this scenario represents a typical binary classification problem, offering a wide range of standardised, established and well documented algorithms (Bonne & Arminger, 2001, p. 199; Hadidi, 2003; Hauschildt, 2000; Hand & Henley, 1997).

For the purposes of this study, we define a loan default as the *response variable*, on the basis of which we can calculate the average probability of default (PD) expressed as the ratio of defaulted loans to all the loans under consideration in a pool. We describe the pools in which we measure PD according to different attributes of the above-mentioned 'operating figures' (*predictors*) specific to the borrower, the transaction and the status of delinquency. The *prognosis period* is set between the observation times of the response variable and the predictors. Its length is specified by the objective of the underlying concept of research (in this case, Basel II requires a one-year period). Since, for the purposes of this classification, we have analysed only *historical data*, the classification techniques determine which predictors out of a given set (in this case, at a minimum the regulatory risk drivers) are most suitable for explaining the attributes of the response variable.

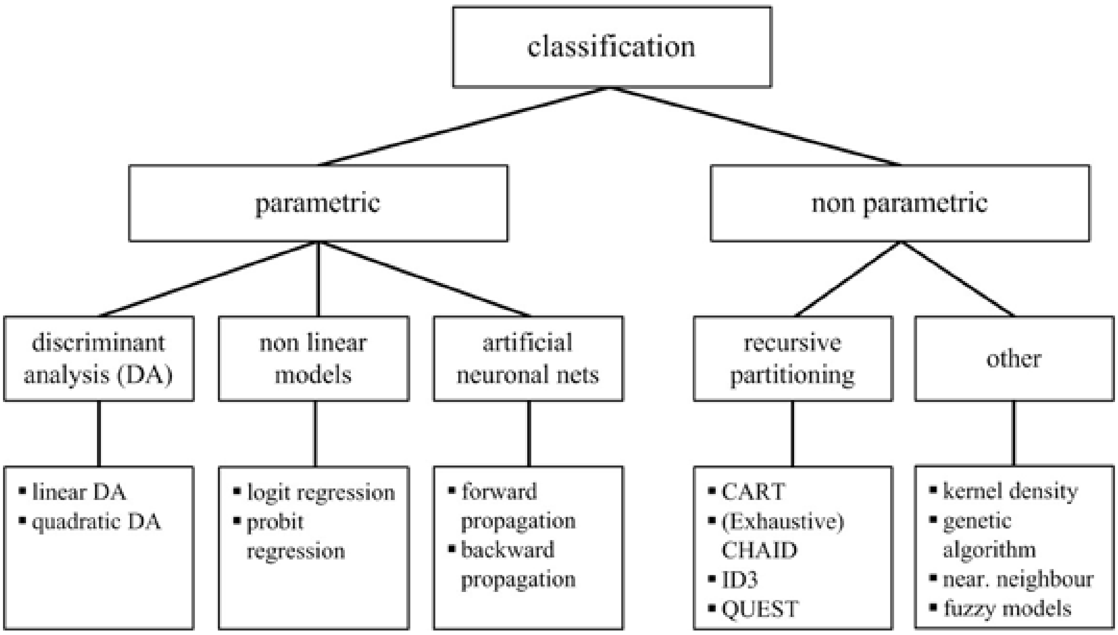
Hence, a suitable approach would be to use discriminant analysis techniques, which, in general, and independently of the way of individual algorithms work, can be applied for the purpose of separating objects (in this instance, we are inferring PD from risk drivers) into more homogenous (sub-)segments (Bonne & Arminger, 2001, p. 199).

The ability of each approach to distinguish between defaulted and current exposures roughly depends on two issues: 1) the power of the specific algorithm applied; and 2) the quality of data included. The latter obviously demands the storage and maintenance of data in a correct, complete and available manner and is not discussed here in detail.

Following Bonne & Arminger (2001, p. 199), Figure 1 organises important methods according to the presence or absence of an underlying distribution assumption with regard to the parameter values of the predictor variables. While *parametric* methods require such assumptions (usually assuming a $N(\mu;\sigma)$ normal distribution), *non-parametric* methods do not and so are not tied to such structures. As distribution assumptions are frequently not met in practice, *non-parametric* methods are deemed the more robust methods (Galindo & Tamayo, 2000, p. 115).

Developed at the end of the 1960s, *multi-variate discriminant analysis* (Altman, 1968; Altman & Saunders, 1998), as well as *logistical regression analysis*, developed a decade later (Martin, 1977; Press & Wilson, 1978), have established themselves as the standard models for single loan default forecasting in banking. Both are based on a mathematical optimisation calculus and can be implemented with minimal data processing effort. Following the expansion of data storage and data processing capacities, ‘recursive partitioning’ (also known as *classification trees*; Kass, 1980; Breiman et al., 1984) and *artificial neuronal nets* (Enache, 1998; Baetge, Kruse & Uthoff, 1996; Lohrbach, 1994; Schmidt-von Rhein & Rehkugler, 1994 and Schumann, Lohrbach & Bährs, 1992) have also gained in importance since the 1980s. If the less conventional *nearest-neighbour methods*, *kernel density estimators*, *genetic algorithms* and *fuzzy-logic-based expert systems* (frequently used for further improving the accuracy of forecasts provided by the previously named methods) are added to the list then this provides the full spectrum of methods used.

Figure 1. Selected discriminant analytical methods and algorithms



Generally speaking, there is no optimal method since every study differs in terms of data structure, data availability and data quality. As a result, there have been calls for individualised handling, taking into account the specific strengths and weaknesses of the different methods (Hand & Henley, 1997, p. 535f.). In addition, the precision, processing speed and comprehensibility of the methods play a major role in the user’s selection process. All methods, under real conditions, are heuristics that do not produce models which perfectly predict the future default status. Every discriminant analytical method merely supplies an optimised model on the grounds of statistical criteria (e.g. minimising the number or cost of misclassification; good methodological overviews of methods to calculate default probability are supplied by Jafar-Shaghaghi, 1996; Caouette, Altman & Narayanan, 1998; Blochwitz & Eigermann, 2001; Hand, 2001 and the Oesterreichische Nationalbank und Finanzmarktaufsicht [Austrian National Bank and Financial Market Supervision], 2004).

2.2 Recursive partitioning as an appropriate procedure for the identification of homogenous risk pools

As already mentioned, one standard method often used for classifying defaults is *linear multi-variate discriminant analysis (LDA)*, which classifies a data set on the basis of the result Z_m of the linear function

$$Z_m = \beta_0 + \beta_u \cdot x_{u,m} + \ldots + \beta_n \cdot x_{n,m} \text{ and} \rightarrow \{c_1, c_2 \mid Z^c\}, \qquad \text{(eq. 1)}$$

where β_u is the u-th coefficient of a predictor X_u and $x_{u,m}$ is the attribute of this predictor for the data record m. If the function value Z_m is greater than a cut-off (Z^c) fixed by the user, it belongs to group c_1 (eg, non-defaulted), or to c_2 (eg, defaulted).

In recent years there has been a shift away from discriminant analysis in favour of *logistical regression (LR)*, which has the advantages of imposing fewer formal statistical requirements on the predictors and producing more robust results (e.g. see Kaltofen, Möllenbeck & Stein, 2004; Ewert & Szczesny, 2002; Jagtiani et al., 2000; Maddala, 1983; Ohlson, 1980; Press & Wilson, 1978; Martin, 1977). Logistical regression provides, through the maximisation of a maximum likelihood function, a probability statement P concerning group classification y_m with categories $\{c_1, c_2\}$ of a data record m in the form

$$P(y_m = c_1) = \frac{1}{1+e^{(\beta_0 + \beta_u \cdot x_{u,m} + \ldots + \beta_n \cdot x_{n,m})}} = \frac{1}{1+e^{-Z_m}} \text{ and } P(y_m = c_2) = 1 - P(y_m = c_1) \qquad \text{(eq. 2)}$$

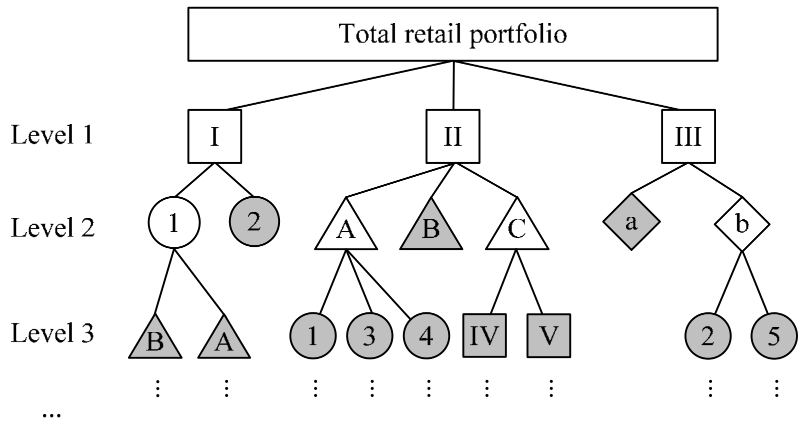
However, as parametric methods, both the discriminant analysis and logistical regression techniques have a disadvantage in that their application can be seriously impaired by outliers, extreme values and missing data for individual predictors (Espahbodi & Espahbodi, 2003; Schewe & Leker, 2000). Further statistical problems may arise from violations of the underlying normality and independence assumptions, the reduction of dimensionality issues, and the interpretation of the relative importance of individual variables rendering the accuracy of the results somewhat questionable (Frydman, Altman & Kao, 1985). Furthermore, the task of grouping single risk measures into homogenous pools has to be solved through the use of additional procedures.

By contrast, classification trees (recursive partitioning) are non-parametric pattern recognition techniques. Although based on the methodology of discriminant analysis, these techniques are not subject to its restrictive conditions (Ripley, 2002; Frydman, Altman & Kao, 1985; Breiman et al., 1984). The *advantages of classification trees* lie in their hierarchical structure and high degree of flexibility. Moreover, extreme values in the predictor attributes usually have no effect on the results because separation rarely takes place in the marginal regions. Classification trees also allow a predictor to be used repeatedly at its various levels, thus taking into account the interaction between different predictors (Espahbodi & Espahbodi, 2003).

The *basic idea underlying this method* is to subdivide in stages, by recursion, a data set on the basis of partitioning rules, such as $x_{u,m} < x_u^C$ and $x_{u,m} \geq x_u^C$, with x_u^C as the cut-off of a predictor X_u, so that the attributes of the response variable in the new sub-segments are more homogenous as in the former united segment (Küsters, 2001, p. 141). Proceeding from the whole database (original pool/node of origin), the first split into at least two new sub-segments is set by the cut-off of the most selective predictor. This procedure is repeated for each sub-segment – masking out the composition and separation of the other segments – until one out of a set of pre-assigned stopping rules takes effect (Hadidi, 2003, p. 256). In this way, it is possible to 'prune' a classification tree to a manageable size. This reduces the risk of overfitting the model, and accordingly the error rate, when applying the process to new data sets (Küsters, 2001, p. 141).

Hence a *pool landscape* is created that resembles the branch structure of a tree (Figure 2). At each level, other *predictors* (represented as geometric shapes) segment the retail portfolio in relation to the default event. The alpha-numerical values represent *cut-offs*, i.e., certain *attributes* of the respective predictors (e.g. age of borrower ≤ 30 years respectively > 30 years). The *loss parameters* PD, LGD and EAD are measured in the final pools (shaded) and apply to all loans in a pool.

Figure 2. Stylised pool landscape



A *possible disadvantage of recursive partitioning* is that binary classification techniques, in which there are always two successor pools, produce their most useful results when applied to data with non-linear structures. Otherwise, the results would be a repeated separation on the basis of the same predictor. However, the technique that we use (CHAID, see below) is able to circumvent this problem through simultaneous multiway splits of a predictor. In addition, it is not possible to assign probabilities of occurrence to the individual exposures, as it is, for example, when using the parametric method of logistical regression. However, we argue that the lack of a single loan risk indicator does not provide a problematic constellation for the purposes of classifying retail exposures under Basel II: the scope of application is to calculate default rates at the pool level (and not at an individual exposure level) and to apply these as default probabilities to another set of data, in which the required parameters are delivered only after classification. Lastly, classification trees are criticised as being data-intensive, as they require a comparatively large volume of data (Espahbodi & Espahbodi, 2003, pp. 553 f.). Steinberg & Colla (1997) call for a minimum of 200 data records in order to carry out a statistically sound classification. With around 413,000 records, our study clearly meets this requirement.

Additionally there is empirical proof of significant statistical superiority of recursive partitioning that favours the recommendation to opt for classification trees versus the demonstrated parametric techniques to evaluate capital treatment for retail loans (see Kaltofen, 2006). Accordingly, in this study we investigate in more detail this new technique in credit risk measurement and corresponding equity capital management, adopting a specific recursive algorithm for a real retail loan portfolio.

2.3 New practice: adopting the CHAID algorithm for a real retail loan portfolio

The literature on recursive partitioning offers a wide *choice of tree algorithms*, which differ in terms of performance, conditions and fields of application (Hadidi, 2003; Loh & Shih, 1997; Quinlan, 1993 and 1986; Biggs, de Ville & Suen, 1991; Breiman et al., 1984; Kass, 1980).

With regard to the results of Kaltofen (2006), we have decided to apply the CHAID (Chi-squared Automatic Interaction Detection) algorithm proposed by Kass (1980). CHAID's main advantage is that it offers the opportunity to 'grow' trees with *multiway* (and not only binary) *splits* compared to the more common algorithms such as CART (Classification and Regression Tree) by Breiman et al. (1984).

The best separation of a categorical target variable (in this case, the default event with binary yes/no attribute) is effected on the basis of χ^2-tests. CHAID carries out two crucial calculation steps: the merging and splitting of categories of predictor variables of a data segment (Levin & Zahavi, 2001; Khoshgoftaar & Allen, 2001; Wilkinson, 1992).

(1) First, CHAID merges the categories of a predictor X, which differ least in regard to the distribution of the target variable Y. Running Pearson's χ^2-test for independence, all categories of X are tested pair-wise by means of 2×2 contingency tables with regard to the null hypothesis (H_0), as to whether the distributions of Y in the two categories of X observed are independent of each other. The measure for the validity of H_0 is the exceeding probability p of the test value χ^2_{emp}. The categories are merged as long as p exceeds the threshold "alpha-to-merge" (α_{merge}) set by the user , i.e., the distribution of Y in both segments does not differ enough to regard them as statistically unequal.

(2) Proceeding from these results, a further χ^2-test is run for all the remaining v categories of each (if appropriate, recoded) predictor with regard to their relation to the target variable in order to create pools that differ significantly from each other in terms of their average PD. The Bonferroni-adjusted p-value p* of the test value χ^2_{emp}* is computed by means of the v×2 contingency table. In the case of values of p* below a user-specified "alpha-to-split" (α_{split}), the predictor with the lowest p*-value is used for segmentation, otherwise we generate a terminal node.

The algorithm ends the search for further sub-segments as soon as the stopping criteria that have been set are reached (in this case, $\alpha_{split} = \alpha_{merge} = 0.01$; minimum size of the final pools of 1.5% of the learning population). Implementation is realised with AnswerTree 3.1 of SPSS.

3. Approaching the problem empirically

3.1 Structure and scope of the database

The database we have used in our research consists of around 1.1 million consumer loans from a specialised German bank between 1999 and 2002, which provide a complete picture of all the bank's (on-balance) motor vehicle finance business (see Table 1, Chapter 3.3). We collected the data as of 31 December of the year in question. Each individual client contract corresponds to one data record. In turn, each record is described by a set of 60 operating figures.

All items are instalment loans for purchasing motor vehicles and apply to the Basel asset class of 'other retail'.

3.2 Definition of non-defaulted and defaulted loans

For the period under investigation (1999-2002), the reference bank has no consistent data on all the regulatory default indicators cited by Basel for the application of the retail IRB Approach (Basel Committee on Banking Supervision, 2006: sections 452 f.). These indicators (loans that are more than 90 days in arrears, provisioning for or restructuring of a loan, debtor insolvency, sale of claim at a loss and interest waiver by the bank) may state the default event either late (e.g. a debtor has become insolvent) or early (e.g. an assumption by the bank that a borrower will not meet the full liability and so creates provisions in its accounts). This results in either more or less conservative estimates of the probability of default, which in turn affects the level of capital requirement.

The first occurrence of any of these events is sufficient for a loan to be classed as defaulted.

For the purposes of the present study, we specify that a loan is in default after it has been in arrears for 105 days (the borrower receives a reminder with a warning that the loan may be called in). Although not fully in line with Basel's 90-day rule, this is a good, consistent match and is as close as we could come to it on the basis of the data collected so far.

We find that a parallel evaluation of the other default criteria is not practical for technical reasons[5].

Furthermore, once borrowers have defaulted, they cannot be registered as 'sound' again, even if they catch up with their payments.

---

[5] To fully apply for the retail IRB approach our reference bank would have to start collecting data that reaches the required standards: i.e. data history spanning at least five years; transitional period three years; parallel monitoring of all (!) events mentioned in sections 452f.

3.3 Choosing predictors to generate the pool landscape

For the purposes of identifying homogenous risk pools in relation to the default event, *in terms of content* we have included as predictors those operating figures supplied by the reference bank which (a) describe the risk drivers that must be taken into account under the new supervisory legislation, or (b) derive from considerations about economic plausibility. *At the formal level*, the following list of requirements applies: restriction on ratio numbers whose numerators and denominators could have negative signs and so produce misleading attributes; minimum availability of attributes of a predictor in 99.5% of all data sets; and a maximum of 98.5% of identical attributes of one predictor, in order to prevent an underrun of the minimum requirement of data records in each final pool of 1.5% of the learning population.

On the basis of these requirements, 12 predictors remain in the data set. These are assigned to the following four risk drivers: (a) transaction characteristics, (b) borrower's characteristics, (c) delinquency of exposure and (d) vintage. Table A 1 and Table A 2 show the interpretation and statistical properties of the 12 predictors. We made a conscious decision not to carry out a time-consuming correction of outlier values of individual predictors (e.g. see the maximum values in Table A 2) for 'age of the borrower (years)' and 'length of employment (months)'. The classification algorithm employed does not set any cut-offs for individual extreme attributes of a cardinally scaled predictor. Also, due to the set minimum size of final pools (amounting to 1.5% of the learning population) outlier and extreme values have a negligible influence on the segmentation result. Because of the transparency of the pool landscape, it would be possible to spot an accumulation of such cases and we could then reassess the data as necessary.

In order to develop a segmentation approach that is optimised according to the default event, we need to use only those predictors whose attributes clearly divide a set of loans, in terms of the default event, into sub-segments (pools) with different Probabilities of Default (PDs).

We therefore collect data at four dates and from this we create three one-year observation periods (1/1-31/12/2000, 2001, 2002). At the end of each observation period we determine the respective delinquency status for each loan (defaulted yes/no), applying the 105-day rule. To this information we add the original attributes of the 12 predictors at the beginning of the observation period. It is important to note that the following items are excluded from this evaluation: (a) those loans that have a history of less than one year with regard to each observation period; and (b) those which are already known as defaulted at the beginning of each period. In this way, over all three periods we obtain a total of 412,757 data records, which represent the input for the CHAID algorithm for calculating the pool landscape.

Table 1. Data input for the CHAID algorithm

Observation period	2000	2001	2002	Total
Data records as of 31 December	288,793	320,536	484,366	1,093,695
Minus less than 1 year under observation	*178,402*	*202,150*	*283,641*	*664,193*
Minus already known as defaulted	*4,781*	*5,483*	*6,481*	*16,745*
Total	105,610	112,903	194,244	412,757

3.4 Calculation of the loss parameters at pool level

Equipped with our pool landscape, we thus compute the regulatory capital requirement for our reference portfolio in accordance with the Basel II risk function for 'other retail', the provisions regarding the treatment of the parameter LGD, as well as the requirements with respect to expected losses. We proceed as follows:

Separately for the above-mentioned periods of 1/1-31/12/2000, 2001, 2002, we determine (one-year) *PD, EAD and LGD values for each individual pool*. For this purpose, the pools that have been created according to the previous steps are filled with the loans present in the portfolio at the end of each year. Here, the characteristics assigned to a loan at the beginning of that year are decisive for its allocation to one pool of the (later on described) pool landscape. Then, the default status of the loans is checked and the loss parameters can be measured at pool level.

As mentioned previously, we calculate *PD* as the ratio of defaulted loans to all loans under consideration in a pool. We determine EAD and LGD for each defaulted borrower, first individually and then at pool level. The loan balance outstanding at the time of default represents *EAD*. In our study, we have chosen to define EAD as the gross balance outstanding at the time of default (GBO) less fees and interest already anticipated at their present value (fees apportionment, FA). In order to avoid values that are difficult to interpret in economic terms, EAD has been assigned a floor of zero.

$$\mathrm{EAD} = \left\{ \mathrm{GBO} - \mathrm{FA} \,\middle|\, 0 \leq \mathrm{EAD} \right\} \qquad \text{(eq. 3)}$$

LGD, on the other hand, quantifies the economic loss ratio in relation to EAD that would be lost if the borrower defaulted, including all cash inflows and outflows after the default event. The data supplied by the reference bank allows us to estimate LGD on the basis of EAD less the net cash flows on marketable motor vehicle collateral (C), manufacturers' guarantees (G) and debt service paid after a regulatory default (DS).

$$\mathrm{LGD} = \left\{ \frac{\mathrm{EAD} - \mathrm{C} - \mathrm{G} - \mathrm{DS}}{\mathrm{EAD}} \,\middle|\, 10\% \leq \mathrm{LGD} \leq 100\% \right\} \qquad \text{(eq. 4)}$$

We record these cash flows only up to the end of the year in question, due to the limitations of our reference data set. We impose an upside limit on LGD with a cap of 100% and a downside limit with a floor of 10%, the latter in order to at least account for default-related processing costs in our calculations. We make conservative estimates of the cash flows from the realisation of collateral securities (C), which are derived from depreciation models on the basis of listings from specialised price information brokers.

Using these individual annual values, the next step is to calculate – again at pool level – the multi-annual default-weighted averages for PD, LGD^e (average expected loss ratio) and EAD as prescribed by Basel II (in our study, the averages each consist of three values).

With regard to the occurrence of systemic risks and their impact on LGD estimates, Basel II requires the calculation not only of the average expected loss ratio LGD^e, but also a downturn 'conditional' loss ratio (LGD^c) to be calculated. However, the revised framework does not yet contain any binding instructions on how banks should evaluate LGD^c. In order to meet this obligation, we set LGD^c as the highest loss ratio (at pool level) observed in our data history raised by another 10% (factor 1.1).

These parameters can now be included in the calculation of the regulatory capital requirement for a forthcoming period: we fill our pools exclusively with *only non-defaulted loans* present in the reference portfolio at T = 1/1/2003 (according to their effective characteristics in relation to the predictors assigned to each pool). We assign to these loans the respective pool estimates for PD, EAD, and LGD, which are based on historical loans with the same characteristics (e.g. the PD to be estimated for T = 1/1/2003 of a given pool i is obtained from the default-weighted sum of the PDs for this pool i measured in t = 2000, 2001 and 2002).

Based on the Revised Framework's UL-only approach, the calculation of the regulatory capital requirement for the unexpected loss (UL) for all pools has the general form:

$$\kappa^p_{UL} = \frac{\sum_{i=1}^{I}\left(\Phi_i \cdot (f[PD]) \cdot LGD^c_i \cdot EAD_i\right) \cdot SF}{GBO - FA}$$

and (eq. 5)

$$f[PD] = N\left(\frac{G(PD_i) + \sqrt{\rho_i} \cdot G(0.999)}{\sqrt{1-\rho_i}}\right) - PD_i$$

i: index for pools with I =1, 2,…,I
p: non-defaulted (*p*erforming) loans
UL: unexpected loss
ρ_i: asset correlation of loans in pool i
PD_i: probability of default of loans in pool i
LGD_i^c: downturn loss ratio of loans in pool i
EAD_i: exposure at default of loans in pool i
Φ_i: number of performing loans in pool i
SF: Scaling-factor 1.06
GBO: gross balance of all outstanding loans
FA: apportionment of fees

If a *loan* is *known to be in default* at the time the regulatory capital requirement is calculated, Basel II requires that these loans are split into a separate pool. This pool needs to be considered in addition to the pool landscape described in Chapter 4.1. By definition, the PD of these loans is 100%. The regulatory capital cushioning for their unexpected loss then works out at:

$$K^{d}_{UL,j} = \left(LGD^{c}_{j} - LGD^{e}_{j}\right) \cdot EAD_{j}$$

$$\kappa^{d}_{UL} = \frac{\sum_{j=1}^{J} K^{d}_{UL,j}}{GBO - FA} \cdot SF \qquad \text{(eq. 6)}$$

K: capital treatment [in currency]
d: defaulted loans
j: index for defaulted loans with j =1, 2,…,J
LGD_j^e: loss ratio of defaulted loan j

Where loan claims have already defaulted, the LGD parameter is determined individually for each claim j, with a distinction made between average LGD and downturn LGD. The EAD also relates to the individual defaulted claim j.

With regard to the *expected loss*, Basel assumes that it has been covered by earned standard risk costs. This has to be approved in a calculation based on the specific and general provisions that have been created. If the specific and general provisions created are less than the expected loss, the bank is required to deduct this shortfall from Tier 1 and Tier 2 capital, with the amount split evenly between the two tiers. The bank may count any excess provisions, up to a maximum amount set by the regulators, as supplementary Tier 2 capital.

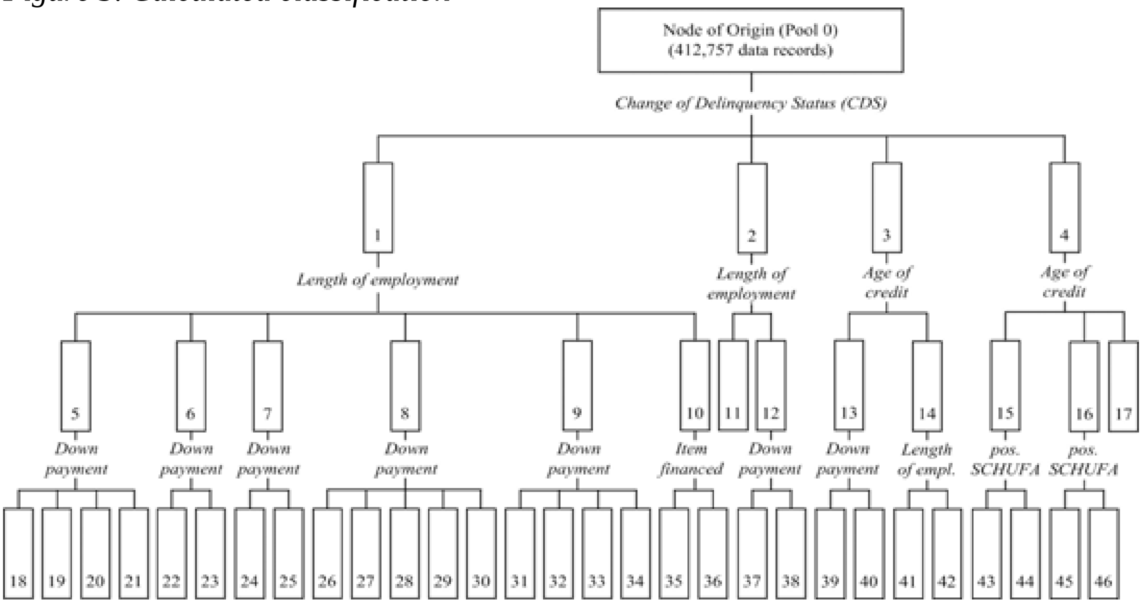
For the reference bank, we found specific and general provisions exceeding the expected loss of the retail portfolio under consideration here. The results are available from the authors on request, but for the purposes of the present study they do not need to be discussed further here as we focus primarily on the UL portion of regulatory capital charge.

4. Results of the empirical study

4.1 The pool landscape

The CHAID algorithm divides the reference portfolio (on the basis of the attributes of our predictors as discussed in Chapter 3.3) into 46 pools, of which 31 are final pools. Figure 3 shows the arrangement of the three segmentation levels in the classification tree we obtained, from which we derived the designations used in the results tables for all 46 pools.

Figure 3. Calculated classification



By far the most effective predictor with regard to credit default is the Change of Delinquency Status (CDS) numerator (evident from the highest test value χ^2_{emp}* in Table 2), followed by 'length of employment' and 'age of credit' twice each. Moreover, CHAID has also shown to be effective in terms of 'positive bureau (SCHUFA) attributes', 'item financed' and 'down payment'. The latter is selected in 7 of 11 splits on the third level. Table 2 lists the criteria for all separations plus the measured test value χ^2_{emp}*, the p*-value derived from it, the numbers and IDs of all successor pools and the corresponding cut-offs for all the splits carried out.

Table 2. Characteristics of the splits

Prede-cessor Pool #	Successor				
	Predictor	**Pool #**	**χ^2_{emp}***	**p***	**Cut-offs**
0	Change of Delinquency Status	1 - 4	33,671.92	$< 10^{-4}$	0; 1; 5
1	Length of employment	5 - 10	1016.26	$< 10^{-4}$	0; 25; 40; 78; 215; >215. „missing“
2	Length of employment	11 - 12	126.88	$< 10^{-4}$	25; >25. „missing“
3	Age of credit	13 - 14	1,336.16	$< 10^{-4}$	280
4	Age of credit	15 - 17	337.31	$< 10^{-4}$	559; 1,028
5	Down payment	18 - 21	109.95	$< 10^{-4}$	0; 2,556.45: 6,340.02
6	Down payment	22 - 23	155.29	$< 10^{-4}$	1,022.58
7	Down payment	24 - 25	97.65	$< 10^{-4}$	1,732.77
8	Down payment	26 - 30	187.25	$< 10^{-4}$	0; 1,022.58; 2,556.45; 4,550.50
9	Down payment	31 - 34	213.22	$< 10^{-4}$	511.29; 1,732.77; 3,067.75
10	Item financed	35 - 36	38.41	$< 10^{-4}$	Used car; new car
12	Down payment	37 - 38	54.12	$< 10^{-4}$	1,732.77
13	Down payment	39 - 40	54.22	$< 10^{-4}$	511.29
14	Length of employment	41 - 42	78.12	$< 10^{-4}$	40; >40. „missing“
15	Positive SCHUFA	43 - 44	57.63	$< 10^{-4}$	1
16	Positive SCHUFA	45 - 46	33.71	$< 10^{-4}$	1

The test values χ^2_{emp}* for all splits in the pool landscape we created are statistically highly significant. The probability of error is at most $4.79 \cdot 10^{-8}$, confirming the robustness of our results. It is evident that CHAID has managed to 'exploit' two of the advantages it has over other classification algorithms. Firstly, it has divided predecessor pools into more than just two successors. Secondly, it has revealed the specific interactions between predictors over the various levels of the segmentation.

In order to create a basis on which we can carry out an economic interpretation of the statistical results, we put all 412,757 data sets of our learning sample into the 31 final pools according to their attributes. We then measure the resulting average (historical, default weighted) PD and LGD values for each pool. (In the next chapter, we calculate the regulatory capital requirement. We must then value those loans that effectively fall into the pools at the time of the respective due date, with the historical loss parameters calculated in line with the provisions of Basel II (see Chapter 4.3).

Proceeding from the PD of the node of origin equal to 3.64%, the segmentation produced homogenous final pools with either much smaller or larger PDs (Table A 3): the measured PDs amount to between 0.23% and 24.96%. In the case of LGDs, we calculated values of between 19.96% and 37.63%.

Pool no. 21 – the pool with a very low PD of 0.69% and simultaneously the lowest LGD (19.96%) – contains, for example, exposures that have never been delinquent (predecessor pool no. 0). It also shows (small) company clients (who qualify for the retail segment), which indicate an employment duration of nil (predecessor pool no. 1) and fall into the segment with the highest down payments (predecessor pool no. 5). In contrast, pool no. 39 exhibits a particularly high PD/LGD combination (19.61% / 35.39%): it contains loans from debtors with a frequent change of the delinquency status (two to five; predecessor pool no. 0), who only recently received their loan (max. 280 days; predecessor pool no. 3) and have made only small down payments (max. €511.29 = 1,000 Deutschmark; predecessor pool no. 13).

The pool landscape does not show an undue concentration in any particular pool. Every end pool contains a minimum of 1.5% and a maximum of 8.1% of all data records.

Our results confirm the minimum risk drivers that should, according to Basel, be taken into account: borrower risk characteristics, transaction risk characteristics and delinquency of exposure. However, the 'delinquency of exposure' risk driver, stipulated for regulatory purposes, needs to be interpreted with caution. Operating figures that describe it tend to (retrospectively) document the non-performance of a loan more than predict its default.

With regard to our CDS numerator, it is worth noting that an additional validation, which separately records upgrades and downgrades, would not improve our results (in any case, the occurrence of downgrades, signalling an improved credit quality, is rare). What matters is the stability of the credit relationship over time, which, in this study, we describe via a higher or lower CDS numerator.

The importance of the 'age of credit' variable for our segmentation is notable because in its second consultation paper (Basel Committee, 2001 [CP 2]), Basel still envisaged a requirement whereby banks should carry out a segmentation on the basis of the vintage of the exposures (CP 2, section 447). However, this requirement does not appear in either CP 3 (Basel Committee, 2003) or the revised framework. Nevertheless, our results (which in this respect match credit agency findings concerning indicators for private debt and debt overload) provide clear evidence that the age of a loan is indeed an important risk driver in the retail portfolio (SCHUFA, 2003). CHAID identifies ‘age of credit’ as a selective predictor exclusively for low-quality loans that have been delinquent at least once (CDS > 1). Regarding these pools (no. 3 and 4 as well as their successors), we observe particularly high probabilities of default in approximately the first nine to 19 months of the credit’s life cycle. An explanation for this comparatively early occurrence of payment difficulties could be that, initially, borrowers are often still financing their principal repayments out of savings, without having adequately adjusted their monthly income to their new borrowing commitments. Irrespective of the regulatory legislation, in their own interests banks should therefore collect enough data on this risk factor to enable an assessment of whether the attributes of the loan life cycle are significant enough to justify segmentation.

Moreover, our result would seem to contradict the phrase “the longer, the riskier”, which Basel applies in other risk categories (e.g. corporates) when assigning risk weights. For instance, in pool no. 13, CHAID assigns a PD of 17.13% to exposures below a vintage of 19 months-on-book, while in pool no. 14 the PD of older loans drops to 5.11%. The volume of exclusive (‘private’) information available to the bank is likely to increase over the life cycle of a lending relationship, thus enabling the bank to obtain a more accurate picture of the borrower and to exercise greater influence on reducing the risk of default.

4.2 Segmentation quality of the pool landscape

In order to measure the discriminatory power of a rating system and the performance of the PD quantification, the Basel Committee´s Validation Group (a subgroup of the Research Task Force reviewing and developing research on validation issues) proposes graphic as well as analytical methods, including:

- Receiver Operating Characteristics (ROC)
- Area-Under-Curve (AUC)
- Cumulative Accuracy Profile (CAP)
- Accuracy Ratio (AR, also: Somers’ D, Gini-Coefficient, Powerstat)

We report in Table 3 on the AUC. It offers the advantage of having its meaning substantiated by a significance measure (p-value) which is derived using the Mann-Whitney-U-Test (equivalent to the Wilcoxon rank sum test). In the case of a perfect segmentation, the AUC equals 1 (100%), for the random model 0.5 (50%), while for realistic models the AUC adopts values ranging between these two figures. Consequently, higher forms of the AUC correspond to an enhanced predictive power (= enhanced segmentation quality; for a comprehensive overview see Basel Committee, 2005; Norden & Weber, 2005, pp. 40-44; Oesterreichische Nationalbank und

Finanzmarktaufsicht, 2004, pp. 99-146; Blochwitz et al., 2004; Deutsche Bundesbank, 2003, pp. 61-74).

Starting with the root node, Table 3 shows that the segmentation quality, i.e., predictive power, of our CHAID classification tree increases with each additional segmentation level. On level 3, we record an AUC rating of 0.84 (84%) and thus reach a value range which characterises good models. The U-Tests secure the indicated AUC values with highly significant p values ($p < 0.001$). Furthermore, this result has proven statistically stable on a ten-fold cross-validation (see Kaltofen, 2006 for more detailed results on model stability).

4.3 The relationship between segmentation quality and regulatory capital

Having obtained our pool landscape and valued its predictive power, we can now illustrate the results regarding the effect of segmentation quality on the regulatory capital requirement for the unexpected loss portion of the credit risk.

In order to highlight the influence of segmentation quality on the required amount of regulatory capital, we value the loans present in the portfolio at T = 1/1/2003 according to the specifications of Basel II. We apply historical multi-year weighted PD averages observed for the period 1/1/2000-31/12/2002 at pool level as estimates for the coming period and apply them to eq. 5. As stipulated by the new supervisory legislation, we also introduce a downturn LGD-scenario: we explicitly consider the uncertainty of the recoverability of the loans if default were to occur. This implies that a UL-risk weight needs to be assigned not only to the unexpected portion of LGD of non-defaulted loans in the retail portfolio (eq. 5) but also to those already defaulted (eq. 6). From the bank regulator's perspective, calculating the capital requirement without taking account of different macroeconomic scenarios could result in a tendency to underestimate the required risk cushioning (in terms of Tier 1 and Tier 2 capital). Unlike what is assumed in Basel's stochastic model (eq. 5), the regulators do not see PD and LGD independently of each other. In the event of an economic downturn, there is a risk of an increase in the probabilities of default (which alone would imply an increase in the regulatory capital requirement). Furthermore, as a consequence of market distortions, the effective loss ratios could also fall short of those expected when collateral (here, motor vehicles) is realised (see Sorge, 2004 with further references).

As a consequence, the Basel Committee called for the banks' internal LGD estimates to be based on conservative economic scenarios. We satisfy this requirement (which has not yet been further specified) by taking as the stress value the worst LGD value in the available data history raised by 10%, measured at pool level (see above Chapter 3.4). This stress value exceeds expected multi-year default-weighted averages LGD^e. In the reference bank's portfolio, this difference between stress LGD and expected LGD leads to a UL-capital requirement for a latent systemic risk for the defaulted loans of κ^d_{UL} = 0.09%.

Table 3 shows the results for all three selective segmentation levels of our CHAID classification tree starting at the level of origin (level 0):

Table 3. UL regulatory capital requirement

	Level of segmentation			
	Level 0	**Level 1**	**Level 2**	**Level 3**
Area-Under-Curve	**50.0%**	**79.3%**	**82.8%**	**84.0%**
p-value		*< 0.001*	*< 0.001*	*< 0.001*
Minimum regulatory capital for unexpected losses:				
Performing (κ^p)	3.56%	3.29%	3.03%	2.92%
Defaulted (κ^d)	0.09%	0.09%	0.09%	0.09%
Total	**3.66%**	**3.38%**	**3.13%**	**3.02%**

As segmentation quality increases through the levels, the UL capital requirement for performing loans in T = 1/1/2003 falls by 18.0% in comparison with the level 0. On level 3, the requirement is 2.92% for the UL-portion of the credit risk of current loans, and 3.56% for the node of origin. At the most selective steps, 1 and 2, a marked reduction in the required capital is registered when jumping from the predecessor pools to these discriminatory levels: 7.8% at the transition from level 0 to 1 and another 7.7% from level 1 to 2.[6] At level 3, 11 separations with declining power take place. Although they lead to a further improvement in segmentation quality, they cause a smaller reduction of 3.6% in capital cushioning.

This effect operates in principle between segmentation quality and the amount of the regulatory capital requirement and thus provides empirical backing for the thesis proposed at the start of this study, that "overall required capital will be lower if PDs are estimated at a finer level of segmentation".

Finally, to interpret the comparatively low level of overall UL-capital requirement calculated, we need to focus on the combination of low LGD values and low PD values within our reference portfolio. To produce our results, we used a portfolio comprising only motor vehicle loans. In the event of a loan default, there is the potential for a bank to recover up to approximately 75% of the exposure by selling the vehicles accepted as collateral in comparatively efficient secondary markets. At the same time, consistent with the findings of risk management literature and credit agency data, borrowers show particular sensitivity when motor vehicles are used as collateral. The anxiety of borrowers (many of whom depend on a car for their jobs) to avoid repossession of their car by the bank means that they are more willing to keep up payments on the loan versus payments on other instalment loans for goods/services that are regarded by the borrower as less important.

[6] The calculation of these figures is based on exact values. Differences may occur when deriving them from the rounded values in Table 3.

5. Conclusion and policy recommendations

In this paper, we have introduced an innovative technique, which adheres to the provisions of Basel II, for grouping retail loans into homogenous pools. To do this, we used non-parametric recursive partitioning supported by the CHAID algorithm. This non-parametric method overcomes many of the shortcomings of parametric techniques, such as discriminant analysis or logistical regression, which have been employed by a number of banks. The empirical results confirm that the ability to separate sound from potentially defaulting borrowers with greater precision through the formation of homogenous pools reduces regulatory capital requirements.

Our process of classifying loans according to selective default predictors lowers the required capital for UL by almost 18% over the different segmentation levels for the period under review. Hence, our analyses imply that significant cost savings can be made in terms of equity capital (or, alternatively, that opportunities can be provided to increase lending capacity).

As a result of our study, we recommend that our findings be used by the managements of banks to initiate a critical review of existing rating and scoring processes. If participants are to benefit fully from the demonstrated effects of reducing capital cushioning, it is essential that they choose not only an efficient and selective algorithm, but that they also ensure that data quality is high and meets the demands of Basel II. Essentially, the algorithm can only be as good as the data it uses.

However, to demonstrate the risk homogeneity of exposures within each segment, Basel II requires that banks treat the loans in each segment in the same manner in their internal risk management processes (Basel Committee on Banking Supervision (2004): section 232). This may involve underwriting and structuring of the loans, economic capital allocation, pricing and other terms of the lending agreement, monitoring, and internal reporting. As a result, there is a trade-off between the potential cost savings on equity capital from an increasingly differentiated pool landscape and the additional costs of managing such a pool landscape (namely in terms of personnel costs).

With the help of the method introduced here, a meaningful differentiation of the credit risk becomes possible. The default probability and the amount of possible credit losses can be consistently assessed with precision. For bank profitability, it is of the utmost importance that banks include their risk assessment for different groups of borrowers in their credit pricing and indeed earn the calculated risk premiums in the market. In this respect, a recent survey carried out by the institute for banking and finance (ikf) shows that, for the German consumer credit market, the vast majority of banks still price their loans across-the-board. Scoring information is predominantly limited to calculate a cut-off-score, below which no credit is issued (Paul & Stein, 2006).

If all borrowers have to pay the same price for a loan regardless of the risk that they present in terms of defaulting, consumers with a good credit score ultimately subsidise those with poor creditworthiness. Ideally, though, the reverse should be true. Those borrowers that are able to keep their finances in good order should be rewarded for doing so. For those consumers with poor credit scores, high interest rates should serve as a warning that increasing their credit burden may not be advisable. In this respect, risk-based pricing signals promote the goal of consumer protection by supporting responsible borrowing decisions, which in view of the growing debt burden of private households, has become increasingly important. Furthermore, empirical studies indicate that access to credit for higher-risk households is improved when banks use advanced methods to better evaluate the credit quality of borrowers *and* differentiate loan prices according to the risk that has been measured (Edelberg, 2003; Berger, Frame & Miller, 2002; Mester, 1997).

With regard to the new Supervisory Review Process (see Chapter 1 above), supervisors will assess and approve the risk measurement and management systems for each bank individually. Decisive in this respect will be the maintenance of fair and equal treatment of different *modi operandi* used in the banks reviewed. Therefore, in the course of the audit process, indications will have to be found as to how far certain structural features of an individual bank's risk management require certain 'stays' (possibly with the development of minimum or maximum requirements). The Committee of European Banking Supervisors (CEBS) has committed itself in this respect to a flexible approach which gives banks considerable room for manoeuvre. Different opinions on the specific ways in which methods are applied in practice are expressly encouraged for the promotion of innovation within this sector. However, the path chosen by a bank must be documented and its consistency with the guidelines issued by the supervisory authorities must be set out in detail. There is no 'one-size-fits-all' approach.

Finally, this highlights an inherent conflict in the new concept of qualitative supervision. If the provisions of the revised framework remain in a general rather than a specific form, they will be left open to interpretation by the supervised institutions and supervising institutions will have considerable margin for manoeuvre in how they evaluate those institutions. On the one hand, this entails a risk that actions will be hidden from view by banks; while on the other there is a danger that the banks will receive unequal treatment. As such, the objective of competitive equality could be jeopardised if, for example, different audit teams apply conflicting standards to internal rating systems. The alternative, i.e., fixing rigid compliance to highly detailed regulations (almost in the sense of quantitative norms) would mean that supervisory guidelines essentially pre-define risk management criteria. If this were to occur, there is a danger that one of the original aims of Basel II, i.e. the promotion of further development and improvement of risk management through competition among financial institutions, would be lost. It will be very important to ensure that the regulatory audit teams are fully equipped with the methodological (and perhaps more importantly, social) skills that will enable them to act as a competent but pragmatic/ 'proportionate' regulator, while recognising the scale and complexity of the activities of the credit institution under review.

REFERENCES

Allen, L. & A. Saunders, eds. (2004), "Special Issue on 'The Path to Basel II Adoption'", *Journal of Financial Services Research*, Vol. 26, pp. 101-191.

Allen, L., G. DeLong & A. Saunders (2004), "Issues in the credit risk modeling of retail markets", *Journal of Banking & Finance*, Vol. 28, pp. 727-752.

Altman, E.I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance*, Vol. 23, pp. 589-609.

Altman, E.I. & A. Saunders (1998), "Credit risk measurement: developments over the last 20 years", *Journal of Banking and Finance*, Vol. 21, pp. 1721-1742.

Altman, E.I., B. Brooks, A. Resti & A. Sironi (2003), *The Link between Default and Recovery Rates: Theory, Empirical Evidence and Implications*, March (retrieved from http://pages.stern.nyu.edu/~ealtman/Link_between_Default_and_Recovery_Rates.pdf).

Baetge, J., A. Kruse & C. Uthoff (1996), "Bonitätsklassifikationen von Unternehmen mit Neuronalen Netzen", *Wirtschaftsinformatik*, Vol. 38, pp. 273-281.

Basel Committee on Banking Supervision (2001), *The New Basel Capital Accord* [CP 2], Consultative Document, Basel, January 2001.

Basel Committee on Banking Supervision (2003), *The New Basel Capital Accord* [CP 3], Consultative Document, Basel, April 2003.

Basel Committee on Banking Supervision (2004), *International Convergence of Capital Measurement and Capital Standards – A Revised Framework*, Revised Framework, June, Basel.

Basel Committee on Banking Supervision (2005), *Studies on the Validation of Internal Rating Systems*, Working Paper No. 14, February, Basel.

Basel Committee on Banking Supervision (2006**),** *International Convergence of Capital Measurement and Capital Standards, A Revised Framework, Comprehensive Version*, July, Basel - http://www.bis.org/publ/bcbs128.pdf

Berger, A.N. & G.F. Udell (1995), "Relationship Lending and Lines of Credit in Small Firms Finance", *Journal of Business*, Vol. 68, pp. 352-381.

Berger, A.N., W.S. Frame & N.H. Miller (2002), *Credit Scoring and the Availability, Price, and Risk of Small Business Credit*, Working Paper 2002-6, Federal Reserve Bank of Atlanta.

Biggs, D., B. de Ville & E. Suen (1991), "A Method of Choosing Multiway Partitions for Classification", *Journal of Applied Statistics*, Vol. 18, pp. 9-62.

Blochwitz, S., A. Hamerle, S. Hohl, R. Rauhmeier & D. Rösch (2004), "Was leisten Trennschärfemaße für Ratingsysteme?", *Zeitschrift für das gesamte Kreditwesen*, Vol. 57, pp. 1275-1278.

Blochwitz, S. & J. Eigermann (2001), "Interne Credit-Ratingverfahren – ein wichtiger Baustein in der Gesamtbanksteuerung", in R. Eller, W. Gruber and M. Reif (eds), *Handbuch Gesamtbanksteuerung: Integration von Markt-, Kredit- und operationalen Risiken*, Stuttgart, pp. 363-386.

Bonne, T. & G. Arminger (2001), "Diskriminanzanalyse", in *Handbuch Data Mining im Marketing*, Wiesbaden, pp. 193-239.

Breiman, L., J.H. Friedman, R.A. Olshen & C.J. Stone (1984), *Classification and regression trees*, Belmont.

Caouette, J.B., E.I. Altman & Paul Narayanan (1998), *Managing Credit Risk: The Next Great Financial Challenge*, New York.

Claessens, S., J. Krahnen & W.W. Lang (2005), "The Basel II Reform and Retail Credit Markets", *Journal of Financial Services Research,* Vol. 28, pp. 5-13.

Cluse, M. & A. Cremer (2006), "Die Umsetzung von Basel II in deutsches Recht", *Zeitschrift für das gesamte Kreditwesen*, Vol. 59, pp. 329-334.

Deutsche Bundesbank (2003), "Validierungsansätze für interne Ratingsysteme", in *Monatsbericht der Deutschen Bundesbank*, September, pp. 61-74.

Dierick, F., F. Pires, M. Scheicher & K.G. Spitzer (2005), *The New Basel Capital Framework and its Implementation in the European Union*, Occasional Paper Series, No. 42, European Central Bank, Frankfurt/M, December.

Edelberg, W. (2003), *Risk-based Pricing of Interest Rates in Household Loan Markets*, Board of Governors of the Federal Reserve System, Working Paper, December, Washington D.C. (retrieved via www.federalreserve.gov/pubs/feds/ 2003/200362/200362pap.pdf).

Enache, D. (1998), "Künstliche neuronale Netze zur Kreditwürdigkeitsüberprüfung von Konsumentenkrediten", in E. Bomsdorf, W. Kösters & W. Matthes (eds.), *Quantitative Ökonomie*, Vol. 86, Lohmar, Cologne.

Espahbodi, H. & P. Espahbodi (2003), "Binary choice models and corporate takeover", *Journal of Banking and Finance*, Vol. 27, pp. 549-574.

Ewert, R. & A. Szczesny (2002), "Risikoindikatoren, Rating und Ausfallwahrscheinlichkeit im Kreditkundengeschäft – Eine empirische Untersuchung vor dem Hintergrund von Basel II", *Betriebswirtschaftliche Forschung und Praxis*, Vol. 54, pp. 574–590.

Frydman, H., E.I. Altman & Duen-Li Kao (1985), "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress", *Journal of Finance*, Vol. 11, pp. 269-291.

Galindo, J. & P. Tamayo (2000), "Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications", *Computational Economics*, Vol. 15, pp. 107-143.

Gross, D.B. & N.S. Souleles, (2002), "An Empirical Analysis of Personal Bankruptcy and Delinquency", *Review of Financial Studies*, Vol. 15, pp. 319-347.

Grunert, J. & M. Weber (2005), *Recovery Rates of Bank Loans: Empirical Evidence for Germany*, Department of Banking and Finance, University of Mannheim and Centre for Economic Policy Research (CEPR), Mannheim and London, March.

Hadidi, N. (2003), *Classification Ratemaking Using Decision Trees*, Data Management, Quality, and Technology Call Papers and Ratemaking Discussion Papers, Casualty Actuarial Society, pp. 253-284, winter.

Jafar-Shaghaghi, F. (1996), *Maschinelles Lernen, neuronale Netze und statistische Lernverfahren zur Klassifikation und Prognose: theoretische Analyse und ökonomische Anwendung*, Berichte aus der Informatik, Aix-la-Chapelle.

Jagtiani, J.A., J.W. Kolari, C.M. Lemieux & G. Hwan Shin (2000), *Predicting Inadequate Capitalization: Early Warning System for Bank Supervision*, Supervision and Regulation Department, Federal Reserve Bank of Chicago, Emerging Issues Series (S&R-2000-10R).

Hand, D.J. (2001): “Modelling consumer credit risk”, IMA *Journal of Management Mathematics*, Vol. 12, pp. 139-155.

Hand, D.J. & W.E. Henley (1997), “Statistical Classification Methods in Consumer Credit Scoring: a Review”, *Journal of the Royal Statistical Society*, Series A, Vol. 160, pp. 523-541.

Hauschildt, J. (2000), “Vorgehensweise der statistischen Insolvenzdiagnose”, in Hauschildt, Jürgen/Jens Leker (eds.), *Krisendiagnose durch Bilanzanalyse*, 2nd edition, Cologne, pp. 119-143.

Kaltofen, D. (2006), *Kreditrisiko- und Eigenkapitalmanagement im Retailportfolio*, Frankfurt/M.

Kaltofen, D., M. Möllenbeck & S. Stein (2004), “Modellierung eines Risikofrüherkennungssystems für das Kreditgeschäft mit mittelständischen Firmenkunden”, in S. Paul & S. Stein (eds), *Brennpunkt Kreditgeschäft*, ff. forschungsfolge 03, Frankfurt/M., pp. 161-174.

Kass, G.V. (1980), “An exploratory technique for investigating large quantities of categorical data”, *Journal of Applied Statistics*, Vol. 29, pp. 119-127.

Khoshgoftaar, T.M. & E.B. Allen (2001), “Controlling Overfitting in Classification-Tree Models of Software Quality”, *Empirical Software Engineering*, Vol. 6, pp. 59-79.

Küsters, U. (2001), “Data Mining Methoden: Einordnung und Überblick”, in Hippner, Hajo/Ulrich Küsters/Matthias Meyer/Klaus Wilde (eds), *Handbuch Data Mining im Marketing*, Wiesbaden, pp. 127-162.

Lang, W.W. & A.M. Santomero (2002), *Risk Quantification of Retail Credit: Current Practices and Future Challenges*, Federal Reserve Bank of Philadelphia, Philadelphia, December.

Levin, N. & J. Zahavi (2001), “Predictive modeling using segmentation”, *Journal of Interactive Marketing*, Vol. 15, pp. 2-22.

Loh, Wei-Yin & Yu-Shan Shih (1997), “Split Selection Methods for Classification Trees”, *Statistica sinica*, Vol. 4, pp. 815-840.

Lohrbach, T. (1994), *Einsatz von künstlichen Neuronalen Netzen für ausgewählte betriebswirtschaftliche Aufgabenstellungen und Vergleich mit konventionellen Lösungsverfahren*, Göttinger Wirtschaftsinformatik (ed. by M. Schumann and J. Biethahn), Vol. 10, Göttingen.

Maddala, G.S. (1983), *Limited-dependent and qualitative variables in econometrics*, Econometric Society monographs, Cambridge.

Martin, D. (1977), “Early Warning of Bank failure – a logit regression approach”, *Journal of Banking and Finance*, Vol. 1, pp. 249-276.

Mester, L. (1997), “What’s the Point of Credit Scoring?”, *Business Review*, Federal Reserve Bank of Philadelphia, September/October, pp. 3-16.

Oesterreichische Nationalbank und Finanzmarktaufsicht (2004), *Ratingmodelle und -validierung*, Leitfadenreihe zum Kreditrisiko, Vienna.

Ohlson, J.A. (1980), "Financial Ratios and the Probabilistic Prediction of Bankruptcy", *Journal of Accounting Research*, Vol. 18, pp. 109-131.

Paul, S. (2006), "Basel II im Überblick", in Hofmann, Gerhard (ed.), *Basel II und MaRisk*, Frankfurt/M., forthcoming.

Paul, S. & S. Stefan (2006), "Ratings, risikoorientierte Preise und Verbraucherschutz im Kreditgeschäft mit privaten Kunden", in WGZ Bank (ed.), *WGZ Bank – Regulierungsmonitor*, Duesseldorf, pp. 3-14.

Petersen, M.A. & R.G. Rajan (1994), "The Benefits of Lending Relationships. Evidence from Small Business Data", *Journal of Finance*, Vol. 49, pp. 3-37.

Press, S.J. & S.Wilson (1978), "Choosing between logistic regression and discriminant analysis", *Journal of American Statistical Association*, Vol. 73, pp. 699-705.

Quinlan, J.R. (1986), "Induction of Decision Trees", *Machine Learning*, Vol. 1, pp. 81-106.

Quinlan, J.R. (1993), *C4.5: Programs for Machine Learning*, San Meteo.

Ripley, B.D. (2002), *Pattern recognition and neural networks*, Cambridge.

Schewe, G. & J. Leker (2000), "Statistische Insolvenzdiagnose: Diskriminanzanalyse versus logistische Regression", in J. Hauschildt & J. Leker (eds.), *Krisendiagnose durch Bilanzanalyse*, 2nd edition, Cologne, pp. 168-178.

Schmidt-von Rhein, A. & H. Rehkugler (1994), "KNN zur Kreditwürdigkeitsprüfung bei Privatkundenkrediten", in H. Rehkugler & H-G. Zimmermann (eds.), *Neuronale Netze in der Ökonomie: Grundlagen und finanzwirtschaftliche Anwendungen*, Munich, pp. 491-545.

SCHUFA Holding AG (2003), *Schulden-Kompass 2003*, Wiesbaden (retrieved from www.schulden-kompass.de/downloads/sk03-gesamt.pdf).

SCHUFA Holding AG (2004), *Schulden-Kompass 2004*, Wiesbaden (retrieved from www.schulden-kompass.de/downloads/sk04_gesamt.pdf).

Schumann, M., T. Lohrbach & P. Bährs (1992), *Versuche zur Kreditwürdigkeits-prognose mit künstlichen Neuronalen Netzen*, Arbeitsberichte der Abt. Wirtschaftsinformatik II, Universität Göttingen, Nr. 2, Göttingen.

Sorge, M. (2004), *Stress-testing financial systems: an overview of current methodologies*, BIS (Bank of International Settlements) Working Papers No. 165, Basel, December.

SPSS (2002), *AnswerTree® 3.1 user´s guide*, Chicago.

Steinberg, D. & P. Colla (1997), *CART: tree-structured non-parametric data analysis*, San Diego (Salford Systems).

Wilkinson, L. (1992), *Tree Structured Data Analysis: AID, CHAID and CART Sun Valley, ID*, Sawtooth/SYSTAT Joint Software Conference, Evanston.

Table A 1. Overview of the predictors

Predictor	Interpretation
Age of borrower	Our assumption is that the income status of younger borrowers is less secure than that of older borrowers. In the case of small corporate clients that qualify for the retail category, the variable relates to the age of the company (SCHUFA 2004; with regard to small enterprises see Berger & Udell, 1995; Petersen & Rajan, 1994).
Age of credit	Data from credit agencies on debt and debt overload in the private household sector show a positively skewed frequency distribution for loans with payment difficulties depending on the age of the loan (SCHUFA, 2003). Thus, we assume that payment problems are more likely in early phases of the loan life cycle than in later phases. In the current study, we measure the age of the loan ('vintage') in days.
Credit score	At the start of the loan period, every private individual is subjected to an application scoring procedure. The higher the score, the higher the estimated credit quality. In this procedure both borrower and product characteristics are taken into account. (In the case of small enterprises that qualify for the retail category, the attribute of this variable is zero.)
Current balance	The amount of the outstanding loan balance is available for each capital calculation date. The balance of the loan is presumed to have an influence on the default risk depending on the combination and composition of other predictors mentioned here (eg, the item being financed).
Down payment	The amount of the down payment indicates the borrower's capacity to build up financial reserves and to finance investment not only by borrowing.
Guarantor	This binary number states whether or not there is a guarantor for the borrower's liabilities. If there is, this should reduce the default rate.
Item financed	Risk management literature suggests causality between the value of an item being financed and the default event: declining values accompany increasing probabilities of default (see Altman et al., 2003). Furthermore, credit agency data on the insolvency of private households show that the willingness to keep up the payments on a loan varies according to the item (or services) being financed – it falls as the importance of the goods/services to the borrower falls (SCHUFA, 2004). In this study, we can only make a distinction between new and demonstration vehicles, and used vehicles.
Length of employment	In the case of private individuals, a long period of employment (here measured in months) implies greater financial stability.
Manufacturer's subsidy	Where the dealer offers financing facilities at a preferential rate of interest, the difference from the usual market rate is borne by the manufacturer. This manufacturer's subsidy is also paid if a borrower defaults. It is therefore a form of loan collateral on the part of the dealer.

Numerator for change of delinquency status (CDS)	Any change of delinquency status (to either a more or less favourable status) increases this numerator. It indicates the stability of the borrower's financial circumstances. A numerator of zero means that no arrears have occurred with this borrower so far; values greater than zero signal that a delinquent payment resulted in a reminder on at least one occasion.
Unanswered enquiries at SCHUFA	When checking creditworthiness as an integral part of the credit application process, the bank obtains information on the payment behaviour of (potential) borrowers from the German credit bureau SCHUFA (Protective Association for Customer Finance and the Safeguarding of Credit). On every enquiry, SCHUFA keeps the borrower file open until the bank tells it that a transaction has been concluded. If a loan request has been rejected or the customer withdraws without telling the bank, the file remains open (unanswered). So the more often enquiries remain 'unanswered', the more often (it is assumed) previous loan applications have been rejected on the grounds of low credit quality.
Positive report from SCHUFA	The more positive attributes SCHUFA has stored about a customer (e.g. former loans that were amortised) the better we expect his/her actual creditworthiness to be.

Table A 2. Statistical properties of the predictors

Predictors (dimension)	Scale	Arithm. mean	Median (mode)	Standard deviation	Min	Max
(A) Item financed (n/a)	N	-	(used car)	-	-	-
(A) Manufacturer's subsidy (Euro)	C	181.47	0.00	384.47	0.00	8,170.44
(A/D) Current balance (Euro)	C	7,717.61	6,082.33	6,109.98	0.51	170,444.26
(A/B) Credit score (points)	O	430.18	598.00	302.85	0.00	937.00
(B) Age of borrower (years)	C	37.49	40.00	17.88	0.00	904.00
(B) Down payment (Euro)	C	2,585.13	1,733.28	3,025.22	0.00	117,523.51
(B) Length of Employment (months)	C	86.03	57.00	95.60	0.00	2,207.00
(B) Guarantor (n/a)	N	-	(no)	-	-	-
(B) unanswered SCHUFA requests (amt.)	C	0.33	0.00	0.60	0.00	9.00
(B) positive SCHUFA data (amt.)	C	2.27	2.00	2.33	0.00	9.00
(C) CDS (amount)	C	2.03	0.00	6.18	0.00	100.00
(D) Age of credit (days)	C	447.82	280.00	437.82	0.00	4,395.00

Scales: C: cardinal; N: nominal; O: ordinal
Risk drivers (A): transaction characteristics, (B): borrower characteristics, (C): delinquency of exposure, (D): vintage

Table A 3. Characteristics of all 31 final pools (averages based on the whole data set of the learning sample)

Pool #	No. of data records Absolute	Relative	PD	LGD	Pool #	No. of data records Absolute	Relative	PD	LGD
11	6,700	1.6%	5.52%	26.96%	32	16,631	4.0%	0.69%	28.01%
17	11,655	2.8%	14.06%	30.82%	33	20,299	4.9%	0.40%	22.67%
18	15,427	3.7%	2.13%	32.04%	34	33,408	8.1%	0.23%	23.83%
19	19,557	4.7%	1.47%	22.09%	35	12,773	3.1%	0.69%	27.64%
20	24,433	5.9%	1.14%	22.15%	36	22,393	5.4%	0.25%	29.47%
21	10,376	2.5%	0.69%	19.96%	37	8,721	2.1%	3.31%	26.94%
22	12,408	3.0%	3.80%	35.30%	38	6,409	1.6%	1.42%	23.82%
23	13,273	3.2%	1.36%	24.48%	39	6,350	1.5%	19.61%	35.39%
24	15,848	3.8%	2.16%	32.94%	40	6,555	1.6%	14.72%	23.10%
25	13,885	3.4%	0.76%	20.64%	41	8,330	2.0%	6.78%	27.33%
26	12,596	3.1%	1.83%	37.63%	42	13,120	3.2%	4.05%	26.68%
27	8,805	2.1%	1.35%	28.17%	43	6,570	1.6%	24.96%	26.88%
28	10,903	2.6%	0.89%	25.31%	44	9,061	2.2%	19.87%	28.92%
29	12,190	3.0%	0.58%	26.11%	45	7,441	1.8%	17.81%	25.86%
30	11,581	2.8%	0.27%	20.51%	46	8,557	2.1%	14.43%	28.84%
31	26,502	6.4%	1.11%	35.84%					

E C R I

European Credit Research Institute

The European Credit Research Institute (ECRI) is an independent research institute devoted to the study of credit markets and the policies affecting these markets. Its principal interests are the macroeconomic and microeconomic implications of credit. The topics we cover in-depth range from banking, retail finance and credit reporting to consumer protection. We are also committed to improving, both in a quantitive and qualitative sense, the collection of statistics related to household credit at the European level. ECRI was founded in 1999 by the Centre for European Policy Research (CEPS) together with a consortium of European credit institutions. It is one of the leading think tanks carrying out research on retail finance in Europe. For further information, visit our website: www.ecri.be

Members

BBVA Finanzia
Cetelem
Citibank Belgium
Cofidis
Cofinoga
Findomestic Banca
Provident Financial
RCI Banque
Santander consumer finance
Sofinco
Visa Europe

Scientific Committee

Prof. Umberto Filotto
Assofin and University of Rome, Italy

Prof. J. M. Hans Groeneveld
Deputy Chief Economist, Rabobank, Netherlands

Prof. Johanna Leskinen
National Consumer Research Centre, Finland

Prof. Michael E. Staten
Credit Research Center, Georgetown University, USA

Prof. Jan Wouters
Institute for International Law, University of Leuven, Belgium

ECRI
Place du Congrès 1
B-1000 Brussels
Tel.: +32 2 229 39 61
Fax: +32 2 219 41 51
E-mail: info@ecri.be
Website: http://www.ecri.be

A priviledged partner of

C E P S