Monetary and relative scorecards to assess profits in consumer revolving credit

Citation for published version:
Sanchez-Barrios, L, Andreeva, G & Ansell, J 2013, 'Monetary and relative scorecards to assess profits in consumer revolving credit', *Journal of the Operational Research Society*, vol. 66, pp. 1-11.
https://doi.org/10.1057/jors.2013.66

Digital Object Identifier (DOI):
10.1057/jors.2013.66

Document Version:
Early version, also known as pre-print

Published In:
Journal of the Operational Research Society

TITLE

"Monetary and relative scorecards to assess profits in consumer revolving credit"

ABSTRACT

This paper presents for the first time a relative profit measure for scoring purposes and compares results with those obtained from monetary scores. The suggested measure is the cumulative profit relative to the outstanding debt. It can also be interpreted as the percentage coverage against default. Monetary and relative measures are compared with both being estimated using direct and indirect methods. Direct scores are obtained from borrower attributes, whilst indirect scores are predicted using the estimated probabilities of default and repurchase. Results show that specific segments of customers ~~with specific attributes~~ are profitable in both monetary and relative terms. The best performing indirect models use the probabilities of default within 12 months on books and of repurchase during 30 months as predictors. This agrees with existing banking practice~~s~~ of default estimation and confirms the significance of the long term perspective on customer value ~~for~~in revolving credit. Direct models outperform indirect models. Relative scores would be preferred under more conservative standpoints towards default because of unstable conditions and if the aim is to penetrate relatively unknown segments. Further ethical considerations justify their use in an inclusive lending context.

KEYWORDS

Profit scoring, return scoring, credit risk, accounting, banking, inclusive lending

Acknowledgements: The authors are grateful to the data provider that chose to remain anonymous and to the provider's contact person for her outstanding help and support throughout the project.

INTRODUCTION

The use of credit scoring models became a usual practice in consumer lending since the 1980's (US Senate, 1979; Rosenberg and Glait, 1994). The literature on credit scoring is extensive and previous studies are devoted to ~~produce~~ default models based on application and/or behavioural characteristics of samples of individuals that ~~were~~have already been granted credit by the lending institution. Such models are then applied to new applicants to decide on their eligibility to be granted credit (Rosenberg and Gleit, 1994; Hand and Henley, 1997).

In the late 1990's, it was suggested that customers should be scored according to their profit profile (Lucas, 2001). Profit scoring is directly related to the concept of customer lifetime value (CLV), which can be quantified as the net present value of the discounted cash flows generated by customers (Berger and Nasr, 1998; Andon et al., 2001; Collings and Baxter, 2005; Pfeifer et al. 2005; Ryals and Knox, 2005). Furthermore, CLV has been used to value companies by using their customer base (Gupta and Lehmann, 2003). Consequently, there has been a progression from default to profit scoring among academics and practitioners. The objective has shifted to accept applicants with higher expected profit rather than with lower probability of default.

In previous studies, individuals were scored according to their expected cumulative profits during the forecast period (Andreeva et al., 2007; Finlay, 2008; Ma et al., 2009; Finlay, 2010; Lieli and White, 2010). The cited papers share a common feature: all of them use monetary based measures. Under that perspective, higher scores are aligned with higher monetary profits, regardless of the investment per customer via their outstanding debt. Two customers may yield the same profits and hence be assigned the same scores; yet in relati~~ve~~on to the amount borrowed~~ terms~~ their ~~situation~~attractiveness might not be the same. Therefore, it is useful to measure the productivity of the funds invested per customer. This rationale agrees with the use of monetary measures and relative ratios to assess the performance of lending institutions (Engels, 2010; Rasiah, 2010).

The aim of this paper is to present for the first time a relative profit measure as an alternative approach to score customers. That is, scaling the CLV by the outstanding debt is useful to provide a fair comparison of customers' profits. Additionally, it is shown that instead of considering default and profits separately, a relative profit measure actually takes into account both aspects. The measure presented in this study is by no means the only alternative to score customers in relative terms. Rather than referring to it as a return measure, which requires a more thorough inclusion of the total assets used to obtain the profits per customer, it should be regarded as a relative profit.

This paper extends the literature on the design of profit scores using direct and indirect methods. Direct scores are obtained by the use of observed attributes from individuals. Indirect methods require the prediction of intermediate variables before producing the final scores. In the case under analysis, sociodemographic, purchase and credit behaviour variables are used to produce direct scores and to predict the probabilities of default and repurchase. These probabilities are also used to obtain indirect profit scores. Similar methods have been used in a credit scoring context (Li and Hand, 2002), but not for profit scoring purposes of revolving credits. It also revisits the incidence of default and repurchase in revolving credits (Andreeva et al., 2005; Andreeva et al., 2007).

An additional motivation to use a relative profit measure for scoring purposes derives from the inclusive lending features of the case under analysis. The sample of customers used in this study belongs to a credit programme launched in 2007 by a Colombian lending institution. The aim of the programme is to offer financial services to segments considered high risks because they lack previous records with the credit bureaus and hence have been financially excluded. These segments usually rely on informal lending, with all the additional financial and safety costs attached to it. This is a common practice for unbanked segments in Latin American emerging economies (Prior and Argandoña, 2009).

The lending institution under analysis decided to use the individuals' previous payment behaviour of utility bills as the sole criterion to grant credit and to set credit limits. This practice is similar to the use of social rent payment data for credit granting decisions, suggested in the UK context (Wilkinson, 2011). The scaling effect of a relative measure would be a more responsible approach to score customers because it accounts for the outstanding debt instead of relying solely on monetary profits that might derive from default or repurchase. This could be useful to tackle ethical issues of responsible lending to vulnerable communities; this has not been considered in previous studies.

This paper is organised as follows: ~~I~~initially, the suggested monetary and relative profit measures are presented. Selected descriptive statistic results are analysed for each measure to provide the rationale of the subsequent sections. Default and repurchase scorecards are presented prior to using them as predictors for indirect monetary and relative scorecards. Direct monetary and relative scorecards are presented and contrasted with indirect scorecards. The impact of monetary and relative profit scorecards at portfolio level is then analysed; recommendations are presented in each case. Finally, a set of conclusions and ideas for further research are presented.

MONETARY AND RELATIVE PROFIT MEASURES

This section presents the suggested measures for an outcome variable in profit scoring models and the results of descriptive statistics for each measure.

Suggested measures

At the portfolio level various measures have been mentioned in the accounting literature to assess the performance of business units: income based measures (e.g. residual income and EVA™) and ratios (e.g.: return on investment and cash flow return on investment) (Drury, 2000; Hirsch, 2000; Fitzgerald, 2007). Two similar types of measures are used in this study to score customers in monetary and relative terms: $EBITACUM_t$ and $ROACUM_t$. Specifically, calculations and further analyses are performed at t=30 months, where t is time from the first purchase onwards. This ~~implied~~is equivalent to adopting a long term perspective, which agrees with the nature of revolving credit. Even though the term "profit" will continue to be used throughout this paper, the suggested measures are cash based. This is the usual practice in a scoring context.

EBITACUM$_t$

This measure, EBITACUM$_t$, is based on the earnings before interests, taxes, depreciations and amortizations (EBITA). It is the cumulative operational profit per customer, after deducting variable and fixed costs from the income generated via interests and commissions. Fixed overheads were allocated equally among active customers, according to the Company's practices. An advantage of using this measure is that it enables a straightforward comparison of customers in monetary terms. Some managers understand better the concept of monetary profits and hence would support its use for scoring purposes; this is the case where credit units are assessed as profit centres instead of investment centres. A natural threshold for this measure is 0. Considering that various segments of customers are usually served through credit programmes, it leads to the comparison of different segments with different purchase capacity and credit payment behaviour. Therefore, a disadvantage of this monetary measure is that customers with higher profits obtain better scores, regardless of the required investment (i.e.: the outstanding balance) and their intrinsic attributes. A relative measure is therefore required to effectively compare customers.

Customers joined the sample in y=0 to 14 consecutive monthly cohorts. Each customer was observed during t=1 to 30 months. First, $EBITA_Z$ was calculated for each z=t+y month (i.e.: z Є {1, 2, 3 ...44}) for each customer. Second, monthly profits excluding those of the base month were deflated according to the monthly inflation:

$$EBITAdef_z = EBITA_z \times df_z \qquad (1)$$

where df_z=monthly deflation factor to express figures relative to the base month

Third, to guarantee a single starting point, deflated profits of each cohort y were discounted during y periods at the opportunity cost *r*. This rate could be interpreted as the cost of capital of the funds invested in the credit programme:

$$\boldsymbol{EBITAdisc_t = EBITAdef_z / (1 + r)^y} \qquad (2)$$

Fourth, discounted values were compounded monthly and accumulated:

$$EBITAcum_t = \sum_{k=1}^{t} EBITAdisc_k * (1 + r)^{k-1} \qquad (3)$$

ROACUM$_t$

In contrast, $ROACUM_t$ measures the profit performance relative to the required investment per customer. It is important to note that even though the term stands for the cumulative return on assets, strictly speaking this measure should be understood as a proxy of such return. The lending institution does not discriminate the fixed assets used in the credit programme and hence assumes that the relevant assets for scoring purposes are the outstanding receivables at t=30 months. It is useful not only to quantify the scaled profit performance per customer but also to actually score them based on their relative performance. Therefore, it enables a fair comparison of the profits generated by customers. This measure can also be understood as the coverage against default if the Company stops its operations at time t=30 months. Under a very conservative standpoint, the minimum value should be 1. This implies complete coverage against default regardless of future payments. A disadvantage of this measure is that it is not always understood by managers that consider their credit units profit centres instead of investment centres.

$$ROAcum_t = {EBITAcum_t}/{FBdef_t} \qquad (4)$$

where ~~*finalbalance*~~*FB*~~*e*~~*def*$_t$= deflated final balance at time t.

Descriptive statistics

This section shows the main results from descriptive statistics of the case under analysis. Values are not displayed for confidentiality reasons.

Initially, monetary and relative measures were obtained monthly per customer (n=35,530) during t=30 months. After excluding from the sample those with missing data at t=12, 24 or 30 months, 33,964 customers were left. This was done based on a rationale of considering those customers that maintained a continuing relationship with the Company during the observation period. Almost all customers were profitable in monetary and relative terms (i.e.: 99.7% of the sample). Figures are not shown for confidentiality reasons.

Initially, the total sample was randomly split (80/20) in training, ~~sample,~~ and holdout, samples~~,~~, 27157 and 6807 customers respectively~~.~~ The distributions of $EBITACUM_{30}$ and $ROACUM_{30}$ had outliers. Difference in results between the training and holdout samples for $ROACUM_{30}$ primarily resulted from extreme outliers. This follows from the relative nature of $ROACUM_{30}$, which magnifies profits or losses in relative terms when the outstanding balance is very low. This might be the result of approximations in the calculation of the outstanding balance, which lead to values that are not significant in economic terms but that define those customers as active in the data base.

In order to visualize the distributions without the effect of outliers, 5% of the total observations were excluded from the training and holdout samples. Such observations exceeded 1.5 times the interquartile range. Specifically, customers with returns less than -0.5 or greater than 1.51 were considered outliers. Such extreme observations were added back when model quality was tested. This was done based on the feasibility that customers with returns less than -0.5 might experience a reduction of approximately half of the credit limit, whereas those that at most were covered 1.5 times against default could have their credit limit increased in the same proportion.

The mean $EBITACUM_{30}$ over the three years did not exceed the minimum monthly wage in Colombia. This confirms the inclusive lending nature of the credit programme. The average $ROACUM_{30}$ shows a significant coverage against default because of the greater interests paid in the long term. **Figure 1** shows the distribution of $EBITACUM_{30}$ for samples without outliers: training$_2$ ~~sample$_2$~~ (label=1) and holdout$_2$ ~~sample$_2$~~ (label=0), 24617 and 6186 customers respectively. Actual ~~F~~figures were ~~scaled~~divided by the greatest values for confidentiality reasons. The distribution of $EBITACUM_{30}$ ($ROACUM_{30}$) is skewed to the left (right). In absolute value, $ROACUM_{30}$ is more skewed than $EBITACUM_{30}$. These features highlight the nature of both measures: ~~M~~more customers may yield higher profits in monetary terms. In relative terms the case is not the same because of the scaling nature of ratios via the outstanding balance. $ROACUM_{30}$ has a higher kurtosis as a result of the higher concentration of observations resulting from the scaling effect. Customers that yield losses are still present, but only a very small number of~~s minor~~ cases. Consequently, lending to those that are financially excluded is a profitable business.

Figure 2 shows the joint behaviour of $EBITACUM_{30}$ and $ROACUM_{30}$ for training$_2$ sample$_2$. Results for holdout$_2$ ~~sample$_2$~~ are similar and available upon request. Overall, it is evident that higher profits lead to higher returns. This follows from the direct relationship between cumulative profits and returns, as defined before. It is evident however, that the relationship is not strictly linear. Customers have the same level of profits (returns) and various levels of returns (profits). Intuitively, it makes sense then to score customers according to each measure and assess the impact on portfolio results.

DEFAULT AND REPURCHASE SCORECARDS

Prior to generating indirect scores, the probabilities of default and repurchase were modelled through logistic regression. Logistic regression is a common classification technique in the banking industry. It produces good results when compared with more sophisticated techniques (Baesens et al., 2003). Models can be evaluated in terms of the Area Under the Receiver Operating Curve (AUROC) (Thomas, 2009). In the case under analysis a stepwise procedure was used, defining a 0.01 significance level entrance for variables. In a default context, bads were defined as those customers that experienced default; goods in repurchase models were customers that made further purchases after the first one. Default and repurchase models were produced accordingly.

Two approaches were taken for modelling purposes: ~~M~~models were developed using training$_1$ ~~sample$_1$~~ and then were tested on holdout$_1$ ~~sample$_1$~~ (~~; these models are identified with the subscript~~approach/models: *a*). Additionally, a ten-fold cross validation (approach/models *b*) was conducted on the complete sample (training$_1$ + holdout$_1$) to verify the robustness of results obtained from ~~the former~~ approach *a*. The sample was randomly divided into 10 subsets, the model was developed on 9 subsets, and tested on the remaining one, the process was repeated 10 times, each time with a different test subgroup. Averages were then calculated for measures of predictive accuracy (Tables 2 and 3). ~~Models were produced for ten random samples that accounted each for 90% of the total sample and then were tested in the remaining observations.~~ In order to obtain one set of parameters, ~~T~~the ~~overall~~final model was then run on the complete sample by using variables that were significant in at least 6 folds. This has been done before in a credit scoring context ((Lin et al., 2012). ~~These models are identified with the subscript: *b*.~~

Predictor variables

A set of sociodemographic, purchase and credit related variables (recorded at the time of the first purchase) were ~~gathered~~used to predict the probabilities of default and repurchase~~. These characteristics were obtained at the time of first purchase~~; see **Table 1**.

Formatted: Font: Not Italic

Formatted: Font: Italic

Formatted: Not Superscript/ Subscript

Formatted: Font: Italic

For prediction purposes, continuous variables were initially classified in decile categories. Categorical variables were classified according to the general categories that were provided by the Company. Further coarse classification was based on the market's knowledge (categorical variables) and collapsing adjacent categories (numerical variables). The same coarse classification was used to produce default, repurchase, direct and indirect models for comparison purposes. The coarse-classes were coded as binary (dummy) variables.

For confidentiality reasons ~~T~~the coefficients and odds ratios of significant variables ~~in the models are~~can not displayed ~~for confidentiality reasons~~, this applies to all models in this paper.

Formatted: Indent: First line: 1.27 cm

Default

A customer was considered to be in default if she was at least 3 months in arrears. Models were developed for Pr (default at t=12 months) according to standard banking practice~~s~~ and for Pr (default at t=30 months), following a long term perspective. In the former case, the standpoint was more conservative; in the latter, it was more profit-oriented. It should be noted, however, that customers can still be at arrears without reaching default. This generates additional interest~~s~~ payments/ charges that result in higher profits.

DEF1: Pr (default at t=12)

Common significant variables in models $\mathbf{DEF1}_{a,b}$ include dummies related with *marital status, education level*, *type of product* and *credit limit usage*. Specifically, cohabitators are more likely to default than singles. The informality of their relationship may be affecting their commitment towards paying the credit. ~~In terms~~As for *type of product*, it does make a difference in terms of default when customers purchase products different to those offered traditionally by the lending institution. This may occur because customers do not relate directly these products with the usual portfolio of products. Moreover, these products may be considered luxury goods that customers would not buy under normal circumstances. These aspects may affect their commitment towards paying loans derived from unusual products. Customers that used their credit limit at an intermediate level are less likely to default than those in the lowest category. These individuals might ~~grant~~attach more

Formatted: Font: Italic

importance to keeping a clean credit record than those in the lowest category and hence are more motivated to avoid default. Compared with customers with missing education level, according to model $DEF1_a$, those with secondary education are more likely to default. In contrast, model $DEF1_b$ shows that customers with primary education are less likely to be in default.

DEF2: Pr (default at t=30)

In the long term, significant variables in models **DEF2*a,b*** were the same. These include dummies related with *age, location, job* and *marital status*. Customers from older age groups are less likely to default. Major investments in education, real estate, and other durables have been made earlier. People at such stage are expected to have less financial commitments and hence have more cash available to pay their loans. Those that live in the capital city are less likely to default than customers from rural areas. This could be related ~~with~~to higher income and greater importance given to building a positive credit record in urban areas. Self-employed customers are more likely to default than those employed. The instability of irregular income under formal or informal conditions affects their ability to pay on time. Finally, married individuals are less likely to default than singles. Formal marital relationships imply regular household financial commitments. This creates the habit of paying financial obligations on time, when compared with singles.

Default in the short and long term

In the long term, it does not make a difference what type of product was purchased first. It seems then that as time goes on, customers relate more the credit programme with the financing of products different to those offered traditionally by the lending institution. The formality of marital status plays an important role in the long term in terms of decreasing the probability of default. The duration of the first loan is not significant to predict default in the short and long terms. This makes sense, since only those customers that were active at t=12, 24 and 30 months were considered. This guarantees that they had outstanding loans throughout the observation period regardless of the duration of the first loan. Finally, stratum and certain occupations that usually do not generate income (e.g.: students and housewives) were not significant in the short and long terms. From a credit risk perspective, these results justify inclusive lending in segments that traditionally are considered high risks. Results for the AUROC of each model are shown in **Table 2.**

Repurchase

Following a similar logic to that of default, the probabilities of repurchase during t=12 and t=30 months were modelled. This is also consistent with the long term nature of revolving credits.

REP1: Pr (repurchase during t=12 months)

Common significant variables in models **REP1**$_{a,b}$ include dummies ~~related with~~representing *stratum, education level, loan duration, credit limit usage* and *type of product.* Specifically, customers from higher stratums and with secondary level studies are more likely to repurchase than those from poor stratums and missing level of studies, respectively. These results follow from a greater purchase capacity associated ~~to~~with wealthier and more educated individuals, especially in developing countries as Colombia. More credit limit usage leads to a lower probability of repurchase because of less available credit limit and/or a more conservative attitude towards spending. Customers with loan durations of at least three years are less likely to repurchase during the first year than those with loan durations of 12 or 31 months. This may result from greater purchases made in the long term, which do not justify additional spending during the first year. A first purchase that includes products that may be considered luxury goods, leads to a lower probability of repurchase during the first year than that of traditional goods. These products may be valued as major investments that hinder customers from making further purchases. According to model REP1a, customers with secondary education are more likely to repurchase than those with missing ~~level of studies~~values. In model REP1b older customers are more likely to repurchase than younger customers, whereas customers with missing marital status are less likely to repurchase than singles. These results might be related with a customer's life stage.

REP2: Pr (repurchase during t=30 months)

According to models **REP2**$_{a,b}$, in the long term customers that take *longer term loans*, use more their *credit limit,* have completed *primary studies* or have more *dependants* are less likely to repurchase. Those variables imply more financial commitments and hence less purchasing power. Likewise, customers with missing marital status may not consider a priority buying new household products. On the other hand, customers from older *age groups*, *urban areas* or higher *socioeconomic stratums* are more likely to repurchase in the long term. These characteristics usually have attached a higher purchasing power resulting from better economic conditions due to life cycle or location. Results for *type of product* are similar to those of the previous section. If a customer works in the manufacturing industry, repurchase is less likely compared with a customer from the services industry. This might result from different economic conditions and wages in both sectors. An additional feature of Model **REP2**$_b$ is that retired customers are more likely to repurchase in the long term than employed customers. This may derive from a greater purchase capacity of this segment at that stage of their lives.

Repurchase in the short term versus long term

A more parsimonious model for the short term compared with that of the long term confirms that additional attributes account for the repurchase behaviour of customers. However, the majority of significant variables in the short term are significant as well in the long term. Attributes such as *stratum, loan duration, credit limit usage* and *type of product* are relatively stable in the long term, unless a major shift in the socioeconomic conditions of a customer occurs. Given the significance of the variable *stratum* in the short and long terms, the effect of socioeconomic differences between individuals is evident. These results suggest that regardless of their lower purchase capacity, belonging to poor stratums not necessarily implies default. These findings provide further support to inclusive lending strategies. Results for the AUROC of each model are shown on **Table 2.**

MONETARY AND RELATIVE SCORECARDS

Revolving credit is a long term product that requires time to yield profits and returns. Profits (returns) accumulate ~~derive~~over time ~~-~~from purchase commissions and interest~~s~~ payments that might be as agreed or in excess as a result of default~~ throughout time~~. Therefore, a long term perspective (t=30 months) was taken to produce monetary and relative scores. Direct scores were obtained by using the predictor variables shown on **Table 1.** The probabilities of default and repurchase from the previous section were used as predictors in the indirect models. Four indirect models were obtained per measure:

$$Y \sim f\,(x, z) \tag{5}$$

where

$$Y = EBITACUM_{30},\ ROACUM_{30} \text{ and } x = DEF1,\ DEF2;\ z = REP1,\ REP2 \tag{6}$$

~~Scores resulted from using~~ Ordinary Least Squares Regression (OLS) was used, ~~-~~ ~~S~~with stepwise ~~OLS~~selection of variables (~~was used initially and subsequent iterations were run considering a~~0.01 significance level of entry ~~of variables of 0.01~~. Given that profits or losses only occur when customers actually purchased products, zero intercept models were used (i.e.: a baseline profit or loss at t=0 cannot be assumed). See **Table 3.** The error rate was calculated per model to compare their accuracy of prediction as:

~~Error rate = [∑(percentage errors) ²]^1/2~~ (7)

$$\mathit{Error\ rate} = \left[\sum_{i=1}^{n}\left(\frac{Y_i - \hat{Y}_i}{FBdef_i}\right)^2\right]^{1/2} \times 100\%$$

where:

~~Percentage error$_x$ = ∑ (actual score – predicted score)/ ∑ (outstanding balance) (8)~~

where:

~~x= overpredicted or underpredicted values~~

Y_i ($\hat{Y}_i$) = actual (predicted) score for customer *i*

*FB*def$_i$ = deflated final balance at t=30.

Comment [g1]: Did you use stepwise for indirect models too?

Comment [g2]: For ROACUM have you also divided errors by balance?

Formatted: HEADING1, Font: Palatino Linotype, 11.5 pt, Lowered by 21 pt

Comment [g3]: if you sum squared errors, it does not matter if under- or over-prediction occurs, does it?

Formatted: Font: Italic, Subscript

Formatted: Lowered by 6 pt

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic, Subscript

This was done to provide a common basis of comparison in both the training and holdout samples. The minimum error rate considered was 0%, which is the ideal case when predicted scores coincide with actual values.

Following a similar rationale to that used to produce default and repurchase models, approaches *a* and *b* were followed for modelling purposes of $EBITACUM_{30}$ and $ROACUM_{30}$.

Under approach *a*, ~~a training sample was used to develop~~ direct and indirect profit (return) scorecards were developed on training$_2$ sample (because of the outliers, as explained earlier). These models were tested on ~~a~~ holdout$_2$ to facilitate the direct comparison of results between samples. ~~sample. Initially, training sample$_1$ was used for modelling purposes of $EBITACUM_{30}$ and $ROACUM_{30}$. However, as explained before, outliers affected significantly the error rate particularly in the case of relative scores. Rather than using transformations and hence making assumptions about the actual values of the predicted scores, two alternatives were used to test accuracy of prediction. First, outliers from both variables (5% of the customers in this sample) were excluded from the training and holdout samples. This resulted in 24,617 and 6,186 customers in training sample$_2$ and holdout sample$_2$, respectively. This facilitated the direct comparison of results between samples.~~ Yet to test the predictive performance of models under extreme conditions in presence of unusual observations (a more realistic scenario), they were applied to holdout$_3$ (holdout$_2$ + all outliers, 9,347 customers). ~~Second, outliers from both variables (5% of the customers in this sample) were excluded from the training sample and were subsequently included in the holdout sample. This avoided excluding any observation and assured testing the models under more extreme conditions. In total, 9,347 customers were left in holdout sample$_3$, respectively.~~

Under approach *b*, a ten-fold process was ~~followed~~repeated to produce overall models for monetary (relative) direct and indirect scorecards on training$_2$ and holdout$_2$ combined (30,803 customers)~~.~~, ~~Outliers were excluded from the overall sample to develop the ten-fold model~~ in order to be consistent with the standpoint taken in approach *a*. This resulted in an overall sample of ~~30,803 customers~~.

~~Coefficients of significant variables are not displayed for confidentiality reasons.~~

Comment [g4]: Have you also trained the models on holdout1? If yes, can we say the results are avaialbr on request?

Formatted: Not Superscript/ Subscript

Formatted: Subscript

Formatted: Not Superscript/ Subscript

MONETARY SCORECARDS

EBITACUM$_{30}$: Direct scorecards

In models $\mathbf{E1}_{a,b}$, monetary profits increase when customers are not *single~~s~~ (*excluding *divorced)*, belong to older *age* groups, live in *urban areas*, belong to higher *socioeconomic stratums*, have a higher *education level*, stay longer in the same *address* or *purchase products* different to the traditional portfolio. Most of these variables can be understood by relating them ~~as well~~ to higher probabilities of default and/or repurchase in the short and/or long term, as explained before. Likewise, profits increase as the *loan duration* increases. Even though they are less likely to repurchase, in the long term they have constantly paid more interests than if they initially took shorter term loans. Customers in the intermediate *credit limit usage* category are less profitable than those from the lowest category. This might result from their lower probability of repurchase in the short term. The two variables with the highest positive ~~economic impact~~effect are *age* and *loan duration*. The least ~~economically~~ significant variable is the *type of product*. Even though the variable stratum does suggest that wealthier stratums yield higher profits, it is not the most ~~economically~~ significant variable. Its significance derives from repurchase rather than default. Furthermore, the variable *type of job* was not significant. Finally, the significance of other variables that apply to any customer regardless of his/her socioeconomic *stratum* impacts as well the overall profit per customer. Therefore, results reflect the inclusive lending nature of the programme.

EBITACUM$_{30}$: Indirect scorecards

Regardless of the time horizon, in models $\mathbf{E2}_{a,b}$ to $\mathbf{E5}_{a,b}$ profits increase as the probabilities of default and repurchase increase. These results are consistent with the nature of the credit programme, as explained before. The coefficients of probability of default vary between 32 and 69 times that of the probability of repurchase. Interest~~s~~ payments from default accumulate throughout time and may occur more than once; commissions from repurchases are received only at the time of each purchase. **Table 3** shows that the best performing models were $\mathbf{E2}_{a,b}$ and $\mathbf{E3}_{a,b}$, which have as predictors the probabilities of default at t=12 and of repurchase during t=12~~,~~(30) months. These results suggest that for profit scoring purposes, it is adequate to follow the payment behaviour of customers during the first year of the observation period; this agrees with usual banking practices to predict default. In contrast, repurchase behaviour should be followed throughout the observation period. This follows from the definition of a revolving credit, which allows repurchases in the short and long terms.

Overall results in **Table 3** show that direct model E1 $_{a,b}$ outperforms indirect models **E2**$_{a,b}$ and E3$_{a,b}$ in terms of the error rate. This might be the result of additional errors included in the probabilities of default and repurchase when used as predictors for indirect models. Therefore, they should be used to score customers based on their monetary profits.

RELATIVE SCORECARDS

$ROACUM_{30}$: Direct scorecards

The following significant variables in direct monetary (**E1**$_{a,b}$) and relative (**R1**$_{a,b}$) profit models were consistent in sign: *location, marital status*, *stratum, studies, loan duration (*between 42 and 55 months and missing *loan durations), credit limit usage and type of product*. Therefore, specific segments of customers are profitable both in monetary and relative terms. Other significant variables increase the return per customer in models **R1**$_{a,b}$: having any type of *contract*, being *self-employed*, having a marital status different to single and *loan durations* between 36 and 37 months. Given the similar coefficients obtained for contract and being self-employed, it seems that the additional risk taken when granting credit limits to individuals that usually would be considered higher risks pays off when compared against that of individuals under more stable conditions. Specific significant variables in model **R1**$_{a}$ that increase relative profits include customers with secondary level of studies and those with more dependants. In the case of **R1**$_{b}$, customers with education level different to missing increase relative profits at a decreasing rate as the education level increases. Finally, customers that work in other industries different to services or not formally employed lead to higher relative profits. This is a positive feature of the inclusive lending programme under analysis.

$ROACUM_{30}$: Indirect scorecards

In Indirect models **R2**$_{a,b}$ to **R5**$_{a,b}$, returns per customer increase when the probabilities of default and repurchase increase. Results for the best performing models (**R2**$_{a,b}$ and **R3**$_{a,b}$ in **Table 3**) are consistent with those for indirect profit scorecards. These results suggest that regardless of the profit measure used, default requires a shorter observation period than repurchase. They confirm as well that current credit policies of the lending institution are very conservative, as the payment behaviour of utility bills is tracked for two years. Depending on the management attitude towards the credit risk implied in this inclusive lending programme, more customers could access it if the observation period for default is reduced to 12 months. **Table 3** shows that direct model $\mathbf{R1}_{a,b}$ outperforms indirect models $\mathbf{R2}_{a,b}$ and $\mathbf{R3}_{a,b}$. Consistent with results for EBITACUM$_{30}$, direct models should be preferred.

IMPACT ON PORTFOLIO RESULTS

Based on the results presented above, models $\mathbf{E1}_{a,b}$ and $\mathbf{R1}_{a,b}$ were chosen to assess the impact of monetary and relative scores on portfolio profits and returns. It would not be reasonable to compare indirect (underperforming) models with direct ~~(overperforming models)~~ones. ~~Figur~~Values are not displayed for confidentiality reasons.

Formatted: Font: Bold

Figure 3 shows the portfolio profits and returns per acceptance band if models E1$_a$, R1$_a$, DEF1$_a$ and DEF2$_a$ are used to score customers in holdout~~ sample~~$_3$. Results for holdout~~ sample~~$_2$ and training~~ sample~~$_2$ showed a similar behaviour and are available upon request. It is clear that portfolio profits (returns) are higher when customers are accepted according to monetary (relative) scores. This was expected, as scorecards were designed based on the same measures at a customer level. The vertical distance between the curves is the opportunity cost resulting from scoring customers with the least beneficial alternative in terms of portfolio profits or returns. As the acceptance rate increases, more customers are accepted and the opportunity cost decreases until it makes no difference to use either monetary or relative scores, as there is no impact on portfolio results. Additionally, results show that monetary (relative) scorecards outperform default scorecards at t=12 and t=30 months. This shows that portfolio results can be improved if monetary (relative) profit scorecards are used instead of traditionally implemented default scorecards. Furthermore, **Table 4** shows that Spearman correlations between direct monetary (relative) profit scores and default scores at t=12 and 30 months are weak. Therefore, monetary (relative) profit scorecards provide an alternative perspective to default scoring. The complete sample including outliers (training$_1$+holdout$_1$) was used to assess the impact on corporate results under approach *b*. Results were similar to those of approach *a*. See **Figure 4** and **Table 4**.

Regardless of the type of score used, the same number of customers is accepted for same values of the acceptance rate. Therefore, it is a question of adopting the scorecard that is more appropriate according to portfolio objectives in terms of profits, coverage against default and the scope of the programme.

In the case under analysis, some reasons would justify the adoption of monetary scores. First, the number of defaulters is very low compared with the total sample. Second, given that all the customers that were granted a credit limit paid on time their utility bills, the perceived credit risk is lower and hence coverage against default is not a priority. Third, the payment behaviour of customers has been stable regardless of adverse ~~climate~~economic conditions that could eventually affect default rates. Finally, the additional cash generated by adopting such strategy results in more funds to grant further credit to new customers from the existing identified segment. If relative scores are used, foregone profits would result in fewer funds available to be granted to additional customers. This would reduce the scope of the credit programme.

Relative scores would be useful for the lending institution under different conditions. First, if the Company's objectives do not consider further growth in the served segment (i.e.: customers with a payment record of utility bills over the last two years) and instead prioritise the allocation of funds among customers that are potentially more "efficient" in relative terms. The portfolio of receivables would be healthier in terms of coverage against default. This would be perceived positively by stakeholders and regulatory authorities that ~~are adverse towards~~may be cautious about inclusive lending, regardless of the customers' previous utility payment record. Second, if more uncertainty arises as a result of socioeconomic or political instability then the aim would be to achieve higher rates of coverage against default. Finally, relative scores could be particularly helpful when the strategy is more ambitious in terms of inclusive lending. That is, when the aim is to diversify the

Comment [g5]: ?

Formatted: Subscript

Formatted: Subscript

existing portfolio and penetrate new segments that are not being currently served and hence which are completely unknown in terms of credit records.

CONCLUSIONS

Based on the sample used in the analysis, it was shown that inclusive lending can be a profitable business (i.e.: almost all customers were profitable in monetary and relative terms). It also shed light on the impact of specific customer attributes on portfolio profits and returns; this may apply to similar contexts where lending institutions have a major role on responsibly increasing the access to credit of traditionally excluded segments.

On the other hand, it showed that using relative and monetary profit scores is useful to further understand profit scoring for revolving credits. From a statistical perspective, it was confirmed that direct models score customers more accurately than indirect models both in monetary and relative terms. This extends the existing literature, as previous studies solely used monetary profit scores. Percentage differences between the error rates of direct and indirect models should not be overlooked as in monetary terms such difference is not negligible given the portfolio value of outstanding debts. Such differences stand for the foregone monetary or relative profits if indirect models are used instead. Therefore, from an economic point of view, direct scores should be preferred. Indirect models are useful, however, as they provide useful insight to understand the significance that the probabilities of default and repurchase have on the generation of both monetary and relative profits. Consistent with the long term perspective of CLV for revolving credits, it was shown that the probability of repurchase should not only be modelled in the short term but also in the long term. An additional reason to predict default and repurchase prior to producing direct profit scores is that they provide useful insight to interpret results from significant attributes in more comprehensible ways to practitioners.

From a financial perspective, additional insight was gained as the majority of variables were consistently significant in direct models for monetary and relative scores. Hence it is possible to identify segments of inclusive lending programmes that are profitable both in monetary and relative terms. This justifies taking the additional perceived risk. This is particularly useful to partially tackle the dilemma between maximising portfolio profits and returns. Choosing between monetary and relative profit scores implies a trade off between portfolio profits and returns depending on the chosen scorecard. Other significant variables account for the difference between portfolio results when using either measure. The choice will be guided by the Company's objectives and assessment policies of the lending institution. If financial results are usually assessed in terms of monetary profits regardless of the required investment (i.e.: as a profit centre), monetary scores will be preferred. This approach is less conservative as it does not account for default, which can eventually occur. It is therefore more convenient to implement it in times of socioeconomic stability and for segments that are similar in terms of purchase and credit payment behaviour. This will not compromise the ethical standpoint of the lending institution in terms of responsible lending. Still there is a foregone coverage against default in the event that it occurs.

Comment [g6]: I'm afraid I do not quite understand the meaning - earlier the paper says that significant variables are similar?

On the other hand, if credit granting policies are more conservative and hence aim at a better coverage against default, relative measures should be preferred. This might be the case of credit programmes that operate under stricter budget restrictions in terms of granted credits and hence are assessed as investment centres. They are also useful when socioeconomic conditions are more unstable. Furthermore, even though relative scores do not solve the ethical dilemma regarding profit scoring, they tackle it better than monetary scores. Customers are ranked considering not only the cash generated as a result of default and repurchase, but also the greater outstanding balance resulting from them. The above considerations apply to any revolving credit. However, they are particularly useful to prevent overindebtness in high risk segments and hence to enhance more socially responsible inclusive lending practices.

Further avenues of research include using mixed methods to complement results from quantitative models with qualitative data analysis. This will allow identifying reasons behind default and repurchase that ultimately drive customers to be more or less profitable in monetary and/or relative terms. A complementary approach could be to design behavioural scorecards that include credit bureau variables once customers start using their credit limit. This offers the benefit of assessing the impact of inclusive lending in the design of monetary and relative profit scores.

Figure 1: Histograms, $EBITACUM_{30}$ and $ROACUM_{30}$ (holdout ~~sample~~$_2$, training ~~sample~~$_2$)

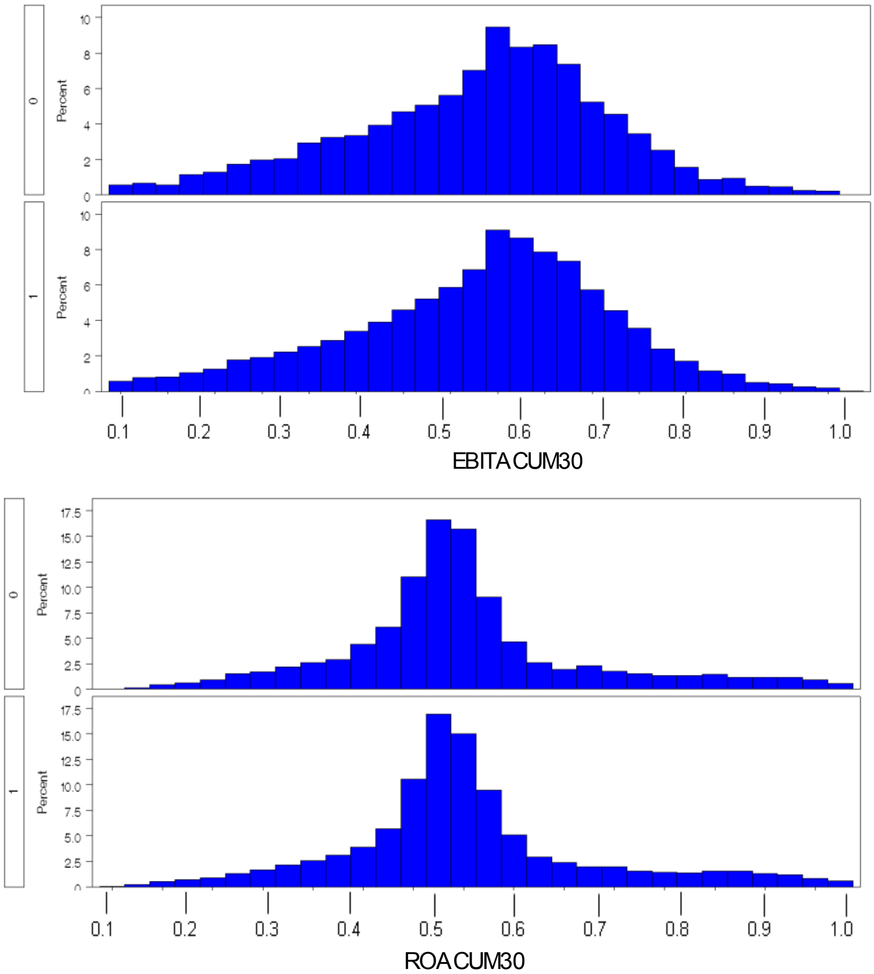


Figure 2: EBITACUM30 versus ROACUM30 (training ~~sample~~ [2])

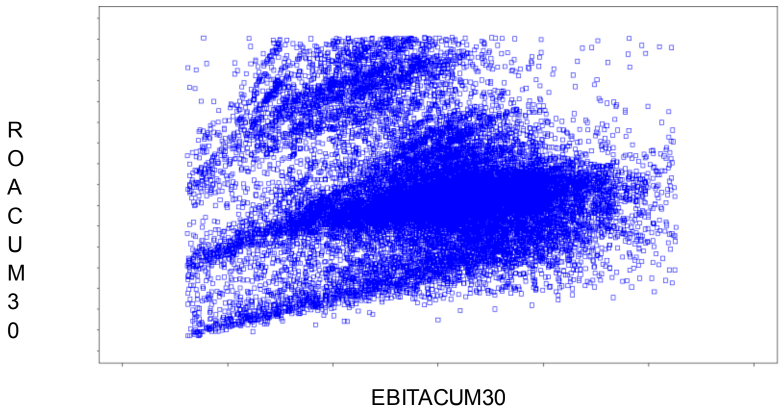
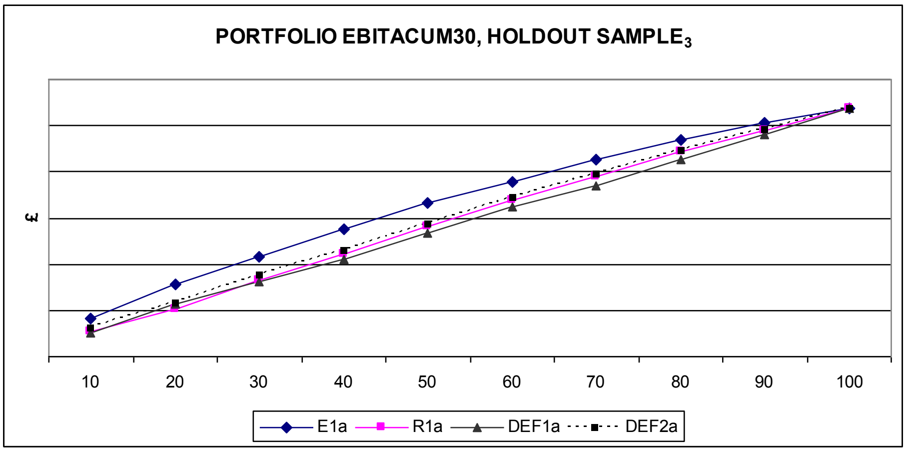


Figure 3: IMPACT OF BEST PERFORMING SCORECARDS ON PORTFOLIO RESULTS, APPROACH *a*



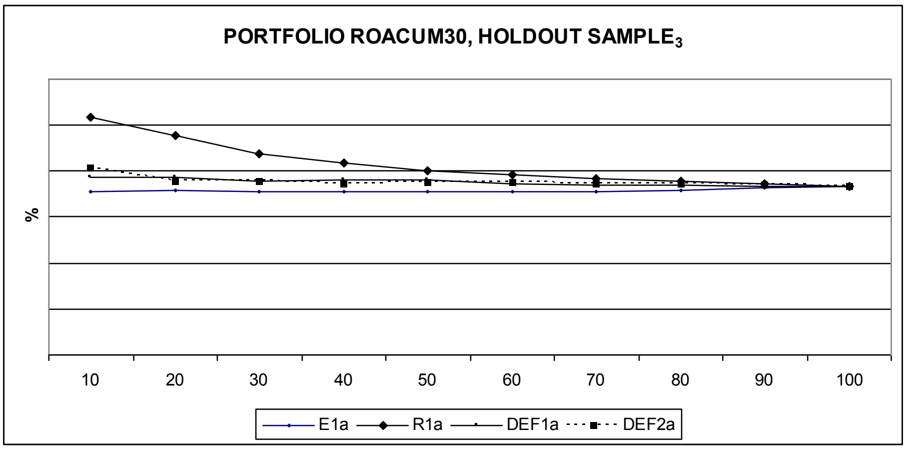
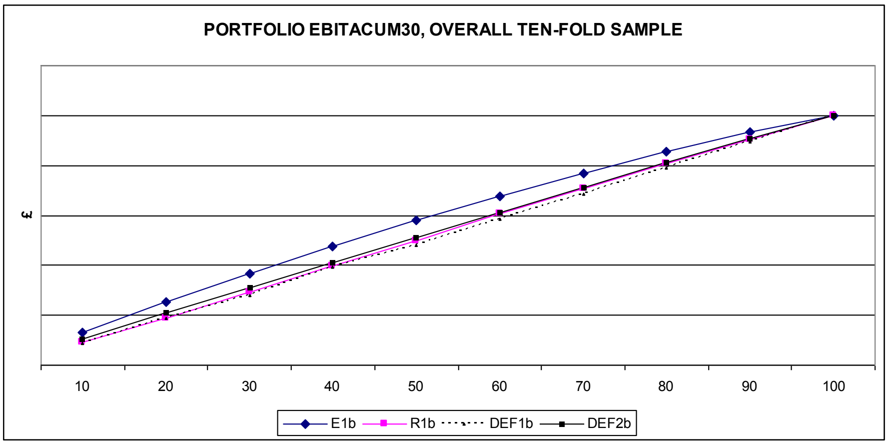


Figure 4: IMPACT OF BEST PERFORMING SCORECARDS ON PORTFOLIO RESULTS, APPROACH *b*



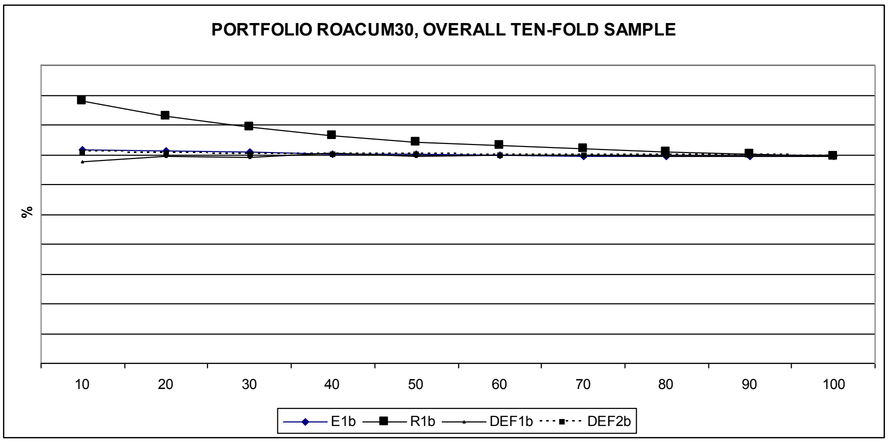


Table 1: Predictor variables at the time of first purchase

VARIABLE	REFERENCE CATEGORY	DUMMY VARIABLES
AGE	18<Age ≤ 35 years	dumAGE3: 35< Age ≤ 43.5 years dumAGE4: 43.5< Age ≤ 52 years dumAGE5: 52< Age ≤ 60.5 years dumAGE6: 60.5< Age ≤ 69 years dumAGE7: 69< Age ≤ 103 years
LOCATION	rural (different to the capital city)	dumCITUR : urban (capital city)
CONTRACT	missing, other, or not applicable	dumCONTCON : Any type of contract (permanent, temporary)
JOB	employed	dumJOBRET : retired dumJOBSELF: self-employed dumJOBNOIN : housewife, student, unemployed, missing
MARITAL STATUS	single	dumMARMAR : married dumMARCOH : cohabitators dumMARWID : widow(er) dumMARDIV: divorced dumMARMIS: missing
STRATUM	stratum 1 (poor segments)	dumSTRA35: stratum>1
EDUCATION	missing	dumSTUPRI : primary dumSTUSEC: secondary dumSTUCOL : college dumSTUHIG: higher
DURATION FIRST LOAN	durloan ≤ 31 months	dumLOAN3637: duration=36 or 37 months dumLOAN4243: duration=42 or 43 months dumLOAN4855: 48≤ duration ≤55 months dumLOAN6061: duration=60 or 61 months dumLOANMIS : missing loan duration
YEARS AT ADDRESS	YAH ≤ 8.5 years	dumYAH2 : 8.5< YAH ≤ 18 years dumYAH3 : 18< YAH≤ 27.5 years dumYAH4 : 27.5< YAH ≤ 37 years dumYAH510 : 37< YAH ≤ 94 years
DEPENDANTS	No dependants	dumDEP1 : 1 dependants dumDEP2 : 2 dependants dumDEP3 : 3 dependants dumDEP4 : 4 dependants dumDEP510 : 5 or more dependants
CREDIT LIMIT USAGE	Low	dumLOANPR2 : intermediate dumLOANPR310 : high
ACTIVITY	Services	dumactNA : Not applicable dumactOTH : Other industries dumactPROD : Manufacturing
FIRST PRODUCT PURCHASED	traditional products	dumprod1 : Non-traditional category 1 dumprod2 : Non-traditional category 2 dumprod3 : Non-traditional category 3

Table 2: Performance statistics for default and repurchase models

	APPROACH a								APPROACH b			
	TRAINING$_1$				HOLDOUT$_1$				10 FOLD COMPLETE SAMPLE			
MODEL	n	goods	bads	AUROC	n	goods	bads	AUROC	n	goods	bads	AUROC
DEF1	27,157	26,567	590	0.60	6,807	6,665	142	0.59	33,964	33,232	732	0.60
DEF2	27,157	26,796	361	0.61	6,807	6,719	88	0.65	33,964	33,515	449	0.62
REP1	27,157	3242	23915	0.70	6,807	809	5998	0.70	33,964	4051	29913	0.70
REP2	27,157	7284	19873	0.71	6,807	1819	4988	0.71	33,964	9103	24861	0.71

Table 3: Direct and indirect models

MODEL	Description	Model composition	ERROR RATE			
			APPROACH a			APPROACH b
			$TRAINING_2$	$HOLDOUT_2$	$HOLDOUT_3$	10 FOLD OVERALL SAMPLE
E1	**DIRECT, EBITACUM30**	Age, location, marital status, stratum, education, loan duration, years at address,credit limit usage and product	9%	9%	17%	9%
E2	INDIRECT, EBITACUM30	Pr(default at t=12), Pr(repurchase up to t=12)	15%	15%	25%	15%
E3	INDIRECT, EBITACUM30	Pr(default at t=12), Pr(repurchase up to t=30)	15%	15%	25%	15%
E4	INDIRECT, EBITACUM30	Pr(default at t=30), Pr(repurchase up to t=12)	19%	19%	28%	20%
E5	INDIRECT, EBITACUM30	Pr(default at t=30), Pr(repurchase up to t=30)	19%	18%	28%	19%
R1	**DIRECT, ROAACUM30**	Location, type of contract, job, marital status, stratum,education, loan duration,dependants, credit limit usage,product and activity	13%	12%	17%	14%
R2	INDIRECT, ROAACUM30	Pr(default at t=12), Pr(repurchase up to t=12)	17%	17%	22%	17%
R3	INDIRECT, ROAACUM30	Pr(default at t=12), Pr(repurchase up to t=30)	17%	17%	22%	17%
R4	INDIRECT, ROAACUM30	Pr(default at t=30), Pr(repurchase up to t=12)	19%	18%	25%	20%
R5	INDIRECT, ROAACUM30	Pr(default at t=30), Pr(repurchase up to t=30)	18%	18%	24%	19%

Table 4: Spearman correlations, profit (return) scorecards and default scorecards

APPROACH *a*				APPROACH *b*	
Scores	Training sample	$Holdout_2$	$Holdout_3$	Complete sample	Sample without outliers
(DEF1,Ebitacum30)	0.2	0.2	0.1	0.2	0.2
(DEF1,ROAcum30)	0.0	0.0	0.0	-0.2	-0.2
(DEF2,Ebitacum30)	-0.1	-0.1	-0.1	-0.1	-0.1
(DEF2,ROAcum30)	-0.2	-0.2	-0.2	-0.3	-0.3

REFERENCES

- Andon P, Baxter J and Graham B (2001). Calculating the Economic Value of Customers to an Organisation. *Australian Accounting Review* **11**(23): 62-72.
- Andreeva G, Ansell J and Crook J N (2005). Modelling the Purchase Propensity: Analysis of a Revolving Store Card. *The Journal of the Operational Research Society* **56**(9): 1041-1050.
- Andreeva G, Ansell J and Crook J (2007). Modelling profitability using survival combination scores. *European Journal of Operational Research* **183**(3): 1537-1549.
- Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J and Vanthienen J (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society* **54**(6): 627-635.
- Berger P D and Nasr N I (1998). Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing* **12**(1): 17-30.
- Collings D and Baxter N (2005). Valuing customers. *BT Technology Journal* **23**(3): 24-29.
- Drury C. (2000). *Management and Cost Accounting*. Business Press: London.
- Engels P (2010). *Mission Drift in Microfinance- The Influence of Institutional and Country Risk Indicators on the Trade-Off between the Financial and Social Performance of Microfinance Institutions.* PlaNet Finance Deutschland e.V.: Stuttgart.
- Finlay S M (2008). Towards profitability: a utility approach to the credit scoring problem. *Journal of the Operational Research Society* **59**(7): 921-931.
- Finlay S (2010). Credit scoring for profitability objectives. *European Journal of Operational Research* **202**(2): 528-537.
- Fitzgerald L. (2007). *Issues in Management Accounting*. Pearson Education: Harlow.
- Gupta S and Lehmann D (2003). Customers as assets. *Journal of Interactive Marketing* **17**(1): 9-24.
- Hand D J and Henley W E (1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **160**(3): 523-541.

- Hirsch M. (2000). *Advanced Management Accounting*.Thomson Learning: London.
- Li H G and Hand D J (2002).Direct versus indirect credit scoring classifications.*The Journal of the Operational Research Society* **53**(6): 647-654.
- Lieli R P and White H (2010). The construction of empirical credit scoring rules based on maximization principles. *Journal of Econometrics* **157**(1): 110-119.
- Lin, S. M, Ansell J and Andreeva, G (2012). Predicting default of a small business using different definitions of financial distress. *Journal of the Operational Research Society* **63**(4): 539-548.
- Lucas A (2001). Statistical challenges in credit card issuing. *Applied Stochastic Models in Business and Industry* **17**(1): 83-92.
- Ma P, Crook J and Ansell J (2009). Modelling take-up and profitability. *Journal of the Operational Research Society* **61**(3): 430-442.
- Pfeifer P E, Haskins M E and Conroy R M (2005). Customer Lifetime Value, Customer Profitability, and the Treatmen of Acquisition Spending. *Journal of Managerial Issues* **17**(1): 11-25.
- Prior F and Argandoña A (2009).Best Practices in Credit Accessibility and Corporate Social Responsibility in Financial Institutions. *Journal of Business Ethics* **87**:251–265
- Rasiah D (2010). Review of Literature and Theories on Determinants of Commercial Bank Profitability. *Journal of Performance Measurement* **23**(1): 23-49.
- Rosenberg E. and Gleit A (1994). Quantitative methods in credit management: A survey. *Operations Research* **42**(4): 589.
- Ryals L and Knox S (2005). Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value. *European Journal Of Marketing* **39**(5/6): 456-472.
- Thomas L C (2009). *Consumer credit models : pricing, profit, and portfolios*. Oxford University Press : Oxford.

- US Senate (1979). *Credit card redlining : hearings before the Subcommittee on Consumer Affairs of the Committee on Banking, Housing, and Urban Affairs, United States Senate, Ninety-sixth Congress, first session on S.15.* US Govt. Print Off.: Washington.
- Wilkinson G (2011). Social Rental Payment Data- Its value in improving credit decisions and scoring processes; moving from proof of concept to implementation.

  http://www.crc.man.ed.ac.uk/conference/archive/2011/Wilkinson-George-Presentation-Social-Rental-Payment-Data-Wilkinson.pdf, accessed 4 January 2012.