Bound and Collapse Bayesian Reject Inference When Data are Missing not at Random

Gongyue G Chen
Department of Management Sciences
University of Waterloo
Waterloo, ON, CANADA, N2L 3G1
g3chen@engmail.uwaterloo.ca

Thomas Åstebro
Rotman School of Management
University of Toronto
105 St George Street
Toronto, ON, CANADA, M5S 3E6
astebro@rotman.utoronto.ca

September, 2003

Abstract

We describe how reject inference in credit scoring can be considered a special case of dealing with data Missing Not At Random (MNAR). We argue that if data are missing completely at random or missing at random there is no need for reject inference. We propose a bound and collapse Bayesian reject inference technique to solve the problem of data MNAR. We test the method's performance relative to other methods using real data. Results show that this technique improves classification power of credit scoring models under MNAR conditions.

Key Words: reject inference, missing data, Bayesian, bound and collapse

† Åstebro acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada operating grant # OGP 0183683.

1. Introduction

Non-random sample selection appears with certainty in credit scoring where complete data is typically only available for those that have gone through a screening process and have been accepted for credit. Comprising a number of statistical and/or mathematical techniques, reject inference is used to impute estimates for the non-selected sub-population from the selected sub-population. The importance of reject inference has been widely recognized in academy and industry. However, most, if not all, of the proposed reject inference techniques are unsatisfactory (Hand and Henley, 1993/4).

According to Chen and Åstebro (2001), current reject inference techniques can be grouped into three classes. The first class is the ideal where the selected sample is known to be representative of the whole population. No reject inference is then necessary. For the second class, reject inference implementations assume that the proportion of good accounts is the same for the rejects as for the accepted. In the third class it is assumed that the distributions over the accepted and rejected regions are different. Techniques in the first class completely avoids selection bias, but are expensive to implement. Unfortunately, conclusions regarding other reject inference techniques[1] are discouraging. For example, Hand and Henley (1993/4) show that the methods employed in industry are problematic as they typically rest on very tenuous assumptions. Crook and Banasik (2002) for example conclude that the usefulness of a re-weighting method is questionable and methods of extrapolation appear to be useless. Empirical research of Heckman's bivariate two-stage model for credit scoring shows little promise (e.g., Jacobson and Roszbach, 1999; Chen and Åstebro, 2001; and Banasik, Crook and Thomas, 2001). More fundamentally, Poirier (1980) argues that quantifying the efficiency loss associated with sample selection bias is not possible without reference to a particular data set, and Meng and Schmidt (1985) claim that the cost of partial observability is very high.

The reject inference problem can be formulated as one of missing data. Feelders (2000) proposes a method based on mixture modeling that can be estimated using the EM algorithm. Chen and Åstebro (2001) argue that this reject inference method belongs to the second class of techniques where the distributions of the accepted and rejected regions

[1] In this paper we do not review various reject inference techniques. interested reader may refer to papers by Crooks and Banasik (2002), Chen and Åstebro (2001), Ash and Meester (2002), Hand and Henley (1993/4) and Hand (1998).

are assumed similar. This assumption is rather questionable. While Feelders (2001) points out that this restrictive assumption may need to be relaxed the only method proposed by him to solve the problem is the bivariate two-stage model, which has previously shown to be ineffective.

An intuitive approach is to use supplementary information about the rejected region for reject inference. For example, Ash and Meester (2002) suggest obtaining bureau data on accepts and rejects at the end of the observation period, and using the performance with other creditors over the observation period to infer how the rejects would have performed had they been accepted. Hand and Henley (1993/4) consider using calibration samples as input of supplementary information. They also propose to use a mixture decomposition approach by assuming some parameterized forms for the component distributions of good and bad. Besides issues of the quality of data and costs for obtaining the supplementary information, one major shortcoming of these methods is an *ad hoc* treatment of data while adding more untested assumptions. Basically, one of the large disadvantages for many of the proposed reject inference techniques is the lack of theoretical support.

In this paper we propose a new reject inference technique using a Bayesian procedure that builds on a model of missing data developed by Rubin (1976). This new technique not only incorporates the impact of the data source by imputing missing data of the dependent variable based on the estimated probabilities of missingness, but also allows for outside supplementary information about the reject region to be flexibly adjusted in the model. Fundamentally, it is a model-based technique and its theoretical support gives it a major advantage over many other reject inference techniques. We also show the ease by which it can be implemented in practice. Finally, we study the efficiency and power of this method when applied to a specific dataset.

The remainder of the paper is organized as follows: Section 2 discusses the mapping of the reject inference problem to a missing data mechanism while the model is discussed in Section 3. Section 4 outlines some issues related to the application of our new model and compares it to some other techniques. Next we test the efficiency and power of the method on two data sets and in Section 6 we discuss some related topics. We finish with a brief conclusion.

2. Mapping Reject Inference to a Missing Data Mechanism

The fundamental problem in reject inference is the non-randomness of the sample, a consequence of missing data. This non-randomness removes the grounds for the use of most statistical methods (Corps and Li, 1997). Thus, intuitively, one can think of applying some classical methods of missing data analysis for reject inference.

Before any further investigation, we need to clearly define the missing data mechanism. The current classification of missing data mechanisms is based on the work of Rubin (1976), Little and Rubin (1987) and Gelman et al. (1995). Consider a sample classified according to the values of two categorical variables X and Y, in which X, the independent variable, is always observed and Y, the response variable, is subject to missing data due to either non-response or censoring[2]. The classification of missing data mechanisms depends on whether the probability of missing data of Y depends on the state of Y and/or X. If the probability depends on X but not on Y, then data are said to be *missing at random* (*MAR*), and observed values of Y are not representative of the complete sample as a whole, but they are so, when considered within categories of X. A special case of this situation is realized when the probability of missing data of Y is neither dependent on Y nor on X. In this case, data are said to be *missing completely at random* (*MCAR*) and the observed values of Y are a representative sub-sample of the complete but unknown original sample, since observed and unobserved entries are generated by the same mechanism. When data are MAR or MCAR, the missing data mechanism is ignorable, in the sense that inference does not depend on it. When the probability of missing data of Y depends on both X and Y, the missing data mechanism is said to be *not ignorable* (*NI*), and the resulting incomplete sample is no longer representative under any condition. Sometimes nonignorable is defined as Missing Not at Random (MNAR) (e.g., Feelders, 2000).

Let's define the missing data of Y as $Y = ?$. Suppose we have a vector X of independent variables $X_1, X_2, \ldots, X_i, \ldots, X_p$, which is completely observed for each observation. By applying a credit scoring model we assign each observation (credit applicant) k a credit score as $S_k = f(X_k)$. There is a threshold or cutoff point h such that when $S_k \geq h$, credit is

[2] Censoring refers to missingness that occurs because an individual cannot respond to the question of interest.

granted to observation *k*, otherwise no credit will be granted. Denote $a = 1$ if *k* is granted credit, *0* otherwise. Credit risk, for example credit default/non-default, can be observed if $a = 1$, but is missing if $a = 0$. Denote this observed credit risk as *Y* ($Y = 1$ if default, *0* otherwise). In such a case the missing values are on the dependent variable (outcome).[3] For simplicity, assume credit scores are bound by $S_{max} \geq S_k \geq S_{min}$. Then the dataset can be set in a matrix form as figure 1 shown:

	X_1	X_2	…	X_p	S	Y	a
1					S_{max}	.	*1*
2					.	.	*1*
.					.	.	.
.					.	.	.
.					*h*	.	*1*
.					*h-1*	?	*0*
.					.	?	*0*
.					.	.	.
n					S_{min}	?	*0*

Figure 1

The pattern of missing data is defined as a univariate pattern where missing values occur on an item with credit score smaller than *h*, but a set of (*p*+1) other items, $X_1, X_2, \ldots, X_i, \ldots, X_p$, and *S*, is completely observed. The indicator variable *a* identifies what is known and what is missing. Therefore the variable *a* is referred to as an auxiliary variable for the missingness (Feelders, 2000). According to Schafer and Graham (2002), we may not have to specify a particular distribution for *a*, but we must agree that it has a distribution. Denote the complete data as Y_{com} and partition it as $Y_{com}(Y_{obs}, Y_{miss})$ where Y_{obs} is the observed sample and Y_{miss} is the missing sample. Defined by Rubin (1976), if the distribution of missingness does not depend on Y_{miss}, i.e., $P(a|Y_{com}) = P(a|Y_{obs})$, missing data are MAR. When the distribution does not depend on Y_{obs} either, data are MCAR. If the missing data for reject inference are MCAR, modern missing data theory provides a straightforward solution: one will obtain unbiased estimates by using the complete cases.

[3] We assume that a credit application is recorded for all applicants.

Given the univariate pattern for missing values in reject inference, case deletion (i.e., delete all observations with missing values) is valid for the MAR missing data mechanism. That is, the parameters of the regression of *Y* on any subset of $X_1, X_2, \ldots, X_i, \ldots, X_p$ can be estimated from the complete cases and the estimates are both valid and efficient under MAR (e.g., Graham and Donaldson, 1993). However, this result does not extend to other measures such as correlation coefficients between *Y* and *X*, and parameters of the marginal distribution of *Y*. But using the complete cases we are guaranteed to establish an unbiased credit scoring model under MAR.

Under MNAR data the problem is quite different. One must specify a model for the missingness that is at least approximately correct. In this case most reject inference techniques may not be valid as they improperly ignore specifying the missingness function. In credit scoring, suppose the model has sufficient classification power. Then it is likely that a missing datum occurs for the reason closely related to the outcome being measured. That is, observations with higher probabilities of default are most likely to have higher probabilities of having lower credit scores. This is clearly a case of a nonignorable missing data mechanism (Schafer and Graham, 2002; Little, 1995.) And this is the case a reject inference technique must be able to handle.

Notice that as we move from MCAR to MAR to MNAR, the observed values of *Y* become an increasingly select and unusual group relative to the population and the problem of sample selection exacerbates (Schafer and Graham, 2002). In the world of reject inference we therefore have two polar situations, none ideal. One is where the original credit score has no classification power for granting credit. Data are then MCAR and reject inference is not important to apply. However, there is a likely to be a credit quality problem.[4] The second situation is where the original credit score has good classification power causing significant differences in the distributions between the accepted and rejected regions. There is no credit quality problem but a large problem when using the selected sample to update the model.[5]

[4] Ironically, such problems of ineffective credit scoring models typically appear because the original model was developed on a select sample of accepted applicants.
[5] It is possible that override or other selection rules will also be applied to screen applicants. We believe this is just an extended case of MNAR where there are more than one credit score variable. We leave this for future research.

Assuming that the credit scoring model is well specified and has sufficient classification power we can safely conclude that most reject inference techniques can be mapped to a MNAR missing data mechanism. In the following section we describe one such mechanism.

3. Bayesian Analysis for Reject Inference

Define Y as the credit risk outcome and S as the credit score. We assume that in credit scoring the missing data of Y is dependent on S and the missing outcome itself -- the MNAR missing data mechanism. Define $Y = 0$ representing good accounts and $Y = 1$ bad accounts. Denote $S = i$, ($i = 1, \dots r$), and $Y = j$, ($j = 0, 1$). In some scoring applications the credit score may be a continuous variable with an upper and lower bound. In this case we simply group the score ranges into buckets that are represented by $i = 1$ to r.

As Sebastiani and Ramoni (2000) discuss, when the missing data mechanism is MNAR, the posterior distribution of the conditional probability of Y given S is still conjugate and therefore estimates of the probabilities of (S, Y) and Y as well as their posterior variance can be computed. However, posterior distributions of the joint probabilities of (S, Y) and of the marginal probabilities of Y do not have simple expressions. A popular method currently used is to resort to MCMC (Markov Chain Monte Carlo) methods, such as Gibbs Sampling, and treat missing entries as unknown quantities of interest, from which empirical estimates and credibility intervals can be computed.

Sebastiani and Ramoni (2000) however suggest another methodological framework called Bound and Collapse (BC). The intuition is that it is possible to set a bound for possible estimates of missing data within an interval defined by some extreme distribution, no matter what the missing data mechanism is. The complete set of data will provide constraints on the interval. When information about the missing data mechanism is available, it is encoded in a probabilistic model of non-response and used to select a single estimate. The second step of BC collapses the interval to a single value of missing data. By this method a randomly imputed datum will be generated to replace the missing datum.

We suggest that this Bound and Collapse method can be applied for reject inference. For this purpose we assume (S, Y) has a multinomial distribution with probabilities $\theta_{ij} = p(S = i, Y = j|\theta)$, so that $\theta= (\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \ldots, \theta_{r0}, \theta_{r1}) = (\theta_{ij})$, ($\theta_{ij} \geq 0$, for all i and j, and $\boldsymbol{\Sigma}_{ij}\theta_{ij} = 1$) parameterizes the joint distribution of (S, Y). The standard conjugate prior for θ is a Dirichlet distribution $D(\alpha)$, with $\alpha = (\alpha_{10}, \alpha_{11}, \ldots, \alpha_{r0}, \alpha_{r1})$, whose density function is:

$$p(\theta) = \prod_{i=1}^{r} \prod_{j=0,1} \frac{\Gamma(\alpha)}{\Gamma(\alpha_{ij})} \theta_{ij}^{\alpha_{ij}-1} \qquad (1)$$

with $\alpha_{ij} \geq 0$ for all i and j, and $\alpha = \Sigma_{ij}\alpha_{ij}$.

If the data sample is complete, exact Bayesian analysis presents no difficulties. For reject inference the entries on the variable Y are reported as unknown when the corresponding credit score S is smaller than the threshold value h. Similar to Sebastiani and Ramoni (2000), let $A = (A_o, A_m)$, where A_o and A_m respectively denote the sample with complete observations and the one with unknown entries on Y. Also let A_k be a possible completion of A_o. Thus $A_k = (A_0, A_{dk})$, where A_{dk} is a possible distribution of the unclassified cases in A_m. Figure 2 shows two possible completions of A obtained by distributing the incomplete case ($S = i, Y = ?$). Denote n_{ij} the frequency of complete cases and let m_i be the frequency of cases ($S = i, Y = ?$). Thus $n = \Sigma_{ij}n_{ij}$ is the number of cases completely observed, $m = \Sigma_i m_i$ is the number of cases partially observed, and $n + m$ is the sample size.

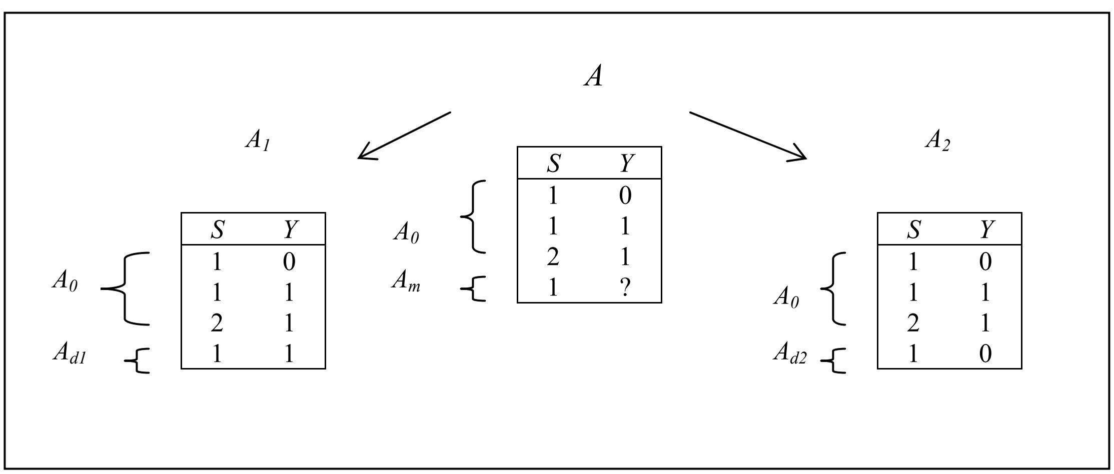


Figure 2

Following Little and Rubin (1987), we can represent the incomplete sample in an $r \times (c+1)$ contingency table (Table 1) where $c = 2$. The $(c + 1)$ - th column represents the frequency of unknown cases for each category of S.

	Y		Y
S	0	1	?
1	n_{10}	n_{11}	m_1
2	n_{20}	n_{21}	m_2
.	.	.	.
.	.	.	.
h	n_{h0}	n_{h1}	m_h
.	.	.	.
.	.	.	.
r	n_{r0}	n_{r1}	m_r

Table 1

As Sebastiani and Ramoni (2000) argue, by the Total Probability Theorem the exact posterior distribution of θ is a mixture of Dirichlet distributions weighted by the probabilities of possible completion of A, and can be computed if information about the missing data mechanism is available. However, for the nonignorable missing data mechanism the simplicity of conjugate analysis is lost. Fortunately, they show that, even without the exogenous information on the pattern of missing data, the incomplete sample is still able to induce bounds on the possible estimates consistent with the information available in the sample. When information about the missing data mechanism becomes available, it can be used to select a single estimate within the set of possible ones.

Define φ as the missing data mechanism. The possible estimates of $p(Y = j | S = i)$ given sample A are bounded as

$$p_{\bullet}(j \mid i) = \frac{\alpha_{ij} + n_{ij}}{\alpha_{i+} + n_{i+} + m_i} \leq p(Y = j \mid S = i \mid A) \leq \frac{\alpha_{ij} + n_{ij} + m_i}{\alpha_{i+} + n_{i+} + m_i} = p^{\bullet}(j \mid i) \qquad (2)$$

where $\alpha_{i+} = \Sigma_j \alpha_{ij}$, $n_{i+} = \Sigma_j n_{ij}$.

Assume that some external information on the missing data mechanism is available from which a probabilistic model for non-response can be deduced:

$$p(Y = j \mid Y = ?, S = i, \varphi, \theta) = \varphi_{j|i} \qquad (3)$$

where $\Sigma_j \varphi_{j|i} = 1$, for all i. This information can then be used to identify a point estimate within the probability interval $[p_\bullet(j \mid i), p^\bullet(j \mid i)]$ via a convex combination of the extreme probabilities:

$$\hat{p}_{j|i} = \varphi_{j|i} p^\bullet(j \mid i) + (1 - \varphi_{j|i}) p_\bullet(j \mid i) = \frac{\alpha_{ij} + n_{ij} + \varphi_{j|i} m_i}{\alpha_{i+} + n_{i+} + m_i}. \qquad (4)$$

To use this technique for reject inference, we have to explicitly specify the nonignorable missing data mechanism of equation (3). We believe that this work is a data dependent, and therefore is not the main focus of this research. To demonstrate the general methodology we will discuss some general methods for describing the proper nonignorable missing data mechanism for reject inference. We will also implement the methodology on a set of data and run several tests. Assuming we can specify the expression for equation (3), we can use equation (4) to impute data for missing data. The solution is empirically feasible and the computational requirement is tolerable. This solution incorporates the credit score into the missing data imputation.

A simple application of this method is when all credit screening is based on the original credit score *S* with cutoff value *h* and there are no other selection rules. In this case Table 1 becomes:

	Y		*Y*
S	0	1	?
1	*0*	*0*	m_1
2	*0*	*0*	m_2
.	.	.	.
.	*0*	*0*	.
h	n_{h0}	n_{h1}	m_h
h+1	.	.	*0*
.	.	.	.
r	n_{r0}	n_{r1}	*0*

Table 2

and equation (4) becomes:

$$\hat{p}_{j|i} = \varphi_{j|i} p^\bullet(j \mid i) + (1 - \varphi_{j|i}) p_\bullet(j \mid i) = \frac{\alpha_{ij} + \varphi_{j|i} m_i}{\alpha_{i+} + m_i}. \qquad (5)$$

for those score bands without observed outcome data. Equation (5) implies that the sample data does not provide valid information for missing data imputation of *Y* when the credit score *S* is in the groups *1* to *h*. The reason is that the estimated probability for missing *Y* is determined by the prior distribution of *Y as well as* the missing data mechanism.

4. **Issues Related to Applying the Model**

Although the application of the model is straightforward, some issues such as how missing data mechanism should be estimated and how the prior distribution should be selected needs to be discussed. In this section we discuss these extended issues related to the application of our model before we are able to fully describe the model application in the next section.

Recent research analyzing MNAR missingness functions reveals two fundamentally different approaches: selection models and pattern-mixture models. These two approaches exploit the conditional probability rule in opposite fashions. Selection models model $Prob(Y_{miss}, Y_{obs}) = Prob(Y_{miss}|Y_{obs})Prob(Y_{obs})$. A typical example is Heckman's bivariate two-stage model. Though selection models have intuitive appeal, the likelihood for these methods can be oddly shaped, with large, flat regions indicating that parameters are poorly identified (Schafer and Graham, 2002). As an alternative, first proposed by Little (1993), pattern mixture models model $Prob(Y_{miss}, Y_{obs}) = Prob(Y_{obs}|Y_{miss})Prob(Y_{miss})$. Pattern mixture models classify missing data by their missingness and describe the observed data within each missingness group. Without information on Y_{miss} we would not know any characteristics of its distribution and the model would be unidentified. Therefore, for this model some identifying constraints are required from information on missing data.

Our model for reject inference is closely related to pattern mixture modeling, where observations with missing data are grouped into different missingness patterns through a latent variable – the credit score *S*. With the identifying information for the missingness distribution we then obtain the distribution of complete data conditional on the missingness distribution using an imputation method. This model describes the observed data and missingness distribution for each missingness group and extrapolates aspects of this behavior to unobserved data.

To apply the proposed model, we do not posit strong theory about the missingness mechanism; rather, we suggest using internal and external information to extrapolate the probability of being missing. For this application we suggest using the probability of being bad as a proxy for the missing data mechanism because the probability of being bad is equal to the probability that the credit application is rejected. Therefore, the score is taken as the probability of being missing. In effect, the original credit score contains all "external" information useful for estimating the missing data mechanism in this simple application.[6] On the other hand, "internal" data can also provide inference for the missingness mechanism. Since we assumed that the credit score has sufficient classification power, the probability of being bad in the observed groups can also be used to infer the missingness mechanism. Some possible methods for this inference include regression, linear extrapolation or exponential extrapolation.

A better solution may be to estimate the missingness mechanism by computing a weighted average of group specific estimates using both external and internal information.[7] Once the estimation of missingness is obtained, we are able to simulate a value for each missing datum based on equation (4). Finally we can impute the missing data to generate a complete sample for future credit scoring development.

Another issue to consider is the selection of the prior distribution. The model uses the multivariate generalization of the Beta distribution, the Dirichlet, as the conjugate prior distribution (see equation (1)). There are several plausible, noninformative Dirichlet prior distributions. A uniform density is obtained by setting $\alpha_{ij} = 1$ for all *i* and *j*; this distribution assigns equal density to any vector θ satisfying $\Sigma_{ij}\theta_{ij} = 1$. Setting $\alpha_{ij} = 0$ for all *j* results in an improper prior distribution that is uniform in the $\log(\theta_{ij})$'s. The resulting posterior distribution is proper if there is at least one observation in each category of *i* and *j*. If the number of observations in the study is relatively large, one can

[6] In a more complex credit screening process additional information must be included to describe the missing data mechanism.

[7] Using combined external and internal information is superior to using one information source since one can always assign a weight of 1.00 to one information source. In the following application we assume that external and internal information have equal values.

expect no large differences in results between these two prior densities.[8] In the following test we will set $\alpha_{ij} = 0$.

5. Statistical Procedures and Model Performance

Using the 1993 and 1998 National (U.S.) Surveys of Small Business Finances (NSSBF) datasets[9] we design an experiment to test the power and efficiency of the proposed model. First, we develop a credit scoring model to predict the probability of credit delinquency using the complete 1993 development sample. Applying this credit score to the 1998 sample, we simulate a credit granting policy. A selected sample is obtained by applying a credit cutoff policy.

To investigate the robustness of the model we simulate two credit granting policies by applying two cutoff scores so that the degree of missingness is different across two selected samples. This also means that where there is potential sample selection bias the degree of bias will be different. A cutoff score is selected such that the acceptance rate is equal to the good rate (or 1- bad rate) of the complete 1998 NSSBF sample. For convenience we name this case "weak selection." The second case is "strong selection" where the cutoff score was chosen so that approximately twice the number of applicants was rejected.

We then apply the decision rules outlined in Section 4 to specify the missingness function as linearly dependent on observed bad rates. We apply two sources of data to estimate the missingness function: "external" information of the bad rate in the complete 1993 sample and "internal" information of the bad rate in the selected 1998 sample. The missingness mechanism is computed as a weighted average of both external and internal information[10]. Based on equation (4) a value for each missing datum is simulated and imputed to generate a complete sample.

---

[8] A natural extension to improve estimations further is multiple imputation [Rubin (1987), Gelman *et al.* (1995)]. One advantage of the proposed Bayesian method is that multiple imputation can easily be applied. We leave this, however, for future research.

[9] For more details please see http://www.federalreserve.gov/pubs/oss/oss3/nssbftoc.htm. We assume that the samples obtained from the Surveys are representative of the true distributions of the underlying universes.

[10] Here the selection of equal weights for external and internal information is arbitrary. In reality the weighting system is determined by the belief to what degree the missingness can be represented by these

The proposed reject inference technique is compared against an ideal situation where the missing outcome data for the 1998 sample are known. The technique is also compared against a two-stage bivariate selection correction model and against the case where no adjustments are made for missing data and where a new model is instead created based only on the selected sample.

5.1 Performance Measurement

Measuring performance of a credit scoring model is key. We are interested in how the classification power of a model for the targeted population distribution is affected by missing data and potential corrections for missing data. For this we use four measures: the Kolmogorov-Smirnov (KS) test, the "bad rate" which is computed as the proportion of observed "bad" among the accepted obligors,[11] the Brier score and the logarithmic score.[12] Note that the lower the Brier and logarithmic scores, the better the classification power.

5.2 Credit Scoring Model

The NSSBF surveys collected information on the availability of credit to small and minority-owned business. The surveys provide detailed information on firms' characteristics, owners' characteristics, business performance and the firms' financial status. Based on the 1993 survey we create a credit scoring model to predict the probability that a small firm will be delinquent 60 or more days on at least one business obligation. We define the dependent variable $Y = 1$ if a firm has been delinquent on at least one business obligation within three years, and 0 otherwise. Definitions of variables are presented in appendix A.

The 1993 NSSBF database contains 4637 observations where all enterprises operated under the current ownership during 1992 with fewer than 500 full-time equivalent employees[13]. To make the sample applicable for potential bank financing we drop

---

two information sources. A cautious procedure may be to first analyze and compare the distributions of external and internal information.

[11] We set the classification rule such that the acceptance rate (e.g., the percentage of applicant granted credit) is equal to the proportion of goods in the population.

[12] Hand (1997, 2001) reviews classification rules.

[13] The database excludes agricultural enterprises, financial institutions, not-for-profit institutions, government entities, and subsidiaries controlled by other corporations.

observations where the firm has zero total expenses or zero total liability in 1992. The development sample size is therefore 4589 observations. In this sample 20% of the population has been delinquent on at least one business obligation within three years. Appendix B presents summary statistics for all variables. We use logit regression to estimate the delinquency probability. Detailed model information is displayed in Appendix C.

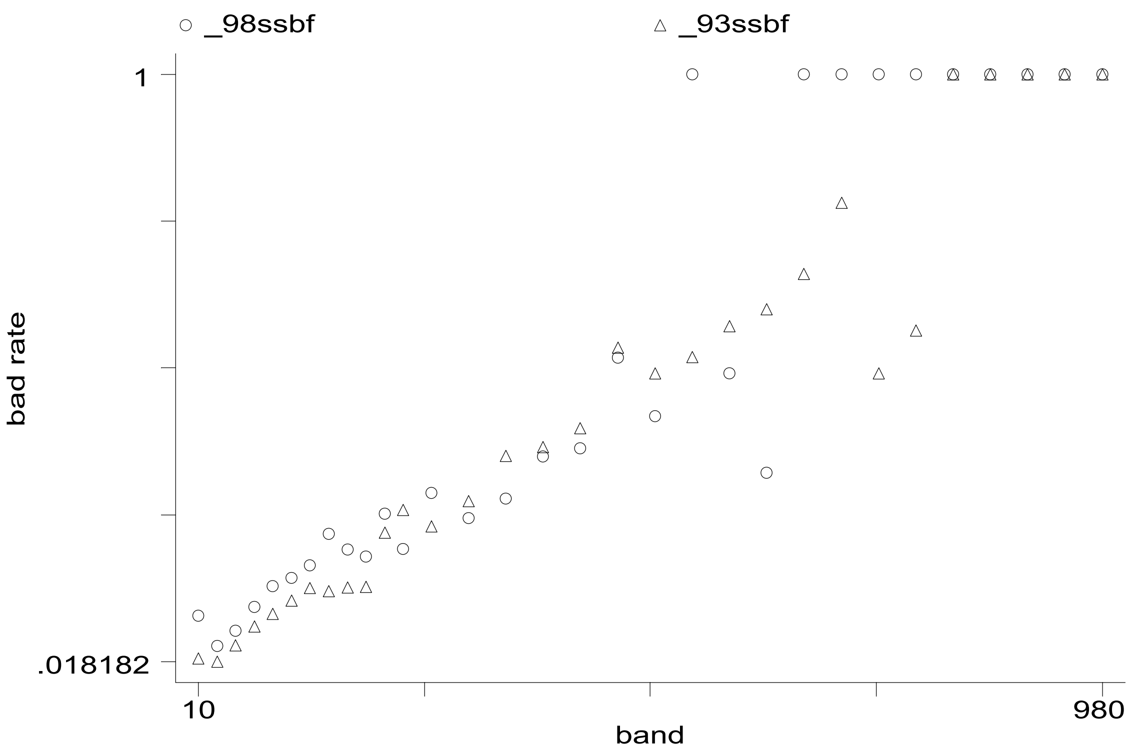
Using the model we calculate the predicted probability of being delinquent for each observation. The credit score is defined as this predicted probability times 1000 with a range from 0 to 1000. The smallest observed score is 0 and the largest is 865. The KS for the model sample is 0.35. We simulate an acceptance rate of 80%. The cutoff score *h* representing this acceptance rate is 297. Based on this cutoff score we obtain a selected sample where the bad rate is 14.2%. By randomly selecting 80% of the applicants the expected bad rate will be 20%. Therefore, the credit score will improve credit quality by 29% over random choice. The Brier score for the model is 0.28 and the logarithmic score is 0.44.

We further test the performance of the original scoring model out-of-sample using the 1998 NSSBF. The 1998 NSSBF contains 3561 observations of small businesses that were in operation during December, 1998. The survey structure is similar to that of the 1993 NSSBF. Applying the same sampling procedure as for the 1993 sample, the final size of the test sample is 2805. 493 (17.6%) of these observations have been delinquent on at least one business obligation.

The KS statistic for this sample is 0.23, a decrease of 36.4% compared to the development sample. If we choose the cutoff so that 493 observations would be denied credit, the cutoff score is 218 and the bad rate is 15.1%. Although this bad rate implies that applying the credit score will improve the credit quality by 14% over random selection, it is clear that the classification power is weaker in out-of-sample tests. The Brier and Logarithmic scores are 0.28 and 0.46 respectively, higher than those obtained for the development sample. This further indicates that the populations are different and that re-development of the scoring model may be necessary to maintain classification accuracy.

Before removing outcome data that are screened out due to the credit policy and for illustrative purposes we graph the bad rates for each score band for both the development and selected samples in Figure 3.[14] As seen, the bad rate is increasing pretty much linearly with the scores for both the development and test samples. However, there is larger variation of bad rates at higher score bands in the test sample illustrating the fact that the scoring model has less predictive accuracy out-of-sample.

Figure 3



5.3 Estimation of the Missing Data Mechanism

To implement the proposed reject inference technique the distribution of the missing data mechanism needs to be specified. As discussed in Section 4, the missing data mechanism can be represented by estimates of the bad rate. Two information sources are

[14] We split the credit score into 31 bands. For the score range from 0 to 240, we use 12 bands, all with 20 as band width. For the score range from 241 to 1000, we use 19 bands all with 40 as band width. The starting range is from 0 to 20 so that the value of band width is actually 21 for the first band.

useful to infer the missing data mechanism. First, the development sample provides a complete distribution of bad rates conditional on credit score bands ("external" information). Second, for the sample with credit screening applied, information of bad rates in the accepted region is available ("internal" information). This information can be extended to estimate the bad rate distribution in the rejected region. We weight these two sources of information equally to estimate the missing data mechanism (equation (4)).

For the external information the relationship between the bad rate and the credit score is supposed to be linear (Figure 3). We therefore use linear regression where the bad rate is the dependent variable and the score is the independent variable. However, Figure 3 shows that the distribution of 1993 bad rates have larger errors for scores above 720. Therefore we only use data from score 0 to 720 to estimate bad rates in these bands.

By applying credit screening to the 1998 sample the bad rates are not available for bands below the cutoff. Our "internal" information approach uses a linear regression model to estimate the bad rates in the reject region based on data from the accepted region for the screened 1998 sample.

To simplify the implementation we select a non-informative prior distribution so α_{ij} = 0 for all i and j. The main reason to select this prior distribution is simplicity. Also, for the case that in some bands there are no or few observations equation (5) can be assumed and simplified as $\hat{p}_{j|i} = \varphi_{ij}$. This treatment is harmless because for the band without any observations (including missing data) no missing data will be imputed. This prior distribution implies that the expected probability of being bad will be determined by the missing data mechanism when there is no observation in the bands. For score bands where there are observations the complete model (i.e., equation (4)) is implemented. Appendix D shows an example of how the estimated expected probabilities of being bad are calculated for each band having missing data. Note that in this simple case the bound of equation (2) becomes [0, 1].

Based on the estimated expected probabilities of being bad for each band we randomly simulate a value for the missing observation, either 0 or 1. Thus we obtain complete data samples where missing data are imputed. A new credit scoring model could be estimated based on these imputed complete samples.

We apply two credit screening policies: weak and strong selection. Setting the acceptance rate equal to the good rate in the complete 1998 sample the cutoff score 218 is obtained. The result shows 17.6% or 493 observations have been screened out. The second case is "strong selection" where the cutoff score is 160 such that 33.3% or 934 observations have been screened out.

To make sure that the assumption of missing not at random is valid, we also perform within-sample tests. These tests are presented first. For the case of weak selection we drop 20% (or 918 observations) of the 1993 sample based on the credit score. The cutoff score for this acceptance rate is 297. For the strong selection case, we select the cutoff score 200 such that 38.5% (or 1767 observations) of the 1993 sample is dropped. To estimate the missing data mechanism for the 1993 sample we estimate the bad rates of the rejected region using the data obtained in the accepted region. All the estimated results for missing data mechanisms are presented in Table 3.

Table 3 presents estimations for missing data mechanisms of the 1993 and 1998 samples [cols. (4), (7), (8) and (9)] over the score bands [col. (1)]. There are two missing data selection biases for each sample: weak and strong selection. To estimate the missing data mechanism for the 1998 sample we use two information sources that are equally weighted. External information contains the bad rate distribution obtained from the 1993 sample [*e.g.* col. (2)]. Internal information contains the bad rate distribution of the accepted region in the 1998 sample [*e.g.* col. (3)]. The bad rate distribution for missing data score bands is predicted using a regression model based on data from the accepted region. To estimate the missing data mechanism for the 1993 sample we assume the only available information is the bad rate distribution from the accepted region of the 1993 sample. Since data on bad rates for the 1993 sample for scores over 721 are scarce a regression model is used to predict these values.

Table 3: Estimation for missing data mechanisms.

Band	*1998 NSSBF sample*						*1993 NSSBF Sample*	
	Strong Selection			*Weak Selection*			*Strong Selection*	*Weak Selection*
	1993 NSSBF bad rate	*1998 NSSBF bad rate*	*Missing data mechanism*	*1993 NSSBF bad rate*	*1998 NSSBF bad rate*	*Missing data mechanism*	*Missing data mechanism*	*Missing data mechanism*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
161-180	14.19%	23.26%	18.72%					
181-200	14.34%	25.56%	19.95%					
201-220	23.42%	27.87%	25.64%	23.42%	26.52%	24.97%	18.15%	
221-240	27.12%	30.17%	28.64%	27.12%	27.07%	27.09%	19.73%	
241-280	24.41%	33.63%	29.02%	24.41%	29.98%	27.20%	22.10%	
281-320	28.63%	38.23%	33.43%	28.63%	33.87%	31.25%	25.26%	28.63%
321-360	36.22%	42.84%	39.53%	36.22%	37.75%	36.98%	28.42%	33.21%
361-400	37.72%	47.45%	42.59%	37.72%	41.64%	39.68%	31.58%	37.08%
401-440	40.87%	52.06%	46.46%	40.87%	45.52%	43.20%	34.74%	40.94%
441-480	54.32%	56.67%	55.49%	54.32%	49.41%	51.86%	37.90%	44.81%
481-520	50.00%	61.27%	55.64%	50.00%	53.29%	51.65%	41.06%	48.68%
521-560	52.73%	65.88%	59.30%	52.73%	57.18%	54.95%	44.22%	52.54%
561-600	57.89%	70.49%	64.19%	57.89%	61.06%	59.48%	47.38%	56.41%
601-640	60.71%	75.10%	67.91%	60.71%	64.95%	62.83%	50.54%	60.27%
641-680	66.67%	79.71%	73.19%	66.67%	68.83%	67.75%	53.70%	64.14%
681-720	78.57%	84.31%	81.44%	78.57%	72.72%	75.65%	56.86%	68.01%
721-760	76.30%	88.92%	82.61%	76.30%	76.60%	76.45%	60.02%	71.87%
761-800	80.45%	93.53%	86.99%	80.45%	80.49%	80.47%	63.18%	75.74%
801-840	84.61%	98.14%	91.37%	84.61%	84.37%	84.49%	66.34%	79.61%
841-880	88.77%	100.00%	94.38%	88.77%	88.26%	88.51%	69.50%	83.47%
881-920	92.92%	100.00%	96.46%	92.92%	92.15%	92.53%	72.66%	87.34%
921-960	97.08%	100.00%	98.54%	97.08%	96.03%	96.55%	75.82%	91.21%
961-1000	100.00%	100.00%	100.00%	100.00%	99.92%	99.96%	78.98%	95.07%
Weight	0.5	0.5		0.5	0.5			

(1) A regression model is used to estimate the bad rates in the range 721 –1000 for the 1993 sample.
Bad rate = -0.00589 +0.001039*score + error
(-0.557) (36.017)
where value in ( ) is the *t* ratio. The R square is 0.983, and adjusted R square is 0.983.

(2) In the case of weak selection for 1998 NSSBF sample, the regression model to estimated the bad rates of the rejected region of 1998 NSSBF sample is
Bad rate = 0.04728 + 0.0009713*score + error
(2.909) (7.583)
The R square is 0.865, and adjusted R square is 0.850.

(3) In the case of strong selection for 1998 NSSBF sample, the regression model to estimated the bad rates of the rejected region of 1998 NSSBF sample is
Bad rate = 0.03673 + 0.001152*score + error
(1.975) (5.712)
The R square is 0.919, and adjusted R square is 0.845.
In the case of weak selection, the bad rate for band 201-220 is estimated from the observed data directly.

(4) In the case of weak selection for 1993 NSSBF sample, the regression model to estimated the bad rates of the rejected region of 1993 NSSBF sample is
Bad rate = 0.003456 + 0.0009666*score + error
(0.286) (13.292)
The R square is 0.936, and adjusted R square is 0.931.

(5) In the case of strong selection for 1993 NSSBF sample, the regression model to estimated the bad rates of the rejected region of 1993 NSSBF sample is
Bad rate = 0.01558 + 0.00079*score + error
(1.504) (8.792)
The R square is 0.906, and adjusted R square is 0.894.

5.4 Performance

The proposed reject inference technique is compared against an ideal situation where the missing outcome data are known. The technique is also compared against a two-stage bivariate selection correction model and against the case where no adjustments are made for missing data and where a new model is instead created based only on the selected sample. Four different measures of classification power are used for these tests. Before presenting out-of-sample results we present within-sample tests on the 1993 data.

Results of within-sample tests on the 1993 sample show that the model performance of the Bayesian reject inference technique is better than that using the censored sample (Table 4). The superiority of the proposed reject inference over "ignorance" is more obvious in the case of strong selection where more applicants are screened out. In this case we know for sure that data are missing not at random and there is lots of bias in the remaining sample. The Bayesian technique then works well.

Table 4: Model performance for the 1993 NSSBF sample.

		Weak selection			*Strong selection*		
	Ideal	RI	Bivariate	Censored	RI	Bivariate	Censored
KS	0.354	0.354	0.346	0.358	0.332	0.257	0.314
Bad rate	14.2%	14.5%	14.3%	14.4%	14.6%	16.3%	15.4%
Brier score	0.281	0.283	0.289	0.284	0.288	0.314	0.296
Logarithmic score	0.443	0.445	0.452	0.446	0.451	0.493	0.462

1. Ideal = model based on complete sample; RI= model using reject inference; Bivariate = bivariate two-stage probit model; Censored = model based on censored sample (accepted).
2. For model estimation results see Appendix A, B and C.
3. The bad rate is calculated based on the sample reject rate at 20% (918 observations screened out).

Table 5: Model performance for the 1998 NSSBF sample.

		Weak selection			*Strong selection*		
	Ideal	RI	Bivariate	Censored	RI	Bivariate	Censored
KS	0.361	0.355	0.347	0.350	0.327	0.337	0.353
Bad rate	12.5%	12.5%	12.4%	12.3%	13.0%	13.0%	12.6%
Brier score	0.251	0.254	0.255	0.252	0.261	0.264	0.255
Logarithmic score	0.402	0.408	0.404	0.405	0.418	0.424	0.409

1. Ideal = model based on complete sample; RI = model using reject inference; Bivariate = bivariate two-stage probit model; Censored = the model based on the censored sample.
2. The new credit model for the 1998 sample is different from the model based on the 1993 sample (Appendix C). Some independent variables are the same, but many are different. In particular, the financial ratios used in the 1993 model are not significant in the 1998 model, and are therefore not included. In appendix E we present variables used for the 1998 sample. For simplicity we neither present the details for this new model, nor for the bivariate two-stage model since it is not the focus of this research.
3. The bad rate is calculated based on the sample reject rate at 17.6% (493 observations screened out).

Table 5 shows that in out-of-sample tests the proposed reject inference is not better than the model based on the (censored) sample of accepted credits. In the case of strong selection the reject inference model is clearly inferior to the model using the censored sample when judging on its classification power. We believe the results shown in Table 5 do not imply that the reject inference technique does not work. The poor classification power is instead caused by underlying assumption not being satisfied – in this case we believe some data in the 1998 sample to be missing at random in addition to being missing not at random. Since, as implemented, the RI technique cannot account fully for MAR data its classification power is reduced. There is evidence to support this claim. First, although the ideal model is still the most efficient, the difference to other models is not large, especially in the case of weak selection where the MNAR data constraint is less severe. This means that at least in the case of weak selection, the reject inference, the bivariate two-stage method and the model using the censored sample are quite similar, and also very close to that of the ideal model. Second, the time lapsed in between the two surveys (five years) is quite long and cover different stages of the business cycle causing the models based on the 1993 and 1998 samples to be quite different (as evidenced from Appendix A and E). Many of the independent variables used in these two models are different. Especially, the financial ratios that are significant in the 1993 sample do not have any predictive power in the 1998 sample. This implies that the credit score based on the 1993 sample is not a strong classifier for the 1998 sample. Among all three tested models that based on the censored sample is most efficient. All this evidence imply that in this out-of-sample test, the efficiency gained by using the censored sample overwhelms the weak bias caused by the inefficient credit screening when applying the 1993 model. Therefore, it is inconclusive to judge the efficiency of the proposed reject inference technique based on the 1998 data. Rather, it supports the claim that if the missingness is missing at random, a model based on the censored sample is efficient and effective.

The following results presented in Table 4 and 5 are consistent across the two datasets and demonstrate evidence that are consistent with theoretical expectation: First, Heckman's bivariate two-stage model does not work well for reject inference. This model is generally inferior to the model based on the censored sample as well as the reject inference technique. One explanation could be model specification problems. In practice one is unable to determine the true model specification (a necessary assumption is Heckman's model is that it true) and the strong sensitivity of the model to model specification makes the usefulness of the bivariate two-stage model for reject inference questionable. Second, going from weak selection to strong selection the power of using reject inference techniques decreases across all techniques tested here. This is understandable since more data loss means that any method to recover lost information will be more unreliable. However, from the test under the condition of MNAR, it seems that the Bayesian model that we propose decreases least in power while Heckman's model decreases most in power.

6. Conclusion

In this paper we describe how reject inference in credit scoring can be considered a special case of dealing with data Missing Not At Random (MNAR). We argue that if data are missing completely at random (MCAR) or missing at random (MAR) there is no need for reject inference. Research shows that under MAR or MCAR most missing data imputation models may not be superior to simply dropping missing data (so-called list-wise deletion) (e.g., Roth and Switzer, 1995; Chen and Åstebro, 2003).

When the missing data mechanism is MNAR a commonly suggested reject inference method is Heckman's two-stage model. However, most empirical research shows that this model may neither be efficient nor useful for reject inference. We instead propose a new reject inference technique for missing credit quality data using a bound and collapse Bayesian imputation procedure. This method is comparatively simple to implement and can easily be extended to allow for multiple imputation.

Using the 1993 and 1998 NSSBF surveys databases we test the proposed model. Results show that the model will improve classification power when there is sample selection bias caused by data missing not at random. This reject inference technique requires estimating the missingness distribution. We demonstrate that by using available censored information on the bad rate distribution from the accepted sample as well as information on the complete bad rate distribution from the original credit scoring development sample to impute missing data one obtains improvements in credit score classification power..

Rsults also show that the proposed reject inference technique is better at compensating for data MNAR than Heckman's bivariate two-stage model. However, when the problem is that data are MAR the proposed model may not be efficient when compared to using the censored sample and just ignoring the sample selection bias.

Appendix A: Definition of Variables for the 1993 Sample

Name	Definition
Deliq	Dependent variable. = 1 if a firm had been 60 or more days delinquent on at least one business obligation within the past three years, else 0.
trad	=1 if the business operates in wholesale trade, retail trade, or services, defined by the U.S. Bureau of the Census as the following SIC 2-digit, else 0.
othliab	= 1 if the firm owned any other assets that were not listed in the survey, else 0.
payable	= 1 if the firm have any accounts payable, else 0.
ratio1	The ratio of sales 1992 over total expense
ratio3	The ratio of profit 1992 over total expense
ratio6	the ratio of sales 1992 over total liability
ratio7	the ratio of profit 1992 over total liability
small	= 1 if the firm had less than 20 equivalent full-time employees, else 0.
isminor	= 1 if more than 50 percent of the firm owned by blacks or African Americans, Asians, Pacific Islanders, American Indians, or Alaskan Natives, else 0.
young	= 1 if the age of the principal owner was smaller than or equal to 40, else 0.
workexp2	= 1 if the owner had more than 5, but smaller than and equal to 10 years work experience, else 0.
workexp3	= 1 if the owner had more than 10, but smaller than and equal to 20 years work experience, else 0.
workexp4	= 1 if the owner had more than 20 years work experience, else 0.
selfdo	= 1 if this business was founded by the current owner(s) , else 0.
newfirm	= 1 if the firm age was smaller than or equal to 5 years, else 0.
local	= 1 if the area of sales was the same area as the firm's main office, else 0.
wide	= 1 if the area of sales was notional or international, else 0.
export	= 1 if the firm exported outside of the United States, else 0.
lineloan	= 1 if during 1993 the firm had business line of credit, else 0.
lease	= 1 if during 1993 the firm had capital leases from financial institutions or other sources, else 0.
motor	= 1 if during 1993 the firm had loans on motor vehicles used primarily for business purposes, else 0.
equloan	= 1 if during 1993 the firm had loans secured by equipment, or was the firm financing any purchases of equipment by installment payments, else 0.
othloan	=1 if as of year end 1993, excluding trade credit or credit with suppliers and loans from banks, the firm have any other loans from financial institutions or from any other sources, else 0.
trade	=1 if during 1993 the firm purchase any goods or services on account during 1993 rather than pay for the purchases before or at the time of delivery, else 0.
short	=1 if during 1993 the firm had ever required financing for seasonal or unexpected short-term credit needs, else 0.
newequ	= 1 if during the last three years, the firm has obtained additional equity capital from existing owners, their relatives, or from new or existing partners, else 0.
inven	= 1 if the firm had an inventory of merchandise or production materials
isassets	= 1 if firm held any bonds; held any stocks for short-term investment; or had any prepaid expenses or other current assets, else 0.
invest	=1 if the firm was owed any money for mortgages or real estate, or did the firm have any other investments, else 0.
land	= 1 if the firm owned any land, else 0.
typeprop	= 1 if firm type is proprietorship, else 0.
typepart	=1 if firm type is partnership, else 0.
typecor	=1 if firm type is corporation, else 0.

Appendix B: Variables Statistics

Variable	Mean	Std. Dev.	Min	Max
deliq	0.200	0.400	0	1
trad	0.594	0.491	0	1
othliab	0.430	0.495	0	1
payable	0.748	0.434	0	1
ratio1	1.336	2.740	0	160
ratio3	0.353	2.743	-1	159
ratio6	69.041	2956.558	0	200000
ratio7	7.556	266.704	-360	18000
small	0.642	0.479	0	1
isminor	0.170	0.376	0	1
young	0.210	0.407	0	1
workexp2	0.165	0.371	0	1
workexp3	0.375	0.484	0	1
workexp4	0.383	0.486	0	1
selfdo	0.713	0.453	0	1
newfirm	0.206	0.404	0	1
local	0.538	0.499	0	1
wide	0.153	0.360	0	1
export	0.123	0.328	0	1
lineloan	0.360	0.480	0	1
lease	0.159	0.365	0	1
motor	0.270	0.444	0	1
equloan	0.501	0.500	0	1
othloan	0.143	0.350	0	1
trade	0.681	0.466	0	1
short	0.265	0.441	0	1
newequ	0.203	0.402	0	1
inven	0.623	0.485	0	1
isassets	0.280	0.449	0	1
invest	0.217	0.412	0	1
land	0.269	0.443	0	1
typeprop	0.326	0.469	0	1
typepart	0.076	0.265	0	1
typecor	0.368	0.482	0	1

Appendix C: Credit Scoring Model Estimation

Number of obs = 4589
LR chi2(33) = 527.22
Prob > chi2 = 0.0000
Log likelihood = -2033.014
Pseudo R2 = 0.1148

deliq	Coef.	Std. Err.	z	P>|z|
trad	-0.2071	0.0853	-2.429	0.015
othliab	0.2580	0.0906	2.848	0.004
payable	0.5768	0.1097	5.258	0.000
ratio1	1.2796	0.6223	2.056	0.040
ratio3	-1.4006	0.6196	-2.261	0.024
ratio6	-0.0023	0.0012	-1.928	0.054
ratio7	-0.0063	0.0035	-1.803	0.071
small	0.2061	0.1037	1.987	0.047
isminor	0.3300	0.1039	3.176	0.001
young	0.2178	0.1058	2.059	0.039
workexp2	0.3392	0.1748	1.940	0.052
workexp3	0.1645	0.1704	0.966	0.334
workexp4	0.1534	0.1838	0.835	0.404
selfdo	0.1662	0.0925	1.797	0.072
newfirm	0.2607	0.1077	2.420	0.016
local	0.0129	0.0920	0.141	0.888
wide	-0.3946	0.1385	-2.850	0.004
export	0.4923	0.1344	3.662	0.000
lineloan	-0.2273	0.0948	-2.397	0.017
lease	0.4716	0.1037	4.547	0.000
motor	0.1702	0.0882	1.930	0.054
equloan	0.1724	0.0918	1.878	0.060
othloan	0.4918	0.1042	4.719	0.000
trade	0.5396	0.0998	5.404	0.000
short	0.8761	0.0843	10.388	0.000
newequ	0.5384	0.0913	5.897	0.000
inven	0.1661	0.0911	1.824	0.068
isassets	-0.2454	0.1048	-2.342	0.019
invest	-0.2522	0.1064	-2.371	0.018
land	-0.2118	0.0982	-2.156	0.031
typeprop	0.2946	0.1215	2.424	0.015
typepart	0.1768	0.1712	1.033	0.302
typecor	0.2929	0.1061	2.760	0.006
Constant	-4.8164	0.6885	-6.996	0.000

Appendix D: Estimation for expected probabilities being bad

Band	$a_{i,1}$	a_{i+}	$n_{i,1}$	n_{i+}	m_i	$\varphi_{1,\iota}$	expected $p_{1\|i}$
201-220	0	0	27	113	20	24.97%	24.06%
221-240	0	0	0	0	92	27.09%	27.09%
241-280	0	0	0	0	147	27.20%	27.20%
281-320	0	0	0	0	93	31.25%	31.25%
321-360	0	0	0	0	57	36.98%	36.98%
361-400	0	0	0	0	34	39.68%	39.68%
401-440	0	0	0	0	16	43.20%	43.20%
441-480	0	0	0	0	19	51.86%	51.86%
481-520	0	0	0	0	7	51.65%	51.65%
521-560	0	0	0	0	0	54.95%	54.95%
561-600	0	0	0	0	0	59.48%	59.48%
601-640	0	0	0	0	0	62.83%	62.83%
641-680	0	0	0	0	0	67.75%	67.75%
681-720	0	0	0	0	0	75.65%	75.65%
721-760	0	0	0	0	0	76.45%	76.45%
761-800	0	0	0	0	0	80.47%	80.47%
801-840	0	0	0	0	0	84.49%	84.49%
841-880	0	0	0	0	0	88.51%	88.51%
881-920	0	0	0	0	0	92.53%	92.53%
921-960	0	0	0	0	0	96.55%	96.55%
961-1000	0	0	0	0	2	99.96%	99.96%

Note: Weak selection case, 1998 NSSBF sample.

Appendix E: Definition of Variables for the 1998 Sample

Name	Definition
Deliq	Dependent variable. = 1 if a firm had been 60 or more days delinquent on at least one business obligation within the past three years, else 0.
Ownman	= 1 if the owner was managing the firm, else 0.
edu1	= 1 if owner's education level is below highschool, else 0.
Edu3	= 1 if owner has college education, else 0.
Edu4	= 1 if owner has university undergraduate education, else 0.
Edu5	= 1 if owner has university graduate or higher education, else 0.
Mortgage	= 1 if as the end of 1998 the firm had any mortgage used for business purposes, else 0.
Equloan	= 1 if during 1998 the firm had loans secured by equipment, or was the firm financing any purchases of equipment by installment payments
Nowliab	= 1 if the firm had any current liability excluding loans and account payable.
Home	= 1 if owner owned a home, else 0.
Fstate	= 1 if the firm used financial statement/accounting reports as written records, else 0.
Saving	= 1 if the firm had saving account, else 0.
Transer	= 1 if as the end of 1998 the firm used transaction services for business purposes, else 0.
Trustser	= 1 if as the end of 1998 the firm used trust services for business purposes, else 0.
Brokser	= 1 if as the end of 1998 the firm used brokerage services for business purposes, else 0.
Traddeny	= 1 if any supplier that offers trade credit to business customers denied a request by the firm for trade credit, else 0.
Payable	= 1 if the firm have any accounts payable, else 0.
Small	= 1 if the firm had less than 20 equivalent full-time employees, else 0.
Selfdo	= 1 if this business was founded by the current owner(s) , else 0.
Newfirm	= 1 if the firm age was smaller than or equal to 5 years, else 0.
Lease	= 1 if during 1998 the firm had capital leases from financial institutions or other sources, else 0.
Othloan	=1 if as of year end 1998, excluding trade credit or credit with suppliers and loans from banks, the firm have any other loans from financial institutions or from any other sources, else 0.
Trade	= 1 if during 1998 the firm had loans secured by equipment, or was the firm financing any purchases of equipment by installment payments, else 0.
Newequ	= 1 if during the last three years, the firm has obtained additional equity capital from existing owners, their relatives, or from new or existing partners, else 0.
Isasset	= 1 if firm held any bonds; held any stocks for short-term investment; or had any prepaid expenses or other current assets, else 0.
Land	= 1 if the firm owned any land, else 0.

REFERENCES

Ash, D and Meester, S. (2002). Best practices in reject inference. Presentation at Credit Risk modeling and Decision Conference, Wharton Financial Institutions Center, Philadephia, May, 2002.

Banasik, J. B., Crook, J. N. and Thomas, L. C. (2001). Sample selection bias in credit scoring models. Working paper 01/5, Credit Research Centre, University of Edinburg, England.

Bradley, Andrew P. (1996). "ROC Curves and the chi^2 Test," Pattern Recognition Letters, Vol. 17, No. 3, Pages 287--294, March 1996.

Chen, Gongyue and Thomas Åstebro (2001). The Economic Value of Reject Inference in Credit Scoring. Presented at the conference of credit scoring and credit control, Credit Research Centre, University of Edinburgh, England, September, 2001

Chen, Gongyue and Thomas Åstebro (2003). How to Deal with Missing Categorical Data: Test of a Simple Bayesian Method. *Organizational Research Methods*, Vol. 6, No. 3, 309-321.

Copas, J. B. and H. G. Li (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society*, B, 59, 55-95.

Crook, Jonathan and John Banasik (2002). Does reject inference really improve the performance of application scoring models? Working paper 02/3, Credit Research Centre, University of Edinburg, England.

Feelders, A. J. (2000). Credit scoring and reject inference with mixture models. *International Journal of Intelligent System in Accounting, Finance and Management*, 9, 1-8.

Feelders, A. J. (2001). An overview of model based reject inference for credit scoring. Working paper, Institute for Information and Computing Sciences, Utrecht University, The Netherlands.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin (1995). *Bayesian Data Analysis*. Chapman & Hall.

Graham, J. W., and Donaldson, S.I. (1993). Evaluating interventions with differential attrition: the importance of nonresponse mechanisms and use of followup data. *Journal of Applied Psychology*, 78, 119-128.

Hand, D. J. and W. E. Henley (1993/4). Can reject inference ever work? *IMA Journal of Mathematics Applied in Business & Industry* 5, 45-55.

Hand, David J. (1997). *Construction and Assessment of Classification Rules*. Chichster: Wiley.

Hand, David J. (1998). Reject inference in credit operations, in *Credit Risk Modeling: Design and Application* (ed. E. Mays), 181-190, AMACOM.

Hand, David J. (2001). Measuring Diagnostic Accuracy of Statistical Prediction Rules. *Satistica Neerlandica*, 53, 3-16.

Jacobson, Tor, and Kasper F. Roszbach (1999). Evaluating bank lending policy and consumer credit risk, in *Computational Finance 1999* (edited by Yaser S. Abu-Mostafa et al.) the MIT Press, 2000.

Little, R. J. A. and D. B. Rubin (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Little, R.J.A (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88, 125-134.

Little, R.J.A (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90, 1112-1121.

Meng, Chun-Lo, and Peter Schmidt (1985). On the cost of partial observation in the bivariate probit model. *International Economic Review*, Vol. 26, No. 1, February, 71-85.

Poirier, Dale J. (1980): "Partial observability in bivariate probit model," *Journal of Econometrics*, 12, 209-217.

Roth, Philip L., and Fred S. Switzer III (1995). A Monte Carlo analysis of missing data techniques in a HRM setting. *Journal of Management*, Vol. 21, No. 5: 1003-1023.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63: 581-592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Schafer, Joseph L., and John W. Graham (2002). Missing data: our view of the state of the art. *Psychological Methods*, Vol. 7, No. 2: 147-177.

Sebastiani, Paola, and Marco Ramoni (2000). Bayesian inference with missing data using bound and collapse. *Journal of Computational and Graphical Statistics*, Vol. 9, No. 4: 779-800.