COUNTRY RISK:

MULTIVARIATE MODELS AND HUMAN JUDGEMENT

RICHARD ANDREW SOMERVILLE

Thesis submitted for the degree of Ph.D.

The City University

The City University Business School

June 1991

Volume 2

SUMMARY CONTENTS

VIII Early warning models: logit analysis

VIII.1 Introduction

A maintained hypothesis of linear discriminant analysis is that the data have a multivariate normal distribution. In contrast, the procedures of this chapter that do not rely on normality. In section VIII.2, an early warning logit model is specified, estimated, and tested. In section VIII.3 this model is compared with the discriminant models of chapter VII. Finally, section VIII.4 presents conclusions.

Logit and discriminant analyses are closely related. Let y be the (0,1) dependent variable, and x be a vector of predictors. Using discriminant analysis, we assume that the conditional distribution of x given y is multivariate normal, with common covariance matrix for both values of y. Under these assumptions, the conditional distribution of y given x is logistic: thus,

$$P(y=1|x) = \exp(b+cx)/(1+\exp(b+cx))$$

where b is a scalar and c is a vector of coefficients. However, the converse does not hold in general. Thus, logit analysis, which is applicable when the cdf of (y|x) is logistic, is more generally applicable than the discriminant model.

If the cdf of (x!y) is multivariate normal, then the discriminant analysis estimator is more efficient than the logit estimator, and is in fact asymptotically efficient. When normality does not hold, then the discriminant analysis estimator is inconsistent, while the logit estimator retains consistency in a wide class of distributional settings (Lo, 1986).

In section VIII.2, a logit model is fitted using the BMDP procedure LR and the SPSS-X procedure PROBIT, and in section VIII.3 the fitted model is compared with the discriminant models of chapter VII.

When a logit model is estimated on panel data via a standard maximum-likelihood procedure, difficulties arise that have been addressed by Maddala (1987). This thesis uses panel data with a cross-section of 55 countries, and a time-series of 8 years. Possibly the intercept of the logit function, while non-stochastic, has different values for different members of the cross-section. This is the case of 'fixed effects', which according to Maddala will yield inconsistent maximum-likelihood estimates of the logit parameters when the time-series is short and the cross section is large. Maddala argues that a conditional maximum likelihood approach (conditioning on the fixed effects) will produce consistent estimates in this case. However, a feature of this procedure is that cross-sectional cases are deleted from the estimating sample if they never change states. In the present context, this would involve dropping countries that are either always arrears or always non-arrears, and would require the discarding of too much data to be acceptable. The primary concern here is with classification, to be tested on a holdout sample, rather than with the coefficients. Consequently, and given that the conditional maximum-likelihood procedure is not usable, the standard maximum-likelihood method will be used to estimate the logit coefficients.

There remains the possibility that 'random effects' may be present, whereby the intercept of the logit function is a random variable. In this case, Maddala argues that the correct approach is to assume that the regression residuals have a multivariate logistic distribution. Moreover, because of certain inflexible features of that distribution, Maddala argues for the use of the probit model in this case.

The development of a random effects probit model would be a detour away from the primary concerns of this thesis, and will not be pursued. Moreover if a univariate probit model is estimated from the pooled data, the estimated coefficients will be consistent, although not efficient.

For the probit model, it is assumed that the conditional distribution of (y¦x) is normal. The normal and logistic distributions only differ markedly in the tails. Consequently, similar results may be expected from the two procedures, particularly in view of the Winsorization of the data. The SPSSX procedure PROBIT estimates probit and logit models at the same time, and after the stepwise logit analysis had been carried out (see section VIII.2), PROBIT was used to estimate both probit and logit models, using the specification indicated by the stepwise logit run. As expected the two approaches yielded very similar results. The estimated coefficients were numerically very similar, and the coefficient of correlation between the estimated probabilities was very high: 0.997. In the middle of the listing, the region covered by Table VIII.2, there was little difference. For example, taking a cutoff probability of 0.52, as at point (a) in Table VIII.2, the number and identities of both types of error were the same.

Since probit analysis added nothing to the results of the logit model, it is not explored further.

VIII.2 The logit model

VIII.2.1 Estimation

The dependent variable in these analyses is binary, and has the values unity or zero according as a case is or is not 'arrears' in year t+1. Using the primary data set of 70 variables along with the categorical variables, the BMDP stepwise logit programme LR was applied interactively to the set of 302 cases (i.e. with weak-year cases excluded). The maximum likelihood option was chosen as being a superior technique to an alternative procedure based on the asymptotic covariance estimator.

In order of entry, the first four variables were:

NARY (Net assets/GDP)
DCPI (Consumer price inflation)
DGDP (GDP growth rate)
INPS (Interest payments/debt service)

These were also the first four variables selected by the first discriminant analysis of chapter VII, with the order of entry of DGDP and INPS reversed. None of the categorical variables was selected; when the dummy AS (Asia) that appeared in discriminant model II was forced into the equation, neither the Wald statistic nor the improvement chi-square was significant at the five percent level.

The estimated logit model is set out in Table VIII.1. Because the data are essentially the population of values for the 55 countries used in the study for 1979-1986, rather than a sample, the two groups are being sampled at the same rate (i.e. 100 per cent). When disproportionate sampling is used, a correction must be applied to the constant term (Maddala, 1988: pp. 275-276). Clearly this adjustment is not required here.

At each step, the programme yields a chi-squared test of the null hypothesis that the entering variable makes no improvement to the estimated logit function. At each stage up to the fourth, the value of this statistic, which is based on the likelihood ratio, exceeds the tabulated chi-squared value with 1 degree of freedom at the 5 per cent level. Thus, at each of the first four stages the null is rejected, while not so at the fifth stage and beyond.

Under the null hypothesis for any of the regression coefficients, the asymptotic distribution of the Wald statistic is standard normal. Clearly, the null would be rejected in each case. However, Hauck and Donner (1977) have examined the performance of the Wald test and found it to be unsatisfactory in many settings, often failing to reject a false null. On this basis, they prefer the likelihood ratio test.

Table VIII.1: Logit model

Prob(arrears) = exp(u)/(1+exp(u))

Estimated logit function		Chi-sq.(1) to remove	Coeff/SE (Wald stat.)
u = -9.011300		102.6	-6.773
-0.081388	NARY (net assets/GDP)	122.3	-6.852
+0.082195	DCPI (inflation)	33.4	3.754
-0.227070	DGDP (GDP growth)	19.6	-4.403
+0.067161	INPS (int./debt serv.)	13.6	3.404

Improvement chi-squared (1): 13.6
Goodness-of-fit: Hosmer-Lemeshow chi-sq.(8): 14.3

The Hosmer-Lemeshow test of goodness-of-fit is based on a grouping of the data determined by the estimated probabilities from the fitted model. The cutpoints are the 'deciles of risk'. The null hypothesis is that there is no significant difference between observed and expected frequencies (of arrears cases, in the present application) - i.e., that the model fits the data. The test statistic is computed from the 2x10 table of observed and expected frequencies. Given a tabulated chi-square(0.05,8) of 15.507, the null cannot be rejected in this case at the 5 per cent level.

For a discussion of the summary statistics displayed in Table VIII.1, see Hosmer and Lemeshow (1989).

The signs of the estimated coefficients are all as would be expected, a priori: the probability of arrears is relative negatively to the net assets ratio and GDP growth, and positively to inflation and the interest burden.

VIII.2.2 Classification within-sample

The within-sample classificatory power of the model is examined by sorting the sample on the predicted value of the dependent variable [i.e. P(arrears)]. The holdout method is utilised: the estimated probability associated with each case is computed from a logit function estimated from the 301-member sample that remains after deleting that case. The central section of the output from this procedure is set out in Table VIII.2.

Lachenbruch and Mickey (1968) introduced the U-method (i.e. 'hold-one-out') in the context of linear discriminant analysis. However, the method has a general applicability: see for example Efron (1982; 1983) and Gong (1986).

Table VIII.2: Logit probabilities
(within sample, holdout method)

Order	Case (t)	ARS (t+1)	Ordered probabilities	
Cases 1 to 76 are all arrears, except for				
69	KS 80	9	0.8891	
77	SE 82	0	0.6967	
78	SU 80	0	0.6790	
79	TN 86	9	0.6657	
80	MW 86	0	0.6380	
81	ZB 81	0	0.6271	
82	ES 81	9	0.5999	
83	SE 86	0	0.5783	
84	EG 80	9	0.5554	
85	SU 82	0	0.5440	
86	SE 81	0	0.5210..	(a) Type I : 21
87	TN 82	9	0.5137	Type II: 4
88	SR 81	9	0.5111	
89	KN 82	9	0.5089	
90	SR 83	9	0.5089	
91	ES 80	9	0.4998	
92	PH 82	0	0.4918	
93	BL 79	0	0.4781	
94	PP 84	9	0.4612	
95	TK 81	0	0.4393	<u>Cutoffs</u>:
96	ES 82	9	0.4258	(a) minimum number of
97	ZI 84	9	0.4155	misclassifications
98	EG 79	9	0.4128	
99	MC 82	0	0.4025	(b) minimum expected
100	CG 81	9	0.3911	misclassification cost
101	EC 82	0	0.3834	
102	PP 85	9	0.3794	
103	SR 82	9	0.3727	
104	GU 86	0	0.3622	
105	HD 80	0	0.3611	
106	JD 85	9	0.3596	
107	TK 86	9	0.3508	
108	DR 84	0	0.3465	
109	UR 82	0	0.3361	
110	KN 85	9	0.3217	
111	YU 82	0	0.3147..	(b) Type I: 11
112	ES 86	9	0.3068	Type II: 19
113	TN 85	9	0.3063	
The remaining 189 cases include 11 arrears cases.				

<u>Note</u>: ARS = 0: arrears; ARS = 9: non-arrears

The problem of choosing a cutoff arises here. As Saini and Bates (1984: p.352) have pointed out

> the results of logit analysis are still difficult to interpret because of the lack of any explicit procedure for selecting the critical probability value.

A natural approach might seem to be to use the prior probability of being an 'arrears' case, and take it that the model classifies as 'arrears' those cases whose computed probability exceeds this value. This would be a mistaken procedure: for a given estimated logit function, the number classified as 'arrears' would be <u>negatively</u> related to the prior probability of being in that group.

In fact, the estimated logit function embodies prior probabilities, in two senses. First, let us write the logit model as

$$P(y=1|x) = \exp(b + cx)/[1 + \exp(b + cx)]$$

where y is the categorical variable, x is a vector of independent variables, c is a vector of coefficients, and b is the constant term. It is clearly the case that:

$$P(y=1|x=0) = \exp(b)/[1 + \exp(b)]$$

so we may regard b as reflecting the probability that y = 1, before taking account of the impact of any of the independent variables.

Secondly, suppose that the predicted probability associated with each sample datapoint were to be computed, by substituting each sample value of x into the estimated logit equation. It is a property of maximum-likelihood estimation that the sum of these predicted probabilities equals the sum over the sample of the criterion variable y (Maddala, 1983: p.26). Interpreting the sum of the probabilities as the expected number of 'successes' (e.g. arrears cases), the implication is that with maximum-likelihood estimation, the expected and observed numbers of successes will always be equal. In this sense, the prior probabilities, as embodied in the sample proportions, will always be reflected in the estimated logit equation.

An alternative procedure is to focus on the <u>ex post</u> classificatory performance of the model. The rationale for this is the same as in chapter VII: if the prior probability of 'arrears' equals the sample proportion, then minimising the <u>ex post</u> and expected costs of misclassification are equivalent.

A cutoff at point (a) in Table VIII.2 minimises the total number of misclassifications, with a cutoff probability in the interval (0.5137, 0.5210). There are 21 type I errors (19.6 per cent) and 4 of type II (2.0 per cent). However, given the differential costs to lenders of each type of error, this equal-weighting scheme is not satisfactory.

An experiment was carried out in which the cost of a Type II error was standardised at 1.0, while the cost of a Type I was assigned successive values in the interval [1.0, 6.0], at increments of 0.1. For each value in this interval, the location of the minimum-cost cutoff was determined, over the field of all possible cutoffs. For values between 1.0 and 1.5, (a) was cost minimising; for all values in the interval [1.6, 6.0] the cost minimising cutoff was at point (b). Compared with (a), this generated 5 additional misclassifications net, but it included only 11 of Type I (compared with 21). Given the stability of point (b) as a cost-minimising choice over a very wide range of misclassification costings (including the value that was used in the discriminant analysis of chapter VII), its associated probability is chosen to be used as cutoff when classifying by the logit model. The precise value assigned to this probability is the midpoint of the interval, 31.075 per cent.

The misclassifications yielded by this cutoff are set out in Table VIII.3. The statistic for the FMM test (see chapter VII) is evaluated at 12.3, so the null hypothesis that the model has no classificatory power is rejected at the 5 per cent level.

Table VIII.3: Logit model: holdout classifications

Actual group:	Total	Predicted group membership 1980-1987: Arrears	Non-arrears
Arrears	102	91 (89.2%)	11 (10.8% type I errors)
Non arrears	200	19 (9.5% type II errors)	181 (90.5%)

Misclassifications

Type II errors		
43	South Korea	1981
33	Tunisia	1987
30	El Salvador	1982
28	Egypt	1981
25	Tunisia	1983
24	Sri Lanka	1982
23	Kenya	1983
22	Sri Lanka	1984
21	El Salvador	1981
18	Papua N.G.	1985
16	El Salvador	1983
15	Zimbabwe	1985
14	Egypt	1980
12	Congo	1982
10	Papua N.	1986
9	Sri Lanka	1983
6	Jordan	1986
5	Turkey	1987
2	Kenya	1986
Total: 19,		9.5%

Type I errors		
134	Nigeria	1986
118	Pakistan	1981
113	Dominican Rep.	1984
90	Dominican Rep.	1982
71	Nigeria	1983
55	Liberia	1980
38	Gabon	1986
30	Venezuela	1983
29	Mexico	1982
28	Senegal	1981
22	Venezuela	1987
Total: 11,		10.8%

The number beside each case is its order from the cutoff: the lower the number, the closer the case is to its correct group.

<u>Notes</u>: Probabilities at date t (Table VIII.2) are used to classify at date t+1.

The cutoff is set at (b) in Table VIII.2.

The type II errors were examined for evidence that they are really cases of advance early-warning. For the discriminant model I, only one case (out of five) could possibly be construed in this way. Here, the proportion is the same, at four in twenty (counting Egypt 1980 and 1981 as one case). Egypt, Congo and Guatemala became arrears cases in 1986, and Jordan did so in 1989. Thus there is little evidence for a re-interpretation of the type II errors.

Support for the final choice of cutpoint is provided by the receiver operating characteristic (ROC) curve for the logit function, which is displayed in Figure VIII.1. (Spiegelhalter and Knill-Jones, 1984). Given 302 cases in the sample, there are 303 possible cutpoints: before the case with the lowest estimated logit, after the case with the highest, and between each of the 301 adjacent pairs in the ordered listing of estimated logits. The ROC curve plots, for each of these possibilities:

$$Y: \frac{\text{no. of arrears cases correctly predicted}}{\text{total arrears cases}} \times 100\%$$

versus

$$X: \frac{\text{no. of type II misclassifications}}{\text{total non-arrears cases}} \times 100\%$$

The ROC curve rises almost vertically at first, and ultimately is almost horizontal. If there is a unique and well-defined sharp knuckle, then the cutpoint associated with that point represents a good tradeoff between the type I and II error rates. Generally, one seeks a knuckle on the ROC curve that is furthest north-west towards the point (0,100) on

---

Figure VIII.1: ROC curve for estimated logit function (holdout within-sample classifications)

```
+----+----+----+----+----+----+----+----+----+----+----+--
¦
+100                                 24445444444444444443
¦                              4444453
¦    Y                 1444544441
¦                454444
¦                2
+90       *5444541
¦         3
¦        52
¦       51    *: indicates cutpoint (b),  Table VIII.2
¦      21
+80   $5      $: indicates cutpoint (a),  Table VIII.2
¦     5
¦    12
¦    2    Vertical axis Y:
¦    2         (True positives/total arrears cases)x100%
+70  2
¦    2    [True positives = total arrears cases less
¦    3                      type I errors.]
¦    2
¦    2    Horizontal axis X:
+60  2    (False positives/total non-arrears cases)x100%
¦    2
¦    2    [False positives = type II errors.]
¦    2
¦    3               ----------------------
+50  2
¦    2    The ROC  curve  plots Y against X for various
¦    2    values of the  threshold  probability  of the
¦    2    estimated logit function,  above which a case
¦    2    is classified as 'arrears'.  This probability
+40  2    ranges between:
¦    2
¦    2    0:  at which all cases are classified as arr-
¦    2        ears,  there  are  no type I errors,  and
¦    2        all non-arrears cases are type II errors.
+30  2        Thus X = Y = 100% (north-east extremity).
¦    2
¦    2    1:  at which all cases are classified as non-
¦    2        arrears,  there  are  no type II  errors,
¦    2        and  all arrears cases are type I errors.
+20  2        Thus X = Y = 0%  (south-west extremity).
¦    2
¦    2
¦    2
¦    2        Numerals indicate frequencies
+10  2
¦    2
¦    2
¦    2
¦    3
+0   1                                                 X
+----+----+----+----+----+----+----+----+----+----+----+--
     0         20        40        60        80        100
```

the vertical axis. It may be seen that the point labelled *, which corresponds to cutpoint (b) in Table VIII.2, fulfills this condition.

VIII.2.3 Out-of-sample tests

For the out-of-sample tests, to be reported in chapter XI, unbiased coefficients are required. These are computed in the same manner as the jackknifed discriminant coefficients of chapter VII, and are summarised in Table VIII.4. (Miller, 1974; Taffler, 1984a.)

Table VIII.4: Summary of logit coefficients

	Raw coeff- icients	Average Pseudo values	Jackknifed coeff- icients
Constant	-9.011300	-9.011366	-8.991424
NARY	-0.081388	-0.083188	-0.083284
DCPI	0.082195	0.082188	0.084257
DGDP	-0.227070	0.227058	-0.230800
INPS	0.067161	0.067165	0.066088

VIII.2.4 The problem of weak-year cases

The classifications of the weak year cases by the jackknifed version of the logit model are set out in Table VIII.5. A similar analysis for discriminant model I is fully described in chapter VII.4, and the conclusions reached there apply equally to the logit model: in view of the evidence presented in the table, it is clearly impossible to partition the range of logit probabilities into separate regions for arrears, non-arrears and weak-year cases.

Table VIII.5: Classification of weak-year cases by logit model: summary, and distribution

CLASSIFIED BY LOGIT	Decile	Actual groups: ARS=1 (numbers of cases)	Actual groups: ARS=2,3,4 (numbers of cases)
ARREARS	1st	19	7
	2nd	11	0
	3rd	15	4
	4th	11	13
	5th	3	7
	Totals	59	31
	Cutoff		
NONARREARS	Totals	18	30
	6th	7	9
	7th	5	8
	8th	4	6
	9th	1	5
	10th	1	2

<u>**Note**</u>:
Deciles are those of holdout probabilities of 440-case sample.

VIII.3 Comparison of discriminant and logit models

The stepwise logit analysis generated a model that is very similar to discriminant model I. Four variables are common to both models, while INVR, fixed investment/GDP, appears in the discriminant model alone. It is the last variable to enter that model, and has a Mosteller-Wallace contribution of only 5.5 per cent.

Table VIII.6: Correlation analysis of discriminant scores and logit probabilities

Pearson product-moment correlation coefficients

	DSI	DSII	LP
DSI	1.00		
DSII	0.96	1.00	
LP	-0.91	-0.90	1.00

Spearman rank-order correlation coefficients

	DSI	DSII	LP
DSI	1.00		
DSII	0.94	1.00	
LP	-0.98	-0.94	1.00

Key: LP: logit probabilities
DSI, DSII: discriminant scores, models I and II

Note:
t-test result: All correlation coefficients are significant at the five per cent level.

A key finding is the similarity of the models in terms of their classification of the estimating sample: the logit model produces a ranking that is not very different from those of the two discriminant models. First, this is revealed by a correlation analysis of the discriminant scores and the logit probabilities, all based on the holdout method, which are set out in Tables VIII.6. It is clear that the classifications of the three models are very close, whether this is measured in terms of the values of the discriminant scores and logit probabilities, or of the rankings implied by them.

Secondly, precise evaluations can be derived from the misclassifications generated by each model. Here, comparisons depend on the differential costs of each type of misclassification. When they are equally weighted, then discriminant model I misclassifies a minimum of 22 cases, compared with 25 for the logit model, and 28 for discriminant model II. However, it has been shown that for a wide range of differential costs, including the value of 3.75:1 that was used in chapter VII, the cost-minimising cutoffs yield the misclassifications that are set out in Table VIII.7. These cutoffs are only least-cost for particular ranges of values of C, the unit cost of a type I error. These are:

Discriminant model I: C in interval [1.0, 4.2]
Discriminant model II: C in interval [2.0, 6.0]
Logit model : C in interval [1.6, 6.0]

Thus the comparisons that are drawn in Table VIII.7

are only applicable for C in the intersection of these three intervals, i.e. in the interval [2.0, 4.2]. However, this covers all plausible values.

The cost functions for each model are graphed in Figure VIII.2.

Table VIII.7: Misclassifications by logit and discriminant models

	Misclassifications Type I	Type II	Total
Discriminant model I	16 (15.7%)	5 (2.5%)	21 (7.0%)
Discriminant model II	11 (10.8%)	20 (10.0%)	31 (10.3%)
Logit model	11 (10.8%)	19 (9.5%)	30 (9.9%)

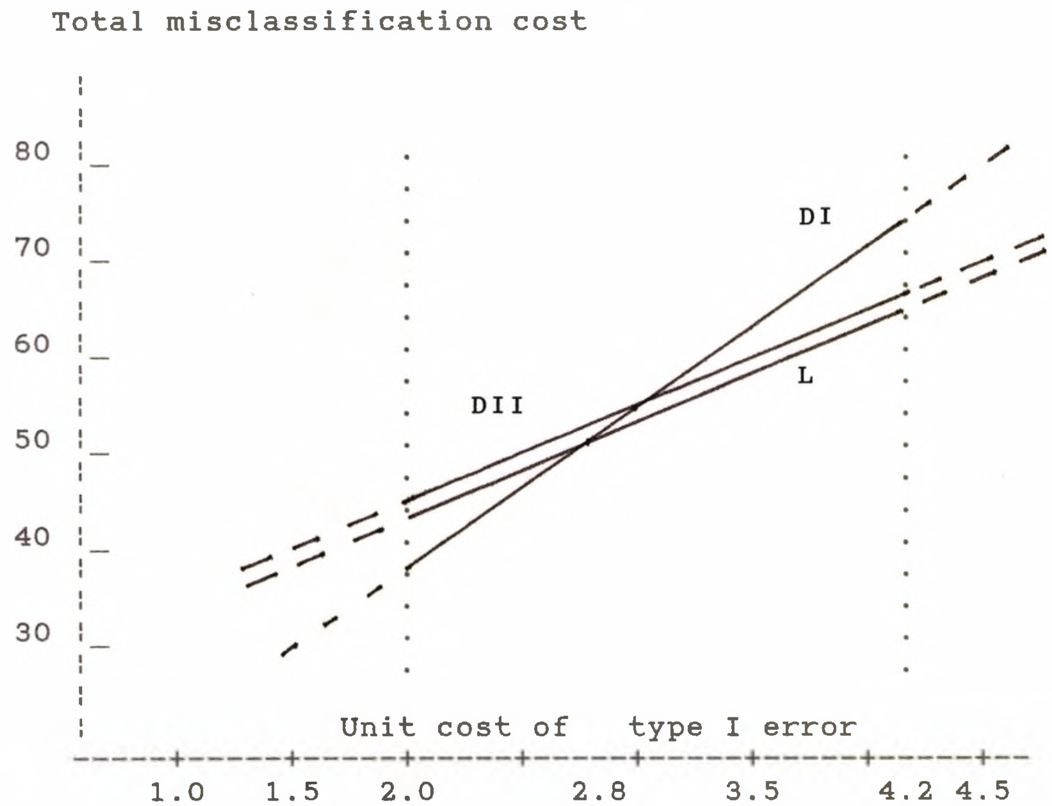
Sources: Tables VII.4, VII.7, VIII.3

Notes: Cutoffs used for this table are optimal for cost per type I error in [2.0, 4.2].

Holdout within-sample misclassifications of 102 arrears and 200 non-arrears cases.

It is clear that discriminant model II is always dominated by the logit model. For values of C up to 2.8, discriminant model I is dominant, while the logit model is optimal for larger values.

Figure VIII.2: Total cost of misclassification by logit and discriminant models, for unit cost of a type I error in the interval [2.0, 4.2]



Key:
DI and DII: discriminant models I and II; L logit model

Note: Holdout within-sample misclassifications

These results conform with the recent findings of Gessner et al (1988), who present a set of conditions, any of which may lead to lack of congruence between the results obtained by applying logit and discriminant analysis (among other techniques) to binary data. The conditions include:

departure from multivariate normality within groups, multicollinearity between the predictors, and the use of many predictors. In the research reported here, discriminant model I uses predictors that are multivariate normal, the largest number of predictors is five, and multicollinearity has been avoided.

Taking these classificatory results together, they compare favourably with the performance of earlier models. For example of the papers cited in Saini and Bates (1984), only the model of Feder and Just (1977) is clearly superior, with error rates of 5 and 2.5 per cent, and in any case this model relates to a period that contained very few arrears cases. Apart from Taffler and Abassi (1984) who report a type I error rate of 10 per cent, all the other reported type I error rates range from 12.5 per cent to 33 per cent, and the total error rates range from 7 per cent to 19 per cent.

Saini and Bates (1984: p.348) offer the general caution that

> comparison is misleading...these studies differ from each other by their respective sample sizes, periods covered, variables examined, and techniques employed.

and in particular, this thesis uses a dependent variable whose definition differs from that used by the authors cited by Saini and Bates.

VIII.4 Conclusion

In this chapter, an early warning four-variable logit model has been developed using an interactive stepwise procedure. The model and the individual coefficients are statistically significant. Using a non-parametric test of the holdout classifications, the null hypothesis that the logit model has no classificatory power cannot be rejected.

The variables appear also in discriminant model I, and three of them appear in model II. This stability of specification suggests that the models are correctly identifying factors that are significant in terms of making arrears/non-arrears classifications. All of the coefficients have the signs that would be expected, *a priori*. In terms of holdout misclassifications, the logit model always dominates discriminant model II, by a margin of one type II error, and it also dominates model I for unit costs of a type I error exceeding 2.8.

A defect of the logit model, which it shares with the discriminant models, is its inability to handle the weak-year cases.

Because it requires neither multivariate normality within groups nor equality of the dispersion matrices, the logit approach is theoretically superior to the discriminant model. The final word on these comparisons must await the out-of-sample investigations of chapter XI. However, up to this point it has not been possible to establish a clear superiority for the logit approach in the development of early warning models for this thesis.

IX A non-parametric approach to early-warning: automatic interaction detection

IX.1 Introduction

IX.1.1 Interaction effects and AID

According to Morgan and Sonquist (1963: p.416)

> Particularly in the social sciences...it is a mistake to assume that the various influences are additive. In the first place, there are many known instances of powerful interaction effects...Second,...we may have interaction effects not because the world is full of interactions, but because our variables have to interact to produce the theoretical constructs that really matter.

In the early warning context, it is at least plausible that interaction effects are present: for example, the significance of a given level of debt might depend on the values of certain other economic and financial variables. This relates to the first of the two points cited above. As regards the second, Morgan and Sonquist are pointing to the use of variables in a data set as proxies for other things, and frequently as proxies for more than one construct. This also may well be a feature of the data that are used in this thesis: for example, they contain no variables that explicitly measure political risk, although they may include proxies for it.

The modelling approaches of chapters VI to VIII took no account of the possibility of interactions. This deficiency provides the motivation for this chapter, in which the automatic interaction detector (AID) technique of Sonquist and Morgan (1964) is used. The purpose of the technique is the detection of interaction effects among configurational data, and the AID algorithm proceeds by seeking a classification of cases, in terms of predictor variables, in which the mean of the criterion variable differs significantly across cells. The criterion variable may be either dichotomous or interval scaled, while the predictors must be categorical. Interval-scaled data may be transformed into new categorical variables, in order to utilise them within an AID analysis.

AID is a sequential splitting process, at each stage of which an optimal split is determined in terms of one of the independent variables - the splitting criterion being based on variance analysis techniques. The objective is to divide the sample into a series of sub-samples so as to maximise the potential for correctly predicting values of the dependent variable, for given values of the independent variables.

AID is described in Green (1978), and the implementation used for this research was the AID programme contained in the PC-MDS software package. In this chapter the criterion variable is the categorical variable ARS(t+1), which is binary valued for arrears and non-arrears classifications, while the predictors are based on financial and economic variables, as described in appendix IX. The model itself is described in IX.2, and conclusions are drawn in IX.3.

IX.1.2 Recursive partitioning: an alternative method

Sequential branching techniques have been utilised in empirical studies of corporate failure: for example, Marais et al (1984), Frydman et al (1985), and Srinivasan and Kim (1987) all use the recursive partitioning method (RPM). However, as far as the author is aware AID has not been used either in that context, or to investigate country risk.

In RPM, the chosen split of the sample is determined on an optimality criterion that involves prior probabilities and relative costs of each type of misclassification. In contrast, AID splits on the basis of maximising the between-groups sums of squares in the predictor. In that it employs a statistical criterion rather than a cost function to determine optimal splits, AID bears a resemblance to discriminant analysis. In this thesis AID is used in preference to RPM, because it does not require a cost function to be fully specified prior to estimation.

IX.1.3 Distributional questions

The linear discriminant model of chapter VII is based on the assumption that the conditional probability density function (cdf) of (x|y) is multivariate normal, where x is a vector of predictors and y is the criterion variable. The logit model of chapter VIII assumes that the cdf of (y|x) is logistic. AID is a distribution-free procedure, and this is a subsidiary reason for using it. The cluster and proximities analyses of chapter VI are also distribution-free, but they have not yielded satisfactory results.

IX.2 The AID analysis of debt restructurings

IX.2.1 Preparation of data

The PC-MDS AID programme requires integer independent variables, and this necessitated a transformation of the ratio-scaled financial and economic variables into integer-valued categorical format. The transformations were made as follows, where z denotes a typical new integer-valued variable:

For x:
less then 1st quartile) z=0

not less than first quartile,)
less than median) z=1

not less than median,)
less than second quartile) z=2

not less than second quartile) z=3

The parameters of the PC-MDS software package limited the number of variables that could be used. The final selection included the categorical variables listed in Table IV.7, along with integer-valued transforms of the 30 financial and economic variables that are listed in Table IX.1. Further details of the integer transformations are given in appendix IX, along with an account of how the variables in Table IX.1 were selected.

Table IX.1: Variables used in AID runs

AGRP	EXIM	RY
BHPCA	INPS	RYPC
CARA	INVR	SIRA
CARAG	MCOV	STPD
DCPI	MCOVG	TDRA
DFIN	MYRA	TDSR
DGDP	NARY	TDPY
DPOP	NBTT	XPM12
EFIR	PDSPX	XPP12
EFMT	PSBR	XYRA

Note: See also Table IXA.1

The criterion variable for the AID analysis of this chapter is ARS(t+1), taking the values 0 for arrears years, 9 for non-arrears years, and values from 1 to 4 for weak-year cases. In order to make the analysis of this chapter comparable with those of chapters VII and VIII, the weak-year cases were excluded from it, leaving a set of 302 cases. While the criterion variable was restricted to the values 0 and 9, the precise values are of no consequence: for example, 0 and 1 would produce identical results.

IX.2.2 The AID analysis

The algorithm requires the user to specify (i) the percent of the total sum of squares that must be contained in any group, for it to be eligible for splitting and (ii) the proportion by which the best split on a candidate group must reduce the unexplained sum of squares, in order for that split to take place. Both of these proportions were set at 0.01 (i.e. 1 per cent). In addition, the minimum group size was set at 15: splits would take place only if the groups thus created were no smaller than this. In all cases, the analyses were confined to monotonic splitting - i.e., all binary splits were restricted to preserving the ordering of the groups. (Green, 1978; PC-MDS, 1988).

The analysis uses the set of 302 observations in which the arrears variable ARS(t+1) has the values 0 or 9 (i.e. excluding weak-year cases), and the tree diagram for it is set out in Figure IX.1. All splits were on four variables:

NARY net assets/GDP
DCPI growth rate, consumer prices
INPS interest/debt service
INVR fixed investment/GDP

that also had a major role in the discriminant analyses of chapter VII, while the first three also emerged in the logit analysis of chapter VIII. DGDP (growth rate of GDP) which had an important place in the discriminant and logit analyses, is absent.

```
------------------------------------------------------
Figure IX.1:  Tree diagram  for AID model
              - dependent variable ARS(t+1)

-------------------------------------------------------
                           (1)
                          N=302
 N: group size            NARY         *: group with larger
                         t=11.9           mean ARS at split
                      |         |
                     |           |
               0,1  |             | 2,3
                   |               |
                  (2)             (3)*
                N=133            N=169
                 INVR             DCPI
                t=9.4             t=6.6
                |   |             |   |
        0,1,2  |     | 3       3 |     | 0,1,2
              |       |         |       |
             |         |       |         |
            (4)         |     (6)       (7)*
           N=99          |   N=22      N=147
           INPS           |  ----      -----
           t=5.5           |  The splits of the predic-
         |     |            | tor variable at each step
 2,3    |       | 0,1        |are indicated  by numbers
       |        |            |beside the branches.
      |         |           (5)*
     (8)       (9)*        N=34   t: t-ratios, with N-2 df,
    N=65       N=34               associated with null hyp-
    ----       DCPI               othesis of no significant
               t=3.5              difference between  group
               |    |             means. The null is rejec-
       2,3    |      | 0,1        ted in each case at 5 per
             |       |            cent level (2-tailed).
            |        |
           |       (11)*          For given splitting crit-
          |        N=15           erion (see text), no fur-
         |         ----           ther splits occur.
        |
       |                          Ordering of groups,  from
     (10)                         highest  to  lowest value
     N=19                         of mean of criterion var-
     ----                         ble,  is: 7 5 11 6 10 8.
---------------------------------------------------------
```

<u>Note</u>: Sample of 302 cases with ARS(t+1) = 0 or 9

The terminal groups, ordered from strongest to weakest, are:

Strongest group:

(7) high NARY
low DCPI.
arrears : 7 cases (4.8 per cent)
non-arrears:140 (95.2 per cent)

(5) low NARY
high INVR
arrears : 4 cases (11.8 per cent)
non-arrears: 30 (88.2 per cent)

(11) low NARY
low INVR
low INPS
low DCPI
arrears : 4 cases (26.7 per cent)
non-arrears: 11 (73.3 per cent)

(6) high NARY
high DCPI
arrears : 10 cases (45.5 per cent)
non-arrears: 12 (54.5 per cent)

(10) low NARY
low INVR
low INPS
high DCPI growth rate, consumer prices
arrears : 15 cases (78.9 per cent)
non-arrears: 4 (21.1 per cent)

Weakest group:

(8) low NARY net assets/GDP
low INVR fixed investment/GDP
high INPS interest/debt service
arrears : 62 cases (95.4 per cent)
non-arrears: 3 (4.6 per cent)

In this listing, 'low' means that the group contains values of the splitting predictor of 0 only, or of 0,1 or of 0,1,2 - as set out in the tree diagram.

These results are summarised in Figure IX.2. For ease of interpretation, INPS has been replaced by 100-INPS, and DCPI by -DCPI. The first of these yields the share of amortization (rather than interest) in debt service. The second is the negative of the inflation rate, and will be high for countries with stable prices. In each case, when the original variable is low, the transform is high, and conversely.

Figure IX.2: AID model: composition of terminal groups

Variable: (Representing:-)		NARY (Solvency)	INVR (Fixed Investment)	100-INPS (Amortization)	-DCPI (Price Stability)
Group	Cases%				
(strongest) 7	49%	High	-	-	High
5	11%	Low	High	-	-
11	5%	Low	Low	High	High
6	7%	High	-	-	Low
10	6%	Low	Low	High	Low
8 (weakest)	22%	Low	Low	Low	-
	100%				

Note: Predictor variables as headed above.

<u>Note</u>:
Creditworthiness is associated with a high value of a variable (or transform, in the case of INPS or DCPI).

No group is completely 'pure', but groups (8) and (7), accounting for 212 cases (70.2 per cent), come close. Group (8), 95.4 per cent arrears, has a profile of low solvency, low fixed investment and low amortization. If instead of the last characteristic, a case has high amortization, but this is offset by high inflation (low price stability), it falls into group (10), of which 78.9 per cent are arrears.

At the other end of the scale, high solvency and price stability define group (7), 95.2 per cent of whose members are non-arrears. Group (5) countries have low solvency mitigated by high fixed investment: 88.2 per cent are non-arrears.

Finally, in group (6) a high net assets ratio is in tension with a high inflation rate (low price stability): the group is almost equally divided.

Table IX.4 sets out the composition of each of the groups defined by the AID model.

Table IX.2: AID model: within-sample composition of groups

		1980-1987		
Group	Possible Cutoff	Arrears	Non-arrears	All
		(numbers of cases)		
7		7	140	147
	- a			
5		4	30	34
	- b			
11		4	11	15
	- c			
6		10	12	22
	- d			
10		15	4	19
	- e			
8		62	3	65
Totals:		102	200	302

<u>Note</u>: The cutoffs a to e are defined in this table for use in due course.

The column 'cutoff' in Table IX.4 indicates the possible cutoffs, a to e, that could be chosen for classification purposes. The groups in the table above any given cutoff will be the non-arrears group, the groups below will be arrears. The choice depends on misclassification costs, which we take to be unity for a type II error and C for a type I, where C > 1. If we assume that prior probabilities are reflected by sample proportions, then as before we may take it that minimising total _ex post_ costs is equivalent to minimising expected costs.

Table IX.3 sets out the ranges of values for C for which each of the possible cutoffs is cost-minimising.

Table IX.3: Cutoffs for AID model

Range of C	Cutoff
1.00 < C < 1.20	d
1.20 < C < 2.75	c
2.75 < C < 7.50	b
7.50 < C	a

Notes: Each cutoff is least-cost for classifying 302 cases for given range of parameter C. Cutoffs labelled according to Table IX.2.

In view of the likely range of misclassification costs (see chapter VII), cutoff b is likely to be appropriate. It yields the cost function:

Total cost = 11C + 30

where the coefficients are respectively the numbers of type I and type II errors, obtained from Table IX.2 using cutoff b (groups 11,6,10 and 8 classified as 'arrears'; 7,5 classified as 'non-arrears').

The within-sample classification of the AID model, using cutoff b, is summarised in Table IX.4. The statistic associated with the FMM test (see chapter VII) is evaluated at 11.3, so the null hypothesis, that the model has no classificatory power, is rejected at the five per cent level. However, this application of the FMM test is not based on hold-one-out classifications, and consequently it is prone to bias (Frank et al, 1965).

Table IX.4: AID model: summary of classifications using cutoff b

Actual group	Total Cases	1980-1987 Predicted group: Arrears	Non-arrears
Arrears	102	91 cases	11 cases (10.8%)
Non-arrears	200	30 cases (15.0%)	170 cases
Total	302	121	181

<u>Notes</u>:
Sample of 302 cases, ARS(t+1) = 0,9 only.

Arrears groups 8 10 6 11; non-arrears 5 7.
Membership based on predictors 1979-1986.

IX.3 Conclusion

The AID analysis supports the conclusion of chapters VII and VIII, that the variables:

NARY net assets/GDP
DCPI growth rate, consumer prices
INVR fixed investment/GDP
INPS interest/debt service

have classificatory ability on the estimating sample, in terms of the arrears/non-arrears status of the members of the sample.

The AID approach goes beyond the discriminant and logit analyses, in revealing interaction effects. As Green (1978: p.200) points out:

> as the tree departs from symmetry, interaction is at work.

INVR has an explanatory role only for cases that have values of NARY below the median. In turn, INPS is only relevant for these cases, and in fact only for that subset whose value of INVR lies below the upper quartile. Moreover, while DCPI appears in both the main branches of the tree, on the left-hand branch (i.e. low NARY), DCPI interacts with INVR and INPS.

Using the FMM test, the null hypothesis of no classificatory ability is rejected. However, if Table IX.4 is compared with Table VIII.7 and Figure VIII.2, it may be seen that with cutoff b, AID yields more type II errors than the logit and discriminant models, and the same number of type I errors as the logit model and discriminant model II. It is thus clear that on the basis of within-sample misclassification, the AID model is always dominated by one of the other multivariate models, for the unit cost C of a type I error in the interval [2.75, 7.50]. To exhaust all reasonable values of C, it is worth exploring the misclassification performance for cutoff c in Table IX.2. This is least cost when C takes values in [1.20,2.75] (see Table IX.3), and gives rise to 15 type I errors (14.7 per cent) and 19 of type II (9.5 per cent). Again, by referring to Table VIII.7 it is easy to see that the AID model with cutoff c is dominated by one of the other models, for values of C in the interval [1.20, 2.75].

The misclassification rates in chapters VII and VIII were computed using the holdout method, while those for AID have been computed using re-substitution and are therefore likely to be biased towards over-optimism. Moreover, the developers of AID have recommended a minimum sample size of 800-1000 cases (compared with the 302 cases used above). Referring to this, Green (1978: p. 200] reports that samples of 200 to 300 may lead to data splits that fail to hold up under cross-validation. In view of these points, it is clear that the comparisons made here probably under-estimate the extent to which the AID model is dominated by the other approaches.

In conclusion, the AID model supports the findings of chapters VII and VIII, in terms of the identification of variables that are useful to assess country risk. Moreover, the approach reveals interactions between the variables that are not apparent from the other methods. However, AID does not itself provide a superior route to assessing country risk, as the comparison of misclassification rates has confirmed. This may be due partly to the loss of information that is inherent in the transformation of ratio scaled data into categorical form. Green (1978) argues for the use of AID as an adjunct to other methods, and not as a replacement for them, and the findings of this chapter should be viewed in that light.

Appendix IX

Preparation of data

The PC-MDS AID programme requires integer independent variables, and this necessitated a transformation of the ratio-scaled financial and economic variables into integer-valued categorical format.

First, quartile points were computed for each continuous variable. Where zero fell in a class interval thus defined, the end point closest to zero was shifted to zero. The data were then transformed by defining a new variable z for each variable x:

For x:	less then 1st quartile)	z=0
	not less than first quartile,) less than median)	z=1
	not less than median,) less than second quartile)	z=2
	not less than second quartile)	z=3

Data based on octiles was also assembled, but to minimise the number of possible splits, to facilitate interpretation, and to retain a reasonable number of cases in each cell, most of the research, including all results presented here, was based on quartiles.

The PC-MDS AID programme allows the user to make the transformation from ratio to integer variables through an equal-intervals subdivision of the range. The difficulty with this approach is that it ignores the shape of the sample frequency distribution completely, and so loses information.

The maximum number of variables that may be included in the input data set for the AID programme is 60, and in any one run a maximum of 39 of these may be submitted to AID. Since the nine categorical variables of Table IV.6, and also the Institutional Investor ratings, are in any case required in the data set, there is only room for 50 other financial and economic variables. To economise on file storage space and processing time, it was found necessary to reduce the input data set below the maximum of 60.

Variable selection was undertaken in two stages. Of the original set of 70 variables (see Tables IV.2 and IV.3), a reduced set of 50 was chosen first. These are listed in Table IXA.1. A further 20 were eliminated at the second stage, leaving 30 marked *.

Table IXA.1: Variables used in AID runs

AGRP*	INDP	RYPC*
BHPCA*	INPS*	RYPCG
CARA*	INPX	SIRA*
CARAG*	INVR*	STPD*
COPC	LABF	STPY
DCPI*	MCOV*	TDRA*
DCPY	MCOVG*	TDSR*
DDCR	MGRO	TDPX
DFIN*	MYRA*	TDPY*
DGDP*	NARY*	TSPD
DMN2	NBTT*	TSPX
DPOP*	OCTD	TSPY
EFIR*	PDSPX*	XGRO
EFMT*	PSBR*	XPM12*
EXIM*	RDPC	XPP12*
FRQPY	RTD	XYRA*
GDPX	RY*	

Notes:
The final data set included 30 variables marked *, along with categorical variables and the Institutional Investor rating. Those 30 appear in Table IX.1.

Each variable is a transformation of the original continuous variable as described in ch. IV, and has integer values in the interval [0, 3].

At the first stage, 20 variables were excluded from the full set of 70 financial and economic variables on condition that each satisfied both of the following criteria: it had played no part in the preceding statistical analyses, and it represented dimensions of the data that were well represented by other included variables. The principal components analyses and the correlation matrix were used as aids to this choice.

For the second stage of the selection procedure, initial exploratory runs were undertaken of the AID analyses that are reported in this chapter and chapter XII. After this exploration, a further 20 variables (those not marked * in Table IXA.1) were deleted from the data set. The remaining 30 variables (marked *) are representative of all the others in terms of a correlation analysis and the principal components analysis; they include all those that have been found in other chapters to be important; finally, the excluded variables did not feature in the preliminary runs of AID.

After this second stage, the complete set of variables remaining to be submitted as input to AID included these 30, plus 8 of the categorical variables, plus the Institutional Investor ratings - i.e. 39 in all, exactly the maximum for an AID run.

X Stability analysis

X.1 Introduction

The results of the multivariate analyses are encouraging. However, further statistical testing of the robustness of the derived models is necessary. This chapter will investigate model stability, while out-of-sample performance will be assessed in chapter XI. Model stability is evaluated by exploring the sample-specificity of the statistical significance of the models, their estimated parameter values, and their classificatory performance.

This question will be investigated using three techniques. Each involves re-estimation of the multivariate models, after modifying the estimating sample in some respect. The first approach uses intertemporal partitioning of the sample, the second involves bootstrapping, while the final approach augments the sample by returning the weak-year cases to it.

(i) <u>Intertemporal partitioning of the sample</u>

Here, the stepwise selection methods are applied to intertemporally defined sub-samples, in order to discover whether model specification and coefficient estimates change over time. In addition, the total-sample equations from earlier analyses are re-estimated on sub-samples of the data. This procedure will pick up inter-temporal instability, and will also reveal whether intertemporal partitioning of the data set is appropriate, for the purpose of estimating the model. The latter point was touched on in chapter VII. These procedures will not be applied to the AID model, in view of the requirement of AID for large estimating samples.

(ii) Bootstrapping

Bootstrapping is described in Efron (1979, 1982, 1983), and Marais et al (1984). It repeatedly re-uses a given sample, via random sampling from that sample, with replacement. In this way, many 'pseudo-samples' are generated, which may be regarded as representative of the underlying population. These pseudo-samples have the same size as the original sample, and the method provides an alternative to repeated sampling of the population.

The data set on which the models of chapters VII, VIII and IX were estimated covers 55 countries for 1979-1986. In one sense, it may be seen not as a sample, but as the population of values for those countries during that period. Moreover, the excluded countries are all small in terms of economic size and external debts. Given that in 1986 the 55 members of data set accounted for 83 per cent of the total external debt of all the countries covered by the 1988-89 edition of the World debt tables, they may be seen to approximate closely to the population of all less developed countries for the sample period.

However in the inter-temporal sense, the data set is clearly a sample, with each country represented by eight observations, drawn from the population of values for that country. Bootstrapping may therefore be used as an alternative to repeated re-sampling, where the repetitions would involve varying the sample period. One purpose of this approach is to test the sensitivity of the results of the earlier chapters to the inclusion or exclusion of particular country-year cases. A second is to obtain parameter estimates, for the distributions of classification rates of the multivariate models.

(iii) Inclusion of weak-year cases

The implications of different treatments of weak-year cases are explored, to reveal the impact on the estimated models of the precise manner in which these cases were treated in earlier chapters.

X.2 Intertemporal partitions

Mensah (1984) has addressed the problem of coefficient stability with respect to sample partitioning, in the context of logit and discriminant analyses of corporate distress. He argues (p.383) that

> A reason for suspecting nonstationarity is that the characteristics of external economic environments which might be expected to affect the financial condition of firms change over time. Three of these external macroeconomic factors are investigated in this study

and goes on to consider inflation, interest rates, and the business cycle. Counterparts to these variables for sovereign debtors are included in the data set of this thesis: import and export prices, nominal and real LIBOR, and the growth rates of world trade and of GNP in the industrial countries (Table IV.4). As reported above in VII.3.3 and VII.4.3, in no case has it been possible to construct a discriminant function that includes any of those variables, or categorical variables based on them, with significant coefficients. However, there still remains the possibility that partitions of the sample should be treated distinctly, and this matter is now addressed.

X.2.1 Discriminant analysis

In earlier chapters it has been argued that the period 1981-1982 was a watershed in the recent history of LDC debt. There are no *a priori* grounds for determining exactly where the partition is to be drawn, so the data were partitioned in two ways.

First, the data were partitioned into two subsamples 1979-1982 and 1983-1986, and interactive stepwise discriminant analyses were run on each sub-sample separately. The best fitting equations that resulted are summarised in Table X.1.

Table X.1: Summary of best-fitting discriminant models: for 1979-1982 and 1983-1986

1979-1982:			
PDSPX	(-)	predicted debt service ratio	(1)
DCPR	(+)	growth rate of private consumption	(5)
INPS	(-)	interest/debt-service	(3)
1983-1986:			
NARY	(+)	net assets ratio	(1)
DCPI	(-)	growth rate of consumer prices	(2)
INPS	(-)	interest/debt service	(3)

Notes:
Variables shown in order of entry; (-) or (+) indicate signs of estimated coefficients; 'negative' group is arrears. Numerals (1) to (5) are principal component on which variable is maximally loaded, both in the full data set and in each of the partitions.

The signs of the coefficients accord with prior expectations, except for the positive sign on DCPR in the 1979-1982 model. A high value of DCPR might be expected to yield early warning of debt-servicing difficulties in certain circumstances, whereas the positive coefficient implies the converse. However, in the principal components analysis of the full data set, this variable was highly loaded on the component 'economic growth'. This holds also within all the subsets of the data that are analysed here, and this component is a dimension of the data that has been found elsewhere (for example, in discriminant model I) to be positively related to creditworthiness.

The partitioning was then revised, and the stepwise analysis was repeated for the periods 1979-1981 and 1982-1986. The same set of discriminating variables was selected for the earlier sub-period as under the previous partitioning, and this also occurred for the later sub-period.

Four points are noteworthy:

-first, the variables PDSPX and NARY (which enter first) are both highly loaded on the first principal component of the data set;

-secondly, the variable DCPR, which enters second into the discriminant functions for the earlier periods, is highly loaded on component 5. While this component is not represented in the discriminant functions for the later periods, it is represented in the full-sample function by the variable DGDP (GDP growth rate);

-thirdly, the variable INPS appears in all the cases examined;

-finally, the variables NARY, DCPI and INPS are the first three variables to enter the stepwise discriminant analysis of the full sample, and in that analysis they have a combined Mosteller-Wallace contribution of 88 per cent.

The preceding suggests that the best specifications for the full sample and the various sub-samples are not too different, and that therefore the applicability of the 5-variable model I to the full sample should be further investigated. A procedure is adopted here that has been widely used in the context of multiple regression (Gujarati, 1970a,b), and it is in effect an extension of the analysis of section VII.4.2. Intercept and slope dummies were defined, for the two different partitions: (i) 1979-1982 versus 1983-1986; (ii) 1979-1981 versus 1982-1986. In each case, this procedure enables tests to be made of the null hypotheses that there are no significant differences between the parameters of the model within each sub-period. Using the interactive stepwise procedure, the five variables of discriminant model I entered in the same order as before. The only dummy variable to enter the discriminant function was the slope dummy defined on DCPI, with the sample partitioned according to (ii) above. However, with a partial F-statistic of 1.34, the null hypothesis cannot be rejected at the five percent level. Otherwise, the significance level of the additional variables was too low for them to be selected by the stepwise procedure.

The full-sample five-variable discriminant function was re-estimated on each of the sub-samples 1979-1982 and 1983-1986. In each case, the coefficients were significant at the 5 per cent level and had the correct signs. In terms of group separation, the model performed better on the later period (see Figures X.1 and X.2). The coefficients of discriminant model I are set out in Table X.2, for the various estimating samples.

Table X.2: Discriminant model I coefficients estimated on different samples

	1979-82	1979-86	1983-86
NARY	0.0836	0.0747	0.0822
DCPI	-0.0687	-0.0670	-0.0680
INPS	-0.0226	-0.0684	-0.1233
DGDP	0.1466	0.1237	0.1087
INVR	0.0731	0.0743	0.1204
Intercept	4.0251	6.8384	10.0040

Figure X.1: Discriminant model I: estimated on 1979-1982; all-groups histogram of discriminant scores

---

```
  ¦
  ¦      Symbols: 1=arrears        2
F ¦               2=non-arrears    2
R + 12                             2
E ¦                                2
Q ¦                             2  2
U ¦                            22 22 2
E +  8                         22 22 2 2
N ¦                          2 22 22 2222
C ¦                          2 22 22 2222
Y ¦                 11       21 2222222222
  +  4              11   1 221222222222222
  ¦                 111  121212222222222222
  ¦              1  1111111121222212122222222
  ¦  1       11 111111111111111111111222222222 2  Scores
  +-------+---------+---------+---------+---------+------
-6.0    -4.0      -2.0       0.0       2.0       4.0
Group:  11111111111111111111111222222222222222222222222
Centroids:              1           2
```

Figure X.2: Discriminant model I: estimated on 1983-1986; all-groups histogram of discriminant scores

---

```
  ¦
  ¦                              2
F ¦                              2
R + 12                           2
E ¦                          22  2
Q ¦                          22  2
U ¦                          222 2 2
E +  8                       222 2 2
N ¦                          2222222
C ¦                        2 2222222
Y ¦           1     1      2 2222222  222
  +  4        11  1 1 1    2 22222222 222
  ¦         11111 11111    2222222222 222 2
  ¦    1  1 11111111111 112222222122 222 2
  ¦    1 11 11111111111 1111122221122222222  2   Scores
  ----+---------+---------+---------+---------+----------
    -4.0      -2.0       0.0       2.0       4.0
Group:   1111111111111111112222222222222222222222222222
Centroids:        1              2
```

The coefficients on NARY, DCPI and DGDP are quite stable while the others vary widely between the samples. Discriminant weights are unique only up to a scale transformation (Green, 1978), but when account is taken of this by normalising within each set, the conclusion is still valid. However, the primary concern here is on prediction: thus, staying within the sample period, the predictive ability of each sub-period function is tested. The function estimated on data for 1979-82 is applied to the data for 1983-86, and the 1983-86 function is applied to the data for 1979-82 (Mensah, 1984: p. 391). The misclassifications by this procedure are set out in Table X.3, along with those derived from the full-sample discriminant model I of chapter VII.

The full-sample model is superior in the earlier period, provided that the cost of a type I error is no smaller that 0.75 that of a type II error: this is certainly the case. The same is true for the later period, provided that the cost ratio exceeds unity. Consequently, if we add the two periods then we conclude that the full sample model outperforms forward and backward prediction by the split-sample models. Provided that the cost of a type I error is more than 0.8 of the cost of a type II error (which is highly likely), then the full-sample model has lower misclassification costs within the sample.

Table X.3: Discriminant model I estimated on sub-periods: forward and backward prediction

Classification of:	No. misclassified Type I	Type II
Model I estimated on 1983-86 - applied to 1979-1982 data:	15/48 (31.2%)	1/102 (1.0%)
Full-sample model I - applied to 1979-1982 data:	11/48 (22.9%)	4/102 (3.9%)
Model I estimated on 1979-82 - applied to 1983-1986 data:	6/54 (11.1%)	0/98 (0.0%)
Full-sample model I - applied to 1983-1986 data:	5/54 (9.3%)	1/98 (1.0%)
Summed results for 1979-1986 (102 arrears, 200 nonarrears)		
- **Backward/forward prediction:**	21/102 (20.6%)	1/200 (0.5%)
- **Full-sample model I:**	16/102 (15.7%)	5/200 (2.5%)

Note:
Misclassifications by full-sample model I are based on the Lachenbruch holdout method (Table VII.4).

X.2.2 Discriminant analysis: conclusions

Two conclusions follow from the results of section X.2.1. First, in terms of model specification, the sub-period discriminant functions resemble those for the full sample period, in that the principal components represented in the sub-periods are all represented in the full-sample model. Secondly, as regards the sample-specificity of the parameter estimates, it turns out that the full-sample model has greater classificatory power than models having the same specification but estimated on sub-samples. This point was established by comparing the hold-one-out misclassifications by the original model with forward and backward prediction by the alternative models, as set out in Table X.3.

X.2.3 Logit analysis

The discriminant and logit models of chapters VII and VIII are rather similar. Therefore, and given the conclusions that are set out in X.2.2, a full analysis of the logit model will not be carried out here. Attention will be limited to stepwise estimations of logit functions on the sub-samples 1979-82 and 1983-86, and to re-estimations of the original model on those sub-samples.

The stepwise logit analyses yielded estimated functions that are summarised in Table X.4.

Table X.4: Summary of best-fitting logit models: 1979-1982 and 1983-1986

1979-1982:			
PDSPX	(+)	predicted debt service ratio	(1)
DCPR	(-)	private consumption, growth rate	(5)
1983-86:			
NARY	(-)	net assets ratio	(1)
DCPI	(+)	consumer prices, growth rate	(2)

<u>Notes</u>:
Variables shown in order of entry; (-) or (+) indicate signs of estimated coefficients, giving direction of association between value of variable and probability of arrears. Numerals (1) to (5) indicate the principal component on which the variable is maximally loaded.

Models with additional variables were rejected by the improvement and Hosmer-Lemeshow chi-squared tests. Compared with the corresponding discriminant analyses (Table X.1), exactly the same variables entered at the first and second stages. In each case, and at the five percent level: all variables had significant coefficients according to the Wald test; the null hypotheses that the models fit the data could not be rejected (Hosmer-Lemeshow test); and at each stage, the value of the improvement chi-square indicated a rejection of the null hypotheses that the added variable made no improvement to the fit.

As regards the re-estimations of the original four-variable model on the partitioned data set, the results are in Table X.5. Table X.6 shows the misclassifications by these logit models, using backward and forward prediction, with cutoffs based on a cost analysis.

It will be seen that all the sub-period coefficients lie within the 95 per cent confidence interval around the full sample estimate. On this basis, we cannot reject the null hypothesis of coefficient stability, at the five per cent level. This conclusion is reinforced by the results of the stepwise analyses, which indicate no dramatic change in specification when optimal logit functions are sought on sub-samples of the data.

The misclassification rates shown in Table X.6 reveal that on a summed basis, backward and forward prediction by the sub-period models yields one fewer type than the full-sample model, and one more of type II. This is not strong evidence of intertemporal instability.

Table X.5: Coefficients of logit model: estimated on various sub-samples

	1979-82	1979-86	1983-86	(95 per cent) Confid. Interval
NARY	-0.0817	-0.0832	-0.0878	-0.1075, -0.0589
DCPI	0.0782	0.0822	0.0797	0.0778, 0.0866
DGDP	-0.2149	-0.2271	-0.2745	-0.3405, -0.1136
INPS	0.0397	0.0672	0.0905	0.0277, 0.1066
Inter.	-7.0752	-9.0113	-10.7365	-11.6713, -6.3513

Note:
The confidence interval lies around the estimates from the full-sample (1979-1986) data, and is based on the t-statistic.

Table X.6: Logit model estimated on sub-periods: forward and backward prediction

Classification of:	No. misclassified Type I	Type II
Model estimated on 1983-86 - applied to 1979-1982 data:	4/48 (8.3%)	19/102 (18.6%)
Full-sample model - applied to 1979-1982 data:	7/48 (14.6%)	11/102 (10.8%)
Model estimated on 1979-82 - applied to 1983-1986 data:	6/54 (11.1%)	1/98 (1.0%)
Full-sample model - applied to 1983-1986 data	4/54 (7.4%)	8/98 (2.0%)
Summed results for 1979-1986 (102 arrears, 200 nonarrears)		
- Backward/forward prediction:	10/102 (9.8%)	20/200 (10.0%)
- Full-sample model:	11/102 (10.8%)	19/200 (9.5%)

Note:
Misclassifications by full-sample model are based on the hold-one-out method: see Table VIII.4.

X.3 Bootstrapping

The method and its purpose are described in X.1. A random number generator was used to generate 100 random samples **with replacement** from the data. The data comprised 302 cases that were not weak years, and each random sample was thus of size 302.

Table X.7 sets out sample statistics for the 100 estimates of discriminant model I, based on this procedure. The coefficients on the net assets ratio and inflation are the most unstable, as reflected in high coefficients of variation. The coefficients (other than the intercept) are shown in declining order of Mosteller-Wallace contribution. In this order, the variability of the coefficients, as reflected in their coefficients of variation, increases.

In large samples, the mean parameter estimate is approximately normal, by the central limit theorem. If the sample standard deviation s is taken as an estimate of the population value, then a confidence interval around the parameter is determined by the distance

ts/SQRT(n)

where n is the sample size and t is the tabulated t

Table X.7: Discriminant model I: bootstrap analysis

NARY	mean estimate	0.080486
	standard deviation	0.008286
	coefficient of variation	0.10295
	confidence interval	(0.0788, 0.0821)
	raw coefficient	0.074704
	jackknifed coefficient	0.073185
DCPI	mean estimate	-0.072117
	standard deviation	0.010544
	coefficient of variation	0.14620
	confidence interval	(-0.0742, -0.0700)
	raw coefficient	-0.067022
	jackknifed coefficient	-0.065477
INPS	mean estimate	-0.071325
	standard deviation	0.016549
	coefficient of variation	0.23203
	confidence interval	(-0.0746, -0.0680)
	raw coefficient	-0.068361
	jackknifed coefficient	-0.066927
DGDP	mean estimate	0.127140
	standard deviation	0.041813
	coefficient of variation	0.32886
	confidence interval	(0.1188, 0.1355)
	raw coefficient	0.123680
	jackknifed coefficient	0.121504
INVR	mean estimate	0.074751
	standard deviation	0.028924
	coefficient of variation	0.38693
	confidence interval	(0.0690, 0.0805)
	raw coefficient	0.074338
	jackknifed coefficient	0.073267
INT.	mean estimate	7.506290
	standard deviation	1.358560
	coefficient of variation	0.18099
	confidence interval	(7.2346, 7.7780)
	raw coefficient	6.838430
	jackknifed coefficient	6.671480

value at the 2.5% significance level with n-1 degrees of freedom (Meyer, 1970: p.306).

In Table X.7 the raw and jackknifed coefficient estimates may be compared with confidence intervals around the bootstrapped estimates. The coefficients on NARY, DCPI, the jackknifed coefficient on INPS, and the intercept (INT.) are all outside; in each case they are smaller in absolute value than the lower (absolute) limit, at the 5 per cent level. If bootstrapping produces unbiased estimates, then at the 5 per cent level we reject the hypothesis that these other estimates are unbiased. For NARY and the intercept, the effect will be to over-classify as arrears, i.e. it will raise the type II error rate. For DCPI and INPS, the coefficients are negative and the effect is to raise the type I error rate.

Table X.8 sets out summary statistics for the classification of the original sample by the bootstrapped discriminant functions. The within-sample classification of chapter VII is based on the Lachenbruch U-method (hold-one-out). This yields error rates that are both biased below the 95 per cent confidence intervals, based on the usual t-statistic and assuming that that bootstrapping approach produces unbiased estimates.

Table X.8: Classifications by bootstrapped discriminant functions (model I)

	1980-1986 Predicted group	
Actual group	Arrears	Non arrears
Arrears	82.7%	Type I mean : 17.3% range: (13.7,22.6) s.d.: 1.9 c.i.:(16.9, 17.7) U : 15.7%
Non arrears	Type II mean : 3.4% range:(0.0, 9.5) s.d.: 2.4 c.i.:(2.9, 3.9) U : 2.5%	96.6%

<u>Key</u>: c.i. = 95 per cent confidence interval.
U = holdout classification rate (ch.VII).

<u>Note</u>: Predictors dated 1979-1986 are used to predict debt-servicing status in 1980-1987.

The key question, however, is out-of-sample testing. This belongs properly in chapter XI. In summary, the results to be reported there include a classification of the sample countries for 1988 to 1990 by discriminant model I, using jackknifed coefficients, that is <u>identical</u> with the classification based on the mean coefficients from Table X.8. Furthermore, using discriminant functions whose coefficients are the end-points of the confidence intervals of Table X.7, we get almost the same classification. A set of coefficients formed from the upper limits yields a bias towards classification as 'non-arrears', compared with classification by the mean coefficients, while the lower limits bias the classification towards 'arrears'. In only seven out of 165 cases do either of these procedures yield a different classification from that of the jackknifed coefficients. Given the great similarity of the bootstrapped and jackknifed classifications, the conclusion is that serious bias through sample specificity is absent. Moreover, given this finding, a bootstrap investigation of the logit and AID models will not be pursued.

X.4 The weak-year cases added back

Discriminant model I was estimated on a sample of 302 cases, 138 weak-year cases having been deleted from the full sample of 440. This procedure was arbitrary, and must be examined further. The approach taken here is to use the full sample, partitioned in various ways, and to subject each partitioning to a stepwise discriminant analysis.

The partitions were determined as follows by the values of the categorical variable ARS(t+1) (see chapter III.7 and Table IV.6):

(i) ARS(t+1) = 0 versus ARS(t+1) > 0;
(ii) ARS(t+1) = 0 or 1, versus ARS(t+1) > 1;
(iii) ARS(t+1) = 0, 1 or 2, versus ARS(t+1) > 2;
(iv) ARS(t+1) < 9, versus ARS(t+1) = 9.

In each case, a discriminant function was sought, having a parsimonious set of discriminating variables. The number of variables selected in each case was either three or four. In each case, all F-statistics were significant at the five per cent level. Moreover, either the inclusion of a further variable into the discriminant function yielded a reduction of no more than one point in Wilks's lambda (e.g. from 0.60 to no less than 0.59), or no other variables with statistically significant coefficients could be found.

The best-fitting discriminant function in each case included the variables set out in Table X.9. For comparison, details of the original model I, estimated on the 302-case data set, are shown first. It is immediately noticeable that no new components of the data set are introduced through this treatment of the weak-year cases.

Table X.9: Best-fitting discriminant models, from 440-case data set: partitions (i) to (iv)

Original model I estimated on 302-case data set:				
	NARY	(+)	net assets ratio	(1)
	DCPI	(-)	growth rate, consumer prices	(2)
	INPS	(-)	interest/debt service	(3)
	DGDP	(+)	growth rate, GDP	(5)
	INVR	(+)	investment/GDP	(4)
Alternative models: 440-case data set, partitions (i), (ii), (iii), (iv) respectively:				
(i)	NARY	(+)		(1)
	DMN2	(-)	growth rate of money supply	(2)
	DGDP	(+)		(5)
(ii)	NARY	(+)		(1)
	DCPI	(-)		(2)
	DGDP	(+)		(5)
(iii)	NARY	(+)		(1)
	DCPI	(-)		(2)
	OCTD	(+)	official creditors/total debt	(4)
	DIND	(+)	growth rate, industrial prodn.	(5)
(iv)	NARY	(+)		(1)
	DCPI	(-)		(2)
	RYPC	(-)	real income per capita	(4)

Notes:
Variables shown in order of entry. Sign of estimated coefficients indicated - or +; 'negative' group is arrears. Numerals (1) to (5) indicate principal component on which variable is maximally loaded.

NARY appears in each case, and is always the first variable to enter, while a variable loaded on component 2 always enters second. In all cases but one, that variable is DCPI.

To these five discriminant functions, a further eight were added as follows: first, the four new specifications were each re-estimated on the 302-case reduced data set; secondly, the original five-variable model I was re-estimated on the 440-case data set, after partitioning it in each of the four ways listed above. In each case, the estimated coefficients had the signs that are listed above. Given that the 'right-hand' group consists of 'non-arrears' cases, it may be seen that the coefficients have their correct sign, except possibly in one case. A 'non-arrears' classification is associated with:

High values of NARY	(net assets ratio)
Low values of DCPI or DMN2	(inflation)
Low values of INPS	(interest burden)
High values of DGDP or DIND	(economic growth)
High values of INVR or OCTD	(development)

The possible exception is RYPC, real income per capita, which has a negative coefficient. This may perhaps be explained as reflecting the fact that large-scale lending has been extended disproportionately to the higher-income LDCs. Consequently it is those countries that tend to have debt-servicing difficulties.

We now have 13 estimated discriminant functions. Since the primary focus is on predictive ability, they are compared in terms of their within-sample classification, using the holdout method.

First, the discriminant scores were computed. The matrix of correlation coefficients between these 13 sets of scores reveals high, and significant, correlation: in all cases, the correlation coefficient exceeds 90 per cent, and in 42 (of 78) cases it exceeds 95 per cent.

Secondly, the classificatory performance of each function was examined. The baseline for this was the performance of model I: 16 type I errors and 5 type II (see Table VII.4). In each case, the classificatory performance depends on where the cutoff between the groups is drawn - equivalently, on the value assigned to the intercept, using a zero cutoff. The simplest procedure is to take the ordered discriminant scores for each function, and by varying the cutoff, to find:

- (a) the minimum number of type I errors
  - given 5 of type II;
- (b) the minimum number of type II errors
  - given 16 of type I.

In each of these ways, the new discriminant functions may be compared with the model I baseline, and the results are shown in Table X.10.

Table X.10: Classification by discriminant models, derived to test treatment of weak-year cases

Model Specification	Estimated on	1980-1987 Misclassifications of arrears/non arrears cases: Minimise: Type I, for 5 type II	Type II, for 16 type I
NARY) DCPI) INPS) DGDP) INVR)	302 cases:	16	5
	440 cases (i):	18	12
	(ii):	17	6
	(iii):	19	6
	(iv):	19	12
NARY) DMN2) DGDP)	440 cases (i):	21	23
	302 cases:	22	24
NARY) DCPI) DGDP)	440 cases (ii):	24	18
	302 cases:	23	20
NARY) DCPI) OCTD) DIND)	440 cases (iii):	29	17
	302 cases:	28	16
NARY) DCPI) RYPC)	440 cases (iv):	35	19
	302 cases:	32	18

<u>Notes</u>: (i) to (iv) indicate treatment of weak-year cases in estimating sample: see text.

Holdout classifications of 302 cases, 1980-87: predictors dated 1979-1986; weak-year cases unclassified.

It is clear that the new specifications of the discriminant function perform significantly worse than model I - whether asked to minimise the number of type I errors for a given number of type II, or conversely. Three conclusions follow. First, the original specification of the discriminant function is stable in the sense that none of the other specifications introduces a new dimension of the data set, although they use fewer than five apiece.

Secondly, the best-performing model, as reflected by the within-sample holdout classification, is the original five-variable function. However, there is a near-equivalence between the original model I, and the same specification re-estimated on the 440-case dataset, partition (ii) - the latter produces one more misclassification of each type.

Thirdly, this near-equivalence may be rationalised as reflecting a robustness of the model towards the original treatment of the weak-years cases. A natural way to bring the weak-year cases into the estimating sample is to assume that cases one year either side of an arrears-year case are likely to have characteristics close to that of an arrears year, and that the further away a weak-year case lies, the closer its characteristics will be to that of a non-arrears year. Thus partition (ii) reflects a limited, and natural, broadening of the definition of an arrears year, which has been shown to have little effect on the classifying performance of the five-variable discriminant model.

X.5 Conclusion

In the preceding sections, the stability of the estimated discriminant model I was examined in three ways: through intertemporal partitioning of the data set, through bootstrapping, and through expansion of the data set to 440 cases, with the weak-year cases reallocated to the arrears and non-arrears categories. The overall conclusion from these investigations is that the original model I is stable and robust.

Section X.2.3 reports a similar conclusion for the logit model, based on an intertemporal data partitioning only. In view of this, and given the similarity between the discriminant and logit analyses of chapters VII and VIII, it is fair to assume that the conclusion of stability and robustness holds for the logit model also.

The grounds for transferring the conclusions to the AID model are weaker: bootstrapping is not technically feasible, and the sample is too small for intertemporal partitioning.

XI Out-of-sample testing

XI.1 Introduction

XI.1.1 Structure of chapter

Out-of-sample performance is the key to evaluating the multivariate models of chapters VI to IX. In developing econometric models, it is good practice to hold out some observations for this purpose. This is one of the pervasive themes of a recent paper by Gilbert (1986) on British econometric methodology: for example (p. 296)

> Econometric models are tested in the wind tunnels of sample data, and in good professional hands, are also validated against immediate post-sample data.

For this reason, the data for 1987 to 1989 were held out of the estimating sample, and they are utilised in this chapter to yield model forecasts. Predictions for 1988-1990 are generated by each model on the basis of data for 1987-1989, and these are then compared with each other and with the actual outturn.

Section XI.2 is concerned with difficulties in determining the *ex post* classification of countries in the period 1988 to 1990. Sections XI.3 to XI.6

deal with the forecasting performance of the cluster/proximities, discriminant, logit, and AID models, respectively. The classificatory power of the models is tested in XI.7, and conclusions are drawn in section XI.8.

XI.1.2 The treatment of outliers and missing data

Where for any out-of-sample case a variable had a value outside the range determined by the Winsorizing procedure (see chapter IV), that value was replaced by the upper or lower bound to that range, as appropriate.

Data were missing for NARY for Nicaragua for 1987 and 1988, and in these cases the mean sample value for Nicaragua was substituted.

XI.2 Classifications out-of-sample

Information on the emergence of debt-service arrears and the dates of restructurings in 1988, 1989 and 1990 (the 'holdout period') was obtained from the most recent sources, including the IMF (1989a), the World Bank (1989-1990), and The Economist Intelligence Unit's Country Reports and Country Profiles. Data for 1990 are provisional.

A classification based on this data is not immediately usable as the basis for out-of-sample testing. This is because the weak-year assignments are incomplete, in that the *ex post* classifications as 'weak-year cases' in the holdout period will depend partly on events from 1991 onwards.

It will not be satisfactory simply to ignore cases from 1987 to 1989 that are weak (i.e. close to an arrears case). Even if the multivariate models were capable of making more than binary classifications, with weak-year cases isolated in one or more distinct groups, it would not be possible to assess model performance accurately in this respect on the holdout period: again, the problem is the incompleteness of the weak-year assignments. In any case, chapters VII to IX have established that the multivariate models are not capable of isolating weak year cases, using either the two-group models, or multiple discriminant analysis. Therefore, an alternative approach must be used.

Since 1986, most of the weak-year cases involve countries whose creditworthiness is still doubtful. In most cases, even after the events of arrears and restructuring, voluntary access to the capital markets on commercial terms has not been attained, and further restructurings are likely. The simplest resolution of the problem would be to make an arbitrary assignment: for example, to assign all weak year cases to the arrears category; alternatively, given the values of ARS(t+1) from 0 to 3, cases with the value 1 could be recoded 0 (arrears), and the remainder recoded 9.

The problem with either of these approaches is that they are arbitrary, and make no use of country-specific information that is available. A minority of countries have shown marked improvements after recent restructurings, and yet would be classified in the arrears group during this improving phase. For example, in 1988 and 1989 Chile had essentially no difficulties in servicing its debts: these were in any event shrinking through the use of debt-equity swaps and other methods of debt reduction, and the Chilean economy was booming. On these grounds, it is plausible that 1988 and 1989 should be designated as non-arrears, yet in the wake of the 1987 restructuring, each of the assignments proposed above would put at least one of them (1988) into the arrears category.

For this reason, the approach taken here is to assign doubtful cases in the holdout period to one of the two groups, arrears and non-arrears, on the basis of information derived from external sources. The primary sources consulted were the quarterly Country Reports and the annual Country Profiles of the Economist Intelligence Unit, supplemented with information from the World Debt Tables and the financial press. Table XI.1 sets out the codings of the countries for 1987-1990 that are yielded by this process, while the initial classification (before assignement of weak-year cases) is shown in appendix XI.1.

Table XI.1: Debt servicing status: 1987 to 1990

Key: ARS(t+1) coded 0 (arrears cases), 9 (non-arrears)

	ARS(t+1)				ARS(t+1)		
t+1 =	1987	1988	1989 1990	t+1 =	1987	1988	1989 1990
AL Algeria	9	9	9	AR Argentina	0	0	0
BD Bangladesh	9	9	9	BL Bolivia	0	0	0
BR Brazil	0	0	0	CB Colombia	9	9	9
CG Congo	0	0	0	CH Chile	0	9	9
CM Cameroon	0	0	0	CN China	9	9	9
CR Costa Rica	0	0	0	DR Dom. Rep.	0	0	0
EC Ecuador	0	0	0	EG Egypt	0	0	0
ES El Salv.	9	9	9	GA Gabon	0	0	0
GH Ghana	0	0	0	GU Guatemala	0	0	0
HD Honduras	0	0	0	ID Indonesia	9	9	9
IN India	9	9	9	IR Iran	9	9	9
IV Ivory Cst.	0	0	0	JD Jordan	9	0	0
JM Jamaica	0	0	0	KN Kenya	9	9	9
KS S. Korea	9	9	9	LB Liberia	0	0	0
MA Malasia	9	9	9	MC Morocco	0	0	0
MW Malawi	0	0	0	MX Mexico	0	0	0
NC Nicaragua	0	0	0	NG Nigeria	0	0	0
PE Peru	0	0	0	PG Paraguay	0	0	0
PH Phil'nes	0	0	0	PK Pakistan	9	9	9
PN Panama	0	0	0	PP Papua N.G.	9	9	9
SE Senegal	0	0	0	SR Sri Lanka	9	9	9
SU Sudan	0	0	0	SY Syria	9	9	9
TH Thailand	9	9	9	TK Turkey	9	9	9
TN Tunisia	9	9	9	TT Trinidad	9	0	0
TW Taiwan	9	9	9	UR Uruguay	0	0	0
VZ Venezuela	0	0	0	YU Yugoslavia	0	0	0
ZA Zaire	0	0	0	ZB Zambia	0	0	0
ZI Zimbabwe	9	9	9				

Note:
Weak-year cases recoded as described in the text.

XI.3 Cluster and proximities analysis

Cluster analyses were run for predictor variables dated 1987, 1988 and 1989, using standardised values of the variables listed in Table V.2 - i.e. the variables that were used to form cluster I for 1979 to 1986. The dendrograms for the three clusters are reproduced in Chart XI.1.

Two points are revealed by the dendrograms. First, a comparison with Table XI.1 indicates that this particular clustering does not isolate the arrears and non-arrears cases, looking one year ahead in each case. Secondly, the original eighteen countries that were clustered together during the sample period are now dispersed. In view of this, there is no purpose in proceeding with a proximities analysis relative to the group centroid of cluster I.

The poor performance is not surprising, given that the analysis failed to cluster the arrears and non-arrears cases consistently and distinctly within the sample period. As Schmidt (1984: p.365) writes, of a similar model:

> the ability for early warning seems to be unsatisfactory.

Figure XI.1: Dendrograms for cluster I

```
Case +---------+---------+---------+---------+---------+
 7HD -+        5        10        15        20        25
 7SE -+-+
*7KN -+ |
 7PP ---+-----+
*7JD -+ |     |
 7BL -+-+     |
 7JM -+       |
 7IV ---+-----+-------------+
 7MC ---+     |             |
*7TN -+       |             |
 7PH -+-------+             |
*7ID -+                     +-+
 7MW -+                     | |
 7SU -+                     | |
 7NG -+-+                   | |
 7EG -+ +---------+         | |
 7ZB ---+         +---------+ |
 7CR -+---------+ |           |
 7CH -+         +-+           |
 7GH ---+-+     |             +---+
 7PE ---+ +-----+             |   |
 7DR -----+                   |   |
*7ES -+                       |   |
*7ZI -+---+                   |   |
 7GU -+   +---------+         |   |
*7SR -+   |         |         |   |
*7CM -----+         |         |   |
*7BD -+             +---------+   +-------------------+
 7PK -+---+         |             |                   |
*7IN -+   |         |             |                   |
*7TH -+   +---------+             |                   |
 7PG -+   |                       |                   |
*7CN -----+                       |                   |
 7LB -+---+                       |                   |
 7CG -+   +-----------+           |                   |
 7EC -----+           +-----------+                   |
 7ZA -----------------+                               |
*7CB ---+---+                                         |
 7VZ ---+   +---------+                               |
 7UR ---+---+         |                               |
 7MX ---+             +-----------------------+       |
 7BR ---+---+         |                       |       |
 7TK ---+   +---+     |                       |       |
 7AR -------+   +-----+                       |       |
*7AL -----------+                             +-------+
*7KS ---------+-+                             |
*7MA ---------+ +-------+                     |
 7PN -----------+       +---------------+     |
 7IR -------------+-----+               |     |
*7TW -------------+                     +-----+
*7SY -----+---------------+             |
 7YU -----+               +-------------+
 7GA -------+-------------+
 7TT -------+
```

Rescaled distance

Year: 1987

Dendrogram from clustering using Ward's method.

The input variables are shown in Table V.2.

Original members of cluster I are marked *.

```
Case +---------+---------+---------+---------+---------+
 8HD -+        5        10        15        20        25
 8SE -+-+                              Rescaled distance
*8SR -+ +---+
*8KN -+ |   |                          Year: 1988
 8PK -+-+   |
 8GU -+     +-+                        Original members
*8ZI -+     | |                        of cluster I are
 8PP -+     | |                        marked *.
 8PH -+     | |
 8BL -+-----+ |
*8ID -+       |
*8ES -+-+     |
*8TN -+ +-----+-------+
 8GH -+-+     |       |
 8PG -+ |     |       |
*8BD ---+     |       +-------+
 8JM ---+-----+       |       |
 8MC ---+             |       |
*8AL ---------+-------+       |
*8CM ---------+               +---------------+
*8IN -+                       |               |
*8TH -+---------------+       |               |
*8CN -+               |       |               |
 8CR -+-+             +-------+               |
 8CH -+ +-------+     |                       |
 8MW ---+       +-----+                       +-------+
 8EC -----+-----+                             |       |
*8SY -----+                                   |       |
 8ZB -+-+                                     |       |
 8EG -+ |                                     |       |
 8DR -+-+-------------------+                 |       |
 8NG -+ |                   |                 |       |
 8SU ---+                   +-----------------+       |
 8IV -+                     |                         |
 8CG -+-----------------+   |                         |
*8JD -+                 +---+                         |
 8LB -------+-----+     |                             |
 8ZA -------+     +-----+                             |
 8GA -------------+                                   |
 8BR -+                                               |
 8MX -+---+                                           |
 8TK -+   +---------+                                 |
 8AR -----+         +-------------------+             |
*8CB -----+-+       |                   |             |
 8VZ -----+ +-------+                   |             |
 8UR -------+                           +-------------+
*8KS ---+-------+                       |
*8MA ---+       +-----------------+     |
*8TW -----------+                 |     |
 8IR -----+-------+               +-----+
 8TT -----+       +-------+       |
 8PE -------+-----+       +-------+
 8YU -------+             |
 8PN ---------------------+
```

```
Case +---------+---------+---------+---------+---------+
*9TN -+        5        10        15        20        25
 9BL -+-+                             Rescaled distance
 9PP -+ |
*9ES -+-+                             Year: 1989
*9ZI -+ |
*9BD -+ +---+                         Original members
*9SR -+ |   |                         of cluster I are
 9GU -+-+   |                         marked *.
*9KN -+ |   +-----+
 9SE -+ |   |     |
 9HD ---+   |     |
*9CM -------+     |
 9PH -+           +-------+
 9PK -+-+         |       |
*9ID -+ +---------+       |
 9GH -+ |         |       |
 9MW -+-+         |       |
 9PG -+           |       +-------------+
 9JM -+-----+     |       |             |
 9MC -+     +-----+       |             |
 9CR -----+-+             |             |
 9GA -----+               |             |
*9CN -+-------+           |             |
*9IN -+       +-----------+             |
*9MA -+-+     |                         |
*9TH -+ +-----+                         +-----+
 9CH ---+                               |     |
 9BR ---+-------+                       |     |
 9TK ---+       +---+                   |     |
*9AL ---+---+   |   |                   |     |
 9MX ---+   +---+   +-------------+     |     |
*9CB -------+       |             |     |     |
 9AR ---------------+             |     |     |
*9JD -+---+                       |     |     |
 9NG -+   +-------------+         +-----+     +-------+
 9ZB -+   |             |         |           |       |
 9EG -+---+             +---+     |           |       |
 9EC -+                 |   |     |           |       |
 9DR -+---+             |   |     |           |       |
 9SU -+   +-------------+   +-----+           |       |
*9SY -+---+                 |                 |       |
 9YU -+                     |                 |       |
 9PN -----------------------+                 |       |
 9LB -------+---------+                       |       |
 9ZA -------+         +-----------------------+       |
 9IV -----+-----+     |                               |
 9CG -----+     +-----+                               |
 9NC -----------+                                     |
 9PE ---+-----+                                       |
 9VZ ---+     +-----------------+                     |
 9UR ---------+                 |                     |
 9IR -----+---+                 +---------------------+
*9KS -----+   +-----+           |
 9TT ---------+     +-----------+
*9TW ---------------+
```

XI.4 Discriminant analysis

Two discriminant models were developed in chapter VII: model I included the net assets ratio (NARY), inflation (DCPI), interest/debt service (INPS), GDP growth rate (DGDP) and the investment ratio (INVR), while in model II the last two of those variables were replaced by the Asia dummy (AS). The discriminant scores for model I for 1987-1989 are listed in Table XI.2, while Table XI.3 summarises the classifications for 1988-1990, based on those scores.

Over the three years, the misclassification rates of type I and type II are 11 and 17 per cent respectively. However, there is no sign of accuracy in classification falling seriously as time proceeds - although the type II error rate rises over the period, the more costly type I rate falls. However, according to Mensah (1984: p.391), in the context of prediction models for corporate insolvency:

> Overall...one may be better off to estimate the model as closely as possible to the prediction period and hope that no fundamental changes occur in the economic environment to cause a structural shift.

This point is likely to apply also to the multivariate models of this thesis.

Table XI.2: Discriminant scores: model I

	1987		1988		1989	
TW	9.85		7.24		7.31	
KS	6.67		6.25		5.74	
CN	5.91		5.04		3.77	
IN	4.54		4.85		4.51	
IR	4.63		2.04		-0.04	T2
TH	4.07		4.65		4.54	
AL	3.96		3.60		2.45	
MA	3.02		4.50		5.00	
CM	2.81	T1	1.56	T1	1.85	T1
PK	2.38		2.18		1.63	
TT	1.55	T1	-0.81		-0.66	
TN	1.43		1.23		0.93	
PP	1.41		2.44		2.34	
ZI	1.32		2.77		1.58	
ES	1.16		1.78		1.19	
CB	1.16		1.18		1.58	
PG	1.16	T1	2.02	T1	-0.41	
ID	0.92		1.90		2.38	
GU	0.85	T1	2.72	T1	0.99	T1
BD	0.42		0.70		0.70	
SR	0.30		-0.23	T2	-1.08	T2
KN	0.13		0.55		0.19	
GA	-0.03		1.46	T1	-0.18	
TK	-0.28	T2	-2.08	T2	-1.60	T2
SY	-0.80	T2	1.52		-0.26	T2
HD	-0.67		-0.82		-1.21	
JD	-0.76		-3.21		-7.52	
VZ	-0.84		0.16	T1	-8.67	
DR	-0.90		-4.47		-4.14	
SE	-1.15		-0.88		-1.85	
PH	-1.23		-1.23		-1.29	
GH	-1.33		-0.13		-1.59	
YU	-2.31		-2.14		-2.15	
SU	-2.69		-10.75		-5.51	
BR	-3.06		-3.69		-1.79	
MC	-3.26		-1.86		-2.53	

Note:
Misclassifications are marked T1 or T2 - based on actual status in 1988, 1989 and 1990 respectively. See Table XI.1.

.../

.../

Table XI.2: (continued)

	1987		1988		1989	
PE	-3.33		-6.11		-6.87	
BL	-3.37		-2.12		-1.87	
CG	-3.43		-3.28		-3.83	
UR	-3.45		-2.16		-4.04	
PN	-3.48		-8.64		-9.53	
CR	-3.74		-2.74		-2.03	
EC	-4.63		-4.88		-8.34	
NG	-4.69		-6.47		-8.29	
MW	-4.85		-3.67		-2.48	
JM	-5.02		-5.28		-6.22	
IV	-5.34		-6.28		-6.61	
CH	-5.46	T2	-2.30	T2	-6.70	T2
MX	-5.88		-3.79		-0.96	
AR	-7.34		-7.74		-10.91	
EG	-7.64		-8.27		-8.30	
LB	-8.11		-7.82		-8.26	
ZB	-8.73		-9.86		-10.91	
NC	-8.90		-9.43		-9.52	
ZA	-10.59		-10.29		-10.50	

The misclassifications require explanation. Two lines of inquiry may be followed here: the values of the discriminating variables may be examined, and additional information may be utilised.

Within-group sample statistics for the discriminating varibles were computed for each year 1987 to 1989. The sample values for each misclassified case were compared with the appropriate mean and standard deviation, but no evidence of a pattern could be found. There was no evidence, for example, of type I errors consistently having particularly extreme values of particular discriminating variables.

Table XI.3: Discriminant models: classifications

Actual category		Classifications: Arrears	Classifications: Non-arrears
		1988: model I	
Arrears	34	30 (88%)	4 (142)
Non-arrears	21	3 (14%)	18 (86%)
		1989: model I	
Arrears	34	29 (85%)	5 (15%)
Non-arrears	21	3 (14%)	18 (86%)
		1990: model I	
Arrears	34	32 (94%)	2 (6%)
Non-arrears	21	5 (24%)	16 (76%)
		1988 to 1990: model I	
Arrears	102	91 (89%)	11 (11%)
Non-arrears	63	11 (17%)	52 (83%)
	165	102	63
		1988 to 1990: model II	
Arrears	102	94 (92%)	8 (8%)
Non-arrears	63	13 (21%)	50 (79%)
Totals	165	107	58

The second approach is more fruitful, but also suggests a note of warning against the mechanical use of multivariate models. If the individual error cases are examined, it is clear that these countries are mostly going through a sharp transition in their creditworthiness.

The type I errors are:

Cameroon	(1988,1989,1990)	Trinidad	(1988)
Guatemala	(1988,1989,1990)	Gabon	(1989)
Paraguay	(1988,1989)	Venezuela	(1989)

As late as 1988, the EIU (1988a: p.17) referred to Cameroon as 'a respectable debtor', and the arrears that emerged in 1987 appear to have been essentially delayed payments to overseas suppliers. Cameroon's first ever restructuring was agreed in 1989. Guatemala first moved into arrears in 1986, when repayments were postponed (EIU, 1989c: p.26). Paraguay first developed arrears in 1987 (EIU, 1990c: p.45). Trinidad moved very quickly from relative creditworthiness, to a position where it had to restructure its bank debt in 1988. Only Gabon and Venezuela have histories of debt problems earlier than 1986.

The type II errors are:

Chile	(1988,1989,1990)	Syria	(1988,1990)
Turkey	(1988,1989,1990)	Sri Lanka	(1989,1990)
		Iran	(1990)

Chile and Turkey have histories of debt-servicing problems, and both are recovering creditworthiness. Both economies have been booming, and Chile has been systematically reducing its bank debt through debt-equity swaps. However, in certain respects the economic profiles of both countries remain similar to those of problem debtors. Notably, Turkey has a high inflation rate while Chile has a low net assets ratio. Sri Lanka likewise has a low level of net assets, which is mitigated by its high proportion of concessional debt: in 1989, bank debt accounted for only 9 per cent of the total, while in the previous year the average interest rate and maturity on new loans were 1.9 per cent and 35 years respectively (EIU, 1990d: p.33). The correct classification of Syria is a fine point: it is a net creditor to the commercial banks, but arrears to the World Bank have developed since the mid-1980s (EIU, 1989d: p.19). Finally Iran's score is marginally below the cutoff in 1989, for 1990 classification. Iran has a policy of paying off its external debts, and it has avoided new formal borrowings since the early 1980s. However, at the end of 1989 it was unable to meet certain letter-of-credit commitments on trade deals (EIU, 1990b: p.20).

A similar analysis was carried out for discriminant model II. It showed no marked superiority in classification. The differences were:

1988: Trinidad was correctly classified. Kenya was an additional type II error. Overall, there was one fewer type I error, one more type II.

1989: Gabon and Venezuela were correctly classified. Sri Lanka was correctly classified, while Syria and Kenya were additional type II errors. Overall, there were two fewer type I errors, and one more of type II.

1990: The two type I errors were the same as for model I. Sri Lanka was correctly classified, while Kenya was a type II error. Overall, the error rates were the same as those of model I.

Over the three years, model II yields 8 type I errors (8 per cent), and 13 of type II (21 per cent). For any differential cost of a type I error exceeding unity, model II yields a lower misclassification cost. However, model I has a firmer statistical basis: it does not contain a binary variable, and while multivariate normality was rejected for model II, it could not be rejected for model I.

Both models were poor predictors of the rapid deterioration of certain formerly sound countries, including Cameroon and Guatemala (both models), and Trinidad (model I only). This exemplifies the difficulty that 'turning points' frequently pose for econometric models.

XI.5 Logit analysis

In this section, the logit model of chapter VIII is tested on 1987-1989 data. The estimated probabilities are set out in Table XI.4, and using a cutoff probability of 31.075 per cent, the model yields the classifications for 1988-1990 that are summarised in Table XI.5.

The classifications yielded by the logit and discriminant models are very similar. Over the three years, the logit model has the same type I error rate as discriminant model I, although the within-year rates differ slightly. In 1989 and 1990, Ghana is a type I error for the logit model. Ghana's arrears date from 1983; the EIU (1990a: p.12) expects these to be cleared by the end of 1990, and describes Ghana as 'a model with the western financial community' (1989b: p.5). Thus Ghana is a marginal case.

In each year, the logit model yields no fewer type II errors than discriminant model I, with an additional three (Sri Lanka 1988, Kenya 1988,1990) over the entire period. Both models classify Syria correctly in 1989 only, because of higher economic growth and lower inflation, compared with 1988 and 1990.

Table XI.4: Logit model: predicted probabilities and misclassifications

	1987		1988		1989	
TW	0.000		0.010		0.012	
KS	0.046		0.071		0.186	
CN	0.126		0.326		1.878	
IR	0.363		9.525			
IN	0.468		0.229		0.396	
TH	0.643		0.248		0.214	
AL	2.389		2.876		8.018	
MA	2.417		0.415		0.262	
PK	3.256		2.973		6.117	
CM	5.529	T1	31.763	T1	18.141	T2
TN	10.604		16.866		17.918	
PP	10.775		4.885		8.230	
GU	11.866	T1	1.898	T1	11.959	T1
CB	12.006		15.173		10.912	
ES	12.022		7.616		15.875	
ZI	13.770		1.613		7.604	
BD	14.738		14.080		15.021	
PG	19.753	T1	6.328	T1		
ID	23.282		8.475		5.649	
TT	24.055	T1				
			GH 27.295	T1	23.877	T1
			KN 24.907			
			SY 4.888			
Cutoff probability = 31.075%						
					IR 40.920	T2
					PG 51.414	
			TT 65.615		59.303	
KN	34.764	T2			36.925	T2
HD	40.830		45.071		55.393	
SR	42.757	T2	55.549	T2	75.732	T2
SE	46.779		41.307		76.028	
TK	54.568	T2	94.616	T2	91.376	T2
DR	61.983		99.494		99.256	
GH	64.255					
VZ	65.887		38.008		99.996	
PH	66.438		54.114		51.943	
JD	72.217		98.936		99.976	
SY	89.702	T2			50.545	T2

.../

.../

Table XI.4: (continued)

	1987		1988		1989	
GA	90.427		35.489		70.017	
SU	90.802		99.999		99.729	
UR	93.510		87.038		97.901	
PE	94.663		99.938		99.963	
YU	95.715		96.246		95.481	
PN	95.927		99.989		99.996	
MC	96.652		75.135		88.124	
CR	97.003		92.910		83.337	
BL	97.205		90.699		87.570	
BR	97.248		98.918		88.590	
NG	98.404		99.681		99.971	
MW	99.203		97.088		85.531	
CH	99.330	T2	81.080	T2	36.404	T2
JM	99.429		99.746		99.867	
IV	99.557		99.908		99.916	
CG	99.681		99.579		99.615	
EC	99.760		99.159		99.994	
MX	99.878		98.994		65.601	
AR	99.954		99.980		100.000	
EG	99.971		99.982		99.986	
LB	99.972		99.966		99.979	
ZB	99.988		99.997		99.999	
NC	99.993		99.998		99.997	
ZA	99.999		99.999		99.999	

Note: T1 and T2 errors are based on actual status in 1988, 1989 and 1990.

In each year, model II outperforms the logit model, and for the entire period it has three fewer errors of type I, and one fewer of type II.

The conclusion that may be drawn from all this is that despite the theoretical superiority of the logit model over the discriminant models, in practice the discriminant approach yields somewhat better out-of-sample results.

Table XI.5: Classifications by logit model

Actual category		Classifications: Arrears	Classifications: Non-arrears
		1988	
Arrears	34	30 (88%)	4 (12%)
Non-arrears	21	5 (23%)	16 (77%)
		1989	
Arrears	34	30 (88%)	4 (12%)
Non-arrears	21	3 (14%)	18 (86%)
		1990	
Arrears	34	31 (91%)	3 (9%)
Non-arrears	21	6 (29%)	15 (71%)
		1988 to 1990	
Arrears	102	91 (89%)	11 (11%)
Non-arrears	63	14 (22%)	49 (78%)
Total	165	105	60

Note:
Cutoff probability = 31.0755 per cent

XI.6 AID analysis

The variables that were included in the AID model are:

NARY net assets/GDP
INVR fixed investment/GDP
INPS interest payments/debt service
DCPI inflation

Given the values of these variables, each country-year case in 1987, 1988 and 1989 is allocated to one of the six groups derived in the AID analysis of chapter IX, and these groupings are used to classify the countries in 1988, 1989 and 1990. For the purpose of classification, the 'non-arrears' group is taken to include terminal groups 7 (high NARY, low DCPI) and 5 (low NARY, high INVR), and the other four groups cover the 'arrears' classification. The classifications are summarised in Table XI.6.

The very high error rates, particularly of type II, are immediately apparent, and bear out the conclusions of chapter IX that the logit and discriminant models are superior to AID for classificatory purposes.

Table XI.6: AID model: classification

Actual group		Classifications: Arrears (groups 8,10, 6,11)	Classifications: Non-arrears (groups 5,7)
		1988	
Arrears	34	28	6 (18%)
Non-arrears	21	13 (62%)	8
		1989	
Arrears	34	28	6 (18%)
Non-arrears	21	12 (57%)	9
		1990	
Arrears	34	30	4 (12%)
Non-arrears	21	11 (52%)	10
		1988 to 1990	
Arrears	102	86	16 (16%)
Non-arrears	63	36 (57%)	27
	165	122	43

XI.7 Tests of classificatory power

In this section, the classificatory powers of the multivariate models are compared.

Table XI.7: Discriminant and logit models: summary of classifications, 1988-90

Actual category		Classifications: Arrears	Non-arrears
		Discriminant model I	
Arrears	102	91 (89%)	11 (11%)
Non-arrears	63	11 (17%)	52 (83%)
	165	102	63
		Discriminant model II	
Arrears	102	94 (92%)	8 (8%)
Non-arrears	63	13 (21%)	50 (79%)
Totals	165	107	58
		Logit model	
Arrears	102	91 (89%)	11 (11%)
Non-arrears	63	14 (22%)	49 (78%)
Total	165	105	60

Given the very high error rate of the AID model, the primary concern is with the logit and discriminant models. Their classifications for 1988 to 1990 are summarised in Table XI.7.

Simple nonparametric statistical tests are available to test the hypothesis that the multivariate models have predictive power. The null hypothesis is that the predictions of the models are no better than random.

First, the computed values of the statistic for the FMM test (see chapter VII and Frank et al, 1965) are 8.7 for discriminant model I, 8.9 for model II, and 8.2 for the logit model. The critical value of t(0.025, 165) is 1.96, so the null is rejected in each case. For the AID predictions (Table XI.6), the value of the test statistic is 3.8, so the null is rejected for the AID model also, although its performance is much worse than those of the other models.

The second nonparametric test is the runs test (Rohatgi, 1984): it is applicable only to the logit and discriminant models, but it has the advantage of not requiring a cutoff score or probability. To carry out this test, the cases are listed, for any given year, in numerical order of the discriminant scores or logit probabilities. The number of 'runs' of arrears and non-arrears cases are then counted, where the score or probability of a country at date t is associated with its debt-servicing status at t+1. The null hypothesis is that the cases are randomly distributed, and it is rejected if the observed number of runs falls outside a range determined by two critical values. The expression for the test statistic is given in Rohatgi; with more than 20 cases, a normal approximation to the theoretical distribution is used; thus the critical values at the five per cent level are -1.96 and +1.96. The results of the runs test are set out in Table XI.8.

Table XI.8: Runs test: logit and discriminant models

[Z is the test statistic]		1987	1988	1989
Discriminant model I	No. of runs	16	16	16
	Z	-3.24	-3.24	-3.24
Discriminant model II	No. of runs	14	14	14
	Z	-3.81	-3.81	-4.38
Logit model	No. of runs	18	12	10
	Z	-2.67	-4.38	-4.94

In all cases, the null hypothesis of randomness is rejected at the five per cent level. Specifically, since the value of Z is below the critical value in the lower tail, the evidence is that the the models are sorting the cases into arrears and non-arrears clusters, as required.

Finally, we may make direct pairwise comparisons of the models, using a chi-square test (Conover, 1980).

Given the predicted grouping of a given case by each of two multivariate models, the case may be placed in one of four categories:

Both arrears	[A,A]	Arrears/non-arrears	[A,N]
Non-arrears/arrears	[N,A]	Both non-arrears	[N,N]

The partition of the sample derived from this grouping defines a 2x2 contingency table. The cell entries are the numbers of cases in each of the four groups, and given the row and column totals, the expected cell entries may be computed. The test addresses the null hypothesis that there is no significant difference between the observed and expected entries, and for each possible pair of multivariate models, the value of chi-squared leads to rejection of the null at the five percent level, for each possible pairing. For example, pairing the logit model with discriminant model I, the value of chi-squared is 136.63, compaired with the critical value of chi-squared (0.05,1) of 3.84.

Thus at the five percent level, we reject the hypothesis that there is no statistically significant relationship between the classification of cases by these two models, and the same conclusion arises for the other two pairwise comparisons.

XI.8 Conclusion

Tests of statistical models on an out-of-sample period are true unbiased *ex ante* tests of predictive ability. This method of testing models is a key measure of their performance, irrespective of their statistical underpinnings. The objective of this chapter has been to assess the classificatory performance of the multivariate models, derived from 1979-1986 data, in this fashion. In later chapters these models are used as criteria for assessing banker judgement, and to this end it is clearly necessary to be sure that these models do in fact have predictive ability. However, in section XI.4 it was found that most of the cases misclassified by discriminant model I were going through a sharp transition in creditworthiness. Thus the models may have difficulty in identifying turning points, and should not be used in a mechanical manner.

The discriminant and logit models were tested on data from the holdout years, 1987-1989. Satisfactory results were obtained. The out-of-sample misclassification rates were not unduly high: moreover they were similar to the misclassification rates within-sample, and there was a considerable overlap between the three models, in terms of the misclassified cases. Despite the better consistency of the logit model with underlying statistical assumptions, its performance is worse than those of the discriminant models.

Nonparametric tests have been applied to the hypothesis that the classificatory powers of the models are no better than random. At the five percent level, the hypothesis is rejected for all three models by both tests. In addition, there is no evidence of degradation of the predictive power as time extends away from the estimation period. Finally, it was concluded that the classifications by the three models are statistically indistinguishable.

The results of the cluster and AID analyses were less impressive, but not out of line with what the earlier chapters had promised. These results suggest that these methods are inappropriately directed at what is essentially a binary task.

Appendix XI

Table XIA.1: Debt servicing status

Country	ARS(t+1) 1987	1988	1989	Country	ARS(t+1) 1987	1988	1989
AL Algeria	9	9	9	AR Argentina	0	1	2
BD Bangladesh	9	9	9	BL Bolivia	1	0	1
BR Brazil	0	0	1	CB Colombia	9	9	9
CG Congo	1	2	3	CH Chile	0	1	2
CM Cameroon	9	1	0	CN China	9	9	9
CR Costa Rica	2	1	0	DR Dom. Rep.	1	0	1
EC Ecuador	0	0	1	EG Egypt	1	2	3
ES El Salv.	9	9	9	GA Gabon	1	0	0
GH Ghana	9	9	9	GU Guatemala	1	1	0
HD Honduras	0	1	2	ID Indonesia	9	9	9
IN India	9	9	9	IR Iran	9	9	9
IV Ivory Cst.	0	1	2	JD Jordan	2	1	0
JM Jamaica	0	0	1	KN Kenya	9	9	9
KS S. Korea	9	9	9	LB Liberia	3	4	5
MA Malasia	9	9	9	MC Morocco	0	0	1
MW Malawi	0	1	2	MX Mexico	1	0	0
NC Nicaragua	1	0	1	NG Nigeria	0	0	0
PE Peru	3	4	5	PG Paraguay	9	1	0
PH Phil'nes	0	1	0	PK Pakistan	9	9	9
PN Panama	0	1	2	PP Papua N.G.	9	9	9
SE Senegal	0	0	1	SR Sri Lanka	9	9	9
SU Sudan	1	0	1	SY Syria	9	9	9
TH Thailand	9	9	9	TK Turkey	9	9	9
TN Tunisia	9	9	9	TT Trinidad	1	0	0
TW Taiwan	9	9	9	UR Uruguay	2	3	4
VZ Venezuela	0	1	2	YU Yugoslavia	0	0	1
ZA Zaire	0	1	0	ZB Zambia	2	3	4
ZI Zimbabwe	9	9	9				

Note: Header second row reads "t+1 = 1987 1988 1989" for each half.

<u>Note</u>:
Categorical variable ARS(t+1) coded 0 (arrears cases), 1 to 5 (weak-year cases; code indicates distance from arrears), 9 (non-arrears).

XII Banker and expert judgement: paramorphic representations

XII.1 Introduction

In this chapter, selected country risk rating systems are utilised as indicators of banker and expert judgement. The primary purpose here is to discover the objective factors, if any, upon which these judgements are based. This problem is tackled using multiple regression, and also the AID technique. Research on human information processing, which is surveyed in chapter II, suggests that bootstrapped models of human judgement can outperform the human judges on which they are based, in a wide variety of task settings. The models of human judgement that are developed in this chapter are used in chapter XIII to test this hypothesis in the context of country risk assessment. Chapter XIII is thus concerned with whether human judgement is applied consistently. Chapter XIV asks whether human judges can be outperformed by multivariate statistical models, specifically those developed in chapters VI to IX. The question is answered by comparing classifications of creditworthiness by the multivariate models with those made by human judges (as reflected by rating systems) and their bootstrapped models.

This chapter has several subsidiary objectives. One is to analyse how banker judgement has changed over time in response to new information: this is relevant to the judgemental bias issue. Correlation analyses and t-tests are used for this purpose. A second subsidiary objective is to determine the extent to which perceptions of political and economic risk are correlated. Finally, the extent to which the various ratings are correlated will be explored.

Section XII.2 introduces the cognitive framework for this group of chapters. Section XII.3 introduces and describes the rating systems. For the purposes of this thesis, the most important is that of the Institutional Investor: it alone is predominantly based on judgement, and it is available for a much longer time-series than the other systems. Consequently, section XII.4 is devoted exclusively to analysing it. The first concern is to derive a paramorphic representation of banker judgement (Hoffman, 1960), as reflected by the rating system, and the section ends with an analysis of the subsidiary question of how the rating pattern has changed over time. All the other rating systems are grouped for analysis in section XII.5, and a summary and conclusions are set out in section XII.6.

XII.2 The cognitive framework

Figure XII.1: Brunswik's Lens

```
                                  *
                             .         .
                       .        * *        .
                   .          * C *            .
Event  .                     *  U  *      The       . Pred-
to be  .  Environment *  E  *      judge     .iction
pred-         .              *  S  *          .        of
icted             .           *   *      .           judge
                             .  * *  .
                                  *
                                Lens
```

The cognitive framework for this chapter and chapters XIII and XIV is Brunswik's (1952) lens model, which is depicted in Figure XII.1. The decision taker receives multiple and overlapping cues, these being financial and economic variables in the present context. These cues must be combined and read so as to obtain a probabilistic estimate of the event.

Assessments by bankers (or other 'experts') of the creditworthiness of debtors may be represented by the right hand side of the lens diagram, and an aggregation of such assessments, such as the <u>Institutional Investor</u> rating system, may also be so represented. In the latter case, the 'prediction of the judge' is the value of the rating. In the context of sovereign risk, the 'event' on the left hand side is the extent to which debt service commitments have been honoured.

In chapter II, the literature on judgemental biases was reviewed. The conclusion reached there, based on the authorities cited, was that a paramorphic representation of human judgement will be at least as good as human judgement, provided that the model is correctly specified. Thus, a correctly specified model of the judge's decisions will be at least as good as the judge's decisions. Much of the research in this area utilises linear modelling techniques: an implicit assumption of these is that there are no environmental nonlinearities in the relationship between cues and criterion variable, that the human judge takes account of.

If the judge is subject to heuristic biases, then the model will incorporate these, but it will at least avoid the lack of consistency of the human judge. These findings provide the motivation for modelling the judge in this chapter, in order to allow comparisons to be made in chapter XIII of the predictions of judges with those of their bootstrapped models.

In this chapter, ratings of creditworthiness will be modelled using linear regression and AID. These models do not seek to optimise the relationship between cues and the criterion variable, which in this context is the debt-servicing record. Rather, they are paramorphic representations of human judgement. This modelling procedure is known as 'bootstrapping', where the term has a completely different meaning from its usage in the statistics literature (see chapter XI).

The approach adopted here is in the tradition of the quantitative investigations of the ratings of commercial and municipal bonds by rating agencies - see Kaplan and Urwitz (1979) and Foster (1986). Despite the 'robust beauty of improper linear models' (Dawes, 1979), the analysis will not be limited to regression. Human thought processes are essentially interactive, so to allow for the possibility of interaction effects, AID will be used also.

The multivariate models of chapters VII to IX belong on the left hand side of the lens model: they map between the cues and the environmental event. In chapter XIV these are compared with the judges' ratings and their bootstrapped models, thus yielding comparisons between human judgement, paramorphic representations of human judgement, and multivariate statistical models.

XII.3 Measures of country risk: banker and expert judgement

XII.3.1 Risk assessment within banks

An extensive literature exists on the methods of country risk assessment that are used by banks. Bird (1986) and Krayenbuehl (1988) discusses the general issues. Goodman (1977), Burton and Inoue (1983), Mascarenhas and Sand (1985), Bird (1986), DeWitt and Madura (1986), Heffernan (1986), and Nisse (1987) present the results of surveys of banking practice. Merrill (1982), Turk (1985), Senkiw and Johannson (1986), and Pietrabassa (1987) discuss risk assessment pactice in specific banks. Early material of each type is collected in Goodman (1978) and de Saint Phalle (1978).

These authors find that banks differ widely in the organisational structure within which country risk assessment is undertaken, in the extent to which quantitative techniques are used, and in the level of technical sophistication that is applied.

For comparing multivariate models with banker judgement, a quantitative measure of the latter is required; moreover it must be a consensus or aggregate measure. There are two directions from which this may be approached. First, there exist published risk ratings of countries that may arguably be taken to represent bankers' assessments of country risk. Alternatively, there are the assessments of specialised rating agencies, which are used by banks as an input to the banks' own risk assessment.

XII.3.2 Published indicators of country risk: the Institutional Investor rating system

The Institutional Investor country credit rating system is based on assessments of countries' creditworthiness that are drawn from between 75 and 100 banks, and expressed on a scale of 0 to 100. The individual responses are aggregated, using an undisclosed formula that weights banks according to their degree of exposure and the sophistication of their analysis. This rating system has its problems: Taffler and Abassi (1984: p.558) consider that

> the Institutional Investor rating system might actually be measuring something else than banker views on a country's creditworthiness, viz.: "banking sentiment" about a particular country which may be related to other factors than likelihood of rescheduling. The high rating given to Latin American countries in many cases, for example, may reflect more familiarity through geographical and ethnocentric association with the United States, and the large sums already lent, rather than their intrinsic creditworthiness per se.

These misgivings, while valid, are concerned with the rating system's accuracy as an objective indicator of risk, rather than as an indicator of banker judgement per se. However, the system is open to two other criticisms.

First, respondents to the survey may complete it carelessly: this could be a source of random error. A more serious problem would arise if a significant number of heavily exposed lenders deliberately over-rated their debtors, in the hope that other banks might be influenced in their lending decisions by a high published rating. This could lead to systematic errors. These potential difficulties with the Institutional Investor rating system provide a motive for using other measures of banker and expert judgement.

There are at least two other possible measures of banker judgement, apart from the products of specialised rating agencies. First, the Euromoney country risk league table ranks countries that have borrowed during the year in the Eurocurrency markets. Originally, it was based on weighted average spread over LIBOR. However, since then the weighting system has changed, and additional factors have been included in the index. Thus the Euromoney indexes for different years are not strictly comparable.

Apart from this difficulty, Taffler and Abassi (1984) have identified reasons against accepting weighted average spread as necessarily being a true indicator of default risk. Moreover, Euromarket credit is only one source of borrowing. For those countries that rarely or never use the Euromarkets, the Euromoney index will be of little significance.

Secondly, in recent years, a secondary market has developed in LDC bank debt. The discounts in these markets could arguably be taken as reflecting the consensus market view on the riskiness of the borrower. However, many countries never feature in these markets. Of those that do, the debt of many is traded infrequently, and most of the turnover involves the debt of five countries: Argentina, Brazil, Chile, Mexico and Venezuela. Thus secondary market discounts are of limited use as risk indicators.

XII.3.3 The commercial rating agencies

Commercially produced risk estimates do not directly reflect banker judgement. However, they are relevant here in that banks are a major market for the suppliers of these services, and presumably banks would not continue to buy a consistently erroneous forecast. Thus those that survive may arguably be regarded as a component of banker judgement. Moreover, these services may be taken to represent expert judgement.

(i) The **BERI** index

The oldest of these services is the BERI Forelend index, produced since the late 1960s by BERI SA of New York. The following description of the index draws heavily on Krayenbuehl (1988).

The index is a weighted average of three components. The quantitative component, with a weighting of 50 per cent, reflects quantifiable measures of debt-servicing capacity. The qualitative component, with a weighting of 25 per cent, is concerned with aspects of debt-servicing capacity that are not directly measurable, such as the competence of economic management. The final 25 per cent is accounted for by an environment component, that reflects the political and social environment. This component is assessed internally by BERI, while the other two are determined through the panel method, and according to de la Torre and Neckar (1988), Delphi methods are used.

Values of the index for 1986 for the following 31 less developed countries were obtained from

Krayenbuehl:

Algeria	Egypt	Malaysia	Philippines
Argentina	India	Mexico	Taiwan
Bolivia	Indonesia	Morocco	Thailand
Brazil	Iran	Nigeria	Turkey
Cameroon	Ivory Coast	Pakistan	Venezuela
Chile	Jamaica	Panama	Zaire
Colombia	Kenya	Paraguay	Zimbabwe
Ecuador	Korea	Peru	

Unfortunately, it has not been possible to obtain more recent data.

(ii) <u>The</u> **International Country Risk Guide** <u>system</u>

The International Country Risk Guide has produced ratings of more than 130 countries, on a scale 0 to 100, for a number of years, and it has been made available to the author since 1986. The ICRG includes all 55 countries that are analysed in this thesis. The monthly ICRG is published by International Reports Inc. of New York, who also publish a very detailed Key to the guide. The latter source describes the construction of the rating system, which includes a political risk rating (with a weight of 50 per cent in the composite rating), and financial and economic risk ratings (each with a 25 per cent weight). The information is drawn both from published sources and from private contacts, and it is analysed by outside consultants and also in-house. The various dimensions of political, financial and economic risk are scored, and transformed into the three sub-ratings using pre-determined weights, and finally the composite rating is derived.

(iii) The EIU Credit Risk Rating Scores

A newcomer is the Economist Intelligence Unit, which has been producing Credit Risk Rating Scores for the 55 countries that are covered by this thesis since April 1989. The overall rating of a country is a composite of economic and financial factors, with a weighting of 55 per cent, structural and economic policy factors (25 per cent), and political and strategic factors (20 per cent). The first component is based directly on seven financial and economic ratios, while the other two involve judgemental elements.

Two of the financial and economic ratios are debt service/exports and interest/exports. For 1979-1986, these ratios are both quite heavily loaded on the first principal component, with loadings of 0.53 and 0.52 respectively. Moreover, their squared correlation coefficient for the same data is 0.78. This suggests that one of them is redundant.

XII.3.4 Summary

Other rating agencies exist, but their products are not relevant. Frost and Sullivan formerly produced an overall rating of countries, but this service was altered some years ago to a grading of countries, within broad bands, along various dimensions of political and economic risk over 18-month and 5-year time horizons. This service, which is now published by Political Risk Services of Syracuse, N.Y., currently covers 85 countries. It includes members of the OECD, the former Warsaw Pact countries, and certain less developed countries, and the listing in January 1990 included 45 of the 55 members of the data-set that is used in this thesis. The ten absent countries were Bangladesh, Congo, Ghana, Jordan, Liberia, Malawi, Paraguay, Papua New Guinea, Senegal, and Trinidad (Political Risk Services, 1990).

Standard and Poor's country ratings are similar in concept to those of Frost and Sullivan/Political Risk Services, although the coverage is narrower, and both these products resemble bond ratings. Standard and Poor's covers 34 countries: essentially the OECD area, plus the following eight LDCs: China, India Malaysia, South Korea, Taiwan, Thailand, Turkey and Venezuela (Standard and Poor's Ratings Group, 1990). Moody's Investors Services produces a rating service for OECD countries only.

It is clear from the description of the various indices that the Institutional Investor rating system is the best one to take to be an indicator of banker judgement: both because of its design and construction, and because of its availability since 1978. It is therefore necessary to examine it further at some length, as statistical analysis of the rating system may reveal how bankers actually form their credit judgements. This is done in section XII.4, after which the other rating systems are then dealt with more briefly in section XII.5.

XII.4 Analysis of the Institutional Investor

XII.4.1 Introduction

The Institutional Investor ratings were obtained from the September issues of the journal for the years 1979 to 1989. Ratings are published in March also. However, the rating system is to be assessed for predictive ability, and it was therefore appropriate to use the September value, it being closer to the year-end.

The rating for each country-year case is a real number between zero and 100: the higher the rating, the more favourable the aggregate evaluation of creditworthiness.

Paramorphic representations of banker judgement are obtained from the Institutional Investor rating system using regression analysis in sections XII.4.2 and XII.4.3, and AID in section XII.4.4. Finally, in section XII.4.5 a correlation and t-test analysis is used to explore the changes in the rating over time.

XII.4.2 Regression analysis

The data set for the regression analysis consisted of the 70 financial and economic variables augmented by categorical variables, along with the Institutional Investor rating (which is coded II below). The categorical variables included those listed in Table IV.6, along with others constructed from the environmental variables listed in Table IV.4, and one based on membership of cluster I (see chapter V). In order to provide for out-of-sample testing, the estimating sample covered the years 1979-1986. By holding out 1987-1989, the analysis here is consistent with the procedure of chapters VI to IX. After eliminating unrated country-year cases from the sample, 414 remained of the original 440.

In chapter IV, it was argued that the data set includes all major indicators of creditworthiness. That being so, it is reasonable to use the same data for building a paramorphic representation of banker judgement.

Particular problems arise in the application of regression analysis to panel data. One possibility is that properties of the data are such that the least-squares regression disturbances will be cross-sectionally heteroscedastic and timewise autocorrelated. In that case, the least-squares parameter estimates would be unbiased and consistent, but not efficient. In principle the data may be transformed to eliminate the problem (Kmenta, 1971). The time-series in the data set here consists of only eight annual observations, so the transformation is not practicable; moreover, this procedure would be in contravention of the model-building methodology that is proposed in Gilbert (1986). In essence, 'problems' such as autocorrelation or heteroscedasticity arise in incorrectly specified models, and the appropriate remedy is to re-specify. Maddala (1977) surveys some of the alternative specifications that have been proposed for models estimated on panel data, including the fixed-effects and variance components models, and the seemingly unrelated regression model. Since no prior information is available on the appropriateness of any particular alternative specification, the classical linear specification will be retained.

An interactive stepwise procedure was applied to the data, and the best fitting regression equation is set out in Table XII.1. The stepwise procedure was similar to that used in the development of the discriminant and logit models of chapters VII and VIII.

Table XII.1: Institutional Investor rating: regression analysis (1), 1979-1986

Estimated coefficients		t-values (409 d.f.)	N=414
II= 15.573		(2.7)	
- 0.072	GDPX	(-12.2)	(public debt/exports)
+ 6.625*	RTD	(12.3)	(real total debt)
+ 0.868	INVR	(9.9)	(fixed investment/GDP)
+ 1.939	MCOV	(7.4)	(import cover)

$* = x10^{-4}$ F(4,409) = 214.50

R-squared = 0.68. Adjusted R-squared = 0.67

The variables are shown in order of entry. All the estimated coefficients, and the F-statistic, are significant at the 5 per cent level. None of the categorical variables in the data set has significant explanatory power.

The sample period was a time of rapid change in the market for loans to LDCs. In particular bankers should have experienced a rapid learning process during the early 1980s, and then have re-assessed their methods of country appraisal. Therefore, it is likely that the regression equation does not have constant coefficients over the sample period. This hypothesis has been explored using the Chow test (Chow, 1960). Splitting successively after each of the years 1979 to 1985, pairs of regressions were run, specified as above, on partitions of the sample.

The null hypothesis of no significant difference between the sets of coefficients estimated on each of two sub-samples was tested at the 5 per cent level, and rejected for the following pairs of sub-samples:

1979,	1980-1986	F(5,404)=4.21
1979-1980,	1981-1986	F(5,404)=6.37
1979-1981,	1982-1986	F(5,404)=6.71
1979-1982,	1983-1986	F(5,404)=6.39
1979-1983,	1984-1986	F(5,404)=3.82

[The critical value, F(5,404,0.05), is 2.21.]

However, for the remaining two pairs, the null could not be rejected:

1979-1984,	1985-1986	F(5,404)=1.55
1979-1985,	1986	F(5,404)=0.61

Samples with a heavy weighting of early years produced estimated regressions that differed significantly from those estimated on the balance of the total sample. The pattern of F-values suggests that the key date is 1981, and further applications of the Chow test confirmed this: when the above analysis was repeated over the years 1982-1986, splitting successively after 1982, 1983, 1984, and 1985, the null could not be rejected. At each successive split, the F-value was 2.03, 1.50, 1.01 and 0.69 respectively, compared with a tabulated F(5,260) of 2.21 at the 5 per cent level.

Re-estimation of the equation on each of the two sub-periods, 1979-1981 and 1982-1986, confirmed that it fits the data well in each period: all estimated coefficients, and the F-statistic, were significant at the 5 per cent level, and the value of adjusted R-squared was 0.67 in each case. Thus, the specification of independent variables is satisfactory for each sub-period. The Chow test tells us that the coefficients differ between the two sub-periods, without identifying precisely how. This matter is addressed by the Gujarati test. (Gujarati, 1970a,b). The equation was re-specified to include dummy variables relating to the intercept and also to all the slope coefficients. When a stepwise regression was applied to the four variables INVR, MCOV, RTD and GDPX plus the five dummies, the four listed variables plus the intercept dummy were selected, but the significance of the slope dummies was insufficient for them to enter. When forced into the equation, singly and all together, their t-statistics were all below 1.5. On this basis, the null cannot be rejected for the slope dummies. Thus, at the 5 per cent level we conclude that the difference between the two periods is entirely captured by the intercept dummy: that is, a shift in the regression equation.

The estimated regression equation, including the dummy variable DUM that takes the value 1 from 1982 and 0 earlier, is shown in Table XII.2.

Table XII.2: Institutional Investor rating: regression analysis (2), 1979-1986

	Estimated coefficients		t-values (409 d.f.)	N=414
II=	18.824		(6.9)	
	− 0.067	GDPX	(−11.4)	(public debt/exports)
	+ 6.750*	RTD	(12.9)	(real total debt)
	+ 0.831	INVR	(9.7)	(fixed investment/GDP)
	+ 1.946	MCOV	(7.7)	(import cover)
	− 5.253	DUM	(−4.9)	(=0: 1979-81, then =1)

$* = x10^{-4}$ F(4,409) = 186.0

R-squared = 0.70. Adjusted R-squared = 0.69

Correlation matrix for independent variables.

	GDPX	RTD	INVR	MCOV
RTD	0.03			
INVR	−0.50	0.10		
MCOV	−0.23	0.18	−0.02	
DUM	0.27	0.05	−0.21	−0.03

This procedure is superior to sub-period regressions, in that all coefficients are estimated on the full-sized sample.

XII.4.3 Interpretation of regression results

The regression results indicate that high ratings are associated with:

High levels of Fixed investment/GDP INVR
elasticity = 0.52;
Low levels of public debt/exports GDPX
elasticity = −0.31;
High absolute levels of real debt RTD
elasticity = 0.19;
High levels of import cover MCOV
elasticity = 0.15.

The elasticities are calculated at the sample means. They indicate the proportionate change in the Institutional Investor rating, given a unit proportionate change in the given variable, starting from the sample centroid. They indicate that, in absolute terms, the rating is most sensitive to changes in INVR, followed by GDPX, then RTD and finally MCOV.

The function shifted downwards after 1981, with the onset of the debt crisis: for any given vector of variable values, the perception of creditworthiness by bankers would be lower in the later period.

Four of these findings are uncontroversial: we would expect bankers to associate creditworthiness positively with import cover and with the share of fixed investment in GDP and negatively with the debt-export ratio, and to have become more sceptical after 1981. However some explanation is necessary for the conclusion that ratings of creditworthiness rise with absolute levels of real indebtedness (RTD). The implication of the estimated regression equation is that, ceteris paribus, an increase of $1000m (at 1980 prices) in a country's real debt will cause bankers so to revise their opinions that the country's Institutional Investor rating will rise by 0.675 points. In seeking to explain this, several possibilities arise.

First, this observation may be the result of hope on the part of bankers, rather than rational calculation.

Secondly, familiarity with the larger debtors may lead bankers to express relatively favourable opinions about them, and to be rather more cautious about small debtors, with whom they have less contact. Taffler and Abassi (1984) have raised this possibility (see section XII.3.2).

Thirdly, real debt is an indicator of the size of the economy. So also are other variables in the data set, including POPN (population) and RY (real GDP), with which RTD has sample correlation coefficients of 0.63 and 0.73 respectively. Moreover, all these variables are highly loaded on the seventh principal component of the data set (Table V.1). To explore this further, regressions were run with each of the alternative 'size' variables replacing RTD, with the other four variables retained. In both cases, all coefficients were significant at the 5 per cent level. However, the t-statistic on POPN and RY fell to 9.8 in each case, and the R-squared values fell to 0.65. Thus the specification including RTD is retained, and the conclusion is that this variable measures not only size, but other dimensions of banker judgement as well.

The estimated equation indicates that size of real debt must not be viewed in isolation: the presence of the public debt/export ratio with a negative coefficient suggests that a high real debt is acceptable to bankers only where servicing capacity is present, in terms of a high level of exports. The coefficient of correlation between the variable RTD and GDPX is 0.03 (see Table XII.2), which is very low. The possibility arises that there are interaction effects at work, and the next section explores this.

XII.4.4 AID analysis

For reasons that have been cited in chapter IX, Morgan and Sonquist (1963) believe that empirical investigations in the social sciences should allow for the possibility of interaction effects. This provides the motivation for applying AID to the Institutional Investor rating system.

The procedure adopted here is very similar to that of chapter IX: ARS is replaced by the rating as criterion variable, and the set of predictor variables and the parameters for the AID algorithm are all the same as before. The sample size is larger than in chapter IX, at 414 cases.

The results of the AID analysis are a little more complex than before, as the tree diagram (Figure XII.2) illustrates. Table XII.3 lists the terminal groups from Figure XII.2, in decreasing order of within-group mean rating.

The first split is at the median value of INVR (fixed investment/GDP): above the median, we get a group with a high average value for the rating. The next two parallel splits are on real GDP (RY) and real total debt (RTD). The next four parallel splits (into eight groups) are on

EFIR effective interest rate
STPD short term debt/debt
GDPX public debt/exports
XPM12 share of first two export markets

Of the eight groups thus formed, only two split again, in each case on MCOV (import cover).

Figure XII.2: Institutional Investor rating: tree diagram for AID model

```
* = group with             (1)              There are 26 cases
higher mean at            N=414             with missing value
split.                    INVR              of rating, so sam-
                         t=14.0             ple is 414 cases
N = group size        |         |
                      |         |
               0,1 |               | 2,3
                   |               |
                  (2)             (3)*
                N=203            N=211
                  RY              RTD
                t=9.0            t=9.0
                |    |           |    |
        0,1 |        | 2,3   0,1 |      | 2,3
            |        |           |      |
          (4)        |           |     (7)*
        N=115        |           |    N=118
        EFIR         |           |    XPM12
        t=6.8        |           |    t=6.7
        |   |        |           |     |    |
      0|    |1,2,3   |           |  0,1|    |2,3
       |    |        |           |     |    |
     (8)   (9)*      |           |   (14)  (15)*
     N=44  N=71      |           |   N=52  N=66
     ----  ----      |           |   MCOV  ----
  [II=11] [II=21]    |           |   t=5.0 [II=56]
                     |           |     |  |
                     |           |  0,1|  |2,3
                     |           |     |  |
                     |           |   (18)  (19)*
                     |           |   N=27  N=25
                     |           |   ----  ----
                     |           |   [37]  [51]
                     |           |
                   (5)*         (6)
                   N=88         N=93
                   STPD         GDPX
                   t=5.4        t=7.0
                   |   |        |   |
             0,1 |       | 2,3  2,3 |   | 0,1
                 |       |          |   |
               (10)    (11)*      (12)  (13)*
               N=39    N=49       N=23  N=70
               ----    ----       ----  MCOV
             [II=24]  [II=39]  [II=22]  t=5.5
                                         |   |
Numbers in [ ] below                   0 |   | 1,2,3
terminal groups  are                     |   |
within - group means                   (16)  (17)*
of II rating. Sample                   N=23  N=47
mean is 34.2.                          ----  ----
                                    [II=31] [II=43]
```

Table XII.3: Institutional Investor rating: terminal groups of AID analysis

Group	15	19	17	11	18	16	10	12	9	8
Mean rating	57.8	50.6	43.3	38.9	36.9	30.9	24.2	21.8	21.1	11.1

The variables INVR and RTD both feature in the regression analysis, as do GDPX and MCOV, all with statistically significant coefficients. In each case, the sign of the relationship between rating and variable is the same as in the regression model. INVR is the first variable to enter the AID model, and it had the largest elasticity of all the regressors.

The AID groups formed by the split on INVR then split again on RY and RTD, which have a correlation coefficient of 0.79, are both highly loaded on principal component 7 (chapter V), and clearly represent the same or similar dimensions of the data set. The implication is that the primary determinant of the magnitude of the rating is the share of investment in GDP, followed by the size of the economy, with a positive association in each case. For a given investment ratio, large economies are viewed more favourably than small ones. However, as in the regression model, a question mark hangs over the association of high rating values with high levels of real debt.

The splitting beyond groups 4,5,6 and 7 should be interpreted with caution. When group 4 splits on EFIR (effective interest rate), cases with a higher value for that variable go into group 9, which has a higher mean rating. A possible explanation for this is the positive association between bankers' ratings of a country, and the share of its debt that is negotiated on market (i.e. non-concessional) terms - given that the latter variable in turn is positively associated with the effective interest rate.

Such an explanation is not supported by the evidence in this case. Many of the cases in group 4 consist of repeated appearances of particular countries over a number of years. If the proposed explanation were valid, we should expect to find consistent allocation of such repetitions for a given country into one group only, after splitting on EFIR: this is because, for any country, concessional debt tends to be a stable proportion of the total, over time. In fact, such a consistent allocation does not occur. An alternative explanation may be that the value of EFIR is partly a reflection of the extent to which a country is current on its interest payments: this could yield the positive association with bankers' ratings.

Similarly, the splits of group 5 on STPD (short-term debt/debt) and group 7 on XPM12 (share of two largest export markets in total exports) are difficult to explain. In each case, there is a positive association between the rating and these variables, where the opposite would be expected on _a priori_ grounds. In conclusion, it is possible that the model is over-fitted beyond groups 4 to 7.

Taking up a point that was raised in the regression analysis, there is an interaction between RTD (real total debt) and GDPX (public debt/exports). GDPX is important only for cases that have small values of RTD: countries with high investment ratios, if also large (as measured by RTD), get a high rating regardless of the value of GDPX. More generally, the lack of symmetry of the tree diagram indicates that interaction is present.

The AID model is summarised in Figure XII.3. Variables RY and RTD are paired, as representing the same dimension of the data. The rating is positively related to each of the variables except GDPX (public debt/exports), so the negative of that variable is recorded in Figure XII.3. For those variables that, arguably, appear in the AID analysis with the 'wrong sign', the entries are shown in parentheses.

Figure XII.3: AID model: composition of terminal groups

Predictor

Variable: Represents:	INVR (Fixed Investment)	RY, (Size)	RTD	EFIR (interest)	XPM12 (Few mkts.)	-GDPX (Solvency)	MCOV (Import cover)	STPD (Short term debt)
Group								
(strongest)								
15	High	–	(High)	–	(High)	–	–	–
19	High	–	(High	–	(Low)	–	High	–
17	High	–	(Low)	–	–	High	High	–
11	Low	High	–	–	–	–	–	(High)
18	High	–	(High)	–	(Low)	–	Low	–
16	High	–	(Low)	–	–	High	Low	–
10	Low	High	–	–	–	–	–	(Low)
12	High	–	(Low)	–	–	Low	–	–
9	Low	Low	–	(High)	–	–	–	–
8	Low	Low	–	(Low)	–	–	–	–
(weakest)								

<u>**Notes**</u>:
A high rating value is associated with a high value of each variable (or transform, in the case of GDPX).

Entries are shown in () where the variable arguably has the 'wrong sign'.

The AID analysis yields three conclusions. First, interaction effects are present. Secondly, the regressors from Tables XII.1 and XII.2 appear in the AID analysis, and in a similar fashion: in both approaches the rating associates positively with INVR, MCOV and RTD, and negatively with GDPX. This provides confirmatory evidence for the bootstrapped regression model. Thirdly, except for RY (real GDP), the other variables in the AID analysis have the 'wrong sign': these are EFIR (effective interest rate), XPM12 (share of major markets), and STPD (share of short-term debt).

XII.4.5 Bankers' perceptions of country risk: - do they reflect new information?

Over the period since 1979, there have been significant changes in the absolute and relative levels of creditworthiness of the countries in the data set. Information flows freely in financial markets. Events such as the emergence of arrears are quickly reported, while the quality of statistical and factual information from sources such as the IMF and the World Bank is of a high order. This raises the possibility of examining the *Institutional Investor* rating system for evidence that new information is quickly absorbed into it. Contrary evidence would suggest the existence of judgemental bias: for example, the use of the *anchoring and adjustment* heuristic.

Table XII.4 sets out the matrix of Spearman rank-order correlation coefficients between the ratings for each of the sample years for this study, for the countries in the sample. Because Ghana has never been rated, the maximum number of countries is

54, while missing values for other countries further reduced the sample size in the years before 1982. For any pair of successive years, the coefficient is in the range [0.95, 0.99]. As the time interval increases, the correlation between the rankings within each of a pair of years falls. The greatest interval (1979 to 1989) yields the smallest correlation (0.64). However, this is still evidence of high correlation - a qualitative judgement that is born out by examining the t-statistics associated with the correlation coefficients: at the 5 per cent significance level, it is not possible to reject the hypothesis of no significant difference between the rankings of countries in any pair of years within the sample, based on the values of the <u>Institutional Investor</u> rating.

Of course, it is quite possible that bankers' perceptions of countries could have changed, without significant effect on the rank-ordering. This possibility can be explored using the Pearsonian correlation coefficient between the ratings in any pair of years.

Table XII.5 sets out the matrix of these correlation coefficients. The value for any successive pair of years is in the range [0.96, 0.99]. As with rank-order, as the time-interval extends, the correlation coefficient falls, and takes its minimum for the pairing 1979/1989. Even for that interval, the value (0.63) is still quite high.

The evidence of Tables XII.4 and XII.5, suggests that bankers are slow to revise their views about creditworthiness. However, it is not conclusive. It is conceivable that macro events that bear on creditworthiness may be perceived as affecting all debtor countries in a similar way. If the consequence for the <u>Institutional Investor</u> rating system is an additive or multiplicative shift that is constant across all countries, then the correlation coefficients of Tables XII.4 and XII.5 will be unaffected. To examine this possibility a simple t-test is applied to the null hypotheses of equality between the mean rating values across pairs of years. The t-values are set out in Table XII.6.

Table XII.6 shows that the mean rating fell continuously from 1979 to 1985, and recovered a little in 1986. Differences in mean ratings were statistically significant at the 5 per cent level within the period 1979 to 1983, but not within the period 1984 to 1989, except for pairings of 1986 with the three later years. The differences in mean ratings were also significant across the boundaries of the two periods, except for 1983/1986.

Table XII.4: Institutional Investor ratings: Spearman rank-order correlation coefficients

	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
1980	0.96									
1981	0.93	0.98								
1982	0.88	0.93	0.96							
1983	0.79	0.84	0.89	0.95						
1984	0.75	0.79	0.86	0.91	0.97					
1985	0.71	0.74	0.82	0.88	0.95	0.99				
1986	0.70	0.72	0.80	0.86	0.94	0.97	0.99			
1987	0.70	0.71	0.78	0.92	0.95	0.97	0.99	0.99		
1988	0.68	0.69	0.76	0.84	0.90	0.94	0.95	0.98	0.99	
1989	0.64	0.64	0.73	0.79	0.86	0.90	0.92	0.94	0.95	0.98

Table XII.5: Institutional Investor ratings: Pearson product-moment correlation coefficients

	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
1980	0.96									
1981	0.93	0.99								
1982	0.88	0.93	0.97							
1983	0.79	0.83	0.90	0.96						
1984	0.74	0.78	0.85	0.91	0.98					
1985	0.70	0.73	0.81	0.88	0.96	0.99				
1986	0.69	0.71	0.79	0.86	0.94	0.98	0.99			
1987	0.68	0.69	0.77	0.83	0.93	0.96	0.97	0.99		
1988	0.66	0.66	0.74	0.81	0.90	0.94	0.95	0.98	0.99	
1989	0.63	0.63	0.71	0.78	0.87	0.92	0.93	0.95	0.97	0.99

Table XII.6: Institutional Investor ratings: matrix of t-statistics between mean annual ratings

	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
d.f.:	45	46	50	53	53	53	53	53	53	53
mean	44.5	42.1	38.2	34.2	30.4	28.9	28.6	29.1	27.8	27.6
1980	3.2									
1981	4.6	4.7						1989: mean=27.3		
1982	6.4	5.7	5.7					d.f.= 53		
1983	7.6	6.6	6.5	5.3						
1984	7.8	6.7	6.6	5.4	3.7					
1985	7.4	6.4	6.1	4.8	2.7	0.9*				
1986	7.3	6.0	5.5	4.1	1.8*	-0.3*	-1.3*			
1987	7.9	6.5	6.1	4.8	3.1	1.7*	1.5*	4.6		
1988	7.8	6.4	5.9	4.7	2.9	1.7*	1.6*	3.3	0.9*	
1989	7.5	6.1	5.7	4.5	2.8	1.8*	1.7*	2.7	1.0*	1.0*

Key: * = not significant at 5 per cent level (2-tailed)

Source of data for Tables XII.4-XII.6:
Institutional Investor, September 1979 - September 1989 - sample of 54 LDCs, with pairwise deletion of missing cases.

Notes:
t > 0 implies a smaller mean rating in later year of pairing.

In conclusion, the the t-statistics of Table XII.6 indicate a significant decline, year by year, in average ratings from 1979 to 1983. The correlation evidence of Tables XII.4 and XII.5 indicates that the decline was seen as uniform across countries. This may be construed as evidence of the 'anchoring and adjustment' heuristic: according to Tversky and Kahneman (1974) use of this heuristic, which is a symptom of conservatism, causes estimates to be biased towards initial values.

XII.5 <u>The BERI, ICRG and EIU rating systems; comparisons of all four systems</u>

XII.5.1 <u>Introduction</u>

Because of a shortage of observations, it is not possible to run AID analyses of the BERI, EIU or ICRG rating systems. For the same reason, it is not possible to replicate for these systems the intertemporal analysis of the <u>Institutional Investor</u>. The major focus of this chapter is on developing paramorphic representations of human judgement, and to this end regression models of these three rating systems are developed in section XII.5.2. Diverse models emerge, and in section XII.5.3 all the regression models, including that of the <u>Institutional Investor</u> rating system, are compared. Finally, in section XII.5.4 a question that was first raised in Chapter II is addressed: this concerns the extent to which perceptions of political and economic risk are independent, and it is tackled by an examination of the correlations between the various constituents of the rating systems, where these are available.

XII.5.2 BERI, ICRG and EIU rating systems: regression analysis

A stepwise regression analysis was applied to each of the three (composite) rating systems in turn, following the procedure described in XII.4.2 for the analysis of the *Institutional Investor* rating system. The results are set out in Tables XII.7-XII.9.

Shortage of observations precludes the retention of a holdout sample for the BERI and EIU systems. This restriction is also imposed on the ICRG system, despite its greater availability, because a holdout of (for example) 1989 would be over-reliant on very recent data on debt-servicing status. These are very much subject to revision at the time of writing. Consequently, the training samples are 1986 for BERI, 1986 to 1989 for ICRG, and 1989 for EIU.

Table XII.7: BERI index: regression analysis, 1986

	Estimated coefficients		t-values (26 d.f.)	N=31
Y =	42.254		(17.9)	
	- 0.320	PDSPX	(-6.8)	(proj. debt serv./expt.)
	+ 0.580	DGDP	(3.7)	(growth rate, real GDP)
	+ 0.381	RTDG	(4.7)	(growth rate, real debt)
	+ 0.004	RYPC	(4.3)	(real GDP/capita)

F(4,26) = 31.3

R-squared = 0.83. Adjusted R-squared = 0.80

Table XII.8: ICRG rating: regression analysis, 1986-1989

	Estimated coefficients		t-values (215 d.f.)		N=220
Y =	66.520		(45.5)		
	- 0.017	TDPX	(-3.9)	(public debt/exports)	
	- 0.190	OCTD	(-8.1)	(official creds./debt)	
	+ 0.706	DGDP	(6.2)	(growth rate, real GDP)	
	- 0.136	DCPI	(-5.7)	(inflation)	

R-squared = 0.53. F(4,215) = 61.7 Adjusted R-squared = 0.53

Table XII.9 EIU credit risk rating scores: regression analysis, 1989

	Estimated coefficients		t-values (50 d.f.)		N=55
Y =	41.903		(12.0)		
	+ 0.529	EFMT	(6.1)	(effective maturity)	
	+ 0.414	PDSPX	(5.1)	(proj. debt serv./expt.)	
	- 3.208	MCOV	(-5.3)	(import cover)	
	- 1.076	DCPR	(-4.6)	(growth, priv. cons.)	

R-squared = 0.86. F(4,50) = 73.9 Adjusted R-squared = 0.84

XII.5.3 The regression models compared

The results of XII.5.2 reveal three points. The first emerges emerges by comparing Table XII.1 with Tables XII.7 to XII.9. These regressions are all significant in terms of their F-values and coefficients of determination, and all the estimated coefficients are significant at the five per cent level. Nevertheless, there is a great diversity in the specifications: not just in the variables, but in the underlying principal components. This is summarised in Table XII.10. Thus there appear to be significant differences of opinion between the different rating agencies, and between them and the bankers, over the key indicators of country risk.

The second point concerns the differences between the specification of each regression, and the financial ratios that are part of the input to the ICRG and EIU rating systems, for which explicit information is available. In the case of ICRG, the only overlap is the inflation rate. For the EIU, the overlap consists of the debt service ratio and import cover. This possibly overstates the overlap, because the rating system uses the actual debt service ratio, while the regression uses the projected ratio. This incomplete overlap arises because the variables in the regression models have to represent all the judgemental elements of the rating systems, in addition to the explicitly stated financial and economic ratios. For this reason, it is legitimate to use the bootstrapped models as paramorphic representations of human judgement.

The third point concerns the signs of the coefficients. Bearing in mind that creditworthiness is associated with low values of the EIU rating and high values of the others, it may be seen that all the estimated coefficients in Tables XII.7 to XII.9 have the sign that would be expected a priori, except for RTDG (growth rate, real total debt) in Table XII.7. Its positive sign implies that the faster the rate of growth of real debt, the higher will be the value of the BERI index. This may reflect the ex post position whereby countries that are rated favourably are able to obtain additional loans. In view of the comparatively small sample size, perhaps not too much should be made of this.

Table XII.10: Country risk ratings: summary of regressions

Princ. Comp.	Inst. Inv.	BERI	ICRG	EIU
1		PDSPX		PDSPX
2			DCPI	
3	GDPX		TDPX	EFMT
4	INVR	RYPC	OCTD	
5		DGDP	DGDP	DCPR
7	RTD			
11		RTDG		
-	MCOV			MCOV

The diversity of underlying structure is greater than that in the rating systems themselves. This is revealed by the correlations between them, which are set out in Table XII.11.

Table XII.11: Country risk ratings: correlation analysis, 1986-1989

	Inst. Inv.	BERI	ICRG
BERI	0.87 (31)		
ICRG	0.76 (216)	0.86 (31)	
EIU	-0.83 (54)	n.a.	-0.60 (55)

<u>Notes</u>: Numbers in (): numbers of common cases. There are no simultaneous observations on the BERI and EIU ratings.

Comparisons between the correlation coefficients must be made cautiously, because they are not all based on the same sample. The correlations suggest that the rating systems are similar but not identical. The greatest divergence is between the ICRG and EIU ratings - squaring the correlation coefficient to indicate magnitude, we get a value of only 36 per cent. However, the other comparisons reveal much closer agreement: the smallest correlation is between the <u>Institutional Investor</u> and ICRG, with a squared coefficient of 58 per cent. Similar qualitative results are revealed by an analysis of the Spearman rank-order correlation coefficient.

In conclusion, differences exist in the underlying structure of the rating systems. However, the key question concerns their predictions. This will be taken up in chapter XIII: with appropriate cutoffs chosen, the rating systems will be used to classify countries as arrears and non arrears, and comparisons will then be drawn.

XII.5.4 Ratings versus multivariate models: comparison of informational basis

The discriminant and logit models of chapters VII and VIII draw their independent variables from the set:

NARY	(net assets/GDP) [principal component:	1]
DCPI	(inflation)	[2]
INPS	(interest/debt service)	[3]
DGDP	(GDP growth)	[5]
INVR	(fixed investment/GDP)	[4]
AS	(=1 for Asian countries, otherwise 0)	[-]

Comparison with Table XII.10 indicates that the rating systems rest on very different foundations. The bootstrapped model of the ICRG system includes the inflation rate. Those of the BERI and ICRG systems include GDP growth, while the model of the _Institutional Investor_ system includes the investment ratio. Apart from these points in common, the specification of each bootstrapped model differs from those of the multivariate models. The overlap of principal components is greater, but still partial. Of five components that appear in the multivariate models, the numbers represented in the four-variable regression models are: _Institutional Investor_: two; BERI and EIU: three each; ICRG: four.

XII.5.5 Intra-rating correlation analyses

Chapter II raised the question whether perceptions of political risk are significantly related to economic factors. An affirmative answer to this has been given by Mumpower et al (1987), as reported in chapter II. In this section, the question is addressed through a correlation analysis of the ICRG and EIU rating systems. The ICRG ratings and the country risk rating scores of the EIU are weighted sums of sub-ratings, as described in XII.3.3. In the case of the ICRG system, these reflect political, financial and economic factors, while for the EIU, the constituents reflect political factors, economic factors, and economic policy. The constituents of the two rating systems are not directly comparable. However, it is worth exploring whether, in each case, the sub-ratings themselves yield independent information.

First, each of the political risk ratings was regressed on the set of financial and economic data, using the interactive stepwise procedure. In each case, it was impossible to develop a satisfactory model. For the ICRG political risk rating, the first variable to enter was OCTD (official creditors/total debt), but the value of R-squared was only 0.33. For the EIU rating, the first variable selected was TDSR (total debt service ratio), and R-squared was only 0.21. Thus with a linear specification, it is not possible to explain these perceptions of political risk in terms of the variables in the data set.

The possibility remains that the sub-ratings within each index may be highly correlated, so that one or more of them, possibly including political risk, may be redundant.

The correlation matrices are set out in Table XII.12.

Table XII.12: ICRG and EIU ratings: correlation analysis

Pearson correlation coefficients (squared)

220 cases	ICRG rating Political	Financial	Economic
Financial	0.56		
Economic	0.12	0.40	
Composite	0.81	0.86	0.48

Pearson correlation coefficients (squared)

55 cases	EIU credit risk rating scores Political	Economic	Economic policy
Economic	0.08		
Ec. pol'y	0.25	0.14	
Composite	0.32	0.85	0.45

The largest of the correlations has the value 0.56, between the ICRG political and financial ratings. This moderate correlation suggests that the components are all making a distinct contribution.

In conclusion, political risk cannot be 'explained' by a linear regression on variables from the data base that is used in this thesis. The correlation analysis suggests that the political risk rating does make an independent contribution to the EIU rating, where its squared correlation coefficient with 'Economic policy' is only 0.25. In the ICRG rating, the squared correlation coefficient of political and economic risk is 0.56, and the independence of political risk is correspondingly smaller. This is consistent with the political risk element in each of these systems. In the case of the EIU, it reflects

> for example the operation of the political system, alternative regime policies, the degree of enfranchisement, the attitude towards foreign creditors, the regional context. (EIU, 1988b: p. x).

While the ICRG political risk rating includes factors such as these, it also gives a weighting of 12 per cent each to two 'economic' factors, namely 'economic planning failures' and 'economic expectations versus reality'. Consequently, it is to be expected that the independent contribution of the political risk rating to the overall composite will be smaller in the ICRG system, compared with the EIU.

XII.6 Conclusion

In this chapter, paramorphic representations of human judgement have been constructed using multiple regression. Diverse models have emerged from the four different representations of human judgement, as reflected by the four rating systems. However, despite the diversity in the judgements behind the rating systems, the ratings themselves are moderately correlated: except for the pairing of the EIU and ICRG ratings, the squared correlation coefficients lie in the interval [0.58, 0.76].

All of the paramorphic models are statistically significant, as are each of the estimated coefficients. The Institutional Investor rating system, for which a large sample was available, was explored also using AID. The results clearly indicate the presence of interaction. Moreover, they confirm the conclusion of the regression analysis in terms of the variables that are important in banker judgement, and the manner in which they affect perceptions of creditworthiness. High fixed investment/GDP and import cover are viewed favourably, and high public debt/exports is viewed unfavourably. Size is seen favourably, but this is unsatisfactorily represented by real total debt.

There is limited overlap between the sets of variables appearing in the bootstrapped models of country risk ratings, and those appearing in the 'early warning' logit and discriminant models. Overlap in terms of principal components is greater, but still only partial.

Two subsidiary points have been established. First, perceptions of the relative riskiness of countries, as represented by Institutional Investor ratings, have been very stable from year to year: when creditworthiness has been perceived by bankers to be declining, the effect has been broadly uniform across countries. This is interpreted as evidence of the **anchoring and adjustment** heuristic, and this conclusion will support the overall conclusions of chapter XIII. Secondly, in the EIU and ICRG rating systems, the largest squared correlation between **political risk** and any other component is only 0.56, with the ICRG financial risk component. This indicates that, for these systems, political risk is making an independent contribution.

The paramorphic models will be used in the following chapters. In chapter XIII, their predictive powers will be explored, and compared with those of the rating systems themselves, while in chapter XIV the rating systems and their models will be compared with the multivariate models of chapters VII to IX.

XIII The consistency of human judgement:

XIII.1 Introduction

This chapter assesses the consistency of human judgements of country risk. Specifically, paramorphic representations of human judges, based on regression or AID, are compared with the judges themselves as represented by the rating systems. More emphasis falls on the *Institutional Investor's* system than on those of the EIU, ICRG or BERI, because it is the most appropriate system to take to represent banker judgement. Also, it is available over a longer time series than the other systems.

Within-sample tests of fit for all the regression models have been reported in chapter XII. However, for the *Institutional Investor* system only, a holdout sample has been retained (see XII.5.2). This permits statistical tests to be made of the power of the regression model to make true *ex ante* forecasts of rating values. These tests are reported in section XIII.2. The point of this is to determine whether the model is correctly specified and exhibits parameter-constancy out of the sample period (Gilbert, 1986: p.291): if it fails this test, then it is of limited use.

The major topic of this chapter is the prediction of creditworthiness by the ratings and their bootstrapped models, and this occupies sections XIII.3 to XIII.5. In section XIII.3, each rating's classificatory performance is examined, and comparisons are drawn. Next, in section XIII.4 comparisons are made between the classifications of the ratings and those of their bootstrapped models, within the training sample periods.

In chapters VII to IX, multivariate models and least-cost cutoffs were estimated on 1979-1986 data, and then tested on a holdout period in chapter X. This distinction between training period and holdout period has no meaning for the ratings themselves, but it has the same significance as before for the cutoffs that must be estimated before classification can take place, and it is of course crucial to testing the predictive ability of the bootstrapped models. The analysis of section XIII.4 provides least-cost cutoffs for each of the ratings and its bootstrapped model. In the case of the Institutional Investor system alone, a holdout period is available, and in section XIII.5 human judgement and its paramorphic representation are compared in terms of predictive power in that period. Finally, in section XIII.6 conclusions are drawn.

XIII.2 Institutional Investor regression model: tests of forecasting power

This section examines the predicted values for 1987 to 1989 of the bootstrapped regression equation of chapter XII (including four variables plus dummy), and summarises some statistical tests of the extent to which they conform with the actual rating values.

First, for each out-of-sample data point, a **95 per cent confidence interval** may be constructed around the central forecast. This is a confidence interval for the expected value of the rating, conditional on the predictors, and its construction is discussed in Kmenta (1971: p.374). Only 30 of the total of 162 cases lie in the confidence intervals, so that on this criterion the regression equation appears to be a comparatively poor predictor: most of the actual rating values lie outside the 95 per cent confidence intervals.

Secondly, the wide scatter of actual values about the predictions from the regression is borne out by the **root mean square error**. Computing this conventionally in terms of proportionate deviations from actual out-of-sample values (Maddala, 1977), the root mean square prediction error for all 162 out-of-sample cases (1987-1989) is 40 per cent, with values in the individual years of 44 per cent, 37 per cent and 40 per cent respectively. However, part of the problem here is the prevalence of low rating values: for example, the median rating, 1987-1989, was 25.15, and the first quartile was 16.8. This implies that even quite a modest absolute prediction error will translate into a high proportionate error. If the computation is based on absolute prediction errors, then the root mean square errors are eight rating points, for each year and for the period as a whole.

A third measure of the degree of congruence between the actual rating values and the regression predictions is given by the **squared correlation coefficient** between them. In fact, the actual and predicted values are highly correlated: for each year, and for the whole period, the coefficient (squared) is evaluated at 0.74. However, this statistic will not detect any systematic linear bias: for example, in the extreme case of an exact linear relationship between the actual and predicted values, the correlation coefficient would be unity, for any slope and intercept values. If such systematic bias is present, then the actual values will not be randomly distributed around the regression plane. This possibility is addressed by means of a **runs test**, which is applied after ordering the out-of-sample cases on the predicted value of the fitted regression. The values of the test statistic for the three years are -0.54, -0.55, -1.64 respectively. These are all well within the interval [-1.96, 1.96], so the null cannot be rejected at the 5 per cent level: there is no evidence of systematic deviation.

Fourthly a test described by Maddala (1977: p.460) formally addresses the null hypothesis that there is no significant difference between the observed values of the rating out-of-sample, and the predicted values, given the out-of-sample values of the regressors. The test is based on the **residual sums of squares**. The computed F-ratios for this test are: 0.637 for the three years together, and 0.673, 0.602 and 0.681 for the individual years respectively. The critical values of F with (162, 408) degrees of freedom in the first case and (54, 408) otherwise are all 1.00 at the five per cent level, so at that level of significance the null may in no case be rejected.

A final, and informal, part of the appraisal involves inspecting individual cases. A small number have rating values that are far from the fitted regression. 'Far' is a subjective term, but Table XIII.1 lists cases where the difference exceeded 10 points in at least one year.

Table XIII.1: Institutional Investor rating, minus predicted regression value

	Actual rating minus predicted		
	1987	1988	1989
Actual rating understates:			
Brazil	-7	-14	-15
Dominican Rep.	-11	-12	-8
Mexico	-17	-11	-12
Philippines	-13	-9	-5
Venezuela	-16	-11	-15
Yugoslavia	-9	-14	-16
Iran	-19	-19	-13
Actual rating overstates:			
Bangladesh	11	9	14
Cameroon	11	7	7
Trinidad	14	16	8
Ivory Coast	11	10	11
Senegal	8	9	12
Malawi	14	5	9
Malaysia	4	7	10
Taiwan	20	23	23
Thailand	8	7	11

Note: Cases included have at least one difference of more than 10 points, during 1987-1989.

Negative values in Table XIII.1 indicate cases where the rating is pessimistic, relative to the regression. Six are heavily indebted countries, and except for the Dominican Republic all are significant in terms of the global debt problem. However, other major debtors, for example Argentina, are absent. Iran is an exceptional case: pessimism is probably founded on perceived political risk, although Iran has been servicing its bank debts, and has no imminent difficulties. The positive values in Table

XIII.1 indicate an optimistic rating, relative to the regression line. These cases fall into several groups. First, Bangladesh is kept afloat by aid and soft loans. Objectively, it is a poor commercial risk. Secondly, Trinidad and Cameroon have recently encountered serious difficulties for the first time each. Cameroon's rating has been fairly stable (down from 47.6 in 1986 to 45.3 in 1989), while Trinidad's has been falling - but not fast enough, if the regression indicates accurately. Next, Ivory Coast, Malawi and Senegal are all in serious payments difficulties, with few grounds for optimism. Finally, it is plausible that a regression estimated on countries that are mostly in difficulties would tend to underplay the creditworthiness of the three strong Asian economies, Taiwan, Thailand and Malaysia, as these grow and develop in the out-of-sample period.

In conclusion, the actual ratings (out-of-sample) are quite widely scattered about the regression plane, but the problem is high variability rather than systematic deviation. If the model reflects the underlying judgemental processes, then the results indicate that these judgements are subject to high variability; alternatively, the high variability may be a consequence of missing variables: for example, indicators of political risk.

XIII.3 The ratings: classificatory power

XIII.3.1 Initial examination: runs test

If the Institutional Investor rating system reflects banker judgement, and if banker judgement is capable of forecasting creditworthiness one year ahead, then the rating system may be used as a forecasting tool. Likewise, a case may be made for using the other rating systems for forecasting, given that they are based on expert judgement.

First, to test the predictive ability of the ratings, the runs test (Rohatgi, 1984) was addressed to the null hypothesis that the rating system has no predictive power. The advantage of the runs test over tests based on classificatory performance (see XIII.3.2 below) is that unlike the latter, the former does not require the specification of a cutoff rating value between arrears and non-arrears cases. To apply this test, the sample is sorted numerically on the ratings. If the rating system has predictive power (a false null), then this ordering will result in a clustering of arrears cases towards the low-rating end of the ordering, and of non-arrears cases towards the opposite end. Group-membership of a country is not independent across different years:

thus it would be inappropriate to apply the test to the pooled sample containing all out-of-sample country-year-cases, and the correct procedure is to test each year separately.

For the **Institutional Investor** rating system, the null is rejected at the five per cent level in only four years out of eleven, 1980, 1986, 1988 and 1989, and in 1986 and 1989 it cannot be rejected at the one per cent level. This suggests that the Institutional Investor rating system has limited ability to predict creditworthiness over a one-year horizon.

Table XIII.2: Runs test statistic: based on number of runs of arrears and non-arrears cases

Institutional Investor rating system	1979	-1.18	
	1980	**-3.29**	
	1981	0.05	
	1982	-0.45	
	1983	-0.87	
	1984	-1.26	
	1985	-0.59	
	1986:	**-2.02**	Ho (randomness) is rejected at the 5 per cent level if the test statistic (based on the number of runs) falls outside the range [-1.96, 1.96]. For a 1 per cent significance level, [-2.57 ,2.57] is the acceptance region. Two-tailed test.
	1987:	-1.36	
	1988:	**-3.10**	
	1989:	**-2.52**	
ICRG rating	1986	-1.53	
	1987	0.30	
	1988	-0.86	
	1989	-1.43	
BERI index	**1986**	**-3.05**	
EIU credit risk rating	**1989**	**-4.90**	

Note: After ordering sample on each rating, runs relative to ARS(t+1) in each year t were counted.

The ICRG rating system also performs poorly: in no year can the null be rejected. However, the BERI index and the EIU rating system yield significant clusterings of the sample cases, in terms of their debt-servicing status: in each case the null hypothesis is rejected at the five per cent level.

XIII.3.2 The ratings: ex post classificatory power

The next step is to examine the classificatory performance over the period 1979 to 1989. This section is concerned with *ex post* classifications, in that the cutoff rating values have been determined from the same rating values whose classificatory power is being tested. The method of minimising misclassification costs, as described in chapter VII, has been applied, and the results are set out in Table XIII.3. In each case, the cutoff is optimal for a given range of relative unit cost of a type I error. For the ICRG rating system, the lower bound to this range is rather high, at 3.1. However, varying the cutoff for this rating in the direction of lower unit cost makes proportionately little difference to the classifications that it makes, as footnote (2) to Table XIII.3 indicates.

For the *Institutional Investor* rating system, the optimal cutoff value is 40.7, for values between 2.4 and 6.0 for the relative cost of a type I error. At this point, there are 8 type I errors (4.0 per cent), and 115 (49.4 per cent) of type II. The former is low, but the latter is very high. There is no evidence that any particular year or group of years predominates. In case the high rate might be a function of pooling all eight years when determining the cutoff, each year was next treated separately. In each case, a least-cost cutoff was sought, for plausible values for the unit cost of a type I error. When the results for all eleven years were added, the combined result was 10 type I errors (5.0 per cent) and 96 type II (41.2 per cent). This remains a very high error rate, and it remains to be seen if the regression or AID models fitted to the rating can out-perform the rating - arguably, by eliminating the random component in human judgement.

The performances of the other three rating systems are similar to that of the **Institutional Investor** in qualitative terms: each has a low type I error rate at the expense of a high type II rate. This is particularly marked in the case of the **ICRG** and **EIU** rating systems, with type II rates exceeding 50 per cent. The **BERI** index is better, but only relatively: the type II error rate is still fairly high, at 30.8 per cent.

Table XIII.3: Country risk ratings 1979-1989: classifications 1980-1990

Rating system	Actual group		Classified: Arrears	Classified: Non-arrears
			1980-1990	
Institutional Investor (1979-1989)	Arrears	201	193	8 (4.0%)
	Non-arrs.	233	115 (49.4%)	118
Cutoff: 40.7				
			1987-1990	
ICRG (1986-1989)	Arrears	135	135	0 (0.0%)
	Non-arrs.	85	71 (83.5%)	14
Cutoff: 65.0				
			1987	
BERI (1986)	Arrears	18	17	1 (5.6%)
	Non-arrs.	13	4 (30.8%)	9
Cutoff: 46.5				
			1990	
EIU (1989)	Arrears	34	34	0 (0.0%)
	Non-arrs.	21	11 (52.4%)	10
Cutoff: 38.4				

FMM statistics:
Inst. Inv. (434 d.f.): 12.08; ICRG (220 d.f.): 7.74
BERI (31 d.f.): 3.71 EIU (55 d.f.): 4.79

<u>Notes</u>:
(1) The cutoffs are optimal for relative cost C of a type I error in the intervals: II: [2.4, 6.0]; ICRG: [3.1, 6.0]; BERI: [1.5, 5.0]; EIU: [2.0, 6.0].

(2) If the cutoffs are moved, to minimise the total number of errors in each case, those numbers, with the totals from the above cells in parenthesis, are: II: 106 (123); ICRG: 67 (71); BERI: 4 (5); EIU: 6 (11).

(3) if the II and ICRG ratings are treated separately in each year, then the combined 3 - year error rates, using least-cost cutoffs in each year, are:
Inst.Inv.: type I: 10 (5.0%), type II: 96 (41.2%)
ICRG: type I: 0 type II: 66 (77.6%)

For each rating system, the FMM test (see chapter VII) furnishes a nonparametric test of the null hypothesis that the classificatory power of the rating is no better than random. The values of the test statistic are shown at the foot of Table XIII.3. Comparison of these with the tabulated t-values leads to rejection of the null in each case, at the five percent level (two-sided test).

Table XIII.4: Country risk ratings 1986-1989: error rates in each year

Rating system (1986-1989)		Misclassifications (1987-1990) Type I	Type II
(Arrears cases, non-arrears)			
		1987	
1986			
BERI	(18,13)	5.6%	30.8%
Institutional Investor	(32,22)	3.1%	40.9%
ICRG	(33,22)	3.0%	72.7%
		1988	
1987			
Institutional Investor	(33,21)	3.0%	47.6%
ICRG	(34,21)	0.0%	81.0%
		1989	
1988			
Institutional Investor	(33,21)	0.0%	47.6%
ICRG	(34,21)	0.0%	71.4%
		1990	
1989			
Institutional Investor	(33,21)	0.0%	38.1%
ICRG	(34,21)	0.0%	76.2%
EIU	(34,21)	0.0%	52.4%

Note: Debt-servicing status in year t+1 classified by rating in t; *ex post* least-cost cutoff in each year.

The classifications shown in Table XIII.3 for the different rating systems are not strictly comparable, in that they do not relate to a standard time-period. Table XIII.4 avoids this difficulty, by focusing in turn on each of those individual years for which data are available for at least two of the rating systems. The table, which summarises the error rates in each case, reinforces the earlier conclusions: all the ratings yield low type I error rates, but very high type II rates. The Institutional Investor rating system yields the lowest type II rate, except for the BERI index in 1986, which only covers 31 countries.

XIII.3.3 Cutoff for holdout period

To enable the Institutional Investor rating system to be used for *ex ante* classification within the holdout period in section XIII.5, it is necessary to estimate a least-cost cutoff from the period 1979-1986 that formed the training sample for the regression models. In fact, a cutoff of 40.7 is least-cost for that period: this is exactly the same value that is optimal for the period 1979-1989. It yields exactly the same type I errors as in the larger sample, in number (8) and identity, but only 74 type II errors. These represent error rates of 7.8 per cent and 43.5 per cent, respectively.

XIII.4 Ratings and bootstrapped models: within-sample comparisons

XIII.4.1 Introduction

In this section and section XIII.5, the judgementally based ratings are compared with their paramorphic representations. As explained in XII.5.2, holdout periods are not available for the BERI, ICRG or EIU rating systems. Consequently, comparisons using these ratings can only be made for the training sample period.

Classifications of the training samples are reported in section XIII.4. In each case, where a country-year case is not rated, it is excluded from the analysis. Section XIII.5 is concerned with the holdout analysis of the Institutional Investor rating.

XIII.4.2 The Institutional Investor rating system

This section is concerned with the Institutional Investor rating system, 1979-1986, and its paramorphic representation by the five-variable linear regression model (including dummy), and the

AID model, of chapter XII. The optimal (least-cost) cutoff for the estimated regression values is 46.03, and this yields the classifications shown in Table XIII.5.

Table XIII.5: Institutional Investor: training sample classified by predicted regression values, 1979-1986

Actual group		1980-1987 Classification Arrears	Non-arrears
Arrears	102	93	9 (8.8%)
Non-arrears	170	82 (48.2%)	88
	272	175	97

Notes: Weak-year cases excluded.
Based on cutoff predicted rating value of 46.03.

For the AID model, if the interval [2.0, 6.0] contains the unit cost of a type I error, then the cost-minimising cutoff is drawn between groups 17 and 11: i.e. the 'non-arrears' classification includes groups 15, 19, and 17, while the rest are 'arrears'.

This performance of both models is worse than that of the rating itself, but not markedly so: the error rates in the training sample, at 7.8 per cent (type I) and 43.5 per cent (type II) are only marginally lower (see XIII.3.3).

Table XIII.6: Institutional Investor rating: AID model: predicted groupings used to classify training sample, 1979-1986

Actual group		1980-1987 Classification Arrears	Non-arrears
Arrears	102	90	12 (11.8%)
Non-arrears	170	76 (44.7%)	94
	272	166	106

<u>Note</u>: Non-arrears groups: 5,19,17.
Weak-year cases excluded.

XIII.4.3 <u>The BERI, EIU and ICRG rating systems</u>

The results for these systems are reported less exhaustively. First, they are less important as indicators of human judgement: they include a significant formal 'checklist' component. Secondly, they involve shorter time-series: they do not yield a holdout period for section XIII.5, and particularly in the cases of BERI and EIU, the training samples are small enough to give concern about the variability of the regression results. Results are summarised in Table XIII.7, along with those for the ratings. Results from XIII.4.2 for the <u>Institutional Investor</u> are also displayed, for comparative purposes. In each case, least cost cutoffs have been determined within the training sample period.

Table XIII.7: Country risk ratings and bootstrapped: regression models: error rates on training sample

Rating system (Arrears cases, non-arrears)		Misclassifications: one year ahead Type I	Type II
		Numbers and percentages	
BERI, 1986 (18,13)		1987	
	Rating	1 (5.6%)	4 (30.8%)
	Regression model	0 (0.0%)	13 (100%)
Institutional Investor, 1979-86 (102, 170)		1980-1987	
	Rating	8 (7.8%)	74 (43.5%)
	Regression model	9 (8.8%)	82 (48.2%)
	AID model	12 (11.8%)	76 (44.7%)
ICRG, 1986-1989 (135, 85)		1987-1990	
	Rating	0 (0.0%)	71 (83.5%)
	Regression model	0 (0.0%)	72 (84.7%)
EIU, 1986 (34, 21)		1990	
	Rating	0 (0.0%)	11 (52.4%)
	Regression model	0 (0.0%)	19 (90.5%)

Note: Debt-servicing status at t+1 classified by rating or model at t; *ex post* least-cost cutoff in each case.

Compared with the ratings themselves, classification by the bootstrapped regression models yields very similar type I error rates in each case. For the ICRG and *Institutional Investor*, the type II error rates are nearly the same as well, while for the regression models of BERI and the EIU rating system, the type II rate is much higher than for the ratings themselves. However, the ICRG and Institutional Investor regression estimates are based on a larger sample than those for BERI and EIU, and consequently have a greater precision.

XIII.5 Institutional Investor: classification of holdout sample

This section improves on the analysis of XIII.4, by utilising the holdout period: 1987 to 1989 rating and regression values and AID groupings are used to classify cases for 1988 to 1990. All models and cutoffs have been estimated on the 1979-1986 sample.

Table XIII.8: Institutional Investor ratings: classifications, 1988-1990

Actual group		Classification: Arrears	Classification: Non-arrears
		Each year, 1988-1990	
Arrears	33	33	0 (0.0%)
Non-arrears	21	13 (61.9%)	8
		All years combined	
Arrears	99	99	0 (0.0%)
Non-arrears	63	39 (61.9%)	24
	162	138	24

Note: Rating values 1987-1989 classify 1988-1990 cases; cutoff of 40.7 derived from 1979-86.

Table XIII.8 sets out the classifications of 54 countries by *Institutional Investor* ratings for 1987 to 1989; Table XIII.9 presents similar results based on predicted regression values; finally, Table XIII.10 presents AID results.

Table XIII.9: Institutional Investor predicted regression values: classifications 1988-1990

Actual group		Classification: Arrears	Non-arrears
		1987	
Arrears	33	31	2 (6.1%)
Non-arrears	21	12 (57.1%)	9
		Each year, 1988 and 1989	
Arrears	33	32	1 (3.0%)
Non-arrears	21	13 (61.9%)	8
		All years combined	
Arrears	99	95	4 (4.0%)
Non-arrears	63	38 (60.3%)	25
	162	133	29

<u>Note</u>: Cases 1988-1990 classified by predicted regression values 1987-1989, using cutoff rating value of 40.034 derived from 1979-1986 period.

Table XIII.10: Institutional Investor predicted AID groupings: classifications, 1988-1990

Actual group		Classification: Arrears	Non-arrears
		Each year, 1987-1989	
Arrears	33	29	4 (12.1%)
Non-arrears	21	12 (57.1%)	9
		All years combined	
Arrears	99	87	12 (12.1%)
Non-arrears	63	36 (57.1%)	27
	162	123	39

<u>Note</u>: Cases 1988-1990 classified by predicted AID groupings 1987-1989; cutoff between groups 17 and 11, as per analysis of 1979-1986 period.

The cutoffs used here are those that were determined on least-cost grounds in section XIII.4. In no case do they coincide with the <u>ex post</u> least-cost cutoffs for the out-of-sample periods, but it would not be legitimate to use the latter for assessing the performance of the rating for <u>ex ante</u> prediction.

As in section XIII.4, the type I error rates are low, at the expense of very high type II rates. The classificatory abilities of the predicted regression values and the AID model are worse than that of the rating itself, but not markedly so. The similarities between the models and the rating system extend to the identities of the cases placed in each group. Table XIII.11 reveals this for the rating-regression comparison: it lists some cases that are fairly consistently classified as non-arrears by the rating and its regression model; an analysis of the AID model reveals a very similar pattern.

Table XIII.11: Institutional Investor rating and regression model: cases consistently classified as non-arrears 1988-1990

Non-arrears cases (correct classifications):	
China	)
India	)
Indonesia	) By each approach
Korea	) in each year
Malaysia	)
Taiwan	)
Thailand	)
Algeria	) By each approach in 1988
Turkey	) By each approach in 1989 and 1990
Arrears cases (type I errors):	
Mexico	) By the regression model in 1988
Venezuela	) By the regression model in each year

Note: Table excludes countries that are sporadically classified as non-arrears.

The first seven countries in Table XIII.11 are certainly among the most creditworthy, and it would be damning if they were not correctly classified. The two cases that appear as type I errors for the regression model do so largely because their import cover was relatively high in the years in question: for example, in 1987 it was 6.0 months for Mexico and 7.7 for Venezuela, relative to a group mean of 1.9 months. It is noteworthy that these countries are major oil producers.

The countries that are consistently type II errors on each approach are:

Bangladesh	El Salvador	Pakistan	Syria
Chile	Iran	Papua N.G.	Tunisia
Colombia	Kenya	Sri Lanka	Zimbabwe

The question arises whether the misclassification of these cases arises from a conservative cutoff, or from overlap in the distributions of ratings (or regression values) of arrears and non-arrears cases. In fact it is the latter, as may be demonstrated in two ways. First, the <u>ex post</u> least-cost cutoffs may be computed. For the regression, this always moves the cutoff <u>upwards</u>, to eliminate one type I error. For the rating itself, this reduces the type II error rate, which remains high. Alternatively, the view may be taken <u>ex ante</u> that the cutoff value should be adjusted to allow for a fall in the distribution of rating values or regression estimates since the sample period. It makes little difference if the adjustment is based on mean or median: the outcome is that too low a cutoff results, with an unacceptably high type I error rate.

XIII.6 Conclusion

In chapter XII, paramorphic representations of four ratings were constructed using regression and AID modelling techniques. The regression models were found in that chapter to have a good statistical fit to the data in the training samples. In section XIII.2 of this chapter, this finding has been confirmed for the Institutional Investor regression model, in the holdout period 1987-1989.

Section XIII.3 explored the classificatory performance of the ratings, using the largest set of observations that is available for each. In each case, a least-cost cutoff yields a low type I error rate, but a high type II rate.

From section XIII.4, there is no evidence of the bootstrapped models out-performing the ratings themselves within the training samples. The error rates of the ratings and their models are approximately the same for the Institutional Investor and ICRG rating systems, while the regression models of the BERI and EIU rating systems produce markedly higher type II errors than the ratings themselves. In section XIII.5, this conclusion for the Institutional Investor is confirmed using a holdout sample, a technique that yields an unbiased test of *ex ante* predictive ability for rating and model.

If the ratings misclassify because of inconsistency in the application of judgemental rules by human decision-takers, then the regression and AID models should overcome this problem if they are correctly specified. The results of this chapter suggest that the problem lies in the foundations of banker or expert judgement, as reflected in the ratings, rather than in random errors (inconsistency) in the application of that judgement. In other words, there is evidence of systematic inaccuracy in the way in which bankers and other experts use information. Related evidence on this is supplied by the intertemporal analysis of the *Institutional Investor* rating system, in section XII.4.5, which suggests that bankers' perceptions of country risk change slowly, and may be based on **anchoring and adjustment.**

These conclusions will be followed up in chapter XIV with a comparison between the predictions of the rating systems and those of the multivariate models of chapters VII to XI.

XIV Man versus model: ratings versus multivariate models

XIV.1 Introduction

This is a key chapter in the development of the thesis. It draws together the two parallel strands that have run through the preceding chapters: on the one hand, multivariate statistical models of debt-servicing status, on the other, human judgement. The left-hand side of the Brunswik lens is represented by the quantitative models, and the right hand by the human judges or their paramorphic representations.

In this chapter, the human judges are represented by country risk rating systems, and for the bootstrapped models of human judgement, attention is confined to the regression approach, given that it out-performed AID in chapter XIII. As for the multivariate models, only the logit and discriminant models are used here, because they dominated all the other approaches that were considered in chapters VI to IX.

In section XIV.2, some comparisons are made for the period 1979-1986, which is the training sample for the statistical models. However, in order to conduct an unbiased test of true *ex ante* predictive ability, the primary focus of the chapter is on comparisons over the holdout sample. Therefore of the four ratings that have been used in chapters XII and XIII, only the *Institutional Investor* system is used here. In section XIV.3, data or rating values for 1987-1989 are used to generate classifications of cases for 1988-1990, by means of the statistical models and cutoff values that have been estimated on the 1979-1986 data. The chapter is summarised and conclusions are drawn in section XIV.4

XIV.2 Man versus model: the training sample period

The investment ratio INVR is included in discriminant model I and in the bootstrapped model of the Institutional Investor system, and this is the only overlap in variable specification between the bootstrapped model and any of the multivariate models. Given the different informational inputs to the judgemental and quantitative approaches to risk assessment, it is not surprising that 'model' and 'man' yield different classifications of the sample.

In terms of overall accuracy of classification, the discriminant and logit models of chapters VII and VIII outperform banker judgement as represented by the Institutional Investor rating.

A concise measure of the performance of any system of classification is given by the total number of correct classifications (arrears plus non-arrears cases), as a percentage of the sample size. The rate of correct classification is 93 per cent for discriminant model I, 89 per cent for model II, and 90 per cent for the logit model, within the training sample period using the Lachenbruch method. These rates do not change if the cases for which no

Institutional Investor rating is available are deleted from the sample.

By comparison, the correct classification rate for the rating is 70 per cent. Chapter XIII revealed that the bootstrapped model is no improvement on the rating itself, and its correct classification rate is 67 per cent.

Type I and II errors should not be equally weighted, and allowance may be made for this by introducing the differential cost C of a type II error. The misclassifications of the multivariate models are set out in Tables VII.4, VII.7 and VIII.3, and those of the rating are in Table XIII.9. If the total costs of misclassification are computed, it emerges that all of the multivariate models have lower costs than the rating, provided that C does not exceed 8.6, and that even if C is greater than this, but does not exceed 18.3, discriminant model II will have a lower misclassification cost.

Of the eight type I errors of the rating within-sample (see chapter XIII), three cases are also in this category for at least one of the multivariate models. The other five cases comprise Argentina (1981) and Brazil, Chile, Ecuador, and Mexico (1982), which is perhaps revealing of the optimism of bankers as the storm was about to break in the early 1980s. In each case, arrears were to emerge in the following year, and in each case the rating in the year indicated ranged from 43.1 (Ecuador) to 56.4 (Mexico).

XIV.3 Man versus model: the holdout sample period

Section XIV.2 drew comparisons between 'man' and 'model' for the 1979-1986 training sample. This is open to the objection that the assessment of classificatory power for each approach is biased, when the assessment is made on the training sample (Frank et al, 1965). In order to avoid this problem, and to test the true *ex ante* predictive abilities of 'man' and of 'model', this section draws comparisons of predictive performance on the holdout period, where ratings and model predictions for 1987 to 1989 are used to predict debt-servicing status for 1988 to 1990.

Table XIV.1 draws on material from chapters XI and XIII, and presents the combined three-year forecast classifications from the 'man' and 'model' approaches to assessing country risk. Because Ghana has never been rated by the *Institutional Investor* system, it is excluded from the analysis: there are thus 54 countries in the sample.

Table XIV.1: Models versus man: model predictions and ratings 1987-1989 used to classify 1988-1990

Actual category		Classifications 1988-1990: Arrears	Non-arrears
		Discriminant model I	
Arrears	99	88 (89%)	11 (11%)
Non-arrears	63	11 (17%)	52 (83%)
	162	99	63
		Discriminant model II	
Arrears	99	91 (92%)	8 (8%)
Non-arrears	63	13 (21%)	50 (79%)
Totals	162	104	58
		Logit model	
Arrears	99	88 (89%)	11 (11%)
Non-arrears	63	14 (22%)	49 (78%)
Total	162	102	60
		Institutional Investor: rating	
Arrears	99	99 (100%)	0 (0%)
Non-arrears	63	39 (62%)	24 (38%)
Total	162	138	24
		Institutional Investor: regression model	
Arrears	99	95 (96%)	4 (4%)
Non-arrears	63	38 (60%)	25 (40%)
	162	133	29

Note:
All cutoffs are optimal for unit cost of type I error in range [2.0, 4.2]. See ch. XI and XIII.

Although the logit and discriminant classifications are shown separately in Table XIV.1, a chi-square test indicates that they are statistically indistinguishable. Given the predicted grouping of a given case by each of two multivariate models, the case can be placed in one of four categories:

Arrears, arrears	Arrears, non-arrears
Non-arrears, arrears	Non-arrears, non-arrears

The test is based on the 2x2 contingency table whose cell entries are the numbers of cases in each of the four possible categories. Given the row and column totals, the expected cell entries may be computed, and the test addresses the null hypothesis that there is no significant difference between the observed and expected entries. In fact, the value of chi-squared leads to rejection of the null at the five percent level, for each possible pairing. For example, pairing the logit model with discriminant model I, the value of chi-squared is 136.63, compaired with the critical value of chi-squared (0.05,1) of 3.84. Thus we reject the hypothesis that there is no statistically significant relationship between the classification of cases by the two models.

It is clear from Table XIV.2 that the multivariate statistical models have a much higher rate of correct classification than 'man' as represented by the rating, while the bootstrapped model has a slightly lower success rate than the 'man' that it represents.

Table XIV.2: Rates of correct classification: models versus man, 1988-1990

		Per cent
Institutional Investor rating	:	75.9
Bootstrapped model	:	74.1
Discriminant model I	:	86.4
Discriminant model II	:	87.0
Logit model	:	84.6

Note:
The rate shown is: total correct classifications/total (162) cases, x100%. Ghana is excluded.

However, as chapter XIII has already revealed, the problem with the rating is the type II error rate: for the 1988-1990 period, the rate is 61.9 per cent, versus no type I errors (Table XIII.10). In contrast, the multivariate models all exhibit a mixture of both types of error. Consequently, if enough weight is placed on type I errors, it is possible for the rating to achieve a lower misclassification cost than the multivariate models.

Table XIV.1 sets out the numbers of type I and type II errors for the various approaches to classification. In each case, these are based on a least-cost cutoff for the training-sample period. As described in earlier chapters, the per-unit cost C of a type I error is required to take plausible values at the minimum-cost point, while the cost of a type II error is set at 1.0. In each case, there is a range of values for C for which the cutoff is in fact least-cost, and these ranges have as their intersection the interval [2.4, 4.2]. Thus, drawing on Table XIV.1 for the numbers of errors of each type, we have the following total misclassification costs, for C in the interval [2.0, 4.2]:

Discriminant model I:	11C + 11
Discriminant model II:	8C + 13
Logit model:	11C + 14
Institutional Investor rating:	39

It is easy to see that the *Institutional Investor* rating system will yield a lower misclassification cost than any of the multivariate models if C exceeds 3.25, and discriminant model II is its closest competitor.

XIV.4 Conclusion

This is a key chapter, in which formal multivariate statistical models are compared with judgemental assessments. In terms of overall classificatory accuracy, 'model' outperforms 'man'. This has been shown for predictions covering the entire period 1979-1989, and in particular it has been demonstrated using 1988-1990 as a holdout period.

When type I errors are differentially weighted, the picture is less clear. For any plausible value for the relative cost C of a type I error, 'Model' still outperforms 'man' on the training sample, but this does not truly test the *ex ante* predictive ability of each approach. The within-sample classificatory performance of multivariate models may be biased: sampling errors are one source of difficulty, while *search bias* may also be present where variable selection has been based on an intensive searching technique, such as stepwise discriminant analysis (Frank et al, 1965). For this reason, the key test of classificatory performance involves the holdout sample, and there it has been shown that 'man' (i.e. the *Institutional Investor* rating system) has the lowest misclassification cost, compared with the multivariate models, if the cost C per unit type I error exceeds 3.25. However, the basis of this performance is the achievement of a minimal number of type I errors, at the expense of a very high type II error rate.

In the exhaustive discussion of costs in chapter VII, a value of 3.75 was estimated for C. This exceeds the threshold of 3.25. However, the estimate in chapter VII was rather tentative; moreover, chapter VII was concerned with the training-sample period, and in fact C is likely to change over time.

Given the differential risk and personal cost to a banker of making a type I error, versus a type II, it would be rational for bankers to place a high penalty weighting on the former. This approach could cause a small number of 'blue chip' LDCs to be highly rated with, at the same time, little distinction drawn between (a) countries that, objectively, are bad risks and (b) other countries. This pattern would give rise to the observed distribution of ratings.

In conclusion, formal modelling techniques have been shown, on the chosen task-domain, to have a very high level of classificatory accuracy. They outperform 'man', when errors are equally weighted. Even with differential weighting, the formal models may have a useful role in supplementing human judgement. The cost-minimising cutoff as applied to the *Institutional Investor* system 'writes off' a significant number of creditworthy LDCs in the interest of minimising type I errors. It is clear that the multivariate models are more discriminating in this respect, and there is clearly a role for 'man with model'.

XV Conclusions

XV.1 The primary issue

The central issue addressed by this thesis is the performance of multivariate statistical models, relative to that of human judgement. The thesis is rooted in the theory of human information processing, which proposes that 'man as an intuitive statistician' is subject to a number of serious biases. Moreover human judgement, whether biased or not, is likely to be applied inconsistently. These two points provide a rationale for developing formal statistical models as an alternative decision tool. The chosen task-domain for this purpose is the prediction of the debt-servicing status of less developed countries, and the conclusion has been reached that formal models, using discriminant and logit analysis, can out-perform human judgement, represented by the rating systems of BERI, the EIU, the ICRG and the Institutional Investor. This conclusion is a key original contribution of this thesis, and a related original feature is the application of the chosen methodology in the chosen task-domain. Moreover, this is the first attempt to represent banker and expert judgement by this set of rating systems.

The man-model comparisons have been drawn both for the training-sample and holdout-sample periods, and the dominance of the models is conclusive during the former. For reasons of data availability and sample size, a holdout sample has been retained for the Institutional Investor rating alone, and relative to that rating, the dominance of the multivariate models is not conclusive during the holdout period. The rating systems are very conservative. The most conservative approach possible would be to classify all countries as 'arrears', yielding type I and II error rates of zero and 100 per cent respectively. None of the ratings, given the cost-minimising cutoffs, goes to this extreme. However, during the holdout period the Institutional Investor rating classifies 85 per cent of cases as arrears (using an *ex ante* cost-minimising cutoff), compared with a true proportion *ex post* of 61 per cent. Correspondingly, the zero type I error rate is accompanied by a type II rate of 62 per cent. This is very high, but given a large enough penalty on type I errors, this can yield a lower cost of misclassification than its closest competitor, discriminant model II, even though that model's error rates are only 8 per cent and 21 per cent (types I and II respectively). This lower-cost performance by the Institutional Investor is achieved at the expense of, in effect, 'writing off' a significant number of LDCs that turn out, *ex* post, to be creditworthy. Thus this rating system is a blunt instrument compared with the multivariate models. The conclusion to be drawn from this is that there is scope for regarding the two methods as complements, rather than alternatives: man-model interaction is valuable. This conclusion conforms with the recommendation of Makridakis (1986), in a recent survey of the forecasting literature, and provides one of a number of directions for further research that are taken up below.

Paramorphic representations of the judgemental bases of the several rating systems have been derived using the bootstrapping technique. The literature on judgemental bias and on bootstrapping models covers a wide field (see ch. II), but this is the first application in the domain of country risk. Moreover, earlier research, such as the paper by Kaplan and Urwitz (1979) on bond ratings, typically used regression whereas in this thesis AID has also been used.

The bootstrapped models have two purposes in this thesis. First, they have been used to show that the rating systems and the multivariate models have different informational inputs: the sets of variables included in the multivariate models differ from those that underpin human judgement, according to the bootstrapped models, albeit with some overlap. For example, only INVR (fixed investment/GDP) is common to discriminant model I and the regression model of the Institutional Investor rating system, each of which includes five independent variables. Overlap is consistently greater in terms of principal components, but it is still only partial: for example, in the same pairing of models, there are only two principal components in common.

The second issue addressed by the bootstrapped models is the extent to which human judgement is applied consistently. The predictive performance of human judgement, as indicated by a number of rating systems, is poor. However, that of the bootstrapped models is very similar to it. If the bootstrapped models accurately model the underlying judgemental processes, then this result implies that the problem is essentially one of bias in the judgemental process, rather than inconsistency in its application. This conclusion is supported by an intertemporal analysis of the Institutional Investor rating, which suggests that bankers are prone to biased judgement, possibly arising from use of the 'anchoring and adjustment' heuristic. In conclusion, this research supports the case for the use of better techniques, including formal multivariate models, in country credit risk assessment.

XV.2 Further conclusions

XV.2.1 Human judgement

Human judgement is represented in this thesis by several different systems for rating country risk. These systems, which have varying judgemental components, include those of BERI, the EIU, the ICRG, and the *Institutional Investor*.

The major conclusions concerning these rating systems have been detailed in section XV.1. However, two other points arise. First, given the specifications of the best-fitting regressions, the judgemental bases of the four ratings differ, but with some overlap. Secondly, it is established that expert perceptions of political and economic risk are not highly correlated: each is making a separate contribution to the overall ICRG and EIU ratings.

XV.2.2 Multivariate models

Multivariate models of country risk have been developed before, although this is the first application of AID to the problem. However, in the 'early-warning' literature the central concern has not been with comparisons between models and man: for example in the papers surveyed by Saini and Bates (1984) and Heffernan (1986), the focus is essentially on multivariate models alone.

However, even though multivariate models are only an input to the central investigations of this thesis, those reported in chapters VII to IX should nevertheless be of considerable value to lenders and regulators. Models developed independently using discriminant analysis and logit analysis turn out to have almost identical specifications, and their classificatory performance out-of-sample is both satisfactory, and similar. In line with the modelling philosophy of Hendry and Mizon (1985), the derivation of these models has followed a 'general to simple' methodology. They are congruent, and their parameterisations are parsimonious and near-orthogonal. Parameter-constancy has been tested exhaustively. According to Makridakis (1986: p.19),

> Empirical evidence indicates that large, complex or statistically sophisticated methods do not necessarily produce more accurate forecasts than smaller, or relatively simple methods.

and indeed, all the models derived here are simple, and use no more than five independent variables.

A key requirement is true <u>ex ante</u> predictive ability. A necessary condition for this to be met is that the models should not be sample-specific, and this has been demonstrated. Makridakis (1986) suggests that accuracy of within-sample fit is no guide to post-sample predictive ability. For that reason, much emphasis is placed in this thesis on the performance of the models on the holdout sample. For the discriminant and logit models, that performance is satisfactory, and shows no sign of degradation over time within the three-year horizon.

The overall conclusion is that the key 'early-warning' variables are:

NARY	Net assets ratio	(+)
DCPI	Inflation	(-)
INPS	Interest/debt service	(-)

while other important predictors are:

DGDP	GDP growth rate	(+)
INVR	Fixed investment/GDP	(+)
AS	Asian country: yes/no	(+)

where the signs indicate the direction of association with creditworthiness.

The AID model, which confirms the importance of these variables (except for 'Asia'), detects interaction effects between them. The variable INPS is only important for cases that have low values for both NARY and INVR. For countries that have a high value of NARY, the only other variable that matters is DCPI, while for other countries, DCPI is interactive with INPS and INVR.

However, despite picking up these interactions, the AID approach is not itself useful for prediction. This may be due in part to the loss of information that is inherent in the transformation of the predictors into categorical format. In conclusion, the AID approach may be used to supplement the logit and discriminant analyses, but it does not replace them.

The utility of these models is enhanced by the careful choice of dependent variable: this applies both to their use in man-model comparisons, and to their potential use by creditors and regulators. The emergence of 'arrears' is the event of interest, while earlier research has generally focused on the event of 'rescheduling'. 'Rescheduling' is essentially a proxy variable, and by specifiying 'arrears' directly, a source of misclassification error is avoided.

XV.2.3 Statistical considerations

The primary data set of 70 continuous financial and economic variables, augmented by thirteen environmental and categorical variables, covers 55 countries for a training sample 1979-1986 and a holdout sample 1987-1989. Extensive analysis of the distributional properties of accounting ratios has been reported in the accounting literature, but the corresponding research reported here is far deeper than its precursors in the area of country risk. The univariate distributional properties of the data have been investigated, against the null hypothesis of univariate normality. In many cases univariate normality has been rejected, so that multivariate normality for the data set as a whole must also be rejected. However, for the five-continuous-variable version of the discriminant model, multivariate normality could not be rejected for the set of discriminating variables.

Compared with earlier work on country risk, the 83 variables that have been used in this thesis cover a wider spectrum of information. Insight into this has been offered by a principal components analysis, in which nineteen components were shown to have eigenvalues exceeding unity, and cumulatively to represent 78.9 per cent of the total variance within the data set. Most of these components have been reified as representing important economic and financial dimensions. The fact that as many as nineteen components are retained bears out the assertion that the data cover a very wide span of information.

XV.3 The audience for this research

This research has been conducted at the boundary of several disciplines, including human information processing, economics, banking, accounting, and applied statistics. The conclusions should be of interest and use to other researchers in all these fields of enquiry, and moreover they should be of practical concern to lenders and their regulatory authorities, and indirectly to borrowers. Accurate risk assessment is clearly of concern to banks and their shareholders, to bank regulators, and to official lending agencies. It is equally of concern to prospective borrowers. If lenders are unable to distinguish between high- and low-risk countries, then countries in the former category may become burdened with debt that they will be unable to service. Alternatively, low-risk countries may be denied funds that they could use efficiently and profitably.

The research that is reported here has a practical orientation. Lenders and their regulators are concerned with the processing of information about borrowers and debtors, and to that extent this thesis in its entirety should be of interest to them. In particular, the multivariate models may be of practical use as a complement to human judgement. The discriminant and logit models are simple and easy to understand. They have been shown to perform well out-of-sample, without degradation as time passes. A possible difficulty may be the availability or quality of data for the independent variables, but this will arise in any forecasting context. In practice, forecasts will be required for the independent variables, to be used as inputs to the models. However, these variables are mainstream economic and financial variables, for which forecasts are readily available. For example, the EIU regularly produces two-year forecasts for inflation, growth, and the external account (EIU, 1988b). An unsettled question is whether a one-year time horizon is adequate for forecasting purposes, and this is taken up in section XV.4.

XV.4 <u>Directions for further research</u>

There are several respects in which this research should be extended. First, the multivariate models have focused on a rather arbitrary one-year time horizon. Looney et al (1989) have applied the Cox proportional hazards model to the problem of bank failure prediction. This approach yields an estimate of likely time to failure, and has an obvious application to country risk assessment.

Secondly, the analysis would be improved by the use of longer time-series for the BERI and EIU ratings. For BERI, availability is a problem; for the EIU, the rating is of recent vintage, and more data will become available as time goes on.

Thirdly, further analysis of political risk ratings should yield insights into the judgemental processes upon which they are based. A related matter is the possibility of using demographic and socio-political indicators as inputs to the multivariate models.

Fourthly, further research is necessary on the relative cost of the two types of misclassification error, including its behaviour over time.

The fifth issue concerns the analysis of the distributional properties of the data. In particular, a direction for further research is suggested by the discordance between the results of the various univariate tests of normality. This arose in chapter IV, where it was found that the various graphical techniques and formal tests showed no consistent pattern of acceptance/rejection of the normality hypothesis.

Finally, there is scope for combination of different forecasts, drawn from within and also between the two categories of forecast: formal quantitative, versus judgemental. One possibility here is to include the Institutional Investor system's ratings along with financial and economic variables, as input to the multivariate modelling process. The purpose of combining is to improve accuracy; Clemen (1989) gives a recent survey, while Mascarenhas and Sand (1989) describe an application to country risk assessment. Makridakis (1986) has argued that

> People should concentrate their efforts in predicting systematic changes...Quantitative forecasting should be viewed as the best way of identifying established patterns or relationships [but] it cannot predict systematic changes from established patterns. Judgemental forecasts are complementary to quantitative ones and effective ways of integrating the two must be developed.

This conclusion finds further support in Klein (1981) and Makridakis and Wheelwright (1989).

References and bibliography

Aldrich J.H. and Nelson F.D. (1984), Linear probability, logit, and probit models. (London: Sage).

Amemiya T. (1981), 'Qualitative response models: a survey'. Journal of Economic Literature XIX, 1483-1536.

Amemiya T. (1985), Advanced econometrics. (Oxford: Basil Blackwell).

Amemiya T. and Powell J. (1983), 'A comparison of the logit model and normal discriminant analysis when the independent variables are binary'. In: Karlin S., Amemiya T. and Goodman L.A. (eds.), Studies in econometrics, time series, and multivariate statistics. (New York: Academic Press). pp. 3-30.

Ascher W. (1982), 'Political forecasting: the missing link'. Journal of Forecasting 1, 227-239.

Ascher W. and Overholt W.H. (1983), Strategic planning and forecasting. (New York: Wiley).

Aziz A., Emanuel D. and Lawson G. (1988), 'Bankruptcy prediction - an investigation of cash flow based models'. Journal of Management Studies 25, 419-437.

Balestra P. and Nerlove M. (1966), 'Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas'. Econometrica 34, 585-612.

BarNiv R. and Hershbarger R. (1990), 'Classifying financial distress in the life insurance industry'. Journal of Risk and Insurance 57, 110-136.

BarNiv R. and Raveh A. (1989), 'Identifying financial distress: a new nonparametric approach'. Journal of Business Finance and Accounting 16, 361-383.

Barnes P. (1982), 'Methodological implications of non-normally distributed financial ratios'. Journal of Business Finance and Accounting 2, 51-62.

Barnes P. (1987), 'The analysis and use of financial ratios: a review article'. Journal of Business Finance and Accounting 14, 449-461.

Barnett V. and Lewis L. (1984), Outliers in statistical data (2nd ed.). (Chichester: Wiley).

Berg A., Sachs J. and Selowsky M. (1988), 'The debt crisis: structural explanations of country performance'. Journal of Development Economics 29, 271-309.

Berkeley D. and Humphreys P. (1982), 'Structuring decision problems and the bias heuristic'. Acta Psychologica 50, 201-252.

Betts J. and Belhoul D. (1987), 'The effectiveness of incorporating stability measures in company failure models'. Journal of Business Finance and Accounting 14, 323-334.

Bird G. (1986), 'New approaches to country risk'. Lloyds Bank Review (October), 1-16.

Bird G. (1989), Third world debt. (Aldershot: Elgar).

Bowman E.H. (1963), 'Consistency and optimality in managerial decision making'. Management Sciences 9, 310-321.

Box G.E.P. and Cox D.R. (1964), 'An analysis of transformations'. Journal of The Royal Statistical Society Series B 26, 211-243

Brady N.F. (1989), 'A reexamination of the debt strategy'. In Fried E.R. and Trezise P.H. (eds.) (1989).

Brewer T.L. (ed.) (1985), Political risks in international business. (New York: Praeger).

Brunswik E. (1952), The conceptual framework of psychology. (Chicago: University of Chicago Press).

Buchman T.A. (1985), 'An effect of hindsight on predicting bankruptcy with accounting information'. Accounting, Organizations and Society 10, 267-278.

Burton F.N. and Inoue H. (1983), 'Country risk evaluation methods: a survey of systems in use'. The Banker 133, 41-43.

Burton F.N. and Inoue H. (1985a), 'An appraisal of the early warning indicators of sovereign loan default in country risk evaluation systems'. Management International Review 25, 45-56.

Burton F.N. and Inoue H. (1985b), 'The influence of country risk factors on interest rate differentials on bank lending to sovereign borrowers'. Applied Economics 17, 491-507.

Calverly J. (1985), Country risk analysis. (London: Butterworths).

Camerer C. (1981), 'General conditions for the success of bootstrapping models'. Organizational Behaviour and Human Performance 27, 411-422.

Casey C. and Selling T.I. (1986), 'The effect of task predictability and prior probability disclosure on judgement quality and confidence'. The Accounting Review LXI, 302-317.

Chow G.C. (1960), 'Tests of equality between sets of coefficients in two linear regressions'. Econometrica 28, 591-605.

Citicorp (1987), Financial review and form 10Q - second quarter 1987. (New York: Citicorp).

Clemen R.T. (1989), 'Combining forecasts: a review and annotated bibliography'. International Journal of Forecasting 5, 559-583.

Cline W.R. (1983), International debt and the stability of the world economy. (Washington D.C.: Institute for International Economics).

Cline W.R. (1984), International debt: systemic risk and policy response. (Washington D.C.: Institutional for International Economics).

Cochran W.G. (1963), Sampling techniques. (New York: Wiley).

Conover W.J. (1980), Practical nonparametric statistics (2nd ed.). (New York: Wiley).

Cooper J.C.B. (1985), Country creditworthiness: a canonical correlation analysis. (Glasgow: Glasgow College of Technology, working paper).

Cramer J.S. (1991), The logit model - an introduction for economists. (Sevenoaks: Edward Arnold).

Crask M.R. and Perreault W.D. (1977), 'Validation of discriminant analysis in marketing research'. Journal of Marketing Research XIV, 60-68.

Dawes R.M. (1971), 'A case study of graduate admissions: application of three principles of human decision making'. American Psychologist 26, 180-188.

Dawes R.M. (1979), 'The robust beauty of improper linear models in decision making'. American Psychologist 34, 571-582.

Dawes R.M. and Corrigan B. (1974), 'Linear models in decision making'. Psychological Bulletin 81, 95-106.

de la Torre J. and Neckar D.H. (1988), 'Forecasting political risks for international operations'. International Journal of Forecasting 4, 221-241.

de Saint Phalle T. (1978), Country Risk Evaluation. (New York: Allied Bank International).

DeWitt R.P. and Madura J. (1986), 'How banks assess country risk'. World of Banking 5, 28-33.

Dhonte P. (1975), 'Describing external debt situations: a roll-over approach'. IMF Staff Papers 22, 159-186.

Dichtl E. and Kogelmayr H.-G. (1986), 'Country risk ratings'. Management International Review 26, 4-11.

Dillon W.R. and Westin S. (1982), 'Scoring frequency data for discriminant analysis: perhaps discrete procedures can be avoided'. Journal of Marketing Research XIX, 44-56.

Dixon W.J. (ed.) (1988), BMDP statistical software manual. (Berkeley: University of California Press).

Dopuch N., Holthausen R.W. and Leftwich R.W. (1987), 'Predicting audit qualifications with financial and market variables'. The Accounting Review LXIII, 431-454.

Doukas J. (1986), 'Bankers versus bankruptcy prediction models: an empirical investigation'. Applied Economics 18, 479-493.

EIU - The Economist Intelligence Unit (1986), Country Report: Thailand, no.3. (London: EIU).

EIU - The Economist Intelligence Unit (1987), Chile: Country report, no.2. (London: EIU).

EIU - The Economist Intelligence Unit (1988a), Cameroon, C.A.R., Chad: Country Report, no. 2. (London: EIU).

EIU - The Economist Intelligence Unit (1988b), Handbook for use with the country credit risk service. (London: EIU).

EIU - The Economist Intelligence Unit (1989a), Economic Risk Service - Risk Ratings Review. (London: EIU). (October).

EIU - The Economist Intelligence Unit (1989b), Ghana: Country Report, no.3. (London: EIU).

EIU - The Economist Intelligence Unit (1989c), Guatemala, El Salvador, Honduras: Country Profile, 1989-1990. (London: EIU).

EIU - The Economist Intelligence Unit (1989d), Syria: Country Report, no.2. (London: EIU).

EIU - The Economist Intelligence Unit (1990a), Ghana: Country Profile, 1990-1991. (London: EIU).

EIU - The Economist Intelligence Unit (1990b), Iran: Country Profile, 1990-1991. (London: EIU).

EIU - The Economist Intelligence Unit (1990c), Uruguay, Paraguay: Country Profile, 1990-1991. (London: EIU).

EIU - The Economist Intelligence Unit (1990d), Sri Lanka: Country Report, no.1. (London: EIU).

Edwards E. (1968), 'Conservatism in human information processing'. Ch. 25 in: Kahneman, Slovic and Tversky (1982).

Edwards S. (1984), 'LDC foreign borrowing and default risk: an empirical investigation'. The American Economic Review 74, 726-734.

Edwards S. and Larrain F. (eds.) (1989), Debt, adjustment and recovery - Latin America's prospects for growth and development. (Oxford: Basil Blackwell).

Efron B. (1979), 'Bootstrap methods: another look at the jackknife'. Annals of Statistics 7, 1-26.

Efron B. (1982), The jackknife, the bootstrap, and other resampling plans. (Philadelphia: Society for Industrial and Applied Mathematics.)

Efron B. (1983), 'Estimating the error rate of a prediction rule: improvement on cross validation'. Journal of the American Statistical Association 78, 316-331.

Eichengreen B. and Portes R. (1985), Debt and default in the 1930s: causes and consequences. (London: Centre for Economic Policy Research, discussion paper no. 75).

Einhorn H.J. and Hogarth R.M. (1975), 'Unit weighting schemes for decision making'. Organizational Behaviour and Human Performance 13, 171-192.

Einhorn H.J. and Hogarth R.M. (1982), 'Prediction, diagnosis, and causal thinking in forecasting'. Journal of Forecasting 1, 23-36.

Eisenbeis R.A. (1977), 'Pitfalls in the application of discriminant analysis in business finance and economics'. The Journal of Finance XXXII, 875-899.

Eisenbeis R.A. and Avery R.B. (1972), Discriminant analysis and classification procedures. (Lexington, MA.: Heath).

Elton E.J. and Gruber M.J. (1987), Modern portfolio theory and investment analysis (3rd ed). (New York: Wiley).

Emerson J.D. and Soto M.A. (1983). 'Transforming data'. Ch. 4 in: Hoaglin, Mosteller and Tukey (1983).

Ensor R.J. (ed.) (1981), Assessing country risk. (London: Euromoney Publications).

Everitt B. (1980), Cluster Analysis (2nd ed.). (Aldershot: Gower).

Ezzamel M. and Mar-Molinero C. (1990), 'The distribution properties of financial ratios in U.K. manufacturing companies'. Journal of Business Finance and Accounting 17, 1-29.

Ezzamel M., Brodie J. and Mar-Molinero C. (1987), 'Financial patterns of U.K. manufacturing companies'. Journal of Business Finance and Accounting 14, 519-536.

Fama E. (1965), 'Random walks in stock market prices'. Financial Analysts' Journal 21, 55-60.

Feder G. (1980), 'Economic growth, foreign loans and debt servicing capacity of developing countries'. Journal of Development Studies 16, 352-368.

Feder G. and Just R. (1980), 'A model for analysing lenders' perceived risk'. Applied Economics 12, 125-144.

Feder G., Just R. and Ross K. (1981), 'Projecting debt servicing capacity of developing countries'. Journal of Financial and Quantitative Analysis XVI, 651-669.

Feder G. and Ross K. (1982). 'Risk assessments and risk premiums in the Eurodollar market'. The Journal of Finance XXXVII, 679-691.

Fischoff B. (1980), 'For those condemned to study the past: heuristics and biases in hindsight'. Ch. 23 in: Kahneman, Slovic and Tversky (1982).

Foster G. (1986), Financial statement analysis (2nd ed.). (London: Prentice-Hall).

Frank C. and Cline W. (1971), 'Measurements of debt servicing capacity: an application of discriminant analysis'. Journal of International Economics 1, 327-344.

Frank R.E., Massy W.F. and Morrison D.G. (1965), 'Bias in multiple discriminant analysis'. Journal of Marketing Research II, 250-258.

Frecka T.J. and Hopwood W.S. (1983), 'The effects of outliers on the cross-sectional properties of financial ratios'. The Accounting Review LVIII, 115-128.

Fried E.R. and Trezise P.H. (eds.) (1989), Third world debt - the next phase. (Washington D.C.: The Brookings Institution).

Frydman H., Altman E.I. and Kao D-L. (1985), 'Introducing recursive partitioning for financial classification: the case of financial distress'. The Journal of Finance XL, 269-291.

Gessner G., Kamakura W.A., Malhotra N.K. and Zmijewski M.E. (1988), 'Estimating models with binary dependent variables: some theoretical and empirical observations'. Journal of Business Research 16, 49-65.

Gilbert C.L. (1986), 'Professor Hendry's econometric methodology'. Oxford Bulletin of Economics and Statistics 48, 283-307.

Gilbert L., Menon K. and Schwartz K. (1990), 'Predicting bankruptcy for firms in financial distress'. Journal of Business Finance and Accounting 17, 161-171.

Goldberg E.S. and Haendl D. (1987), On edge - international banking and country risk. (New York: Praeger).

Goldberg L.R. (1970), 'Man versus model of man: a rationale, plus some evidence, for a method of improving on clinical inferences'. Psychological Bulletin 73, 422-432.

Goldberg L.R. (1976), 'Man versus model of man: just how conflicting is that evidence?'. Organizational Behaviour and Human Performance 16, 13-22.

Gombola M.J., Haskins M., Ketz J. and Williams D (1987), 'Cash flow in bankruptcy prediction'. Financial Management 16, 55-65.

Gong G. (1986), 'Cross validation, the jackknife and the bootstrap'. Journal of the American Statistical Association 81, 108-113.

Goodman S.H. (1977), 'How the big US banks really evaluate sovereign risk'. Euromoney (February), 105-110.

Goodman S.H. (ed.) (1978), Financing and risk in developing countries. (New York: Praeger).

Gordon A.D. (1987), 'A review of hierarchical classification'. Journal of The Royal Statistical Society Series A 150, 119-137.

Green P.E. (1978), Analyzing multivariate data. (Hinsdale Ill.: The Dryden Press).

Griffith-Jones S. (1989), 'The bargaining position of debtor nations'. Ch. 3 in: Bird (1989).

Gruson M. and Reisner R. (eds.) (1984), Sovereign lending: managing legal risk. (London: Euromoney Publications).

Gujarati D. (1970a), 'Use of dummy variables in testing for equality between sets of coefficients in two linear regressions: a note'. The American Statistician (February), 50-52.

Gujarati D. (1970b), 'Use of dummy variables in testing for equality between sets of coefficients in two linear regressions: a generalization'. The American Statistician (December), 18-22.

Gupta Y.P., Rao R.P. and Bagchi P. (1990), 'Linear goal programming as an alternative to multivariate discriminant analysis: a note'. Journal of Business Finance and Accounting 17, 593-598.

Gutmann P. (1980), 'Assessing country risk'. National Westminster Bank Quarterly Review (May), 58-68.

Hallett M.P.and Taffler R.J. (1982), Balance sheet financing in the U.K.: an empirical analysis. (London: The City University Business School, working paper no. 44).

Hand D.J. (1981), Discrimination and classification. (New York: Wiley).

Haner F.T. (1980), Global business strategy for the 1980s. (New York: Praeger).

Harman H. (1976), Modern factor analysis (3rd ed.). (Chicago: University of Chicago Press).

Hauck W.W. and Donner A. (1977), 'Wald's test as applied to hypotheses in logit analysis'. Journal of the American Statistical Association 72, 851-853.

Hawkins D.M. (1980), Identification of outliers. London: Chapman and Hall).

Heffernan S.A. (1986), Sovereign risk analysis. (London: Allen and Unwin).

Hendry D.F. and Mizon G.E. (1985), Procrustean econometrics: or stretching and squeezing data. (London: Centre for Economic Policy Research, discussion paper no. 68).

Hoaglin D.C., Mosteller F. and Tukey J.W. (eds.) (1983), Understanding robust and exploratory data analysis. (New York: Wiley).

Hoffman P.J. (1960), 'The paramorphic representation of clinical judgement'. Psychological Bulletin 57, 116-131.

Hogarth R. (1987), Judgement and choice (2nd ed.). (Chichester: Wiley).

Hopwood W., McKeon J. and Mutchler J. (1989), 'A test of the incremental explanatory power of opinions qualified for consistency and uncertainty'. The Accounting Review LXIV, 28-48.

Hosmer D.W. and Lemeshow S. (1989), Applied logistic regression. (New York: Wiley).

ICRG (1984), Key to the guide: an introduction to the methodology of the International Country Risk Guide. (New York: International Reports).

IMF - International Monetary Fund (1981, 1986b, 1988b, 1989b), World Economic Outlook. (Washington D.C.: IMF).

IMF - International Monetary Fund (1983), Recent multilateral debt restructurings with official and bank creditors. (Washington D.C.: IMF).

IMF - International Monetary Fund (1985), Recent developments in external debt restructuring. (Washington D.C.: IMF).

IMF - International Monetary Fund (1986a, 1988a, 1989a), International capital markets - developments and prospects. (Washington D.C.: IMF).

IMF - International Monetary Fund (1988c), Multilateral official debt rescheduling. (Washington D.C.: IMF).

IMSL (1989a), User's manual: IMSL Math/Library. (Houston: IMSL).

IMSL (1989b), User's manual: IMSL Sfun/Library. (Houston: IMSL).

IMSL (1989c), User's manual: IMSL Stat/Library. (Houston: IMSL).

International Working Group (1988), External debt - definition, statistical coverage and methodology. Report by an international working group on external debt statistics. (Paris: World Bank, IMF, BIS and OECD).

Jalilvand A. (1987), Euromoney and Institutional Investor country risk indices and predictions of debt servicing problems in developing countries. (Montreal: Concordia University, working paper).

Jarque C.M. and Bera A.K. (1980), 'Efficient tests for normality, homoscedasticity and serial independence of regression residuals'. Economics Letters 6, 255-259.

Jodice D.A. (1985), Political risk assessment. (London: Greenwood).

Johnson K.J. (1980), Pitfalls in applying principal components analysis and factor analysis in finance. (Manchester: University of Manchester, Department of Management Science, Occasional Paper no. 8008).

Johnson R.A., Srinivasan V. and Bolster P.J. (1990), 'Sovereign debt ratings: a judgemental model based on the analytic hierarchy process'. Journal of International Business Studies 21, 95-117.

Johnson R.A. and Wichern D.W. (1982), Applied multivariate statistical analysis. (London:Prentice Hall).

Johnson W.B. (1983), '"Representativeness" in judgemental predictions of corporate bankruptcy'. The Accounting Review LVIII, 78-97.

Jolliffe I.T. (1989), 'Rotation of ill-defined principal components'. Applied Statistics 38, 139-147.

Kahneman D., Slovic P. and Tversky A. (eds.) (1982), Judgement under uncertainty: heuristics and biases. (Cambridge: Cambridge University Press).

Kahneman D. and Tversky A. (1979), 'Intuitive prediction biases and corrective procedures'. Ch. 30 in: Kahneman, Slovic and Tversky (1982).

Kaiser H.F. (1959), The application of electronic computers to factor analysis. (American Psychological Association: symposium on the application of computers to psychological problems).

Kaplan R.S. and Urwitz G. (1979), 'Statistical models of bond ratings: a methodological inquiry'. Journal of Business 52, 231-261.

Kendall M.G. and Stuart A. (1979), The advanced theory of statistics (4th ed.). (London: Griffin).

Kharas H. (1984), 'The long-run creditworthiness of developing countries: theory and practice'. Quarterly Journal of Economics 99, 415-439.

Klein L.R. (1947), 'The use of econometric models as a guide to economic policy'. Econometrica 15, 111-151.

Klein L.R. (1981), Econometric models as guides for decision-making. (London: Macmillan).

Kmenta J. (1971), Elements of econometrics. (New York: Macmillan).

Koziol J.A. (1982), 'A class of invariant procedures for assessing multivariate normality'. Biometrika 69, 423-427.

Krayenbuehl T.E. (1988), Country risk - assessment and monitoring (2nd ed.). (Cambridge: Woodhead-Faulkner).

Krueger A.O. (1987), 'Origins of the developing countries' debt crisis: 1970-1982'. Journal of Development Economics 27, 165-187.

Krzanowski W.J. (1975), 'Discrimination and classification using both binary and continuous variables'. Journal of the American Statistical Association 70, 782-790.

Krzanowski W.J. (1987), 'Selection of variables to preserve data structure, using principal components'. Applied Statistics 36, 22-33.

Kugler P. (1984), 'Economic indicators and debt reschedulings 1977-1981: empirical results from a logit model'. Rivesta Internazionale di Scienze Economiche e Commerciali ,992-1003.

Lachenbruch P.A. (1975), Discriminant analysis. (New York: Hafner).

Lachenbruch P.A. and Mickey R. (1968), 'Estimation of error rates in discriminant analysis'. Technometrics 10, 1-11.

Lachenbruch P.A., Sneeringer C.A. and Revo L.T. (1973), 'Robustness of the linear and quadratic discriminant function to certain types of non-normality'. Communications in Statistics 1, 39-56.

Lau, A. H-L. (1987), 'A five-state financial distress prediction model'. Journal of Accounting Research 25, 127-138.

Libby R. (1976a), 'Man versus model of man: some conflicting evidence'. Organizational Behaviour and Human Performance 16, 1-12.

Libby R. (1976b), 'Man versus model of man: the need for a nonlinear model'. Organizational Behaviour and Human Performance 16, 23-26.

Libby R. and Lewis B.L. (1982), 'Human information processing research in accounting: the state of the art in 1982'. Accounting, Organizations and Society 7, 231-285.

Lloyd-Ellis H., McKenzie G.W. and Thomas S.H. (1989), 'Using country balance sheet data to predict debt rescheduling'. Economics Letters 31, 173-177.

Lo A.W. (1986), 'Logit versus discriminant analysis'. Journal of Econometrics 31, 151-178.

Lomax D.F. (1986), The developing country debt crisis. (London: Macmillan).

Looney S.W., Wansley J.W. and Lane W.R. (1989), 'An examination of misclassifications with bank failure prediction models'. Journal of Economics and Business 41, 327-33.

McLeay S.J. (1987), The statistical properties of financial ratios: European evidence. (University of Lancaster: unpublished Ph.D. thesis).

Maddala G.S. (1977), Econometrics. (New York: McGraw-Hill).

Maddala G.S. (1983), Limited-dependent and qualitative variables in econometrics. (Cambridge: Cambridge University Press).

Maddala G.S. (1987), 'Limited dependent variable models using panel data'. The Journal of Human Resources XXII, 307-338.

Maddala G.S. (1988), Introduction to econometrics. (New York: Macmillan).

Makridakis S. (1986), 'The art and science of forecasting'. International Journal of Forecasting 2, 15-39.

Makridakis S. and Wheelwright S.C. (1987), The handbook of forecasting - a manager's guide (2nd ed.). (New York: Wiley).

Makridakis S. and Wheelwright S.C. (1989), Forecasting methods for management (5th ed.). (New York: Wiley).

Mann F.A. (1984), 'Sovereign immunity under English Law'. Ch.9 in: Gruson and Reisner (1984).

Marais M.L., Patell J.M. and Wolfson M.A. (1984), 'The experimental design of classification models: an application of recursive partitioning and bootstrapping to commercial bank loan classifications'. Journal of Accounting Research 22 (Supplement), 87-118.

Mardia K.V. (1974), 'Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies'. Sankhya B 36, 115-128.

Mascarenhas B. and Sand O.C. (1985), 'Country-risk assessment systems in banks: patterns and performance'. Journal of International Business Studies 16, 19-35.

Mascarenhas B. and Sand O.C. (1989), 'Combination of forecasts in the international context: predicting debt reschedulings'. Journal of International Business Studies 20, 539-552.

Mayo A.L. and Barrett A.G. (1978), 'An early-warning model for assessing developing country risk'. In: Goodman (1978).

Mear R. and Firth M. (1987), 'Cue usage and self-insight of financial analysts'. The Accounting Review LXII, 176-182.

Melvin M. and Schlagenhauf D. (1985), 'A country risk index: econometric formulation and an application to Mexico'. Economic Inquiry XXIII, 601-619.

Mensah Y.M. (1983), 'The differential bankruptcy predictive ability of specific price level adjustments: some empirical evidence'. The Accounting Review LVIII, 228-246.

Mensah Y.M. (1984), 'An examination of the stationarity of multivariate bankruptcy prediction models: a methodological study'. Journal of Accounting Research 22, 380-395.

Merrill J. (1982), 'Country risk analysis'. Columbia Journal of World Business 17, 88-91.

Meyer P.L. (1970), Introductory probability and statistical applications. (Reading, Mass.: Addison Wesley).

Milivojevic M. (1985), The debt rescheduling process. (New York: St. Martin's Press).

Miller R.G. (1974), 'The jackknife - a review'. Biometrika 61, 1-15.

Mills R.H. and Terrell H.S. (1984), 'How front-end fees on syndicated Euroloans are determined'. The Banker 134, 27-33.

Mistry P.S. (c.1989), African debt: the case for relief for sub-Saharan Africa. (Oxford: Oxford International Associates).

Morgan J.B. (1986), 'A new look at debt rescheduling indicators and models'. Journal of International Business Studies 17, 37-54.

Morgan J.N. and Sonquist J.A. (1963), 'Problems in the analysis of survey data, and a proposal'. Journal of the American Statistical Association 58, 415-435.

Morrison D.G. (1969), 'On the interpretation of discriminant analysis'. Journal of Marketing Research VI, 156-163.

Moses D. and Liao S. (1987), 'On developing models for failure prediction'. Journal of Commercial Bank Lending 69, 27-38.

Mumpower J.L., Livingston S. and Lee T.J. (1987), 'Expert judgements of political riskiness'. Journal of Forecasting 6, 51-65.

Mundlak Y. (1978), 'On the pooling of time series and cross section data'. Econometrica 46, 69-85.

Nichols B.W. (1984), 'Sovereign debtors under US immunity law'. Ch.7 in: Gruson and Reisner (1984).

Nisse J. (1987), 'Political risks: the benefits of hindsight'. The Banker 137, 35-39.

Nunnenkamp P. (1988), The international debt crisis of the third world: causes and consequences for the world economy. (Brighton: Wheatsheaf).

OECD - Organisation for Economic Co-operation and Development (1983, 1987), Financing And External Debt Of Developing Countries. (Note: dates given here and cited are those of the title.)

Ozler S. (1989), 'On the relation between rescheduling and bank value'. American Economic Review 79, 1117-1131.

PC-MDS (1988), Multidimensional statistics package. (Provo: Institute of Business Management, Brigham Young University).

Peters S.C. and Freedman D.A. (1985), 'Using the bootstrap to evaluate forecasting equations'. Journal of Forecasting 4, 251-262.

Peterson C.R. and Beach L.R. (1967), 'Man as an intuitive statistician'. Psychological Bulletin 68, 29-46.

Pietrabassa E. (1987), 'Management of banks' international lending: country risk analysis and country exposure measurement and control'. European Journal of Operational Research 30, 42-47.

Pindyck R.S. and Rubinfeld D.L. (1981), Econometric models and economic forecasts (2nd ed.). (London: McGraw-Hill).

Political Risk Services (1990), Political Risk Letter. (Syracuse, N.Y.: Political Risk Services). February 1.

Portes R. (1986), Sovereign borrowing, debt and default: the lessons of history for current policies. (Dublin: University College Dublin, Department of Political Economy: policy paper no. 19).

Press S.J. and Wilson S. (1978), 'Choosing between logistic regression and discriminant analysis'. Journal of the American Statistical Association 73, 699-705.

Rice G. and Mahmoud E. (1990), 'Political risk forecasting by Canadian firms'. International Journal of Forecasting 6, 89-102.

Rohatgi V.K. (1984), Statistical inference. (New York: Wiley).

Rubin S.M. (1987), Guide to debt equity swaps. (London: The Economist Publications, special report no. 1104).

Sachs J. and Kyle S. (1985), Developing country debt and the market value of large commercial banks. (Washington D.C.: National Bureau of Economic Research, NBER working paper no. 1470).

Sachs J.D. (ed.) (1989), Developing country debt and the world economy. (Chicago: The University of Chicago Press).

Saini K.G. and Bates P.S. (1978), Statistical techniques for determining debt-servicing capacity for developing countries: analytical review of the literature and further empirical results. (New York: Federal Reserve Bank of New York, research paper no. 7818).

Saini K.G. and Bates P.S. (1984), 'A survey of the quantitative approaches to country risk analysis'. Journal of Banking and Finance 8, 341-356.

Saloman Brothers Inc. (1987), Indicative prices for less developed country bank loans. (New York: Saloman Brothers).

Sargen N. (1977), 'Economic indicators and country risk appraisal'. Economic Review: Federal Reserve Bank of San Francisco 4, 19-35.

Schepanski A. (1983), 'Tests of theories of information processing behaviour in credit judgement'. The Accounting Review LVIII, 581-599.

Schmidt R. (1984), 'Early warning of debt rescheduling'. Journal of Banking and Finance 8, 357-370.

Senkiw R. and Johannson R. (1986), 'Political environment forecasts in an international banking environment'. World of Banking 5, 24-27.

Sevon G. (1984), 'Cognitive maps of past and future economic events'. Acta Psychologica 56, 71-79.

Shanteau J. (1989), 'Cognitive heuristics and biases in behavioural auditing: review, comments and observations'. Accounting, Organizations and Society 14, 165-177.

Singh H. (1988), 'Investigating the compatibility of econometric forecasts and subjective expectations: a suggested framework'. Journal of Economic Psychology 9, 233-249.

Slovic P., Fischoff B. and Lichtenstein S. (1980), 'Facts versus fears: understanding perceived risk'. Ch. 33 in: Kahneman, Slovic and Tversky (1982).

Slovic P., Fleissner D. and Bauman W.S. (1972), 'Analyzing the use of information in investment decision making: a methodological proposal'. Journal of Business 45, 283-301.

Slovic P. and Lichtenstein S. (1971), 'Comparison of Bayesian and regression approaches to the study of human information processing in judgement'. Organizational Behaviour and Human Performance 6, 649-744.

Solberg R.L. (1988), Sovereign rescheduling - risk and portfolio management. (London: Unwin Hyman).

Somerville R.A. (1985), Country risk: an early warning model of debt rescheduling. (London: The City University - unpublished dissertation for M.B.A. degree).

Somerville R.A. and Taffler R.J. (1986), 'Country risk: an early warning model of debt rescheduling'. In: Proceedings of the 13th annual meeting of the European Finance Association.

Sonquist J.A. and Morgan J.N. (1964), The detection of interaction effects. (Ann Arbor, Michigan: Institute for Social Research, University of Michigan).

Spiegelhalter D.J. and Knill-Jones R.P. (1984), 'Statistical and knowledge-based approaches to clinical decision-support systems'. Journal of The Royal Statistical Society Series A 147, 35-77.

SPSS X (1986), User's guide. (London: McGraw-Hill).

Srinivasan V. and Kim Y.H. (1987), 'Credit granting: a comparative analysis of classification procedures'. The Journal of Finance XLII, 665-683.

Standard and Poor's Ratings Group (1990), Standard and Poor's International Creditweek. (New York: McGraw-Hill). November 5.

Sudarsanam P.S. and Taffler R.J. (1985), 'Industrial classification in UK capital markets: a test of economic homogeneity'. Applied Economics 17, 291-308.

Suratgar D. (1984) (ed.), Default and rescheduling - corporate and sovereign borrowers in difficulty. (London: Euromoney Publications).

Suttle P. (1989), 'The problem of forecasting debt servicing difficulties'. Ch. 2 in: Bird (1989).

Taffler R.J. (1982), 'Forecasting company failure in the UK using discriminant analysis and financial ratio data'. Journal of The Royal Statistical Society Series A 145, 342-358.

Taffler R.J. (1984a), 'Empirical models for the monitoring of UK corporations'. Journal of Banking and Finance 8, 199-227.

Taffler R.J. (1984b), 'Statistical techniques in stock selection: cognitive biases and stock market performance'. (Paper read at fourth international symposium on forecasting, London Graduate School of Business Studies).

Taffler R.J. (1986), Z-score analysis: does it work?. (Paper read at the south-east area meeting of the British Accounting Association, September 1986).

Taffler R.J. and Abassi B. (1984), 'Country risk: a model for predicting debt servicing problems in developing countries'. Journal of The Royal Statistical Society Series A 147, 541-568.

Taffler R.J. and Sudarsanam P.S. (1980), 'Auditing the board'. Managerial Finance 5, 127-147.

Taylor C.L. and Jodice D.A. (eds.) (1983), World handbook of political and social indicators (3rd ed.). (New Haven: Yale University Press).

Turk H. (1985), 'How Eximbank analyzes country risk'. ABA Banking Journal 77, 129-132.

Tversky A. and Kahneman D. (1974), 'Judgement under uncertainty: heuristics and biases'. Ch. 1 in: Kahneman, Slovic and Tversky (1982).

United Nations (1988), Handbook of International Trade and Development Statistics. (New York: United Nations).

United States Code 1982 ed.. Title 28, Judiciary and judicial procedure. (Washington D.C.: GPO).

Watson C.J. (1990), 'Multivariate distributional properties, outliers and transformations of financial ratios'. The Accounting Review LXV, 682-695.

Westin R.B. (1974), 'Predictions from binary choice models'. Journal of Econometrics 2, 1-16.

White K.J. (1982), SHAZAM - an econometrics computer program - version 4.5. (Vancouver: Department of Economics, University of British Columbia).

World Bank (1988-89, 1989-90, and 1990-91), World Debt Tables. (Washington D.C.: The World Bank). Note: dates cited and given here are those of the title.

Zmijewski M.E. (1984), 'Methodological issues related to the estimation of financial distress prediction models'. Journal of Accounting Research 22 (supplement), 59-82.

Serial publications consulted

Country Reports (quarterly) and Country Profiles (annual). (London: The Economist Intelligence Unit).

Euromoney (monthly). (London: Euromoney Publications).

Financing and External Debt of Developing Countries (annual). (Paris: Organisation For Economic Co-operation And Development).

Institutional Investor (monthly). (New York: Institutional Investor).

International Banking and Financial Market Developments (quarterly). (Basle: Bank For International Settlements).

International Country Risk Guide (ICRG) (monthly). (New York: International Reports).

International Financial Statistics (monthly). (Washington D.C.: IMF).

International Financing Review (weekly). (London: IFR Publishing).

The Economist (weekly). (London: The Economist).

The Financial Times (daily). (London: The Financial Times).

The Maturity and Sectoral Distribution of International Bank Lending (half-yearly). (Basle: Bank For International Settlements).

World Debt Tables (annual). (Washington D.C.: The World Bank).