Financial performance measures in credit scoring

Robert M. Oliver

Received: 18 December 2012 / Accepted: 3 September 2013 / Published online: 22 October 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract In this paper, we suggest that financial performance of loans and loan portfolios, rather than statistical performance of scores, should be the critical focus for lenders and lending institutions using credit scoring to fund and acquire new borrowers. We propose two measures for measuring relative financial performance of loan accounts when portfolio acquisition decisions are based on risk assessments. These compare with numerous statistical performance measures for response and risk scores that assess borrower response to offers, default, bankruptcy, late payment and fraud. Popular statistical measures often used in scorecard development and validation testing include the K-S statistic, Gini coefficient, AUROC, ROC curves, divergence, and others. The measures that we suggest focus on relative return on equity (ROE) and market share; in our opinion, such financial performance measures are more meaningful to lenders and businesses than statistical measures of scorecard performance.

Keywords Credit scoring · Financial performance · ROC curves · Loan portfolios · Retail lending

List of symbols

$p(B \mathbf{x}), p(B s(\mathbf{x}))$	Probability of default (Bad) conditional on data, $\mathbf{x}$, or score, $s(\mathbf{x})$
$p_G = p(G), p_B = p(B)$	Unconditional population priors
$f(s G), f(s B), f(s)$	Conditional, marginal score densities
$F(s G), F(s B), F(s); F^{-1}(s)$	Conditional, marginal score cumulative distributions; inverse function

R. M. Oliver (✉)
University of California at Berkeley, Berkeley, USA
e-mail: rmo@Berkeley.edu

$F^{(c)}(s G) = 1 - F(s G),$	Tail distribution
$F^{(c)}(s B) = 1 - F(s B),$	
$F^{(c)}(s) = 1 - F(s)$	
$c_1, c_2; C_M$	Unit costs for misclassification in Type I, II errors; random total misclassification cost
r_L, l_D, c_D, E	Loan rate for borrower, loss given default, commercial borrowing rate, equity level for lender
$r_E, r_A; r_F$	Random returns on equity, assets; risk-free rate
$s_c, s_c^*, \hat{s}_c$	Score cutoff, optimal score cutoff, optimal unconstrained score cutoff
η	Lender's financial tradeoff between expected net profit of Good and cost of Bad (default)
$\mathbb{E}[x], \mathbb{E}[x y]; \mathbb{E}_{PI}$	Unconditional, conditional expectations; expectation under perfect information
$V_0; \hat{V} = \mathbb{E}[V(\hat{s}_c)]$	Volume (market share) constraint; expected volume for unconstrained optimal ROE
λ, λ^*	Shadow price for volume (market share) constraint, optimal shadow price on efficient frontier

Introduction

In the early days of retail credit much of the attention of lenders and credit managers was focused on simple “knock-out” rules that denied credit to the lender’s “identification” of a high-risk individual; lenders wanted to avoid large default rates and default losses; little attention was given to improving profit margins or return on equity. As usage of risk scores proliferated, statistical methods for testing and measuring their discriminatory power improved. There is a substantial literature on credit scoring, risk ranking and classification models as well as techniques for comparing and validating their statistical performance. See Hand and Henley (1997) and Thomas (2009). However, there is a much smaller body of literature on the measurement and validation of economic/financial outcomes in portfolios whose acquisition decisions are based on credit scores.

The major interest in the use of scores was to give a lender (decision-maker) a simple and reliable way to rank the odds of payment delinquency or default and to provide ways to discriminate accounts likely to make late payments or default from those that would make timely payments of amounts due. It was found that scores dramatically improved the ability of lenders to rank the likelihood of late payments, delinquencies and default. Even though few lenders had explicitly formulated models of profitability or market share there was general agreement that the better one could discriminate the relative odds of default the more one could control extreme losses. Simple policies based only on lender judgment and experience gave way to more sophisticated acquisition and pricing models based on combining judgment and experience with statistical assessments of default, delinquency, prepayment and response to offers. As scoring techniques were formally

incorporated in models for decision-making under uncertainty, it became clear that policies could be designed to achieve efficient frontiers and even calculate the shadow prices associated with tradeoffs between ROE, market share and default losses. There are at least two reasons why financial performance measures for loan accounts differ from statistical performance of scores used in the decision-making process: the first is that data on costs, losses, revenue, equity, pricing, capital reserves and lender decisions play a critical role in measuring financial performance; the second is that in those situations where adverse selection is an important consideration, risk scores, themselves, are usually affected by financial terms of loans and lender offers.

Consider the lender's asymmetric decision tree in Fig. 1. The Good or Bad outcomes among the accepted borrowers are illustrated in the bottom branches with payoffs to the lender at the end of each branch. In making financial decisions on which borrowers to accept or reject, the relevant costs/returns are associated with realized returns and default losses among the *Accepts*, i.e. those borrowers booking loans by the lender. Unless one pursues extremely expensive experiments to infer performance of unobservable Goods and Bads among the *Rejects*, known as *Reject Inference*, it is seldom possible to assess the economic consequences of hypothetical Goods and Bads in the top branch.

The traditional decision model used to derive an optimal cutoff policy to maximize expected ROA (return on assets) or ROE (return on equity) for the lending process is well documented in the credit scoring literature. See Lewis (1992), Hoadley and Oliver (1998), Oliver and Wells (2001), Trench et al. (2003), Beling et al. (2005), Stein (2005) and Thomas (2009). For risk-based pricing models see Keeney and Oliver (2005), Phillips and Raffard (2010) and Oliver and Oliver (2012). In Fig. 1, the payoffs to the lender (from top to bottom) for a unit loan are

$$\begin{aligned} &0 && \text{if the borrower is rejected,} \\ &(r_L - c_D) && \text{if the booked loan, at rate } r_L, \text{ does not default, labelled G,} \\ &-(l_D + c_D) && \text{if the booked loan, with LGD } l_D, \text{ defaults, labelled B.} \end{aligned} \quad (1)$$

The commercial borrowing rate for the lender to source the loan is c_D , the loan rate offered the borrower by the lender is r_L with $r_L \geq c_D$, and the loss given default, usually written as LGD, is denoted by $0 \leq l_D \leq 1$. The probability of non-default, $p(G|s)$, and the probability of loan default, $p(B|s) = 1 - p(G|s)$, depend on a risk score, s , and are displayed on each branch corresponding to the random outcome for each accepted borrower (see “List of symbols” for a brief summary of

Fig. 1 The lender's accept/reject decision tree

notation). In the case of a Reject decision by the lender, the loan is not booked and the certain payoff is zero. Note that in the event of borrower default the lender is still obligated to repay (with interest) the commercial loan used to source the retail loan. The assumptions of this one-period decision model can be made more realistic (and complex) in operating environments by including uncertain borrower response, multiple risk level with different pricing tiers, i.e. risk-dependent loan rates, adverse selection from borrowers, the presence of complex competitive environments with pricing pressures from other lenders, the effects of dynamic changes in systemic risk and the inclusion of risk-dependent capital reserve requirements. More realistic assumptions and complexities notwithstanding, this simple model can be used to illustrate some important differences between measures based on the quality and statistical performance of scores with measures based on financial performance of a loan portfolio. A slightly more realistic model in which the objective is ROE rather than ROA necessitates the inclusion of equity capital, the risk-free rate and commercial borrowing by the lender to provide the source of funds for retail loans. Finally, we consider the expected financial performance of score policies that yield optimal expected ROE relative to perfect information at a given level of market share.

Financial performance with perfect information on borrower risk

Without any information or risk assessment of the eventual state of each borrower, a lender can randomly accept or reject each individual. Before we consider the effects that scores have upon optimal policies with different business measures, we describe two baseline cases: one is where we select borrowers at random but are unable to predict the default of individual accounts; in the other case we have perfect information on the eventual status of each borrower and are able to order them as a string of Goods (G) followed by a string of Bads (B).

Any discussion of return on equity for a commercial lender requires a specification of the amount of equity and the risk-free return that can be earned by this equity. The additional notation we use in our models for ROE include E for equity capital and r_F for the risk-free interest rate. More sophisticated models for commercial lending may include multiple risk levels for the loan assets, multiple commercial borrowing rates to recognize different loan asset risks and different rates of return on equity or investor capital; for simplicity in model structure we restrict our attention to a single tier. In deriving an expression for the expectation of the return on equity, we again note that the net return to a lender who borrows money to source a unit loan is $r_L - c_D$ for the Good (no default) borrower and a net cost $-(l_D + c_D)$ for a Bad (default). With default the lender must pay $l_D + c_D$ because obligations for borrowed funds require full repayment with interest; i.e. there is no relief to the debt of a commercial lender or bank because of the default of a retail borrower whose loan was sourced by these borrowed funds.

Assume that we have a large population of N prospective borrowers; a fraction of them, p_G , do not default (G: Goods) and a fraction, $p_B = 1 - p_G$, default on their payments (B: Bads). If we accept and book the entire population the expected net

return from the loan portfolio is easily calculated by subtracting the expected default losses of the Bads from the expected revenue derived from the Goods. Assume each Good account provides a net return of $r_L - c_D$ and each Bad has a loss given default (LGD) of $l_D + c_D$. If $r_L p_G - l_D p_B - c_D > 0$, the expected net return to the lender for each unit loan is positive and shown as the upper black dot vertex in the dashed triangle of Fig. 2. When the inequality is reversed, a negative expected net return corresponds to the lower black dot at the vertex of the solid line triangle. If we randomly select $n \leq N$ from the pool of applicants the expected net return increases (or decreases) linearly with positive or negative slope, as shown on the bottom edges of the dashed and solid triangles. Revenues may be large but net returns may be negative because of the presence of a large number of defaults that result in losses. In the bottom triangle we see that regions of negative profit depend on how many accounts are booked and the relative sizes of p_G and p_B .

If the lender is clairvoyant or fortunate enough to have perfect information on the outcome of each individual borrower he or she could order the list of potential borrowers and accept only Goods while rejecting Bads. How does the expected net return for this hypothetical but desirable case compare with an Accept All or Accept None policy? With no booked borrower accounts (zero market share) we start by only accepting Goods; expected ROE, denoted by $\mathbb{E}_{PI}$, increases linearly with the number accepted (slope $(r_L - c_D)/E$) in Fig. 2 until we reach an expected maximum of $Np_G(r_L - c_D)/E$ from the loan assets. At that point we run out of Goods; were we to continue accepting individual accounts *in order to increase market share* we could then only accept Bads until we reach the end-point described earlier. In other words, once the maximum profit at the top vertex of each triangle has been reached, the only growth in market share available to us is an increase in new Bad accounts that are certain to default. The profit would decrease [slope $-(l_D + c_D)/E$ in Fig. 2]; depending on how large l_D is in comparison with the loan rate and the PopOdds of the borrowing population, the end points (when all N borrowers are accepted) would be either negative (solid line end-point) or positive (dashed line

Fig. 2 Expected ROE versus number booked

end-point). Usually, LGD is very large compared to the return on each Good so that the straight line on the right side of each triangle has a steep negative slope compared to the line with positive slope on the left. The region inside each triangle corresponds to sets of feasible acceptance policies that can be based on discriminating Good/Bad credit scores as well as judgment and experience.

In unusual cases the policy is to reject all or accept all prospective borrowers. With perfect information the highest expected ROE is the fraction of Goods multiplied by the net return of each loan or $Np_G(r_L - c_D)/E$. The expected net return of a Reject All or Accept All policy *relative* to the perfect information case is therefore independent of equity E and the number of borrowers, N . We have

$$\begin{aligned} \text{Relative ROE} &= \begin{cases} 0 & \text{if everyone is rejected} \\ \frac{r_L N p_G - l_D N p_B - N c_D}{N p_G (r_L - c_D)} & \text{if everyone is accepted} \end{cases} \\ &= \begin{cases} 0 & \text{if everyone is rejected} \\ \left(1 - \frac{p_B l_D + c_D}{p_G r_L - c_D}\right) = (1 - \eta^{-1}) & \text{if everyone is accepted} \end{cases} \end{aligned} \quad (2)$$

The parameter η , defined by the ratio,

$$\eta \triangleq \frac{p_G r_L - c_D}{p_B l_D + c_D} \quad r_L \geq c_D, \quad (3)$$

is the lender's financial tradeoff between expected net revenue for a Good and expected loss for a Bad. The decision on whether to reject or accept all borrowers depends on whether η is less than or greater than one. What is interesting is that this same tradeoff plays an important role when the decision-maker uses risk scores to maximize expected ROE and that this ratio is the slope of the ROC curve at the optimal cutoff; it is less than or equal the slope for the optimal ROA cutoff.

A scoring predictor and an acquisition rule based on a score cutoff allow us to select portfolio returns and market share that lie in the interior of the triangle in Fig. 1. When we compare the expected profit and market share of a feasible operating point in the interior of the lower triangle, it is natural to compare it with the expected ROE at the top vertex; but a more useful measure might be to compare it with a point on the perfect information line having the same market share. In what follows, we examine both cases in the context of score-based decision rules.

Background and notation for credit scores

In what follows, we denoted the conditional probability of non-default (Good, G) and default (Bad, B) for a booked borrower as:

$$p(G|\mathbf{x}) = \Pr\{\text{Good}|\mathbf{x}\} = p(s(\mathbf{x})) = p(s); \quad p(B|\mathbf{x}) = 1 - p(G|\mathbf{x}) \quad \mathbf{x} \in \mathcal{X} \quad (4)$$

Consider a population of borrowers being assessed by a lender for the purpose of making loans to a subset of qualified individuals. It is common practice to use a log odds risk score to assess the default risk of each prospective borrower. The score is usually based on borrower-relevant financial, behavioral and demographic data; unfortunately, the decisions, offers or terms of earlier loans are seldom included in that data. We use the notation $s(\mathbf{x})$ to denote a log odds score: We know from Bayes' rule that the total log odds score is the sum of two parts, one being the log of the

population odds, independent of data $\mathbf{x}$, the second being the log of information odds that depends on data $\mathbf{x}$:

$$s = s(\mathbf{x}) \triangleq \ln \frac{p(G|\mathbf{x})}{p(B|\mathbf{x})} = \ln \frac{p_G}{p_B} + \ln \frac{f(\mathbf{x}|G)}{f(\mathbf{x}|B)} = s_0 + s_{\text{INF}}(\mathbf{x}); \quad s_0 \equiv \ln \frac{p_G}{p_B}, \quad (5)$$

where the p_G/p_B ratio is often called the population odds or PopOdds, o_{Pop} , and the constant s_0 is known as the PopOdds score. Our notation for the data-dependent information odds score is

$$s_{\text{INF}}(\mathbf{x}) \triangleq \ln \frac{f(\mathbf{x}|G)}{f(\mathbf{x}|B)}. \quad (6)$$

$p(\cdot|\cdot)$ and $f(\cdot|\cdot)$ denote conditional probability and likelihood, respectively. A great deal of thought and expertise has gone into the models and statistical estimation techniques as well as the testing, validation and calibration of risk and response scores. Once each prospective borrower is assigned his or her risk score, it is possible to calculate the marginal and conditional Good/Bad score densities, $f(s|G)$, $f(s|B)$ and $f(s)$. In this article we use standard notation, $F(s|\cdot)$ to denote the cumulative and $F^{(c)}(s|\cdot)$ to denote the tail distributions. Ideally, a well-calibrated score has the sufficiency property

$$p(G|\mathbf{x}) = p(G|\mathbf{x}, s(\mathbf{x})) = p(G|s(\mathbf{x})),$$

which means that only the scalar score is required to predict G or B —in other words, there is a one-to-one mapping with the required conditional probabilities. With a different but more informative data set, $\mathbf{y} \in \mathcal{Y}$, we can expect a different score, $s(\mathbf{y})$, to provide improved statistical and financial performance. Although the algorithms for calculating and estimating scores from historical data are well developed, there remains the perplexing problems of reject inference and estimation of unrevealed Bads. Another difficulty is that the development sample may not include explicitly recognized acquisition policies in use during data collection periods.

Scores must be clearly identified with well-defined outcomes; although it is often done in practice it does not make sense to rely on a late-payment score to predict defaults or to use a default score to predict late-payment behavior. Obviously, the most up-to-date relevant information should be included in the score—an acquisition score should be based not only on behavioral and demographic data relevant to the borrower but also include data that capture the terms of the loan and policy decisions used by the lender. For example, if adverse selection is known to be present the default score for a prospective borrower may be very sensitive to the offer rate being made by the lender.

With sufficiently large sample sizes and fine binning for the relevant predictor variables in $\mathbf{x}$, it is common practice to approximate the conditional scores as normal distributions and then compare fitted odds of validation samples with the score obtained from a development sample. When the variances of the conditional Good/Bad score distributions are equal, it is easy to show that fitted odds curves in validation samples versus development scores are straight lines; because departures from a straight line are easy to detect and interpret, this test provides yet another

way to measure statistical discriminatory power of scores. Perhaps the most frequently referenced statistical measure that quantifies the degree of discrimination offered by a score is the ROC curve.

ROC curves

The ROC curve is a plot of the cumulative conditional distributions, typically with $F(s(\mathbf{x})|B)$ plotted on the Y -axis, against $F(s(\mathbf{x})|G)$ on the X -axis, i.e. a comparison of the cumulative fraction of Bads against the cumulative fraction of Goods for each cutoff score. If we define a proper score as one where the posterior odds of a Good is strictly increasing in score, then it follows that the ROC curve is concave (Beling et al. 2005).

An illustration of a concave ROC curve is shown in Fig. 3. Iso-contours of expected ROE are parallel lines that point in a northeasterly direction; they have slope η defined earlier in (3) as the financial tradeoff. Expected ROE increases in the northwest direction. Iso-market share contours are shown as dashed lines with increasing market share corresponding to motion in a southwest direction. It is useful to point out that with perfect information, one can discriminate perfectly among Goods and Bads which means that the ROC curve has a right-angled knee with boundaries given by the vertical segment (0,0) to (0,1) on the left-hand side and the horizontal segment (0,1) to (1,1) at the top; without discriminatory information the ROC curve is the straight 45° line connecting (0,0) and (1,1). This is often referred to as a naive scorecard.

In an earlier paper, Beling et al. (2005) show that ROC-dominance is a necessary and sufficient condition for dominance of efficient frontiers in the expected ROA-volume (market share) plane. Thus, there is a fundamental connection between the purely statistical ROC measure that is used to evaluate performance of a scoring predictor and measures of economic performance that include tradeoffs of revenue and profitability with market share or business volume. This result is, in effect, a special case of equivalencies that exist between notions of statistical and utility dominance for binary classifiers studied by Blackwell (1951) and Zhu et al. (2002)

Fig. 3 ROC curve with tradeoffs and iso-contours for expected ROE and market share

and others. It does not apply with statistical measures such as Gini Coefficient, AUROC, K-S, R^2 or many of the popular statistical point measures often mistakenly used to link improved scores and scoring algorithms with improved financial performance. The implication is that measures of financial or economic performance cannot depend solely on statistical performance of scorecards except in unusual situations where dominance exists for all operating regions. For example, if the K-S statistic for one scorecard is larger than that of a second scorecard, we cannot guarantee that efficient frontiers derived from the first scorecard dominate those of the second.

One might argue that if there is a formal connection between dominance in ROC curves (or any other statistic that establishes discriminatory power) and measures defined in terms of profitability to the lender, then it is a simple matter to make the necessary financial calculations to move from a statistical framework to a financial one and, thus, link statistical to financial measures. Even so, we believe it is best to use measures that directly establish relative financial benefit in order to gain some insight on which data variables are important to collect and the sensitivity of financial returns to decision variables. It is useful to tell a lender how far away expected ROA is likely to be from the perfect information case or how an improved scorecard can be expected to provide an 8 % increase in expected ROE or market share rather than a 10 % increase in the K-S or Gini statistic. The latter may be interesting to a statistician, but the former is much more useful for a businessman.

The relative financial contribution of scores

Financial performance for a loan account or portfolio of retail credit accounts usually focuses on quantities such as return on assets (ROA) or return on equity (ROE); we formulate our decision model in terms of ROE because ROA can be viewed as a special case where there is no borrowing and unit investment in a loan asset coincides with $E = 1$ and $c_D = 0$. In the case of ROE we assume that no portion of E can be invested in risky assets, but only in risk-free assets with return rate, r_F . Under this assumption the net return on equity, r_E , for a unit loan with equity level, $E < 1$, is

$$r_E = \begin{cases} r_F & \text{if the borrower is rejected,} \\ r_F + E^{-1}(r_L - c_D) & \text{if the loan at rate } r_L \text{ does not default, probability } p(G|s), \quad r_L \geq c_D \geq r_F, \\ r_F - E^{-1}(l_D + c_D) & \text{if the loan defaults, probability } p(B|s) = 1 - p(G|s), \quad 0 \leq E, l_D \leq 1. \end{cases} \quad (7)$$

In ROE models both sides of the balance sheet play an important role; the loan assets earning the retail lending rates plus equity earning the risk-free rate while the liabilities include the costs of borrowed funds (debt) sourcing asset loans as well as returns on equity, the capital investments providing the leveraged return from loan investments. In ROA models, only the asset side of the balance sheet is included which means that the effects of leverage is not brought into play. The expected portfolio ROE includes expected net returns from all booked risky loan assets with scores above a score cutoff, s_c , i.e.

$$\mathbb{E}[r_E(s_c)] = r_F + \frac{1}{E} \int_{s_c}^{\infty} ((r_L - c_D)p(G|s) - (l_D + c_D)p(B|s))dF(s). \quad (8)$$

Using Bayes' rule one can express the conditional probabilities in (8) in terms of the density function $f(s|\cdot)$ and, after integration, the expected ROE in terms of the tails $F^{(c)}(s|\cdot) = 1 - F(s|\cdot)$ (Oliver and Wells 2001; Thomas 2009). Thus, the expected ROE premium from risky loans can be rewritten in terms of the tail scores, equity level, cost of funds that source loans, net revenues after default losses, and the score cutoff,

$$\mathbb{E}[r_E(s_c)] - r_F = \frac{1}{E} \left(p_G(r_L - c_D)F^{(c)}(s_c|G) - p_B(l_D + c_D)F^{(c)}(s_c|B) \right). \quad (9)$$

The expected ROE *premium* over the risk-free rate depends on the equity level, E , and can be very large under historical or even current Basel rules. For example, a regulatory capital requirement of 8 % for prime paper provides a multiplier of 12.5 to the expected net profit in the numerator of (9) above. The current Basel formulas provide less leverage for riskier loan products and greater leverage for low-risk ones. Nevertheless, the expected financial performance relative to the perfect information case is *independent* of the equity level:

$$\frac{\mathbb{E}[r_E(s_c)] - r_F}{\mathbb{E}_{PI} - r_F} = F^{(c)}(s_c|G) - \frac{p_B(l_D + c_D)}{p_G(r_L - c_D)} F^{(c)}(s_c|B) = F^{(c)}(s_c|G) - \frac{1}{\eta} F^{(c)}(s_c|B). \quad (10)$$

Relative performance of the expected ROE premium for the risky assets now depends on only one parameter: the financial tradeoff, η , defined in (3). An interesting aspect of this result is that it focuses the attention of the lender on solvency rather than control of losses and imposition of capital adequacy constraints. The relative performance of the *optimal* expected ROE now depends on two parameters: the PopOdds and the optimal cutoff score; the latter can be expressed in terms of financial borrowing and lending parameters but independent of equity level or risk-free rate. Thus, optimal cutoff and relative performance of optimal expected ROE are:

$$\frac{\mathbb{E}[r_E(s_c^*)] - r_F}{\mathbb{E}_{PI} - r_F} = F^{(c)}(s_c^*|G) - \frac{p_B}{p_G} e^{s_c^*} F^{(c)}(s_c^*|B), \quad s_c^* = \ln \frac{l_D + c_D}{r_L - c_D} \quad (11)$$

As mentioned earlier, expected return on assets (ROA) is a special case where we neglect the cost of commercial borrowing and $E = 1$. See Hoadley and Oliver (1998). Additional requirements on equity or regulatory capital can be easily included in (7)–(11) as can mixed objectives of the lender that include ROE in combination with expected market share and/or loss constraints.

Expected misclassification costs

While it is tempting to believe that statistical misclassification models are identical to financial acquisition/decision models, this is seldom the case. The former are

formulated on a symmetric tree whose relevant branch costs correspond to two misclassifications among *four observable* outcomes rather than *two observable* Good/Bad outcomes for Accepts in the asymmetric three-branch tree of Fig. 1. Minimum expected cost misclassifications policies have a mathematical cutoff structure that resembles Accept/Reject policies, but they are based on different data, different outcomes and different scores; they should be used with great caution in the support of acquisition policies. Stein (2005) examines some of the relationships between ROC curves and minimum cost classification of four observed loan outcomes: false positives, false negatives, true positives and true negatives. The important effect of adverse selection among the very risky borrowers, i.e. those rejected below cutoff, is not included.

Let c_1 be the cost associated with misclassification of a predicted non-default that turns out to be an actual, observable default (error of Type I), let c_2 be the cost of misclassification of a predicted default that turns out to be a non-default (error of Type II) and C_M the random cost of misclassification. With a single cutoff score, expected misclassification costs can be written in terms of the conditional score distributions as

$$\mathbb{E}[C_M(s_c)] = c_1 p_B F^{(c)}(s_c|B) + c_2 p_G F(s_c|G) \quad c_1 \gg c_2 \quad (12)$$

where the first term on the rhs is the expected cost of unanticipated defaults among the “Accepts” and the second is the expected “cost” of foregone revenue among the “Rejects”. Because all outcomes are assumed to be observable (12) requires conditional non-default score distributions below the cutoff, an estimation intimately connected with the subject of *reject inference* in the loan acquisition model of Fig. 1; note also that (9) only uses the tails of both conditional score distributions.

If one wants to compare the relative economic performance of two classifiers, it makes sense to compare each relative to the expected costs with perfect information on a particular category. If perfect information on Type II errors is the baseline, the relative cost comparison is

$$\frac{\mathbb{E}[C_M(s_c)]}{\mathbb{E}_{PI}} = \frac{c_1 p_B F^{(c)}(s_c|B) + c_2 p_G F(s_c|G)}{c_2 p_G} = F(s_c|G) + \frac{c_1 p_B}{c_2 p_G} F^{(c)}(s_c|B), \quad (13)$$

whose optimal cutoff is deceptively similar to (10), but in loan portfolios, has the unrealistic requirement of observable Type II errors.

Examples of relative financial performance

A plot of (11) for a “strong” (good discrimination) scorecard is shown in Fig. 4a for expected ROA when there is no borrowing by the lender. The horizontal line at 1.0 corresponds to PI. The graph provides a clear indication of the superior profitability achieved by use of a scorecard relative to an at-random selection indicated by the left-most curve; at the same time it shows how dramatically a rare but costly loss given default affects profitability. The PopOdds of the development and validation samples in this application was 10.33 or a PopOdds score of 2.33. For example, if

Fig. 4 **a** Relative ROA performance versus optimal score cutoff. **b** Relative ROE performance versus optimal score cutoff for two scorecards

the optimal cutoff score based on the financial parameters is 2.0 (odds of 7.39, probability of default at cutoff equal to 0.11), the relative ROA performance is approximately 68 % of the PI case, whereas at an optimal cutoff score of 4.0 (odds of 54.5, probability of default equal to 0.018) the relative performance is approximately 33 %.

The scorecard dominates a naïve score with optimal cutoff odds above one (score above zero). Randomly selected acquisitions always represent an inferior policy because, in this case, the discriminatory power of the scorecard provides significant improvements in expected ROA; the rapid decrease of relative ROA is due to the decrease in number of accounts exceeding the optimal cutoff scores. The jagged shape of the curve at extremely high optimal cutoffs (>5.5) is due to the small sample sizes for estimating scores from historical loan data associated with very rare defaults—rates less than a quarter of a percent. In the earlier study by Hoadley and Oliver 1998, comparisons of relative ROA performance of several different

scorecards were made, one being a behavior scorecard, another a scorecard developed exclusively from Credit Bureau data.

A different plot of (11) for two “weak” (poor discrimination) scorecards is shown in Fig. 4b where the left-most curve again corresponds to the at-random selection case; as previously noted, this corresponds to the diagonal (naïve score) ROC curve. The two top curves compare the *optimal* expected ROE versus the *optimal* score cutoff for two different scorecards, S1 and S2. The data used in constructing the two distinct scorecards were identical except that the scorecard for the relative performance of the more discriminatory rightmost curve, S1, includes, in addition to the behavioral and financial data used in the lower curve, loan offer terms such as size of down-payment. In the left-most and lowest curve where no scorecard is used but borrowers are accepted at random, we find that it is optimal to reject all borrowers once the optimal cutoff score is larger than the PopOdds (see 2). In all cases the X-axis represents the *optimal* cutoff score determined from the financial parameters of the lender, but independent of the scores. The PopOdds is 11.5:1 which equals the (natural) log odds score of 2.44 and coincides with the point on the X-axis for the left-most curve, where the optimal policy shifts from “Accept” to “Do Not Accept”. What is noteworthy is that the curves display almost identical performance in the optimal score interval $(-1, 1.5)$; the three curves diverge in the interval $(1.5, 4)$ and it appears that there is significant improvement in the relative economic performance with optimal cutoff scores between 2 and 3.4. For example, at an optimal cutoff score of 1.5 there is no need to use a scorecard, at 3.0 the relative improvement in ROE performance for the scorecard S1 that uses loan terms in its predictors is about 20 % of the perfect information case but 30 % higher than that derived from the less discriminatory scorecard S2. We mention again that this measure is independent of equity levels; besides the conditional score distributions we only require the Good/Bad odds of the population under study and the optimal cutoff. Of course, systemic risk (PopOdds) affects all curves in Fig. 4a, b. In this example with high PopOdds and rare defaults, only a highly discriminatory ROC curve is helpful.

Although we have not provided an illustration of a case where there are non-dominant ROC curves, there is no difficulty in doing so. The relative financial performance curves may cross each other one or more times but relative performance for each is easy to identify.

Multiple objectives

When lender objectives combine profit and market share, different measures may be more suitable. Consider Fig. 5 where ROE under perfect information, $\mathbb{E}_{PI}$, is plotted as straightline segments versus market share. The curve underneath the triangle is a plot of the expected ROE versus expected volume of bookings using a given scorecard; high cutoffs, low volumes are near the origin, while low score cutoffs lead to larger volumes and losses. The use of a different scorecard based on different data and/or a different technology would change the shape and location of the lower curve. As cutoffs are decreased, a maximum expected ROE is obtained which can be compared, as we have already done, with the top of the triangle representing

Fig. 5 Comparing scorecard and perfect information efficient frontiers

maximum ROE in Fig. 2. These occur at different values of market share, denoted by $V(s_c)$. Another proposal for a financial measure is to make ROE and market share comparisons jointly. Expected ROE premia with perfect information is

$$\mathbb{E}_{PI}(s_c) - r_F = \frac{1}{E} \begin{cases} (r_L - c_D)\mathbb{E}[V(s_c)] & 0 \leq \mathbb{E}[V(s_c)] \leq p_G \\ p_G(r_L - c_D) - (l_D + c_D)(\mathbb{E}[V(s_c)] - p_G) & p_G < \mathbb{E}[V(s_c)] \leq 1, \end{cases} \quad (14)$$

where fractional expected volume (market share) at each cutoff is the fraction of all booked borrowers above cutoff:

$$\mathbb{E}[V(s_c)] = p_G F^{(c)}(s_c|G) + p_B F^{(c)}(s_c|B) = F^{(c)}(s_c). \quad (15)$$

As already mentioned, the maximum expected ROE without a market share constraint occurs at a smaller market share than the point where the expected perfect information ROE attains its maximum; market share for the latter always occurs at p_G . Each point in the rightmost shaded portion of the curve in Fig. 5 represents an optimal operating point on the efficient frontier which maximizes expected ROE subject to a lower bound requirement on expected market share,

$$\underset{s_c}{\text{Max}}(\mathbb{E}[r_E(s_c)] - r_F) \quad \text{subject to } \lambda \leq 0 : \quad \mathbb{E}[V(s_c)] - V_0 \geq 0, \quad (16)$$

with V_0 the required market share and λ the inequality constraint. Along the efficient frontier (except at the single ROE maximizing solution described earlier) the equality holds, the optimal cutoff depends on V_0 and the optimal shadow price is strictly negative.

On the efficient frontier, one can show that the optimal (superscript*) cutoff and shadow price can be written as

$$s_c^* = s_c^*(V_0) = F^{-1}(1 - V_0) = \ln \frac{l_D + (c_D + E\lambda^*)}{r_L - (c_D + E\lambda^*)} \leq \ln \frac{l_D + c_D}{r_L - c_D}; \quad \lambda^* < 0 \quad (17)$$

where we use the standard convention that the inverse of $y = F(x)$ is written as $x = F^{-1}(y)$. Because we know that the market share inequality in (16) is satisfied as

a strict equality on the efficient frontier, one can use a simple procedure to compare returns. Select the desired market share, V_0 , (X-axis in Fig. 5) and find the optimal cutoff from tables or inverse calculations. Calculate the expected return on the PI boundary, the expected optimal ROE premium and the relative ROE ratio. In what follows, denote the expected volume at optimal cutoff for the unconstrained problem by $\hat{V}$ with $\hat{s}_c = F^{-1}(1 - \hat{V})$. We are interested in cases where $V_0 > \hat{V}$ and $s_c^* < \hat{s}_c$. Because of the two PI edges, relative ROE performance is

$$\frac{\mathbb{E}[r_E(s_c^*)] - r_F}{\mathbb{E}_{PI} - r_F} = \begin{cases} p_G V_0^{-1} (F^{(c)}(s_c^*|G) - \eta^{-1} F^{(c)}(s_c^*|B)) & \hat{V} \leq F^{(c)}(s_c^*) \leq p_G, \\ \frac{F^{(c)}(s_c^*)|G - \eta^{-1} F^{(c)}(s_c^*)|B}{1 - (p_B \eta)^{-1} (V_0 - p_G)} & p_G \leq F^{(c)}(s_c^*) \leq 1. \end{cases} \quad (18)$$

This result is comparable with (11), but described as a function of expected market share rather than optimal cutoff. At $V_0 = p_G$ the result is

$$\frac{\mathbb{E}[r_E(s_c^*)] - r_F}{\mathbb{E}_{PI} - r_F} = (F^{(c)}(F^{-1}(p_B)|G) - \eta^{-1} F^{(c)}(F^{-1}(p_B)|B)) \quad s_c^* = F^{-1}(p_B)$$

The tradeoff between increases in market share and decreases in expected ROE at each optimum on the (shaded) efficient frontier is the negative shadow price solving (16):

$$\lambda^*(V_0) = \frac{r_L - c_D}{E} \frac{e^{s_c^*} - e^{\hat{s}_c}}{1 + e^{s_c^*}} < 0 \quad \text{with } \hat{s}_c > s_c^* \quad (19)$$

With efficient frontier solutions, the volume constraint leads to a negative shadow price; this is equivalent to lowering the lender's commercial borrowing rate and reducing the optimal cutoff score. This policy leads to larger expected volumes, greater losses and smaller expected ROE than one obtains in the unrestricted case. Optimal shadow prices on the efficient frontier can then be compared with the tradeoff value in (3).

Summary and conclusions

There is a large literature documenting statistical methods to estimate, test, calibrate and validate the ability of credit risk scores that rank and predict outcomes such as fraud/non-fraud, default/non-default, late/on-time payments, or borrower response/non-response to offers. This is seldom the case for measures of financial performance that are directly influenced by multiple scores and score-based acquisition policies. In validating a score, two of the statistical by-products of the analytical process are the conditional score distributions $f(s|G)$ and $f(s|B)$ whose estimates can be compared with actual default performance of individual loans. This is a sensible and useful comparison, but does not go far enough. Fortunately, these same conditional score distributions, along with financial information and cost and revenue data for the lender as well as equity and capital reserve requirements, can be used to predict and validate measures of expected financial performance. Thus, it is possible, with little extra effort, to measure financial performance that can be used by the businessman and link it to score performance measured by the statistician.

In the credit risk industry, enormous resources are spent processing and mining vast amounts of data to improve statistical performance of risk and response scores, much of which may be irrelevant to the financial performance of booked portfolios. Although it is seldom done in practice, the prediction of financial measures is easily implemented. Predicted financial performance can be compared with actual financial outcomes and summarized for a portfolio manager or risk officer; they should be monitored and tracked as is the practice with risk and response scores. The prediction/decision model in this paper has illustrated simple ways to compare the financial performance of traditional ROA and ROE measures by comparing perfect information predictors as well as those without any discriminatory power. In addition to making financial performance comparisons with what is theoretically possible, we can also assess the relative financial performance of two different scorecards in the same or different economic settings under the same or different cutoff policies. It is a straightforward exercise to perform a sensitivity analysis of different PopOdds (systemic risk) along with different commercial borrowing rates and borrower preferences for different offers.

Although the algebraic details become more complex these ideas can be extended to include the effects of risk-based pricing, adverse selection, the preferences of borrowers, multiple objectives of lenders, multiple time periods and even judgmental rules that are desired or practiced. Although there are exceptions, optimal operating rules generally depend on financial data associated with the lending institution and the loans, not exclusively on the scorecard itself. Many financial performance measures that quantify relative performance are distribution-free and are based on empirical results easily obtained from commercial or privately developed scorecards that have been tested and validated. This also means that one does not have to estimate means, variances or shape parameters of named distributions. In summary, we encourage lenders, lending institutions and financial advisors to place much more emphasis than they have on the evaluation of financial benefits that directly influence good decisions and business performance in retail credit scoring. Good credit decisions depend on high-quality predictions of financial performance that can be obtained from a coherent integration of statistical scores, lender decisions and financial data. In our view, this is much more important than focusing exclusively on statistical performance of the scores themselves.

Acknowledgments The author is most appreciative of the perceptive questions and critiques made by reviewers and the many constructive recommendations made by them and the guest editor.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Beling P, Covaliu Z, Oliver RM (2005) Optimal scoring cutoff policies and efficient frontiers. *J Oper Res Soc* 56:1016–1029

- Blackwell D (1951) Comparison of experiments. In: Proceedings of second Berkeley symposium on mathematical statistics and probability. University of California Press, pp 93–102
- Hand DJ, Henley WE (1997) Statistical classification methods in consumer credit scoring. *JRSS Ser A* 160:523–541
- Hoadley B, Oliver RM (1998) Business measures of scorecard benefit. *IMA J Math Appl Bus Ind* 9:55–64
- Keeney RL, Oliver RM (2005) Designing win–win financial loan products for consumers and businesses. *J Oper Res Soc* 56:1030–1040
- Lewis EM (1992) An introduction to credit scoring. Athena Press, San Rafael, CA (republished by Fair Isaac)
- Oliver BV, Oliver RM (2012) Optimal ROE loan pricing with or without adverse selection. *J Oper Res Soc.* doi:[10.1057/jors.2012.87](https://doi.org/10.1057/jors.2012.87)
- Oliver RM, Wells ER (2001) Efficient frontier cutoff policies in credit portfolios. *J Oper Res Soc* 52:1025–1033
- Phillips R, Raffard R (2010) Theory and empirical evidence for price-driven adverse selection in consumer lending. Columbia University Business School, Working Paper
- Stein RM (2005) Relationship between default prediction and lending profits. *J Bank Financ* 29:1213–1236
- Thomas LC (2009) Consumer credit models. Oxford University Press, New York
- Trench MS, Pederson SP, Lau ET, Wang H, Nair SK (2003) Managing credit lines and prices for bank one services. *Interfaces* 33:4–21
- Zhu H, Beling P, Overstreet G (2002) A Bayesian framework for the combination of classifier outputs. *J Oper Res Soc* 53(7):719–727