AN ANALYTIC MODEL FOR CREDIT ANALYSIS

Daniel E. O’Leary

ABSTRACT

Credit decisions generally are based on judgments about the applicant derived from a report about a set of applicant attributes. Credit management does not directly observe applicant attributes. Instead the applicant provides a report of the relevant attribute set in the credit application. Unfortunately, those reports are not always perfectly reliable. In addition, attempts to clarify the reports are also not perfectly reliable. As a result, this chapter extends the credit model to include reliability.

Reliability is introduced by distinguishing between the report and the actual attribute information. It is found that credit decisions can be quite sensitive to changes in reliability. Some important properties of the model, including monotonicity are explored in some detail. In addition, the analytic model is compared to some heuristics. As a result of the critical point nature of the problem, it is found that the heuristics may provide misleading recommendations.

1. INTRODUCTION

Building models of the installment credit analysis decision has received much attention over the last 60 years. Recent efforts to analyze this problem include
expert systems (e.g. Srinivasan and Kim (1988)). However, initial analysis began with Durand (1941) and basic model building was initiated with Myers and Forgy (1963) and Smith (1964).

Credit (or installment credit) is characterized by an application process. The credit applicant provides a ‘report’ on their current situation for a number of attribute variables (attributes). Ultimately, it is from this report that credit management decides whether to grant credit. Management does not directly observe the credit applicant’s attributes, so it must depend on the report.

Unfortunately, there are a number of reasons to assume that those reports are not perfectly reliable. Errors in the reports that credit management receives can arise from a number of sources for a number of reasons. It may be that applicants report differently than actual because they anticipate that the actual attributes will not result in them getting the installment credit they are pursuing. It may be that the credit applicant has misunderstood or forgotten. In addition, errors can occur in the recording and transcription of information, either by the applicant or by employees (resulting in so-called ‘computer errors’). Even if credit management goes through the process of verifying the attribute information, recent news releases indicate that those verification sources often are in error.

Unfortunately, previous decision models have assumed that the actual process and the report were the same, i.e. that the reports were perfectly reliable. Thus, the purpose of this study is to introduce reliability of attribute information into credit models to study the impact on the optimal strategy.

Reliability is defined roughly as the probability that the report and the actual state of the attribute are the same. If they are the same there is perfect reliability and the probability equals one. It is found that even small changes in reliability can have a significant impact on decisions. Further, when the report is at its minimum reliability, it is found that the credit report should have no impact on the decision. In some situations it is found that the credit decision functions are monotonic in the reliability parameters. Since the installment credit decision is a critical point process, the impact of changes in reliability can be investigated.

If the models that management uses do not account for the reliability of the attribute information, management interested in incorporating reliability might use some heuristic adjustment of an installment credit model. Unfortunately, as noted in this work, such an adjustment can yield misleading results.

The focus of this study is on consumer installment credit analysis. However, this model can be directly applied to other financial decisions, such as the granting of a credit card or limits on credit cards.

2. A DECISION THEORETIC MODEL OF CREDIT

This work assumes a simple decision theory approach to the analysis of installment credit decisions. I couch the analysis in the context of this model so that the impact of reliability can be studied in detail. Throughout the interest is in the behavior of credit decisions as a function of attribute reliability. More complex models might be used but they could obscure the basic focus on reliability.

There are two primary issues that result from that model developed in this section, beyond the model formulation. First, the credit analysis decision is a ‘critical point’ problem. Second, one of the primary parameters in the credit decision analysis is a likelihood ratio.

2.1. Assumption of Independent Attributes

In this study applicant attributes gathered for the credit process are assumed to be independent. If the attributes are independent then the interaction effects can be largely ignored and reliability discussed directly. Interaction effects also could be ignored if the credit decision was based on a ‘winner take all’ strategy. Such a strategy might be represented, for example, if the attribute with the smallest critical point was used as the sole criteria for credit granting. The strategy of ‘winner take all’ might be used in those credit situations where there was very tight control on credit, where one criteria dominated other criteria (e.g. conviction for fraud) or where one attribute was found to be a particularly strong predictor. Alternatively, the heuristic approach used by Smith (1964) of...
assuming the interaction effects away could be used. Although that approach has been criticized in the literature (Cohen & Hammer (1966)), it may be found to be empirically appropriate in specific applications.

Finally, there are economic and legal incentives to ensure that the attributes are independent. If the attributes are not independent then that might indicate that the attribute gathering effort includes redundant effort. As a result, for economic reasons, we might anticipate that the gathering and review efforts could be reduced if the attributes were independent. Further, if some attributes are found to be highly correlated with attributes such as race, religion, etc. then it may be that there were legal reasons for ensuring the attribute space is relatively independent.

2.2. Critical Point Model

The original model and notation used by Smith (1964) will be the starting point in this work. It is assumed that management will grant installment credit if the expected value of granting the credit exceeds the expected losses. Let \( L \) be the expected loss on bad accounts and let \( R \) be the expected return on good accounts. Throughout it is assumed that a marginal analysis approach is appropriate.

Let accounts for which a loss is experienced be denoted \( B \) (bad accounts). Let all other accounts be denoted as \( G \) (good accounts). The entire set of accounts is represented by the sets \( B \) and \( G \).

Let there be a set of \( n \) independent attributes, \( A_i \), associated with each applicant. Since attributes are independent, we can examine the impact of individual attributes, independently. Thus, for the remainder of the paper the index will be dropped and simply referred to as \( A \) or \( x \).

Let \( p \) be the probability of a bad account given that we know some attribute \( A \). The expected return on the good accounts can be obtained by multiplying \( (1 - p) \) by the average return on good accounts, \( R \). The expected return on the bad accounts can be obtained by multiplying \( p \) by the average loss on bad accounts, \( L \). Thus, we would not grant installment credit if

\[
pL > (1 - p)R, \text{ or } p > R/(L + R).
\]

Using the example of Smith (1964), if \( L = 400 \) and \( R = 20 \), then \( p = 0.0476 \). Thus, if \( p \) is greater than 0.0476 then installment credit management would not grant credit. As a result, this model is called a critical point model, since at or under the value of 0.0476 credit is granted, but over that value, credit is not granted. The critical point nature of installment credit decisions will be used later in this chapter.

2.3. A Likelihood Ratio Formulation

Let \( x, y \) and \( z \) be random variables. Let \( Pr(x | y, z) \) be the probability of \( x \) given \( y \) and \( z \). Let \( q \) be the prior probability of a bad account \( q = Pr(B) \). Using Bayes theorem, we can relate the posterior \( p \) to the prior probability of a bad account, \( Pr(B | x) \), where \( x \) is attribute information. Thus,

\[
p = Pr(B | x)
\]

\[
Pr(B | x) = Pr(B, x)/Pr(x)
\]

\[
Pr(B | x) = q/[q + (1 - q)L(x)],
\]

where

\[
L(x) = Pr(x | G)/Pr(x | B).
\]

\( L(x) \) is a likelihood ratio and the only parameter in the right-hand side equation for \( Pr(B | x) \) that includes attribute information. The likelihood ratio is the basis of the generation of the reliability arguments throughout the rest of the chapter.

The actual attribute information changes the prior \( q \) to yield \( p \). \( Pr(B | x) \) signals one of two potential actions. If the critical value is below \( Pr(B | x) \) then the account is considered a good account and credit is granted. If the critical value is greater than \( Pr(B | x) \) then the attribute signals a bad account, and the credit is not given.

3. A MODEL OF INSTALLMENT CREDIT WITH ATTRIBUTE RELIABILITY

One limitation of the model in Section 2 is that it does not explicitly account for the reliability of the attribute information. \( L(x) \) assumes that the actual attribute information is provided, not a report of the information. That model does not differentiate between the actual applicant attributes and the application or report on the applicant. This section adapts the model of \( Pr(B | x) \) to depend instead on the report. As a result, the concern is with \( Pr(B | y) \), where \( y \) is the report version of \( x \). This is done through the one parameter in \( Pr(B | x) \) that is concerned with the attributes, \( L(x) \) and is summarized in Lemma 1. Let \( x' \) represent 'not \( x \).

**Lemma 1**

\[
L(y) = [Pr(y | G, x)Pr(x | G) + Pr(y | G, x')Pr(x' | G)]/[Pr(y | B, x)Pr(x | B) + Pr(y | B, x')Pr(x' | B)]
\]
Proof

\[ L(y) = \frac{\text{Pr}(y \mid G)\text{Pr}(G)}{\text{Pr}(y \mid B)\text{Pr}(B)} \]
\[ L(y) = \frac{[\text{Pr}(y, x, G) + \text{Pr}(y, x', G)]\text{Pr}(G)}{[\text{Pr}(y, x, B) + \text{Pr}(y, x', B)]\text{Pr}(B)} \]
\[ L(y) = \frac{[\text{Pr}(y \mid G, x)\text{Pr}(x \mid G)\text{Pr}(G) + \text{Pr}(y \mid G, x')\text{Pr}(x' \mid G)\text{Pr}(G)]/\text{Pr}(G)}{[\text{Pr}(y \mid B, x)\text{Pr}(x \mid B)\text{Pr}(B) + \text{Pr}(y \mid B, x')\text{Pr}(x' \mid B)\text{Pr}(B)]/\text{Pr}(B)} \]

3.1. Relationship of \( L(y) \) and \( L(x) \)

The \( \text{Pr}(y\mid\ldots) \) are referred to the reliability of the reported evidence by Schum and DuCharme [1971]. If \( \text{Pr}(y \mid G, x) = \text{Pr}(y \mid B, x) = 1 \) and \( \text{Pr}(y \mid G, x') = \text{Pr}(y \mid B, x') = 0 \) then the report is perfectly reliable. In that situation, \( L(y) \) reduces to \( L(x) \).

3.2. Symmetric and Asymmetric Reliability

The model in lemma 1 has four different reliability parameters. In order to make the discussion more tractable and to focus on the impact of reliability, two special cases of \( \text{Pr}(y\mid\ldots) \) will be investigated. This is done to reduce the number of parameters in order to facilitate the discussion.

The first assumption is one of symmetric reliabilities, where \( \text{Pr}(y \mid G, x) = r \). In this case lemma 1 reduces to the following theorem.

Theorem 1

If reliability is symmetric then

\[ L(y) = \frac{[r\text{Pr}(x \mid G) + (1 - r)\text{Pr}(x' \mid G)]/\text{Pr}(x \mid B) + (1 - r)\text{Pr}(x' \mid B)]}{[\text{Pr}(y \mid B, x)\text{Pr}(x \mid B)\text{Pr}(B) + \text{Pr}(y \mid B, x')\text{Pr}(x' \mid B)\text{Pr}(B)]/\text{Pr}(B)} \]

The second assumption is that there is asymmetric reliability, where \( \text{Pr}(y \mid G, x) = r_1 \) and \( \text{Pr}(y \mid G, x') = r_2 \). In this case \( L(y) \) can be written as follows.

Theorem 2

If reliability is asymmetric then

\[ L(y) = \frac{[r_1\text{Pr}(x \mid G) + r_2\text{Pr}(x' \mid G)]/\text{Pr}(x \mid B) + r_2\text{Pr}(x' \mid B)]}{[\text{Pr}(y \mid B, x)\text{Pr}(x \mid B)\text{Pr}(B) + \text{Pr}(y \mid B, x')\text{Pr}(x' \mid B)\text{Pr}(B)]/\text{Pr}(B)} \]

Thus, in order to develop \( \text{P}(B \mid y) \), rather than using \( L(x) \) we would use \( L(y) \). It will be seen in the example, later in the chapter that \( L(y) \) (and thus \( \text{P}(B \mid y) \)) is very sensitive to minor changes in reliability.

4. MONOTONICITY RESULTS

This section develops results that indicate that the functions \( L(y) \) and \( \text{P}(B \mid y) \) are monotonic in reliability. Since the decision for installment credit is a critical point decision, this result allows us to study the impact of different levels of reliability on that decision.

Definition 1 (Gaughan (1968))

Let \( R \) be the set of real numbers. Let \( D \) be a subset of \( R \). Let \( f : D \rightarrow R \). The function \( f \) is said to be increasing (decreasing) if and only if for all \( x \) and \( y \) in \( D \) with \( x \leq y \), \( f(x) \leq f(y) \). If \( f \) is either increasing or decreasing then \( f \) is said to be monotone.

4.1. Symmetric Case

The direction of the monotonicity result is dependent on the relationship between \( \text{Pr}(x \mid G) \) and \( \text{Pr}(x \mid B) \). The following lemma is developed for the situation where it is likely that the account is \( B \), i.e. \( \text{Pr}(x \mid G) \leq \text{Pr}(x \mid B) \). However, the result can be easily extended to the opposite case.

Lemma 2

Assume that reliability is symmetric. If \( \text{Pr}(x \mid G) \leq \text{Pr}(x \mid B) \) and \( r \geq r' \), then \( L(r', y) \leq L(r, y) \); that is, \( L(r, y) \) is monotone decreasing in \( r \).
Proof—by contradiction
Assume that \( r' \geq r \) and \( L(r', y) \geq L(r, y) \). Let \( p_1 = \Pr(x \mid G) \) and \( p_2 = \Pr(x \mid B) \).

\[
[r'p_1 + (1 - r')(1 - p_1)][r'p_2 + (1 - r')(1 - p_2)] \geq [r_1p_1 + (1 - r_1)(1 - p_1)][r_1p_2 + (1 - r_1)(1 - p_2)]
\]

However, \( r' \geq r \), thus there is a contradiction and \( L(r', y) \not\geq L(r, y) \).

Lemma 2 has direct implications for \( \Pr(B \mid x) \), as noted in the following, Theorem 3.

Theorem 3
Assume the case of symmetric reliability. If \( \Pr(x \mid G) \not\leq \Pr(x \mid B) \) and \( r' \geq r \), then \( \Pr(B \mid r', x) \geq \Pr(B \mid r, x) \), i.e. \( \Pr(B \mid r, x) \) is monotonically increasing in \( r \).

Proof—By Lemma 2
The results in theorem 3 have direct implications for the credit decision. The results indicate that \( \Pr(B \mid x, r) \) is monotonically increasing in \( r \). If a credit model does not directly account for reliability, then \( r = 1 \) is assumed. Thus, \( \Pr(B \mid x) \) is assumed to be higher than it actually should be when \( r < 1 \). Since installment credit is a critical point decision this indicates that an account is classified as \( G \) when it should be classified as \( B \). Thus, by not accounting for reliability, inappropriate decisions can be made.

Theorem 4
Assume the case of asymmetric reliability. If \( \Pr(x \mid G) \geq \Pr(x \mid B) \) and \( r' \geq r \), then \( \Pr(B \mid r', x) \leq \Pr(B \mid r, x) \), i.e. \( \Pr(B \mid r, x) \) is monotonically decreasing in \( r \).

If the installment credit model does not account for reliability, then that implies that \( r = 1 \). In this case, \( \Pr(B \mid x, r) \) is assumed to be lower than it should be when \( r < 1 \). Since installment credit is a critical point decision, this indicates that an account will be classified as \( G \) when it should be a \( B \). Accordingly, by not accounting for reliability inappropriate decisions can be made.
In the situation of perfect reliability \( r_1 = 1 \) and \( r_2 = 0 \). Thus, under perfect reliability \( r_1 r_2 \) approaches infinity. Theorem 5 indicates that as we move away from perfect reliability, it is seen that \( \Pr(B|y) \) will be overestimated, unless we account for reliability. As a result, too many cases will be estimated as B. In that situation, too many cases will be estimated as G. As a result, in either case it is important to consider the impact of reliability in the analytic analysis of installment credit.

5. EXAMPLES AND THE USE OF HEURISTICS

An example can be used to study the sensitivity of \( L(y) \) to changes in reliability and the impact of the use of heuristics rather than the analytic approach presented here.

5.1. Example

The impact of reliability will be illustrated using an example drawn from Smith (1964). In that example, for the attribute \( A = 'rents a room' \), \( \Pr(A|G) = 0.013 \) and \( \Pr(A|B) = 0.088 \).

The results are summarized in Table 1. As seen in an examination of the data in that table, reliability has a substantial impact on both \( L(y) \) and \( \Pr(B|y) \). For example a change of \( r \) from 1 to 0.99 yields an increase in \( L(y) \) of over 60% and a decrease in \( \Pr(B|y) \) of almost 40%.

At \( r = 1.00 \) \( \Pr(B|y) \) exceeds the critical point established in the example in Section 2. However, as the reliability in the attribute information drops the \( \Pr(B|y) \) drops below the critical point. This would lead to a change in decisions from categorizing the account as a B to a G.

<table>
<thead>
<tr>
<th>Table 1. Example – Symmetric Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>0.99</td>
</tr>
<tr>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
</tr>
<tr>
<td>0.80</td>
</tr>
<tr>
<td>0.70</td>
</tr>
<tr>
<td>0.60</td>
</tr>
<tr>
<td>0.50</td>
</tr>
</tbody>
</table>

5.2. Heuristics

Since the model in Section 2.1 does not accommodate reliability, if management wishes to integrate reliability into the installment credit decision, then management is likely to use any of a number of heuristics, including,
- do nothing, that is, use the existing model,
- increase or decrease the resulting probability, \( \Pr(B|x) \) by a constant factor related to the reliability, e.g. 10% for a 90% reliability, or
- increase or decrease the probabilities \( \Pr(x|I) \) by say 10% to reflect reliability.

The example can be used to illustrate those heuristics can guide the user to misleading decisions. As noted above, if the decision maker were to do nothing then the decision model could be very misleading. Similarly, since the changes in \( \Pr(B|y) \) were around 40%, an a change to \( \Pr(B|y) \) by an amount the same as the reliability would also yield misleading results. Finally, if both \( \Pr(x|B) \) and \( \Pr(x|G) \) are changed by an amount corresponding to the reliability, then that amount will just factor out and the value of \( \Pr(B|y) \) would be unchanged. Thus, in general, such heuristic adjustments do not capture the impact of reliability.

6. SOME IMPLEMENTATION CONSIDERATIONS

In order to implement the approach discussed in this chapter the parameters in the models need to be estimated. Although the implementation problems can be difficult, there are still feasible solutions.

Typically, the credit process generates substantial data. In addition, annual audit of the credit process generally produces substantial analysis of that data. This data can be used throughout the probability estimation process. Unfortunately, there may be a bias in the data, since if an applicant is judged as a B, then there will be no further information available, in general. However, if they are judged a G, then we will generate additional data.

The model developed in this chapter requires the estimation of the reliability parameters, such as \( \Pr(y|x,..) \). The model requires a minimum of one
reliability parameter, in the case of symmetric probability, and up to four reliabilities in the most general case. In some cases, it may be that the single parameter case would be used to provide an estimate of the impact of reliability, before further research was committed to a more general model. In any case, these probabilities may be estimated using data quality information. Some indications of data quality can be generated using complaints from customers on the incorrectness of data used in evaluating them. However, that data is likely to be biased to errors resulting in negative assessments of credit.

There has been substantial research on assessing probabilities for financial applications. Felix (1976) explored a number of methods for developing prior probabilities. Wright (1987) analyzed alternative approaches for reducing the risk of misspecified prior distributions.

Finally, Bayes' theorem can be used to develop some of the probabilities necessary for this investigation.

7. EXTENSIONS AND SUMMARY

This section presents some extensions of the model, additional uses of the reliability model for credit assessment and a brief summary of the chapter.

7.1. Extensions to the Basic Decision Model

There are a number of extensions that can be made to the basic decision model in this work. First, the model might be extended to the situation where there are multiple attributes and the attributes are not assumed to independent.

Second, this study employed a single period model. That model could be extended to a multiple period model (e.g. Bierman and Hausman (1970) and others). If more than a single period model is required then that can indicate that marginal costing is not appropriate. In these situations, rather than a marginal approach, alternative approaches such as total costs or profit maybe more appropriate.

Third, this work used a two state world (B and G). Alternatively, we could have used fuzzy sets or additional states in our probability model to generate 'in-between' states.

7.2. Additional Uses of the Reliability-based Model

This model was used to examine the impact of reliability on heuristic solutions. Since expert systems use heuristics, the approach in this paper may be useful in the investigation of the quality of heuristics in expert systems for the credit decision (Srinivasan and Kim (1988)).

The model developed in Section 2.2 formed the basis of an index used to in the installment credit decision of applicants Smith (1964). Smith's (1964) index model added the posterior probabilities associated with each of the attributes and multiplied by 1000 (an arbitrary index base). Although there have been justified criticisms (Cohen and Hammer (1966)), that index can be extended using the reliability-based approach of this paper. That index can be extended to include the reliability-based estimates \( \Pr(B | y) \), rather than the model that does not include reliability \( \Pr(B | x) \).

7.3. Summary

This chapter developed an approach to introduce reliability into the installment credit decision. It was shown that model was very sensitive to small changes in reliability. In addition, under the situation where the uncertainty about the reliability was at a maximum, the it was found that the reported level of the applicant attributes was not relevant to the decision as to whether the applicant was a good account or a bad account.

The study also investigated embedding reliability in the installment credit decision rather than using heuristic approaches. It was demonstrated that selected ad hoc heuristics may generate misleading results. Because the decision is a critical point decision, the heuristic approaches could result in significantly different credit decisions.

The model was examined in detail for two special cases of reliability: symmetric and asymmetric. It was found that credit decision is monotonic in reliability. Since credit granting is a critical point decision, monotonicity allows the study of the impact of reliability.

REFERENCES


