The Construction of Empirical Credit Scoring Rules Based on Maximization Principles *

Robert P. Lieli[†]
University of Texas, Austin
1 University Station, C3100
Austin, TX 78712

Halbert White
University of California, San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0508

Abstract

We examine the econometric implications of the decision problem faced by a profit/utility-maximizing lender operating in a simple "double-binary" environment, where the two actions available are "approve" or "reject", and the two states of the world are "pay back" or "default". In practice, such decisions are often made by applying a fixed cutoff to the maximum likelihood estimate of a parametric model of the default probability. Following Elliott and Lieli (2007), we argue that this practice might contradict the lender's economic objective and, using German loan data, we illustrate the use of "context-specific" cutoffs and an estimation method derived directly from the lender's problem.

JEL codes: C14, C25, C51, C53
Keywords: Credit scoring, binary variables, profit maximization

1 Introduction

In this paper we examine some econometric implications of the decision problem faced by a profit- or utility-maximizing lender. We make the simplifying assumption that the lending decision is essentially a binary decision—the terms of the contract are exogenously determined from the decision maker's point of view. The potential profit the lender can make by granting the loan

*[†]Corresponding author. Email: rlieli@eco.utexas.edu. The authors thank Nicholas Kiefer, John Relman and two anonymous referees for their advice. All errors are our responsibility. The paper was part of Robert Lieli's dissertation at the University of California, San Diego.

is nevertheless a function of these terms. A similar, and equally important, assumption is that there are essentially two possible consequences of granting the loan. In one state of the world the borrower complies fully with the terms of the contract (i.e. pays the loan back on schedule); in the other the borrower defaults. Again, the loss incurred by the lender in case of default is a function of the terms of the contract.

A lender's ability to generate profits depends fundamentally on how successful they are in predicting default based on observed socio-economic characteristics of the borrower and the terms of the contract. A formal (and widely used) method of relating these variables to the conditional probability of default is known as credit scoring. The method entails assigning a predetermined number of points to the possible values of each covariate. Credit is then granted to applicants with total scores over a fixed cutoff value and denied to those below the cutoff. Such a decision rule is of course intended to ensure that credit is extended to those with a high probability of paying it back. (In fact, credit scores can be regarded as transformed default or compliance probabilities.) For a review of credit scoring methods see, e.g., Hand and Henley (1997) and the references therein.

Hence, there are two aspects to constructing a "good" credit score-based approval rule. First, a "good" estimate of the probability of default must be obtained conditional on the observed covariates. Second, the cutoff must be drawn at an "appropriate" level. But what do "good" and "appropriate" mean? The ultimate goal of the lender is to maximize (expected) profit or utility, and the construction of an optimal approval rule should reflect this goal. We draw on the methodology in Elliott and Lieli (2007) to argue that (1) the optimal (profit- or utility maximizing) cutoff is in general "context-specific", i.e. it varies from contract to contract or borrower to borrower; and (2) the objective (or loss) function used to estimate the conditional probability of compliance should be derived from the lender's economic optimization problem.

In constructing a scoring rule, one must also take into account numerous laws and regulations concerning lending activity. In particular, there is extensive legislation aimed at preventing disparate treatment of certain "protected" or minority groups. The prohibition of disparate treatment has various implications for our framework. Certainly, lenders cannot exhibit or exercise preferences that are disadvantageous for these groups. Moreover, lenders are prohibited from using minority status as a variable in estimating the conditional probability of default/compliance. Nevertheless, even if a score-based approval rule is carefully designed to avoid *disparate treatment*, it may still have an unintended *disparate impact* on a protected group, and lenders have been held responsible for this effect under the law (see Barefoot 1997, Cocheco 1997). We show how our framework can be used to design approval rules that mitigate or eliminate disparate impact.

The plan of the paper is as follows. First (in Sections 2 and 3), we will derive the optimal approval rule under a general formulation of the lender's objective function, where in addition to profits, the lender may care about some characteristics of the borrower and the laws regulating the lending process. We will show that the optimal decision rule is of the form

> "extend the loan if and only if the conditional probability of compliance is greater than a cutoff",

where the the cutoff is determined by the lender's objective function and may vary from person to person or with the characteristics of the loan. This is in contrast to existing practice where it is customary to use a uniform cutoff, which is often chosen according to a simple rule of thumb (e.g. one half or some quantile of the estimated default probabilities; see, e.g., Fortowsky and LaCour-Little 2001).

Second, following Elliott and Lieli (2007), we will argue in Section 4 that the modeling and estimation of the conditional probability of compliance should be based on the lender's economic optimization problem. In particular, we will show that one does not need a fully correctly specified model of this conditional probability in order to consistently estimate the optimal approval rule. Nevertheless, to take advantage of this flexibility, the misspecified model must be estimated by solving the sample analog of the lender's optimization problem, which is not necessarily the same as the maximum likelihood problem. Hence, maximum likelihood-based procedures such as (potentially misspecified) logit or probit regressions may lead to suboptimal decision rules.

Third, in Section 5 we illustrate the proposed methodology by applying it to a data set consisting of records of 1000 customers of a German commercial bank. The results show that the proposed econometric method is indeed capable of producing approval rules in practice that lead to more profitable lending decisions than simple logit regressions. The added gain from the methodology may be enough to compensate for the costlier numerical procedures needed to implement it.

2 A simple view of the lending process

We follow Feelders (2002) in viewing the creditor's problem as consisting of two parts: (*i*) the selection or decision mechanism; (*ii*) the outcome mechanism. The former refers to a decision rule by which the lender decides whether to accept or reject a loan application. The focus of the paper is on this binary decision: the terms of the loan contract (the interest rate, the size and duration of the loan, etc) are assumed to be exogenously given.[1] That is, we view the lender as offering one fixed loan contract or a number of different ones. The prospective borrower then applies for the contract of his choice and the lender merely accepts or rejects the application.

We will assume that each loan contract offered by the lender requires equal monthly installments over the duration of the loan. A loan contract is then completely characterized by the triple $\ddot{X} = (L, D, r)$, where L is the size of the loan, D is the duration of the loan in months and r is the (monthly) interest rate on the loan. The size of the monthly installment I can be determined from the identity

$$L = \sum_{i=1}^{D} \frac{I}{(1+r)^i} = d(r, D)I, \tag{1}$$

where $d(r, D) \equiv \sum_{i=1}^{D}(1 + r)^{-i}$.

The outcome mechanism, on the other hand, determines whether a borrower with a vector of observed characteristics $\tilde{X}$ repays the loan in accordance with the terms of the contract. We assume that there are only two possible outcomes in this regard: the borrower either complies fully with the conditions of the contract or the borrower defaults on the loan in which case only a given percentage of the principal can be recovered at the end of the loan's maturity. If Y is the indicator of default (i.e. $Y = -1$ is a "bad loan" and $Y = 1$ is a "good loan") and $X = (\ddot{X}, \tilde{X})$, then the outcome mechanism can be represented as the mapping

$$x \mapsto p(x) = P(Y = 1 \mid X = x).$$

This is the conditional probability of compliance given the observed characteristics of the borrower and the loan contract.

We make a number of additional simplifying assumptions about the lending process. First, even though we allow for loan contracts of varying lengths, time plays a limited role in the setup. The decision problem under examination involves a one-shot static decision—we do not consider the dynamic consequences of the approval decision for the decision environment. Future

[1] One way to think about this is that the terms of the contract are determined by a competitive market. Another is that the terms of the contract are reviewed at discrete time periods and we focus on optimal decisions in between these periods.

decisions are not contingent on the decision today. Second, no application is rejected because of the lack of loanable funds. Finally, if an application is rejected, we assume that the lender will instead have the option to invest in a risk-free government bond matching the size and the duration of the loan applied for, but paying a lower interest rate.

While the proposed setup may not be realistic in many aspects, it will enable us to formulate an objective function for the lender, defined over the two-by-two matrix of possible actions (approve/reject) and outcomes (default/compliance). We will show that full knowledge of the outcome mechanism $p(x)$ combined with the given objective function of the lender is sufficient to derive an optimal selection mechanism. Nevertheless, the function $p(x)$ is unknown; one must learn about it from historical loan data using statistical methods. We will argue that, in contrast to standard methods, statistical inference about $p(x)$ (or, more precisely, the optimal decision rule) should be guided by the objective function of the lender.

Finally, we caution that statistical inference about $p(x)$ is complicated by the fact that the data available on credit history is generally contaminated by the selection mechanism used by other lenders. In other words, one can observe the outcome Y only for individuals who were able to pass the selection process of a lender in the past. Therefore, one must either model this selection process or at least recognize that inference will be conditional on being in the formerly selected group.[2] Because the current task at hand is sufficiently challenging without also treating the reject inference problem, we shall abstract from this issue in what follows. Suitable modifications of the approach developed here that accommodate the reject inference problem are the subject of future research.

3 The lender's objective

3.1 Profit-maximizing lender

In this section we consider profit-maximizing lenders, who care only about earning a profit on the loans extended. The profitability of a loan can be measured by its net present value (NPV, also denoted by π), defined as the revenue stream from the loan, discounted at an appropriate rate, minus the amount of the loan. By the assumptions made in the previous section, the lender's alternative to accepting a loan application is to invest the amount in

[2]See Feelders (2002) or Crook and Banasik (2004) for a review and evaluation of reject inference methods.

Table 1: The lender's profit (NPV) in four contingencies

	no default ($Y = 1$)	default ($Y = -1$)
approve (A)	$\pi_{A,1}(\ddot{x}) > 0$	$\pi_{A,-1}(\ddot{x}) < 0$
reject (R)	$\pi_{R,1}(\ddot{x}) = 0$	$\pi_{R,-1}(\ddot{x}) = 0$

a (risk-free) government bond of the same maturity; therefore, the applicable discount rate is f_D, the interest rate on the D-month government bond.

The revenue stream from the loan is of course uncertain; it depends on whether the borrower will default on the loan ($Y = -1$) or not ($Y = 1$). We assume $r > f_D$ for each contract (L, D, r), which means that the net present value of each contract is positive in the absence of default (i.e when $Y = 1$). In particular, the NPV associated with the (approve, $Y = 1$) contingency is given by

$$\pi_{A,1}(L, D, r) = d(f_D, D)I - L = d(f_D, D)\frac{L}{d(r, D)} - L > 0, \tag{2}$$

where use is made of equation (1) and the definition following it. By assumption, the NPV the lender incurs when the loan application is rejected is zero, regardless of the hypothetical Y outcome ($\pi_{R,1} = \pi_{R,-1} = 0$). Therefore, if the lender knew with certainty that the borrower was going to honor the contract ($Y = 1$), the loan would be approved.

On the other hand, if $Y = -1$, the assumption is that only a certain fraction $q \in (0, 1)$ of the principal can be recovered at the end of the maturity of the loan. Hence, the NPV associated with the (approve, $Y = -1$) contingency is given by

$$\pi_{A,-1}(L, D, r) = \frac{qL}{(1 + f_D)^D} - L < 0. \tag{3}$$

Again, the NPV of denying the loan is zero, so if the lender knew with certainty that the borrower was going to default on the loan, the loan application would be rejected. The payoffs associated with the four possible contingencies are summarized in Table 1.

Of course, the lender cannot observe the outcome Y at the time the approval decision has to be made. Economic theory postulates that the lender will instead seek a decision rule that maximizes expected net present value conditional on the observable characteristics of the loan contract ($\ddot{X}$)

and the borrower ($\tilde{X}$):

$$\max_{d\in\{A,R\}} E[\pi_{d,Y}(\ddot{X}) \mid \tilde{X}=\tilde{x}, \ddot{X}=\ddot{x}]$$
$$= \max_{d\in\{A,R\}} \big\{p(\tilde{x},\ddot{x})\pi_{d,1}(\ddot{x}) + [1-p(\tilde{x},\ddot{x})]\pi_{d,-1}(\ddot{x})\big\}. \tag{4}$$

That is, a loan application will be approved ($d = A$) if and only if[3]

$$p(\tilde{x},\ddot{x})\pi_{A,1}(\ddot{x}) + [1-p(\tilde{x},\ddot{x})]\pi_{A,-1}(\ddot{x}) > 0,$$

or

$$p(\tilde{x},\ddot{x}) > \frac{1}{1+\pi_{A,1}(\ddot{x})/[-\pi_{A,-1}(\ddot{x})]} \equiv c(\ddot{x}) \in (0,1). \tag{5}$$

Using equations (2) and (3), the cutoff function $c(\ddot{x})$ can be written as

$$c(\ddot{x}) = \left[1 - \frac{d(f_D,D)d(r,D)^{-1}-1}{(1+f_D)^{-D}q-1}\right]^{-1}. \tag{6}$$

Form (5) of the cutoff function $c(\ddot{x})$ has an intuitive interpretation. The quantity $-\pi_{A,-1}(\ddot{x}) > 0$ is the magnitude of the loss resulting from an approved loan "gone bad", while $\pi_{A,1}(\ddot{x})$ is the payoff from an approved loan that is in compliance. The cutoff $c(\ddot{x})$ depends on the relative size of these quantities. A lower value of $\pi_{A,1}(\ddot{x})/[-\pi_{A,-1}(\ddot{x})]$ means that the relative cost of a wrong approval is higher, resulting in a higher cutoff. That is, it becomes "harder" for any particular applicant to get approved.

A noteworthy implication of (5) is that it is not optimal to use a uniform cutoff in making approval decisions. In the current setup, the expected profit maximizing cutoff is a rather complicated nonlinear function of the conditions of the underlying loan contract. Thus, the covariates contained in $\ddot{X}$ play a double role: First, they might provide information about the likelihood of default. Second, they determine the optimal cutoff, i.e. the manner in which the information provided by the conditional probability $p(\ddot{X},\tilde{X})$ should be evaluated by a profit maximizing lender.

In the following section we will discuss conditions under which the vector $\tilde{X}$, the personal characteristics of the applicant related to the probability of default, may also play a similar double role.

[3] If (4) has the same value for $d = A$ and $d = R$, the decision is taken to be "reject".

Table 2: The lender's utility in four contingencies

	no default ($Y = 1$)	default ($Y = -1$)
approve (A)	$u_{A,1}[\pi_{A,1}(\ddot{x}), \tilde{x}]$	$u_{A,-1}[\pi_{A,-1}(\ddot{x}), \tilde{x}]$
reject (R)	$u_{R,1}[\pi_{R,1}(\ddot{x}), \tilde{x}]$	$u_{R,-1}[\pi_{R,-1}(\ddot{x}), \tilde{x}]$

3.2 Utility-maximizing lender

Given equal profits, a lender may strictly prefer to give the loan to applicants who possess a certain characteristic of interest. In fact, the lender may even be willing to forego (expected) profits in order to ensure that applicants in the target group have an easy access to loans (e.g. in case of certain government loan programs). In a similar vein, taste-based negative discrimination, as defined by Becker (1971), "requires that the discriminator pay or forfeit income for the privilege of exercising prejudicial tastes" (Ladd 1998, p. 42).

We can formally capture this idea by replacing the objective function given in Table 1 by the more general one shown in Table 2. The utility functions $u_{A,1}(\pi_{A,1}, \tilde{x})$, $u_{R,1}(\pi_{R,1}, \tilde{x})$, etc. determine the rate at which the lender is willing to trade off profits for the "privilege of exercising his preferences". It is reasonable to assume that these functions satisfy

$$u_{A,1}[\pi_{A,1}(\ddot{x}), \tilde{x}] > u_{R,1}[\pi_{R,1}(\ddot{x}), \tilde{x}] \quad \text{and} \tag{7}$$

$$u_{R,-1}[\pi_{R,-1}(\ddot{x}), \tilde{x}] > u_{A,-1}[\pi_{A,-1}(\ddot{x}), \tilde{x}] \tag{8}$$

for each possible value $(\ddot{x}, \tilde{x})$ of $(\ddot{X}, \tilde{X})$. These assumptions mean that profit (or the outcome Y) is still the primary factor in the lender's objective: if it were known with certainty that a given borrower was going to honor the contract, then the lender would approve the loan, regardless of the characteristics of the borrower. Conversely, if default were a certainty, the loan would always be denied.

The lender's optimization problem can now be written as

$$\max_{d \in \{A,R\}} E\left[u_{d,Y}\left(\pi_{d,Y}(\ddot{X}),\, \tilde{X}\right) \middle| \tilde{X} = \tilde{x}, \ddot{X} = \ddot{x}\right]. \tag{9}$$

Repeating the argument in the previous section leads to a cutoff rule of the same form as in (5). However, the optimal cutoff is now a function of $\tilde{X}$ (the personal characteristics of the borrower) as well as $\ddot{X}$ (the terms of the contract):

$$c(\ddot{x}, \tilde{x}) = \left\{1 + \frac{u_{A,1}[\pi_{A,1}(\ddot{x}),\, \tilde{x}] - u_{R,1}[\pi_{R,1}(\ddot{x}),\, \tilde{x}]}{u_{R,-1}[\pi_{R,-1}(\ddot{x}),\, \tilde{x}] - u_{A,-1}[\pi_{A,-1}(\ddot{x}),\, \tilde{x}]}\right\}^{-1}. \tag{10}$$

The basic interpretation of $c(\ddot{x}, \tilde{x})$ is retained. The net cost of a wrong approval is now given by the denominator term

$$u_{R,-1}[\pi_{R,-1}(\ddot{x}),\, \tilde{x}] - u_{A,-1}[\pi_{A,-1}(\ddot{x}),\, \tilde{x}] > 0.$$

Similarly, the numerator term

$$u_{A,1}[\pi_{A,1}(\ddot{x}),\, \tilde{x}] - u_{R,1}[\pi_{R,1}(\ddot{x}),\, \tilde{x}] > 0$$

can be interpreted as a the net benefit of a correct approval. Once again, the optimal cutoff (10) is determined by the relative magnitudes of these two costs.

As can be seen, in this simple "double-binary" framework there is no essential difference between risk neutral decision makers (maximizing expected profit) and risk averse ones (maximizing expected utility). The optimal decision rule displays the same type of dichotomy in both cases: the conditional probability of default (an unknown object of "nature") is compared with a cutoff completely determined by the decision maker's preferences. Given the cutoff function $c(\cdot)$, the econometric analysis will proceed exactly the same way in both cases (see Section 4).

3.3 Legal restrictions on the lender's objectives

We motivated the generalization of the lender's objective function by alluding to the possibility of positive or negative discrimination in lending. Of course, there is extensive legislation aimed at regulating the former and eliminating the latter. These laws put additional restrictions on the lender's objective function as given in Table 2. In the United States, the Equal Credit Opportunity Act (15 U.S.C. §1691) explicitly prohibits discrimination "against any applicant, with respect to any aspect of a credit transaction... on the basis of race, color, religion, national origin, sex or marital status, or age...".

As usual, the letter of the law is open to a number of interpretations. One intention of the law is to rule out *disparate treatment*, in which the loan approval process purposely involves the consideration of the "protected characteristics" cited above. Nevertheless, allowance is made for special purpose credit programs, administered by the government or non-profit organizations, "for the benefit of an economically disadvantaged class of persons". In this case protected variables can be used to identify this group. For-profit organizations are also allowed to run such programs as long as the "program is established... to extend credit to a class of persons who, under the organization's customary standards of credit-worthiness, would not receive such credit..."(Regulation B, 12 C.F.R. Section 202.8).[4] Thus, the lender is allowed to trade off profit for exercising certain preferences, but not others. This means that protected characteristics can enter $\tilde{X}$, but the law has restrictions on how $\tilde{X}$ can enter the objective function in Table 2.

The prohibition of disparate treatment has another interpretation, which does not have to do with the lender's preferences. This interpretation rules out behavior called "statistical discrimination" (see Arrow (1973) and Phelps (1972)). This means that the lender is prohibited from using the protected characteristics in estimating the conditional probability of default or, more generally, in judging an applicant's credit-worthiness. The prohibition is necessary because the protected characteristics are cheaply observable and are often statistically related to the probability of default, especially in the absence of other relevant conditioning variables that may be costlier to observe. Ladd (1998, p. 43) concludes that "[i]n essence, the law requires that lenders make decisions about... loans as if they had no information about the applicant's race, regardless of whether race is or is not a good proxy for risk factors not easily observed by the lender." Based on this interpretation, protected characteristics are not allowed to enter models of $p(\ddot{x}, \tilde{x})$.

Under a strict interpretation of anti-discriminatory laws, lenders can be held liable not only for disparate treatment of applicants, but also for the (possibly unintended) *disparate impact* of the approval process. That is, protected characteristics may be completely missing from $\tilde{X}$, yet the resulting decision rule may produce a higher than average rejection rate among protected groups. (This typically happens because "unprotected" socio-economic variables on which the approval decision is based can be correlated with minority status.) Under a strict interpretation of the law, such an approval process can only be justified only by a significant "business necessity" (see Fortowsky and LaCour-Little 2001). This interpretation of the law can make lenders fairly vulnerable to legal and political attacks. We will now show how the present framework can be used to devise a formal approval rule with reduced or no disparate impact.

An easy way to construct an approval rule that alleviates disparate impact is to attach some conceptual monetary premium to the approval of minority applicants. Let $\tilde{X}_1 = 1$ if an applicant belongs to a certain minor-

[4] We thank John Relman for guiding us to the appropriate section of the law.

ity and zero otherwise. Then one can, for example, write

$$\begin{aligned}
&u_{A,1}[\pi_{A,1}(\ddot{x}), \tilde{x}] = \pi_{A,1}(\ddot{x}) + \lambda_{A,1}\tilde{x}_1 \\
&u_{R,1}[\pi_{R,1}(\ddot{x}), \tilde{x}] = \pi_{R,1}(\ddot{x}) = 0 \\
&u_{A,-1}[\pi_{A,-1}(\ddot{x}), \tilde{x}] = \pi_{A,-1}(\ddot{x}) + \lambda_{A,-1}\tilde{x}_1 \\
&u_{R,-1}[\pi_{R,-1}(\ddot{x}), \tilde{x}] = \pi_{R,-1}(\ddot{x}) = 0,
\end{aligned}$$

where $\lambda_{A,1} > 0$ and $\lambda_{A,-1} > 0$ give the extra value associated with the approval of "good" and "bad" minority applicants. Formula (10) shows that the result of this specification is a lower compliance probability cutoff for minority applicants. The usage of the minority status indicator $\tilde{X}_1$ in the lender's preferences is legal, because it amounts to the implementation of a special loan program.[5] In particular, the resulting lower cutoff for the protected group makes credit more accessible for them, without excluding anybody who would have been approved previously.

How one determines the size of the parameters $\lambda_{A,1}$ and $\lambda_{A,-1}$ is, however, an open question. In theory, these values should be related to the expected cost of a "disparate impact" violation, e.g. the cost of a trial, fine, restitution, etc. times the probability that such a payment has to be made. In practice, it is probably reasonable to specify these parameters as a fraction of $\pi_{A,1}(\ddot{x})$ and $\pi_{A,-1}(\ddot{x})$.

4 The econometric implications of the optimal approval rule

An important implication of decision rule (5) is that the lender does not need to know the exact value of the function $p(\ddot{x}, \tilde{x})$ to make optimal approval decisions. For any given value $x = (\ddot{x}, \tilde{x})$ of the covariates, all the decision maker needs to know is whether the conditional probability $p(x)$ is below or above the known cutoff function $c(x)$. In other words, the optimal decision is determined by the *sign* of the function $p(x) - c(x)$. Formally, one can rewrite the optimal cutoff rule as

$$\text{approve the loan iff } \operatorname{sgn}[p(X) - c(X)] = 1, \tag{11}$$

where the sign function $\operatorname{sgn}(\cdot)$ is defined as $\operatorname{sgn}(z) = 1$ for $z > 0$ and $\operatorname{sgn}(z) = -1$ for $z \leq 0$. Thus, if $m^*(x)$ is any other function such that

$$\operatorname{sgn}[m^*(x) - c(x)] = \operatorname{sgn}[p(x) - c(x)] \quad \forall x \in \operatorname{support}(X), \tag{12}$$

[5]Nevertheless, $\tilde{X}_1$ cannot be a conditioning variable in estimating compliance probabilities due to the prohibition of statistical discrimination.

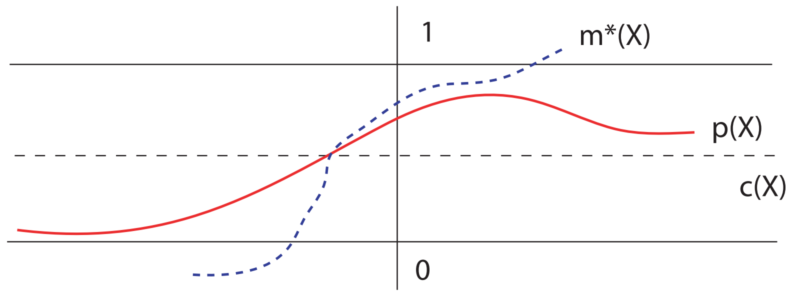

Figure 1: The multiplicity of the representations of the optimal decision rule

then relying on $m^*(x)$ instead of $p(x)$ in decision rule (11) will also lead to optimal loan approval decisions.[6] See Figure 1 for an illustration of condition (12).

This seemingly trivial observation is the basis of the econometric methodology proposed by Elliott and Lieli (2007) for modeling and estimating the unknown conditional probability function $p(x)$, when the primary objective is to ensure that the resulting estimate leads to "good" (i.e. approximately utility- or profit-maximizing) approval decisions. The following two subsections introduce the main elements of this methodology.

4.1 Model specification for $p(x)$

Condition (12) clarifies the extent to which models of $p(x)$ need to be correctly specified if the modeler's objective is to estimate the optimal decision rule (11). In particular, consider the parametric class of functions (i.e. model)

$$\mathcal{M}_\Theta = \{m(\cdot, \theta) : \theta \in \Theta\},$$

where Θ is a given subset of $\mathbb{R}^p$, and for any $\theta \in \Theta$, $m(\cdot, \theta)$ is a real valued function that maps the covariates x into a real number.

For $\mathcal{M}_\Theta$ to be a "good" model of $p(\cdot)$, it is not necessary that it be fully correctly specified for $p(\cdot)$, i.e. there need not exist $\theta^\circ \in \Theta$ such that

$$m(x, \theta^\circ) = p(x) \qquad \forall x \in \text{support}(X).$$

[6]Intuitively, one can think of $m^*(x)$ as a credit score constructed for lenders with cutoff functions $c(x)$. It is not required that $m^*(x)$ coincide with $p(x)$ at every point (in fact, m^* can even be negative or larger than unity), but it has to lead to the same approval/rejection decisions as $p(x)$.

Rather, the model is "good" from the decision maker's point of view if $m(\cdot,\theta) - c(\cdot)$ is correctly specified for the *sign* of $p(\cdot) - c(\cdot)$, i.e.

$$\exists\theta^* \in \Theta: \quad \operatorname{sgn}[m(x,\theta^*) - c(x)] \\ = \operatorname{sgn}[p(x) - c(x)] \qquad \forall x \in \operatorname{support}(X). \quad (13)$$

Obviously, this requirement is much weaker than fully correct specification.

Hence, we observe that misspecified models of $p(\cdot)$ can potentially reproduce the optimal decision rule (11). The key to specifying the class $\mathcal{M}_\Theta$ is to allow for a functional form "flexible" enough so that hopefully all sign changes in $p(\cdot) - c(\cdot)$ can be captured. Suppose, for example, that $c(x)$ is constant in the first component of x, but economic theory predicts that $p(x)$ is U-shaped in that variable. In this case the modeler should ensure that $m(x,\theta)$, as a function of the first component of x, is capable of crossing the constant cutoff at least twice. For example, if $m(x,\theta)$ is specified as a polynomial, it should be at least quadratic in the first component of x. We will briefly return to the issue of model specification at the end of Section 4.3, after introducing the maximum-utility (MU) estimator and establishing its basic properties.

4.2 Utility maximization-based estimation

Specifying a parametric model $\mathcal{M}_\Theta$ for $p(x)$ means that the decision maker faces a restricted set of possible decision rules $d(X,\theta)$ indexed by the parameter θ:

$$d(X,\theta) = \text{"approve the loan iff } \operatorname{sgn}[m(X,\theta) - c(X)] = 1\text{"}, \ \theta \in \Theta. \quad (14)$$

The economic objective of the lender is unchanged: the goal is to choose the value of the parameter θ so as to maximize expected profits or expected utility—now also subject to the additional constraint imposed on the form of the decision rule. The optimal decision rule within the class (14) is obtained by solving

$$\max_{\theta\in\Theta} E_{X,Y}\left\{u_{d(X,\theta),\,Y}\left[\pi_{d(X,\theta),\,Y}(\ddot{X}), \tilde{X}\right]\right\}, \quad (15)$$

subject to

$$d(X,\theta) = \begin{cases} \text{A} & \text{if } m(X,\theta) > c(X) \\ \text{R} & \text{if } m(X,\theta) \le c(X). \end{cases}$$

If condition (13) is satisfied, then solving the constrained optimization problem (15) will produce a decision rule $d(x, \theta^*)$ equivalent to the "first-best" optimal decision rule (11). Furthermore, even if $\mathcal{M}_\Theta$ is misspecified for $p(\cdot)$ to an extent that condition (13) is not satisfied, the optimization problem above still delivers the "second best" optimum for the lender, i.e. the best decision rule *given* the specification $\mathcal{M}_\Theta$.

The maximization problem (15) can be rewritten in a form that better highlights the role of the specification $\mathcal{M}_\Theta$ and is more suitable for theoretical study as well as for practical use. In particular, Elliott and Lieli (2007) show that the following maximization problem is equivalent to (15):

$$\max_{\theta\in\Theta} S(\theta) \equiv \max_{\theta\in\Theta} E\Big\{b(X)[Y - 2c(X) + 1]\operatorname{sgn}[m(X,\, \theta) - c(X)]\Big\}, \quad (16)$$

where $b(x)$ is given by

$$b(x) \equiv u_{R,-1}[\pi_{R,-1}(\ddot{x}),\, \tilde{x}] - u_{A,-1}[\pi_{A,-1}(\ddot{x}),\, \tilde{x}] \\ + u_{A,1}[\pi_{A,1}(\ddot{x}),\, \tilde{x}] - u_{R,1}[\pi_{R,1}(\ddot{x}),\, \tilde{x}] > 0.$$

The function $S(\theta)$ is a generalized version of Manski's (1985, 1986) "population score" and is an affine transformation of the objective function in (15). Hence, $S(\theta)$ can be regarded as a rescaled and recentered measure of expected profit or expected utility.

The exact form of the function $S(\theta)$ is unknown to the decision maker as the expectation in (16) is taken with respect to the unknown joint distribution of the vector (Y, X'). However, if a random sample of observations $\{(Y_i, X_i')\}_{i=1}^n$ is available from this distribution[7], then an estimated decision rule can be obtained by solving the sample analog problem

$$\max_{\theta\in\Theta} \hat{S}_n(\theta) \\ \equiv \max_{\theta\in\Theta} n^{-1} \sum_{i=1}^{n} b(X_i)[Y_i - 2c(X_i) + 1]\operatorname{sgn}[m(X_i,\, \theta) - c(X_i)]. \quad (17)$$

Let θ^{MU} denote a solution of (16) and $\hat{\theta}_n^{\mathrm{MU}}$ a solution of (17). (The superscript MU connotes maximum utility.) If the lender were able to use the decision rule

$$\operatorname{sgn}[m(X, \theta^{\mathrm{MU}}) - c(X)], \quad (18)$$

[7]This is where the reject inference problem is assumed away. If banks use selection rules in granting loans, then the data available from previous loans is not a random sample from the full distribution of (Y, X'); rather, it is a random sample from some truncation of this distribution.

the resulting expected utility (profit) would be measured by $S(\theta^{\mathrm{MU}})$. However, the lender can only use the estimated decision rule

$$\operatorname{sgn}[m(X, \hat{\theta}_n^{\mathrm{MU}}) - c(X)], \tag{19}$$

implying an expected utility (profit) value equal to $S(\hat{\theta}_n^{\mathrm{MU}})$. A relevant statistical question to ask is the following: Is the lender asymptotically as well off relying on the estimated decision rule (19) as if (18) were known? In other words, when can we conclude

$$S(\hat{\theta}_n^{\mathrm{MU}}) \to_{a.s.} S(\theta^{\mathrm{MU}}) \quad \text{as } n \to \infty? \tag{20}$$

This condition is of course weaker than requiring $\hat{\theta}_n^{\mathrm{MU}} \to_{a.s.} \theta^{\mathrm{MU}}$, the traditional question of interest in econometrics. In fact, (20) can easily occur without $\hat{\theta}_n^{\mathrm{MU}}$ converging at all. We do not attribute any economic meaning to θ^{MU} and so we are not particularly interested in its value; we regard the parameterization $\mathcal{M}_\Theta$ as arbitrary to begin with. Instead, the focus is on the welfare of the decision maker.

Elliott and Lieli (2007) give regularity conditions on the distribution of X and on the form of the model $\mathcal{M}_\Theta$ under which (20) holds. Stronger versions of these conditions ensure a convergence rate of $\sqrt{n}$.[8]

4.3 Utility maximization-based estimation vs. maximum likelihood

The parametric model $\mathcal{M}_\Theta$ was introduced as a specification for the conditional probability function $p(x)$. Thus, given a random sample of observations from the distribution of (Y, X'), one could write down the (log) likelihood function for the parameter θ and then maximize it to obtain the maximum likelihood estimate $\hat{\theta}_n^{\mathrm{ML}}$:

$$\hat{\theta}_n^{\mathrm{ML}} = \arg\max_\theta n^{-1} \sum_{i=1}^{n} (1 + Y_i) \log[m(X_i, \theta)] + (1 - Y_i) \log[1 - m(X_i, \theta)].$$

Here we consider the question: how "good" is the decision rule

$$\operatorname{sgn}[m(X, \hat{\theta}_n^{\mathrm{ML}}) - c(X)] \tag{21}$$

[8] Indeed, using a stochastic equicontinuity argument, Elliott and Lieli (2007) show that $\sqrt{n}[S(\hat{\theta}_n^{\mathrm{MU}}) - S(\theta^{\mathrm{MU}})]$ is asymptotically normal. The regularity conditions needed for this result will typically be satisfied if the model is Lipschitz-continuous in the parameter and X has a continuous component.

in terms of fulfilling the lender's objective? How does the expected utility (profit) associated with decision rule (21) compare with that associated with $\mathrm{sgn}[m(X, \hat{\theta}_n^{\mathrm{MU}}) - c(X)]$?

These questions are hard to answer for a finite sample size n. Nevertheless, we can show that given a sufficiently large sample, the lender is always at least as well off relying on $\hat{\theta}_n^{\mathrm{MU}}$ as he would be if he instead relied on $\hat{\theta}_n^{\mathrm{ML}}$. Moreover, $\hat{\theta}_n^{\mathrm{MU}}$ generally does strictly better (asymptotically) than $\hat{\theta}_n^{\mathrm{ML}}$.

The two estimation methods (decision rules) are asymptotically equivalent if the model $\mathcal{M}_\Theta$ is fully correctly specified for $p(\cdot)$. In this case the maximum likelihood estimator (or any other consistent estimator) will asymptotically reproduce the first best optimal decision rule (11). The same is true for the utility maximization based estimator. Nevertheless, in this case the maximum likelihood estimator has optimality properties other estimators cannot in general claim.

Further, we must recognize that $\mathcal{M}_\Theta$ will generally be misspecified for $p(\cdot)$, as economic theory is rarely strong enough to provide detailed knowledge about $p(\cdot)$. The behavior of the two estimators in this case is best understood through an example. Suppose there is only one covariate x and the true conditional probability function $p(x)$ is given by the solid line in Figure 2. For simplicity, also assume that the decision maker's objective function is such that the optimal cutoff $c(x)$ is constant. Next suppose that the proposed model of $p(x)$ is a probit specification based on a linear index: $m(x, \theta) = \Phi(\theta_1 + \theta_2 x)$. Clearly, this model is misspecified, as $p(x)$ has more than one inflection point, and $m(x, \theta)$ has only one as a function of x. However, as $p(x)$ and the cutoff $c(x)$ have only one intersection point, there exists a value of θ that reproduces the first-best optimal decision rule.

Imagine generating a large sample of X values from the uniform distribution on the interval (-3,6) (see the histogram in Figure 2). The corresponding Y outcomes (not shown) are then generated according to the given conditional probability function $p(x)$. Using the sample obtained, we can estimate both $\hat{\theta}_n^{\mathrm{ML}}$ and $\hat{\theta}_n^{\mathrm{MU}}$. The fitted model $\Phi(\hat{\theta}_{1n}^{\mathrm{ML}} + \hat{\theta}_{2n}^{\mathrm{ML}} x)$ is shown in Figure 2 by the dash-dot line, while $\Phi(\hat{\theta}_{1n}^{\mathrm{MU}} + \hat{\theta}_{2n}^{\mathrm{MU}} x)$ is depicted by the dashed line.

The utility-maximization based method produces a fitted model that intersects the cutoff almost exactly at the same point as $p(x)$; furthermore, the fitted model is always on the same side of the cutoff as $p(x)$. Thus, the fitted model succeeds in reproducing the first best decision rule (11) in large samples, despite the fact that the fit is "poor" away from the point where $c(x)$ and $p(x)$ intersect. However, the vertical distance between the fitted model and $p(x)$ is inconsequential from the decision maker's point of view as long as the fitted model is on the "correct" side of the cutoff.

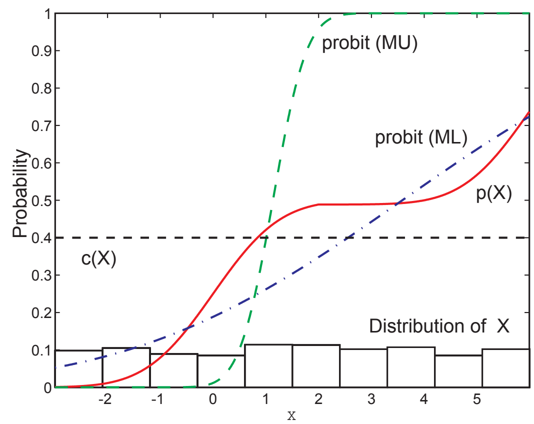

Figure 2: The comparison of utility maximization based estimation and ML

On the other hand, if the probit model is estimated by maximum likelihood, the estimator will, asymptotically, try to produce a *globally* good approximation to the function $p(\cdot)$. Because of misspecification, however, the probit model cannot perfectly reproduce $p(\cdot)$ even as the sample size goes to infinity. Maximum likelihood will nevertheless try its best to optimize the fit of the model even away from the intersection point of $p(x)$ and $c(x)$, where the decision maker does not care about the magnitude (only the sign) of estimation errors. As a result, the fitted model will miss the intersection point between $p(x)$ and $c(x)$ and will lead to a suboptimal decision for a range of x values (from approx. $x = 0.75$ to 2.5). The result is a reduction in expected utility or profit. Intuitively speaking, the objective implicit in maximum likelihood estimation is in general not consistent with the objectives of the utility- or profit-maximizing lender facing a binary approval decision.

It is important to note that the asymptotic optimality property of the MU method demonstrated through this example is *conditional* on the model specification. A strict improvement over ML can be expected to obtain only if $m(x, \theta)$ is misspecified for $p(x)$ as a whole, but $m(x, \theta) - c(x)$ can still capture (most of) the sign changes in $p(x) - c(x)$. While this is a weaker requirement than fully correct specification, one still needs to choose a parameterization, and economic theory is often not specific enough to identify the number of "crossing points". While Elliott and Lieli (2007) provide a consistent model selection criterion, it is not clear how much one should penalize more complex parameterizations in finite samples.

A practically feasible strategy for parameterizing the model $m(x, \theta)$ is to use standard likelihood-based specification tests such as a likelihood ratio test or Lagrange multiplier test; see Davidson and MacKinnon (1984). If one finds a parameterization that is not rejected, the model can be re-estimated by the maximum utility method to obtain a decision rule.

Of course, one possible reason for failing to reject a given specification is that it is (approximately) correct. In this case, MU is asymptotically equivalent to ML, but $S(\hat{\theta}_n^{\mathrm{ML}})$ might exceed $S(\hat{\theta}_n^{\mathrm{MU}})$ with positive probability for small n. The other, quite relevant, possibility is that the model specification is incorrect, but it is not rejected because of lack of power. Still, the sample might be large enough for the "MU-asymptotics" to work, so the MU estimator may well lead to a better (i.e. more profitable) decision rule out of sample.

5 Application to German data

5.1 The data set and its limitations

We will now apply the econometric methodology described in Section 4 to German banking data, publicly available at the following URL, maintained by the Department of Statistics at the University of Munich:

http://www.stat.uni-muenchen.de/service/datenarchiv/kredit/kredit.html

The data set consists of observations on 1000 individuals, customers of a German commercial bank, with outstanding loans in the early 1980's.[9] The loans in question are relatively small consumer loans ranging from DM250 to DM18,000 in size.[10] One of the variables in the data set is a binary indicator of whether the loans were repaid or not (of the 1000 observed borrowers 300 defaulted). In addition, observations on a number of covariates are available, describing the terms of the loan contract as well as the credit history and socio-economic status of the borrowers. The description of some of the more important covariates is shown in Table 3 (for more details, see

[9]The Department of Statistics at the University of Munich was not able to provide more specific information about the period in which the data were collected. The data set is already used in Fahrmeir and Hamerle (1984).

[10]Based on the average USD/DM exchange rate at the time, these loans range from approx. \$140 to \$10,000 in 1980 US dollars. Exchange rate data were obtained from Global Financial Data Inc., at http://www.globalfindata.com/.

the URL given above).

In our application we consider the problem of a profit-maximizing lender described in Section 3.1. The relevant objective function is shown in Table 1 on p. 6. The case of the utility-maximizing lender is not essentially different, either theoretically or in terms of implementation. We caution, however, that the empirical exercise presented below is for purposes of illustration only, for at least three reasons.

First, the lending framework introduced in Sections 2 and 3 is obviously highly stylized; a "serious" application should be based on somewhat more realistic assumptions. Nevertheless, if it follows from whatever assumptions are made that the lender's decision is binary and there are essentially two possible states of the world, then a similar analysis applies.

Second, not all variables required by the theory in Section 2 are available in the data set. In particular, it is not possible to recover the interest rate r charged on a loan. We attempt to make up for this deficiency by constructing a proxy for the interest rate. We take the government bond closest in maturity to the loan and set the loan rate equal to the average annual yield on the bond in 1980, plus a markup of 10 percentage points.[11] (Thus, the interest rate is completely determined by the duration of the loan.) Of course, this proxy is very crude and likely does not capture the true variation in the interest rates charged on the loans recorded.

Finally, a full-fledged application may also need to address the reject inference problem mentioned in Section 2. The essence of the problem is that the available sample is conditioned on the selection rule that the bank used when granting loans. New applicants cannot, in principle, be judged on the basis of an approval rule estimated from samples so obtained.[12] Nevertheless, the approach illustrated here is appropriate for deciding whether to extend a new loan to someone who had previously held one with the bank (or a bank with similar policies).

[11] The bond yields used in the construction of the proxy loan rate were obtained from Global Financial Data Inc., at http://www.globalfindata.com/; see Table 4. Given the available data set, it is natural to take months as a unit of time. Interest rates must then be measured as monthly rates. Given an annual interest rate y, the corresponding monthly rate m is given by the formula $m = (1 + y)^{1/12} - 1$.

[12] Formally, new applicants are a "realization" from the entire distribution of (Y, X), whereas observations in the sample are a realization from some truncated version of this distribution.

Table 3: Variables in the German banking data set

NAME	UNITS	DESCRIPTION
COMPLY	binary	The dependent variable. 1: in compliance with terms; -1 (or 0): in violation of terms
SIZE	DM	The size of the loan (250-18424).
DURATION	months	Duration of the loan (4-72).
INSTLRATE	categorical (1-4)	Installment as a fraction of monthly disposable income. 1: over 35% … 4: below 20%
PURPOSE	categorical (0-10)	Purpose of the loan. E.g. car, furniture, appliances, education, vacation etc.
SAVINGS	categorical (1-5)	Balance of savings account or value of stocks. 1: no savings … 5: over DM1000
CHECKING	categorical (1-4)	Checking account balance. 1: no account; 2: zero balance or debt; 3: between 0 and DM200; 4: over DM200
SAVINGS	categorical (1-5)	Value of savings. 1: NA/no savings … 5: over DM1000
ASSETS	categorical (1-4)	Most valuable fixed asset. 1: no assets … 4: house/land owner
HISTORY	categorical (0-4)	Credit history. Higher numbers indicate better credit history.
COSIGNER	categorical (1-3)	1: no cosigner; 2: co-applicant; 3: guarantor
OLDLOANS	categorical (1-4)	No. of loans at this bank. 1: only current one … 4: 6 or more
OTHLOANS	categorical (1-3)	Further running credits. 1: at other bank; 2: commercial credit; 3: none
AGE	years	Age. (19-75)
SEX–MS	categorical (1-4)	Sex and marital status. 1: male: divorced or separated; 2: female: divorced, separated or married; or single male; 3: male: married or widowed; 4: single female (Very cryptic definition.)
OCCUP	categorical (1-4)	Type of job. 1: unemployed … 4: highly skilled/self employed
CURREMPL	categorical (1-5)	Length of current employment. 1: unemployed … 5: 7 or more years
CURRADDR	categorical (1-4)	Number of years at current address. 1: less than 1 … 4: over 7
HOUSING	categorical (1-4)	1: rented 2: owner occupied 3: no cost to borrower
PHONE	categorical (1-2)	1: no phone line under customer's name; 2: yes
FOREIGN	binary	Foreign worker. 1: yes; 2: no

Table 4: German government bond yields in 1980

Maturity (months)	3	12	24	36	48	60
Yield (%, annual)	7.85	9.04	8.79	8.67	8.61	8.56

5.2 Specifications

The first step is to decide on the vector $X = (\ddot{X}, \tilde{X})$ of covariates to be used in the exercise. Since the main purpose of the exercise is illustration, we will keep the dimension of X relatively low so that the numerical optimization of $\hat{S}_n(\theta)$ does not become excessively burdensome.

By the theoretical considerations in Section 2 and 3, the relevant properties of the loan contract are (1) the size L of the loan; (2) the duration D of the loan; (3) the interest rate r of the loan. Given the limitations of the data set, these characteristics will be measured by the following vector:

$$\ddot{X} = (\text{SIZE, DURATION, INTPROXY})',$$

where INTPROXY denotes the duration-based interest rate proxy described in the previous section.

In the framework of Section 3, the vector (L, D, r), along with the recovery rate q, completely determine the objective function of a profit-maximizing lender and, consequently, the optimal cutoff for the probability of compliance. For the reader's convenience, we restate the formula for the optimal cutoff:

$$c(L, D, r) = \left[1 - \frac{d(f_D, D)d(r, D)^{-1} - 1}{(1 + f_D)^{-D}q - 1}\right]^{-1}, \tag{22}$$

where $d(\cdot, \cdot)$ is defined after equation (1).

The duration of the loan has a direct as well as indirect effect on the cutoff value; the indirect effect is through the risk free rate f_D and the proxy for r, which is also constructed on the basis of D. The risk free rate (the opportunity cost of the loan) is measured by the yield on the German government bond with maturity closest to the duration D (see Table 4). In order to further emphasize the role of a variable cutoff, we make the assumption that the recovery rate q is a decreasing function of the size of the loan L, as shown in Table 5. In this case the size of the loan will also have an (indirect) effect on the value of the cutoff.

In contrast to the components of $\ddot{X}$, the economic theory presented in the paper does not directly say which personal characteristics of the borrower

Table 5: The recovery rate as a function of the size of the loan

L (1000DM)	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-
q_L	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6	0.55	0.5

Note: In case of a tie, the higher recovery rate is assigned.

should be included in $\tilde{X}$. In order to facilitate the specification of $\tilde{X}$, we first estimated a simple logit regression of the default indicator on all the covariates in Tables 3. The results are shown in Table 6. We then selected covariates that appeared to have a significant statistical relationship with the probability of compliance/default. In particular, we choose

$$\tilde{X} = (\text{CONSTANT, HISTORY, CHECKING})'.$$

A number of other variables appear to be relevant as well; most of these were left out for the sake of parsimony. Some variables were, however, excluded for specific reasons. For example, we dropped INSTLRATE because it enters the logit regression with a counterintuitive sign or SEX–MS because of its cryptic definition. Furthermore, as described in Section 3.3, U.S. law would prohibit the lender from using covariates such as SEX–MS and FOREIGN in modeling $p(x)$.

Finally, one must specify a model for $p(x)$. To keep the computational burden to a minimum and to keep the example simple, we chose a linear specification:

$$m(x, \theta) = x'\theta.$$

That is, the lender's decision rule is constrained to be of the form $\text{sgn}[x'\theta - c(x)]$. As far as decisions are concerned, this specification is equivalent to a logit or probit model with a linear index, since the latter is just a monotone transformation of the former.

5.3 Estimation results

As discussed in Section 4.2, the lender can obtain an asymptotically optimal approval rule—conditional on the linear specification for $m(x, \theta)$—by solving

$$\max_{\theta} \hat{S}_n(\theta) = \max_{\theta} n^{-1} \sum_{i=1}^{n} b(X_i)[Y_i - 2c(X_i) + 1]\text{sgn}[X_i'\theta - c(X_i)]. \quad (23)$$

For a fixed realization of the sample $\{Y_i, X_i\}_{i=1}^n$, the function $\hat{S}_n(\theta)$ is a

Table 6: A logit regression of COMPLY on all covariates using the full sample

Name	Coeffs	Std Error	Z-stat	Prob
C	-4.1462	1.0044	-4.1279	0.0000
SIZE	-9.31E-05	4.01E-05	-2.3190	0.0204
DURATION	-0.0245	0.0087	-2.8091	0.0050
INSTLRATE	-0.2940	0.0825	-3.5628	0.0004
PURPOSE	0.0320	0.0301	1.0626	0.2880
CHECKING	0.5812	0.0700	8.3026	0.0000
SAVINGS	0.2376	0.0582	4.0835	0.0000
ASSETS	-0.1834	0.0909	-2.0173	0.0437
HISTORY	0.3847	0.0873	4.4042	0.0000
COSIGNER	0.3449	0.1777	1.9404	0.0523
OLDLOANS	-0.2531	0.1604	-1.5774	0.1147
OTHLOANS	0.2462	0.1109	2.2188	0.0265
AGE	0.0084	0.0081	1.0323	0.3019
SEX–MS	0.2470	0.1146	2.1543	0.0312
OCCUP	0.0256	0.1363	0.1883	0.8506
CURREMPL	0.1479	0.0709	2.0858	0.0370
CURRADDR	-0.0153	0.0773	-0.1977	0.8432
HOUSING	0.2841	0.1672	1.6994	0.0892
PHONE	0.2936	0.1877	1.5638	0.1179
FOREIGN	1.1557	0.6096	1.8957	0.0580

step function of the parameter vector θ with at most 2^n distinct values.[13] Because of its "coarse" nature, gradient-based algorithms cannot be used to maximize $\hat{S}_n(\theta)$. Instead, a search-based algorithm must be used. Simulated annealing has been shown to find the global optimum of functions with a range of unpleasant properties; see Corana et al. (1987) and Goffe et al. (1994) for description. The choice of the "cooling schedule" is important for obtaining good convergence, and there are a variety of other computational nuances that can help the method more easily find the global optimum.

We will now describe the exercise designed to evaluate the performance of the maximum utility (MU) method using the specifications described in Sections 5.1 and 5.2. Formally, the decision rule under consideration is given by

$$\mathrm{sgn}[X_i'\hat{\theta}_n^{\mathrm{MU}} - c(\ddot{X}_i)],$$

where $\hat{\theta}_n^{\mathrm{MU}}$ solves (23). The maximum utility method is compared against

[13]The function $\hat{S}_n(\theta)$ is a sum of n terms, where only the *sign* of each term is affected by the value of θ. There are 2^n different ways one can conceivably assign signs to these n terms and, therefore, the sum of the signed terms has at most 2^n different values.

two benchmarks: (i) a logit model based on a linear index estimated by maximum likelihood and combined with a constant cutoff; (ii) the same logit model combined with the optimal cutoff function $c(\ddot{x})$ derived from the decision maker's problem. Formally, the two benchmark decision rules are given by

$$(i)\ \mathrm{sgn}[\Lambda(X_i'\hat{\theta}_n^{\mathrm{ML}}) - c] \text{ and } (ii)\ \mathrm{sgn}[\Lambda(X_i'\hat{\theta}_n^{\mathrm{ML}}) - c(\ddot{X}_i)],$$

where Λ denotes the c.d.f. of the logistic distribution. In decision rule (i) the constant cutoff c is chosen so as to produce the same in-sample acceptance rate as does decision rule (ii).

The available sample of 1000 observations is divided into two parts: a randomly chosen subset consisting of 600 observations is used to estimate the decision rules in question, while the remaining 400 observations are held out for out-of-sample evaluation. Given an estimated decision rule $d(X, \hat{\theta})$, we calculate the net present value associated with each observation, both in sample and out of sample:

$$NPV_i = \pi_{A,1}(\ddot{X}_i)1_{\{d(X_i,\hat{\theta})=1,Y_i=1\}} + \pi_{A,-1}(\ddot{X}_i)1_{\{d(X_i,\hat{\theta})=1,Y_i=-1\}},$$

where $1_{\{\cdot\}}$ denotes the indicator function. (Recall that $\pi_{R,1} = \pi_{R,-1} = 0$.) Next, four averages (expected values) are calculated: (1) in sample average NPV per applicant; (2) in-sample average NPV per approved application; (3) out-of-sample average NPV per applicant; (4) out-of-sample average NPV per approved application. In addition, for each decision rule we report in- and out-of-sample approval and rejection rates, and the percentage of "correct" decisions, broken down as the percentage of good loans among approved applications, and the percentage of bad loans among rejected applications.

It turns out that the results of the calculations described above are sensitive to the particular subsamples chosen for estimation vs. evaluation. Therefore, the entire exercise is repeated 250 times, each time with a different randomly chosen subsample (of size 600) for estimation (and the remaining observations for evaluation). In Table 7 we report the averages of the statistics described above over the 250 repetitions.

As seen in Table 7, the rejection rates associated with the estimated decision rules are around 55 percent. Given that these decision rules were estimated using data on loans that were actually approved by a banker at some point in the past, one might wonder why the obtained rejection rates are so high. While it is possible that the previous banker's approval process

Table 7: The performance of decision rules based on MU vs ML estimation. Estimation samples=600 obs.; evaluation samples=400 obs. Reported figures are AVERAGES over 250 repetitions.

Method	Cutoff	A	$G\|A$	R	$B\|R$	ENPV per appl.	ENPV per loan
			IN-SAMPLE			(DM)	(DM)
ML, logit	0.785	0.445	0.884	0.555	0.446	10.40	22.92
ML, logit	$c(\ddot{x})$	0.444	0.875	0.556	0.439	19.60	44.08
MU, linear	$c(\ddot{x})$	0.456	0.886	0.544	0.454	51.21	115.70
			OUT-OF-SAMPLE			(DM)	(DM)
ML, logit	0.785	0.447	0.875	0.553	0.443	-4.30	-8.57
ML, logit	$c(\ddot{x})$	0.445	0.869	0.555	0.437	9.41	22.51
MU, linear	$c(\ddot{x})$	0.454	0.857	0.546	0.432	16.12	35.29

Note: A: acceptance rate; $G|A$: proportion of good loans among accepted; R: rejection rate; $B|R$: proportion of bad loans among rejected; ENPV per appl.: expected (average) net present value per applicant in DMs; ENPV per loan: expected (average) net present value per loan approved in DMs.

was suboptimally lenient, another explanation is that the decision environment (and hence the lender's objective) was different from that underlying the cutoff function $c(\ddot{x})$ specified here. What the result says is that *if* one's objectives are adequately captured by the cutoff $c(\ddot{x})$, then it would be optimal to reject over half of these individuals if they applied for a loan again. The high rejection rate simply means that in the decision environment described *here* it is optimal for the lender to be rather conservative.

The theoretical prediction that, conditional on model specification, the MU estimator leads to higher average profits is borne out by the results shown in Table 7. This is, of course, very much expected in the in-sample exercise: unless the numerical procedure used to maximize the empirical score breaks down, the MU-based decision rule should, by construction, do no worse in-sample than either of the logit-based decision rules. In particular, the MU-based decision rule is expected to outperform logit with a constant cutoff for two reasons: first, because the optimal cutoff value varies from contract to contract; and, second, because maximum likelihood estimation might be inconsistent with the goal of profit maximization.

Indeed, the per applicant in-sample average NPV associated with the MU procedure in Table 7 is DM 51.21, which is roughly five times as large as the number corresponding to logit with a constant cutoff.[14] The logit model with a variable cutoff has a per applicant in-sample average NPV of DM 19.60, which is an improvement over the logit model with a constant cutoff, but is still less than 40 percent of the MU value. The comparison of the two logit models suggests that even if estimation is undertaken by ML, the profit-maximizing variable cutoff is worth using.

Favorable in-sample figures notwithstanding, the real test of the MU method is in its out-of-sample performance. Not surprisingly, in the out-of-sample exercise all decision rules do worse on an absolute scale. Nevertheless, when compared with each other, their relative performance is roughly unchanged. The per applicant out-of-sample average NPV associated with the MU-based decision rule is DM 16.12, which is about 70 per cent larger than logit with a variable cutoff (DM 9.41). Logit with a constant cutoff produces a *loss* of DM 4.30 per applicant; the relative performance of this method, when compared with the other two, is actually worse out-of-sample than in-sample.

In sum, the MU-based decision rule apparently continues to outperform the ML/logit-based decision rules out of sample. This suggests that, given the model specifications, the maximum utility estimator succeeds in capturing relevant features of the theoretically optimal decision rule $\mathrm{sgn}[p(x)-c(\ddot{x})]$ that the maximum likelihood estimator does not.

6 Summary and conclusions

In this paper we construct a theoretically optimal loan approval rule when the lending process is regarded as a binary decision/binary outcome problem, and estimate it using the method developed by Elliott and Lieli (2007), taking the lender's economic objective explicitly into account.

We briefly examine the impact of legal regulations on the lender's decision problem. In addition to taking into consideration the prohibition of disparate treatment, we also show how a simple modification of the lender's objective function might reduce the unintended disparate impact of the proposed approval rules.

[14] At first glance, an average NPV of DM 51.21 may seem rather small. To interpret this number correctly, one must keep three things in mind: (1) This is a per applicant (as opposed to per loan) average and the rejected applicants carry a value of zero. Per loan averages are higher. (2) This number measures *economic* profits, i.e. a value of zero would mean that the lender did just as well as if he had invested in risk-free government bonds. (3) There is implicit averaging over the duration and size of the loans. A DM 51.21 excess return on a DM 200 loan over four months would constitute a very good investment.

The analysis implies that it is in general optimal to use a context-specific cutoff for credit scores or, equivalently, the conditional probability of compliance/default. The optimal cutoff can vary from individual to individual, because the relative cost of the two types of errors the decision maker can make in this simple context may not be the same for all individuals or loan contracts.

The estimator used here ensures that the credit scoring model is fit well at those points where the conditional probability of compliance intersects the optimal cutoff. If a *given* parametric model is estimated using this method, the resulting decision rule will generally lead to more profitable lending decisions than if the model were estimated by a traditional method such as maximum likelihood. This property is demonstrated using real world data and the effect seems to hold up out of sample. Although the maximization of the objective function used in the estimation requires a tedious numerical procedure, the method is feasible in practice. A shortcoming of the empirical analysis is that the reject inference problem is assumed away.

References

[1] Arrow, K., 1973, The theory of discrimination, in: O.A. Ashenfelter and A. Ries, (Eds.), Discrimination in Labor Markets. Princeton University Press, Princeton, N.J., pp. 3-33.

[2] Barefoot, J.S., 1997, The next fair-lending Risk: Disparate impact, ABA Banking Journal 89.

[3] Becker, G., 1971, The economics of discrimination, 2nd ed., University of Chicago Press, Chicago, IL.

[4] Boyes, W., D. Hoffman and S. Low, 1989, An econometric analysis of the bank credit scoring problem. Journal of Econometrics 40, pp. 3-14.

[5] Cocheco, S., 1997, Compliance chief details OCC probe of credit-scoring programs, ABA Banking Journal 89.

[6] Corana, A., M. Marchesi, C. Martini and S. Ridella, 1987, Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm, ACM Transactions on Mathematical Software 13, pp. 262-280.

[7] Crook, J. and J. Banasik, 2004, Does reject inference really improve the performance of application scoring models?, Journal of Banking and Finance 28, pp. 857-874

[8] Davidson, R. and J.G. MacKinnon, 1984, Convenient specification tests for logit and probit Models, Journal of Econometrics 25, pp. 241-262.

[9] Elliott, G. and Lieli, R.P., 2007, Predicting Binary Outcomes, Manuscript, Department of Economics, University of Texas, Austin.

[10] Fahrmeir, L. and Hamerle, A., 1984, Multivariate statistische Verfahren, De Gruyter, Berlin.

[11] Feelders, A.J., 2002, An overview of model based reject inference for credit scoring, Manuscript, Institute for Information and Computing Sciences, Utrecht University.

[12] Fortowsky, E. and M. LaCour-Little, 2001, Credit scoring and disparate impact, Manuscript, Wells Fargo Home Mortagage.

[13] Goffe, W.L., G.D. Ferrier and J. Rogers, 1994, Global optimization of statistical functions with simulated annealing, Journal of Econometrics 60, pp. 65-99.

[14] Hand, D.J. and W.E. Henley, 1997, Statistical classification methods in consumer credit scoring, Journal of the Royal Statistical Society, Series A (Statistics in Society) 160, pp. 523-541.

[15] Ladd, H., 1998, Evidence on discrimination in mortgage lending, Journal of Economic Perspectives 12, pp. 41-62.

[16] Manski, C.F., 1975, Maximum score estimation of the stochastic utility model of choice, Journal of Econometrics 3, pp. 205-228.

[17] Manski, C.F., 1985, Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, Journal of Econometrics 27, pp. 313-333.

[18] Phelps, E.S., 1972, The statistical theory of racism and sexism, American Economic Review 86, pp. 25-53.