TI 2000-090/4
Tinbergen Institute Discussion Paper

Scoring Bank Loans that may go wrong – A Case Study

J.S. Cramer

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Scoring Bank Loans That May Go Wrong: a Case Study

J.S. Cramer *

October 20, 2000

Abstract

A bank employs logistic regression with state-dependent sample selection to identify loans that may go wrong. Inspection shows that the logit model is inappropriate. A bounded logit model with a ceiling of (far) less than 1 fits the data much better.

1 Introduction and summary

Each year, a Dutch bank grants many thousands of loans to small and medium-sized firms. At the time the loans are made a number of standard financial ratios of the debtors are recorded. Upon review two years later about 3% of them are found to have moved into the danger zone by making substantial losses or by a large fall in their solvency. These loans are classified as (potentially) *bad loans* and subjected to a regime of close control.

The distinction between good and bad loans (or debtors) is used in a statistical analysis to link the probability of a loan going bad to the initial financial ratios of the debtor. The estimated risk is then used to score all loans and to establish a classification that will help local account managers to concentrate their efforts of supervision and control where they are most

*University of Amsterdam and Tinbergen Institute, Keizersgracht 482, 1017 EG Amsterdam; e-mail mars.cram@worldonline.nl. I have benefited from comments on earlier versions of this paper by Aernoud Boot, Herman van Dijk, Martin Fase, Hans van Ophem and Jan Sandee.

needed. The method of analysis employed is a routine fitting of a logit model to a selected sample composed of all bad loans and an equal number of good loans drawn at random from this group.

In the present instance it is found this method of analysis does not work. The reason is misspecification of the model: the incidence of bad loans does not obey a logit distribution. The introduction of a ceiling or maximum risk provides an easy remedy.

Two lessons can be drawn from this example. The first is that the plain logit model does not always apply; a routine goodness-of-fit test is therefore advisable. The second is that the routine estimation technique of a logit model from a state-dependent selected sample is quite sensitive to misspecification.

The principles of sample selection and their application to the logit model are recalled in section 2. The failure of this method in the present instance is described in section 3. Section 4 presents the simple alternative of a bounded logit function, which works well.

2 Estimation with state-dependent sample selection

We consider a large sample of binary $0, 1$ outcomes Y_i. The focus is on $Y_i = 1$, which has a low incidence, while the complement $Y_i = 0$ is abundant: here, $Y_i = 1$ represents a bad loan and $Y_i = 0$ a good loan. The model specifies the probability that i is a bad loan as a function of regressors x_i as

$$P(Y_i = 1|x_i) = P^*(\theta, x_i) = P_i^*. \tag{1}$$

We wish to estimate θ from a *selected sample*, discarding a large part of the abundant zero observations for reasons of expediency. Suppose that the initial *full sample* is a random sample with sampling fraction α and that only a fraction γ of the zero observations (taken at random) is retained. The probability that element i has $Y_i = 1$ *and* is included in the sample is

$$\alpha P_i^*,$$

but for $Y_i = 0$ it is

$$\gamma\alpha(1 - P_i^*).$$

By Bayes rule, the probability that an element of the selected sample has $Y_i = 1$ is then

$$\tilde{P}_i = \frac{P_i^*}{P_i^* + \gamma(1 - P_i^*)}. \tag{2}$$

The likelihood of the observed sample is easily expressed in $\tilde{P}_i$. Since γ is a known constant, substitution of (1) into (2) gives $\tilde{P}_i$ as a function of θ alone. The parameters of any specification (1) of P_i^* can thus be estimated by standard Maximum Likelihood methods.

Estimation is particularly easy in the special case that (1) is a logit model, for then

$$P_i^* = \frac{exp(x_i^T \beta)}{1 + exp(x_i^T \beta)} \tag{3}$$

so that (2) becomes

$$\tilde{P}_i = \frac{exp(x_i^T \beta)}{exp(x_i^T \beta) + \gamma} \tag{4}$$

or

$$\tilde{P}_i = \frac{1/\gamma . exp(x_i^T \beta)}{1 + 1/\gamma . exp(x_i^T \beta)} = \frac{exp(x_i^T \beta - ln\gamma)}{1 + exp(x_i^T \beta - ln\gamma)}. \tag{5}$$

Thus the $\tilde{P}_i$ of the selected sample also follow the logit model, and apart from the intercept the same parameters β apply as in the logit model for the full sample. Standard programmes for fitting a logit model may therefore be applied directly to the selected sample, and they will give proper estimates of the slope coefficients of the model (3) for the full sample. If needs be, its intercept can be retrieved by adding $ln\gamma$ to the intercept of the selected sample. But if the analyst is primarily interested in odds ratios or in the ordering of observations by the estimated probabilities, the value of the intercept is of no concern. Since the slope coefficients can be estimated if γ is unknown, the sample may then even be constructed by drawing observations on the two outcomes from separate sources with unknown sampling fractions. This is common practice in the *case-control* studies of medical research.

Conditions that call for state-dependent sample selection occur in many fields. A common example is that there is a very large sample with a low incidence of $Y_i = 1$, and that it is expedient to discard a major part of the abundant zero observations. In the case under consideration the bank has files on over twenty thousand bank loans with only six hundred bad loans. Similarly, Palepu (1986) examines 163 firms that have been the subject of of a takeover bid among a total of 2217 eligible firms firms listed on the stock exchange. In direct marketing, mail campaigns yield only a small number of responses from a vast data base of potential customers. In all these cases it can be advantageous to use all observations with $Y_i = 1$ in combination with a fraction of the surfeit of zero observations. In epidemiology, a given number of *cases* of a particular disease (or treatment) is likewise supplemented by *controls* that can be recruited in any number; but here there is no natural framework of a random sample to begin with, and the sampling fractions and their ratio γ are unknown. The same theoretical argument, briefly set out above, applies to all these cases. There is a vast literature on the subject. In statistics, Anderson wrote already in 1972 about combining separate samples; in econometrics, Manski and Lerman introduced the term *state-dependent* or *choice-dependent* sampling in 1977; in epidemiology, the method is known as *case-control studies,* see the surveys by Breslow and Day (1980) and Breslow (1996).

Until recently, a major motive for using only part of the available information was computational expediency, but with present computing power this argument has lost its strength. But if the zero observations still have to be collected, or if additional information for these data must be obtained, cost considerations may dictate restraint. In practice, there is a strong tradition of using equal numbers of observations of the two outcomes, probably for reasons of symmetry. One may of course improve precision by increasing the number of zero observations within reason; as a rule of the thumb there is however little additional precision to be gained from going beyond 3 or 4 times as many controls as cases (Breslow and Day (1980), p.27; Cramer et al (1999)).

3 Logit analysis of bank loans

The original full sample of bank loans consists of 20816 loans from a single year. Two years later, 627 loans are identified as bad loans and 20189 as good loans. The bank performs a routine logit analysis on a selected sample of 1254 loans, with the bad loans supplemented by an equal number of good loans drawn at random from that group. The regressors are five financial ratios of the debtor firm recorded two years ago, suitably scaled, viz.

. *solvency;*
. *rentability;*
. *working capital ratio;*
. *cash flow coverage;*
. *stocks* (a dummy variable)

The initial aim of the present study was to assess the effect on the estimates' precision of adding larger numbers of zero observations or good loans to the 627 bad loans. To do this the 627 bad loans were supplemented by independent samples of increasing size from the good loans, their numbers increasing with multiples K of 627 observations. They thus range from the standard case of 627 (or $K = 1$) to 20819 (all, or $K = 33$). Plain logit analyses were then performed on this series of samples.

The successive selective samples are not independent, for they share the same bad loans. As the sample size increases, the estimates of the slope coefficients will of course converge to the values for the full sample. By the arguments of Section 2 they may be expected to show some erratic fluctuations, but no strong systematic variation. But Table 1, which gives the results, shows that the estimates do not behave in this way: they all exhibit a steady and systematic movement with the factor K that governs sample size, and the coefficients for the standard case $K = 1$ are quite far removed from the final result. In the present instance the state-dependent sample selection technique introduces a severe bias, and this is of more urgent concern than the course of the standard errors.

Table 1. Estimated logit coefficients, selected samples of various sizes
(standard errors in brackets)

K	1	2	3	5	10	20	all
sample size	1254	1881	2508	3762	6897	14617	20816
solvency	-2.32 (.32)	-1.22 (.29)	-1.19 (.25)	-.85 (.19)	-.73 (.15)	-.60 (.12)	-.49 (.09)
rentability	-1.30 (.37)	-.72 (.31)	-.89 (.28)	-.57 (.26)	-.48 (.18)	-.60 (.13)	-.43 (.12)
working capital	-1.38 (.30)	-1.26 (.23)	-1.05 (.23)	-1.30 (.19)	-1.10 (.17)	-.96 (.14)	-.90 (.12)
cash flow cov	-1.72 (1.12)	-2.82 (1.09)	-3.45 (1.29)	-3.69 (1.14)	-3.67 (.82)	-2.45 (.49)	-2.60 (.40)
stocks	.14 (.21)	.25 (.19)	.34 (.17)	.21 (.17)	.27 (.15)	.35 (.15)	.33 (.15)

Since the results for a single standard sample with $K = 1$ may be attributed to chance (though the systematic variation of Table 1 can not), this particular sample has been replicated 100 times, and the size of the bias established from the mean estimate. Table 2 shows that the single example in Table 1 was indeed rather unfortunate; but still the best that can be said about the two sets of estimates from the replicated small sample and the overall sample is that they have the same sign - and these signs agree with what one would expect from the nature of the variables. For four out of five coefficients the estimates from the standard selected sample differ from the final values by a factor 2 or 3.

Table 2. Mean estimates of 100 replications for $K = 1$ [1]

	100 selected samples, $K = 1$	full sample	ratio of estimates
solvemcy	-1.65, .34 (.34)	-.49 (.09)	3.4
rentability	-.93, .25 (.39)	-.43 (.12)	2.2
working capital	-1.41, .22 (.28)	-.90 (.12)	1.6
cash flow cov	-2.96, .86 (1.52)	-2.60 (.40)	1.1
stocks	.13, .07 (.20)	.33 (.15)	.4

The only substantial assumption of section 2 is that the observations in the population or in the full sample satisfy the logit model. The failure of the method can therefore only be due to the failure of this assumption. This can be verified by the goodness-of-fit test of Hosmer & Lemeshow (1980, 1989). The observations are ordered into J classes of equal size by the estimated probability $\hat{P}_i$, and for each class the expected frequency of bad loans is determined (this is simply the sum of the estimated probabilities). This is compared to the actual frequency. The test statistic is

$$C = \sum \frac{(y_j - n_j \bar{P}_j)^2}{n_j \bar{P}_j (1 - \bar{P}_j)} \tag{6}$$

with n_j the number of observations in class j, y_j the actual number of bad loans in the class and $\bar{P}_j$ the mean estimated probability for the class. The null hypothesis is that the observations have been generated by a logit model as used in estimating the $\hat{P}_i$. Under this null, C has a chi-square distribution with $J - 2$ degrees of freedom.

[1] The numbers in brackets are standard errors derived from the mean variance of 100 replications, the numbers after the means are standard errors among the replications.

Table 3 shows expected and actual numbers of bad loans in 10 classes of the full sample of 20816 observations. The test statistic C of (6) is 164.35, which is highly significant at 8 degrees of freedom: the logit model is soundly rejected. Inspection indicates severe discrepancies, with overestimation by the model in the lower ranges and underestimation in the two highest classes.

Table 3. Expected and actual number of bad loans in full sample, ten classes of expected probability according to fitted logit

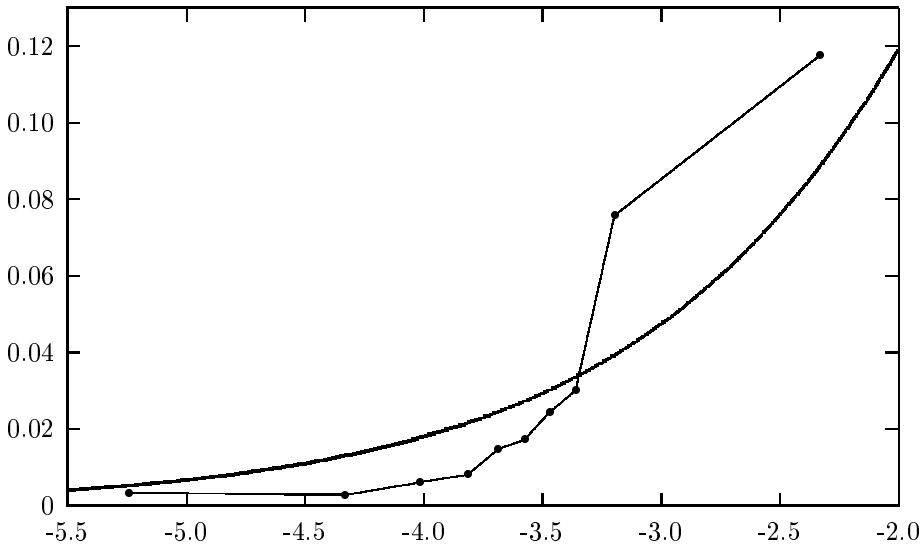
highest $\hat{P}_i$ in class	nr of observations	expected nr of bad loans	actual nr of bad loans
.0100	2082	11	7
.0158	2081	27	6
.0198	2082	37	13
.0229	2081	45	17
.0258	2082	51	31
.0286	2082	57	36
.0317	2081	63	51
.0359	2082	70	63
.0443	2081	82	158
.9148	2082	185	245

This is further illustrated in Figure 1. The expected and actual numbers of bad loans in each class of Table 3 have been converted into frequencies, and these have been plotted against the argument of the logit function, the log-odds ratio of the expected probability, or

$$z_j = log(\bar{P}_j/(1-\bar{P}_j)).$$

This is the term $x_i^T\hat{\beta}$ of (3) that drives the fitted logit function. The expected frequencies therefore follow this logit function by construction; since the overall frequency of the present sample is only 3%, the smooth line represents the far left-hand tail of the logit. In contrast, the actual frequencies appear to trace a far larger and much more central part of the same sigmoid shape, if on a miniature scale.

Figure 1. Expected and actual frequency of bad loans from Table 3



The above test bears on the full sample, that is the logit estimated in the last column of Table 1, but the same goodness-of-fit test can be applied to a selective sample; by the argument of Section 2 this should also obey the logit model if the random sample does. Table 4 gives the result for the standard selective sample of 1254 observations of the first column of Table 1. The expected probabilities are of course much higher than in the full sample, precisely because the number of observations has been reduced selectively. The total number of bad loans (both actual and expected) is still 627, but the sample proportion is 50% as against 3% in the full sample. The discrepancies in Table 4 are less pronounced than in Table 3, and the value of the test statistic C is only 65.62, but this is still quite significant. The logit model is also rejected on the basis of the sample constructed by state-dependent selection.

Table 4. Expected and actual number of bad loans in selected sample, $K = 1$, ten classes of expected probability according to fitted logit

highest $\hat{P}_i$ in class	nr of observations	expected nr of bad loans	actual nr of bad loans
.1617	125	12	9
.2731	126	28	17
.3496	125	39	30
.4228	126	49	43
.4938	125	57	50
.5545	125	65	72
.6296	126	75	94
.7280	125	85	102
.8618	126	100	106
.9999	126	118	164

4 A bounded logit function

The failure of the logit function is a matter of the functional form of the probability function (1). In the search for something better we may take a second look at the pattern of the incidence of bad loans of Figure 1, even though it must be realized that the classification of the observations has been determined by the plain logit probability which turned out to be a misspecification. As we remarked before the actual frequencies appear to trace out a large part of a logit curve, if on a miniature scale. This suggests that they will be adequately described by a simple modification of the model, namely

$$P_i^o = \omega \frac{exp(x_i^T \beta)}{1 + exp(x_i^T \beta)}. \tag{7}$$

In this model, the logit function is bounded by an upper limit ω of the probability that a loan is a bad loan. Conversely, a fraction $1 - \omega$ of all loans is always a good loan, regardless of the values taken by the regressor variables. The logit mechanism applies only to a fraction ω of the population.

This model can be estimated by standard Maximum Likelihood methods, though the routine is not generally available in statistical packages (in contrast to the plain logit) and the analyst will have to do some programming. It has been fitted to the full sample, and the estimates are given in Table 5. The slope coefficients are of course not directly comparable to those of the plain logit, but the elasticities at the sample mean are; apart from the rather erratic coefficient of cash flow coverage they show no great differences. A likelihood ratio test of the bounded logit vis-a-vis the plain logit confirms that the additional parameter ω makes a difference: the loglikelihood of the bounded model is -2425 against -2594 for the common logit model of Table 1. The single extra parameter leads to a quite significant increase in likelihood.

Table 5. Estimates of bounded logit model, full sample

	coefficient bounded logit	elasticity bounded logit	elasticity plain logit
ceiling ω	.14 (.01)		
solvemcy	-4.20 (.41)	-.17	-.14
rentability	-1.80 (.50)	-.05	-.09
working capital	-2.97 (.34)	-.03	-.06
cash flow cov	-3.27 (2.11)	-.05	-.26
stocks	.27 (.23)		

Table 6 gives Hosmer and Lemeshow's goodness-of-fit test for the bounded logit. For the computation of the test statistic C the first three classes are merged, for otherwise the numbers would be too small; C is then 5.38, and this has six degrees of freedom. The bounded logit model is not rejected.

Table 6. Expected and actual number of bad loans in full sample, ten classes of expected probability according to fitted bounded logit

highest $\hat{P}_i$ in class	nr of observations	expected nr of bad loans loans	actual nr of bad loans loans
.0009	2082	1	5
.0030	2081	4	4
.0060	2082	9	7
.0100	2081	17	20
.0155	2082	26	21
.0233	2082	40	37
.0337	2081	59	51
.0510	2082	86	80
.0870	2081	135	152
.1374	2082	247	250

The bounded logit model can also be estimated (and tested) from state-dependent selected samples, provided the constant γ in (3) is known; these calculations demand some further programming by the analyst. In the present case it was found that the estimates for a series of samples of varying size, as in Table 1, do not show the systematic variation that was observed for the common logit. The model also passes the goodness-of-fit test for the selected sample for $K = 1$.

In short, the bounded logit passes the tests that the common logit failed. In the present case it is a superior descriptive device. But it is not so easy to think of a theoretical or intuitive justification. The present model is a special case of the general extension of a discrete bivariate model like the logit or probit with fixed probabilities for either or both outcomes. These are sometimes introduced to allow for specific errors of measurement or recording, sometimes because it is believed that other factors are at work besides the stochastic mechanism under scrutiny. In the classic bio-assay studies the probit model describes the effect of an insecticide; an extra constant probability term may be added to represent natural death, not induced by the poison. This example of Finney (1964) is quoted by Hausman et al (1998), who introduce a similar additive probability to take care of errors of observation.

And there may of course be two such terms, one for each outcome; see the application to marketing data of Hsiao & Sun (1999). But these arguments, usually adduced for quite small constant additive probabilities, do not apply in the present case.

In the present case we must have recourse to the argument that the standard screening of applicants by the bank ensures that no debtor ever has a probability of more than ω (here only 14% !) to drift into the bad loan category. It can, however, be objected that the effect of this screening is already apparent in the favourable values of the regressor variables, as reflected by the positioning of the entire sample at the lower tail of the logit function (see Figure 1). The screening argument must mean that the probability is confined to values below .18 because of *other* characteristics of the debtor, like business talents or innate honesty. It is well known that bank managers liberally employ criteria of this nature rather than accounting ratios. Still, I find this justification of an upper bound of only .14 hard to accept.

5 Concluding remarks

There are two lessons to be drawn from this case study. The first is that some simple binary attributes do not follow the logit model distribution. The goodness-of-fit test of Hosmer and Lemeshow will detect this, and this test should be employed as a matter of routine. But while the bad loans of a Dutch bank do not obey the logit model, other binary outcomes still do; in an earlier paper (Cramer (1999)) I examined two marketing data sets along with the bank loans, trying out several other flexible variants of the logit model as well, and bank loans were the only case where the plain logit failed.

The second lesson is that the selected sample method for the logit model gives quite wildly erroneous results in case of misspecification. This is already known from much earlier studies by Scott and Wild (1986) and by Xie and Manski (1989). It does not mean that the principles of the selected sample method do not apply, but that one should not use the logit short-cut if the logit model does not apply. Xie and Manski show that the use of the logit as a general approximation to *any* bivariate model is much better served by the weighted Maximum Likelihood technique; moreover any model and in particular any modification of the logit model can be estimated from a selected sample by the route sketched in Section 2.

In theory the practical conclusions are clear: test the fit of the logit model

(if necessary on a selected sample); if it is rejected try a modified model (if necessary on a selected sample), or, failing this, estimate the logit by weighted Maximum Likelihood. But these straightforward prescriptions cannot be implemented in the standard statistical packages like SPSS, STATA or LIMDEP. They either demand programming from scratch in languages like Gauss or Ox or at least the specification of the loglikelihood (and its derivatives) for a ready-made Maximum Likelihood routine. And this deters most analysts from probing further.

6 References

. Anderson, J.A. (1972). Separate sample logistic discrimination. *Biometrika*, **59**, 19-35.

Breslow, N.E., and N.E. Day (1980). *Statistical Methods in Cancer Research.* IARC, Lyon.

Breslow, N.E. (1996). Statistics in Epidemiology: The Case-Control Study. *Journal of the American Statistical Association*, **91**, 14-28.

Cramer, J.S. (1999). Twisting the logit curve around. Paper for the European Meeting of the Econometric Society, Santiago de Compostella.

Cramer, mars, Philip-Hams Franses, and Erica Slagter (1999). Censored regression analysis in large samples with many zero observations. Econometric Institute Report EI-9939/A, Erasmus University Rotterdam.

Finney, D.J. (1964). *Statistical Method in Biological Assay.* Havner, New York.

Hausman, J.A., Jason Abrevaya, and F.M. Scott-Morton (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, **87**, 239-269.

Hosmer, D.W., and S. Lemeshow (1980). Goodness of Fit Tests for the Multiple Logistic Regression Model. *Communications in Statistical-Theoretical Methods*, **9**, 1043-1069.

Hosmer, D.W., and S. Lemeshow (1989). *Applied Logistic Regression.* New York: Wiley.

Hsiao, Cheng, and Bao-Hong Sun (1999). Modelling survey response bias. *Journal of Econometrics*, **89**, 15-39.

Manski, C.F., and S.R. Lerman (1977). The Estimation of Choice Probabilities from Choice-based Samples. *Econometrica* **45**, 1977-1988.

Palepu, K.G. (1986). Predicting Takeover Targets. *Journal of Accounting and Economics* **8**, 3-25.

Scott, A.J., and C.J. Wild (1986). Fitting Logistic Models under Case-Control or Choice-based Samples. *Journal of the Royal Statistical Society, series B*, **48**, 170-182.

Xie, Y., and C.F. Manski (1989). The logit model and response based samples. *Sociological Methods and Research*, **17**, 283-302.