



ELSEVIER

Journal of Banking & Finance 19 (1995) 281–297

Journal of
BANKING &
FINANCE

Banker judgement versus formal forecasting models: The case of country risk assessment

R.A. Somerville ^{a,*}, R.J. Taffler ^b

^a *Department of Economics, Trinity College, University of Dublin, Dublin 2, Ireland*

^b *City University Business School, London, UK*

Received August 1992; final version received September 1993

Abstract

The allocation of credit to less-developed countries depends upon lenders' judgements of country risk. Research in psychology suggests that human judgement may be prone to bias. This paper uses the *Institutional Investor* country credit ratings as indicators of banker judgement. In terms of their ability to predict the emergence of arrears on external debt-service, it is shown that bankers are overly pessimistic about the creditworthiness of less-developed countries. Multivariate statistical models have a higher overall predictive accuracy, although they may arguably be outperformed by banker judgement when allowance is made for the differential costs of type I and type II errors.

Keywords: Country risk; Forecasting; Logit; Discriminant; Institutional investor; Banker judgement

JEL classification: F34

1. Introduction

Bankers assess country credit-risk using a range of techniques, from formal statistical models to informal judgemental methods – for example, see Calverley (1990). These assessments are a crucial part of the process of credit-allocation to less-developed countries and therefore their accuracy is a matter of major impor-

The authors are grateful to Mina Toksoz of *The Economist Intelligence Unit* for supplying data.

* Corresponding author. Tel. (353-1-) 608 1824, Fax (353-1-) 677 2503.

tance, both to borrowers and lenders. The use of multivariate statistical models for predicting debt-servicing difficulties among less-developed countries (LDCs) is well established in the academic literature,¹ and the central concern of this paper is whether formal statistical models can outperform judgementally based methods of country-risk assessment, and thereby enhance the efficiency of credit allocation.

A large body of psychological research is concerned with the hypothesis that 'man as an intuitive statistician' (Peterson and Beach, 1967) is liable to incur significant biases in the use of information.² Tversky and Kahneman (1982) describe several heuristics (i.e. 'rules of thumb') that are commonly used in the making of judgements about the probabilities of occurrence of uncertain events. These heuristics are arguably a source of bias, in that each may be less than perfectly correlated with the determinants of the events in question. Clearly, there are dangers in basing general conclusions on evidence from one, or a few, task environments. Similarly, evidence gained from laboratory experiments may not necessarily yield valid conclusions about the decision-making ability of skilled and experienced agents making real decisions. However, evidence to support the existence of judgemental biases comes from studies of actual decision processes, as well as laboratory experiments, and both types of evidence are drawn from a wide range of task domains, including finance and accounting: e.g. Slovic (1972), Libby and Lewis (1982), and Shefrin and Statman (1985). Nonetheless, recent research, e.g. Shanteau (1989), suggests that the 'heuristics and biases' conclusions are not necessarily as clear-cut as was previously believed. In fact, in their recent survey Bunn and Wright (1991) conclude that the quality of expert judgemental forecasts may be higher than many previous researchers have found, while Evans (1989) argues that bias may exist, but not according to some general principle that is broadly valid for all decision-takers and in all task domains.

The rest of this paper is concerned with the ability of bankers and multivariate models to predict the emergence of arrears on external debt-service among less-developed countries. Comparisons are made in terms of true *ex ante* predictive ability. The approach taken here differs from that of most research in this area, which generally focuses on the event of rescheduling. The event of major interest to bankers is the emergence of arrears, to which the event of rescheduling is generally only a lagged reaction.

Section 2 of this paper summarizes the methodological approach, and describes the dependent variable. Section 3 outlines the techniques of country-risk appraisal that are used by banks, introduces the *Institutional Investor* rating system (hereinafter *II*) as a measure of banker judgement, and examines the predictive

¹ See for example Frank and Cline (1971), Dhonte (1975), Feder and Just (1977), Sargen (1977), Mayo and Barrett (1978), Saini and Bates (1978), Feder et al. (1981), Kugler (1984), Schmidt (1984), Taffler and Abassi (1984), Cooper (1985), Morgan (1986), Lloyd-Ellis et al. (1989), Lee (1991), Moghadam and Samavati (1991) and Moghadam et al. (1991).

² For example, the papers in Kahneman et al. (1982).

power of that system. Section 4 discusses two multivariate statistical models. Section 5 compares the models and the *II* rating, in terms of the true ex ante predictive abilities of each approach. Finally, in Section 6 the results of this research are discussed and conclusions are drawn.

2. Methodology and the dependent variable

2.1. Methodology

The first forecasting method is based on banker judgement. This is represented here by the *Institutional Investor* rating system, which is described fully in Section 3. The second method utilizes multivariate statistical models, using alternatively the discriminant and logit methodologies. In all cases, the paper is concerned with one-year-ahead forecasts of country creditworthiness. The data period is 1979–1989 for both the *II* rating and for the predictors in the multivariate models, while for the criterion variable y (reflecting creditworthiness), it is 1980 to 1990. Each data point consists of either (II_{it}, y_{it+1}) or (x_{it}, y_{it+1}) , $t \in [1979, 1989]$, where x is a vector of predictors and i is a country subscript.

The sample is divided into an estimation period and a forecast period. In the former, for all i , and all $t \in [1979, 1986]$, the sets of data $\{(II_{it}, y_{it+1})\}$ and $\{(x_{it}, y_{it+1})\}$ are used respectively to estimate the cutoff value of the *II* rating, the parameters of the multivariate models, and the cutoff values of the discriminant score and the logit probability. The cutoffs are used for forecasting out-of-sample.

The remaining data are used for testing the forecasting power of the *II* rating and of the two statistical models, on a one-year-ahead basis. Actual rating values and predicted discriminant scores and logit probabilities for year t , $t \in [1987, 1989]$, are used to forecast the criterion variable y for year $t + 1$, $t + 1 \in [1988, 1990]$, using the parameters and cutoff values derived from the estimation sample.

2.2. The dependent variable

Creditworthiness is represented by a binary dependent variable y , which takes the value 1 for a country-year-case if arrears of external debt-service emerge, for any reason, for the country in the given year, on debt owed to commercial banks, governments, or multilateral institutions. Otherwise, the variable y takes the value zero. Comprehensive quantitative data on arrears are not available for the period covered by this paper, and the classification of each country-year-case is determined primarily with reference to descriptive material in successive annual issues of the *World Debt Tables*, supplemented by IMF (1986, 1988, 1989). In virtually all cases, countries that have gone into arrears have subsequently restructured their debts. While restructuring has been commonly used as the salient event in earlier research on country risk, this paper focuses on the emergence of arrears, it being

normally the event that first impinges on creditors' cash-flows. It is in effect the macroeconomic counterpart to corporate failure, while the event of restructuring is only a lagged reaction to it. For example, the Argentinian crisis broke in August 1982, but it took until December 1985 before a restructuring agreement was finally signed.

In the few cases where no arrears emerged prior to an involuntary restructuring, then the date of the latter event is used to determine an 'arrears' classification, given that such an event has an effect on creditors' cash-flows similar to that of emergence of arrears.

The period 1980–1987 is the training sample for the dependent variable. In this period, distinct groups are required for the purposes described above, and therefore country-year-cases up to three years either side of an arrears classification are deleted from the sample, on the basis that they are likely to share characteristics with both arrears and non-arrears cases [see Taffler and Abassi (1984), who term such cases 'weak-years']. All other cases are classified non-arrears, and coded $y = 0$. For further details, see Somerville (1991, Ch. III).

3. Banker judgement

3.1. How banks assess country risk

Banks differ widely in their methods of assessment of country risk. Mascarenhas and Sand (1985), Heffernan (1986) and Nisse (1987) survey banking practice. Calverley (1990) discusses the various systems in use in banks. Many analysts use informal systems; however, various more formal systems are also in use. Checklist systems are among the less quantitative of the latter, and may either form an annex to a country report, or may form the entire report. Scenario analysis is also widely used. The popularity of scoring systems, which generate a numerical rating, appears to have declined in recent years. To the extent that banker judgement is based on multivariate techniques, the potential exists for correlation between judgementally based forecasts and those of formal models, and this will reduce the power of statistical tests to find significant differences between the forecasting power of formal models and pure judgement (i.e. excluding the model-based component). Calverley (p. 168) believes multivariate techniques, essentially logit and discriminant analysis, to be 'widely used' by banks, but as adjuncts to other techniques rather than as replacements for them. Finally, the most sophisticated quantitative techniques consist of formal econometric models of developing countries, but Calverley expresses doubt about the credibility of this approach.

The judgemental component of any mode of analysis can be enhanced by visits to the country by the analyst, and Calverley argues from experience that an analyst who has not made such visits will lack credibility.

3.2. *The Institutional Investor ratings*

The focus of this paper is on banker judgement in general, rather than the performance of any specific banker: therefore an average measure is required. Moreover an average will have the property, advantageous in this context, that combining multiple forecasts generally leads to increased forecast accuracy (Clemen, 1989). This paper uses country credit ratings, published regularly in the *Institutional Investor* since 1979, as measures of banker judgement. The other data that are used in this paper are exclusively annual, so a single *II* rating value is required for each country in each year. This paper is concerned with the power of forecasts, formed in year t , of debt-servicing status in year $t + 1$. A rating value is required that best reflects the judgement of bankers over the year as a whole, and hence the September ratings are used, rather than the alternative March values.

Other ratings of country risk are available: for example, those that are published in *Euromoney* and the *International Country Risk Guide (ICRG)*.³ During the early 1980s the *Euromoney* rating was essentially a measure of spread above LIBOR on syndicated loans, which is not a direct measure of average banker judgement. Since 1987, it has become a form of weighted checklist, similar to the *ICRG* rating. In each case, the published ratings are the results of combining objective indicators with ratings produced by a survey of economists and country experts. The judgemental component forms only part of these ratings, and in any case 1986–87 is too recent a starting date for the purposes of this paper.

The *II* ratings “are based on information provided by leading international banks [who] grade each of the countries on a scale of zero to 100 The sample . . . ranges from 75 to 100 banks The individual responses are weighted using an *Institutional Investor* formula that properly gives more weight to responses from banks with greater worldwide exposure and more sophisticated country-analysis systems.”⁴ (Shapiro, 1989, p. 135).

A number of authors use this rating system as an indicator of country risk, including Feder and Ross (1982), Taffler and Abassi (1984), Burton and Inoue (1985), Cosset and Roy (1991), and Oral et al. (1992). Taffler and Abassi (1984)

³ The *ICRG* ratings are not available for the years of publication earlier than 1986. Other ratings are not usable, mainly because of shortage of observations. Those of the *Economist Intelligence Unit* only date from 1989. The authors have been able to obtain only one year’s data for the *BERI* index. *Standard and Poor’s* only covers 18 of the 54 countries that are included in this paper, and is in bond-rating format. *Political Risk Services* (formerly *Frost and Sullivan*) covers 44, but the format (like *Standard and Poor’s*) is not suitable for use here.

⁴ To the extent that ‘more sophisticated systems’, involving multivariate models, are part of the armoury of some of the banks that report to the *Institutional Investor*, the rating is not completely independent of the modelling approach of Section 4 below. However, Mascarenhas and Sand (1989, p. 540) believe that the rating is “composed mostly of line bankers’ opinions.”

examine the power of the *II* ratings to predict debt-servicing problems, and find high error rates. Given the purpose to which the ratings are to be put in this paper, it is necessary to test their power to predict debt-servicing difficulties in recent years, and the rest of this section is concerned with this matter.

The ratings cover over 100 countries: in 1986, there were 109, including 79 LDCs as well as 30 from the OECD and CMEA areas. This paper is concerned with LDC debt, and focuses on 54 less-developed countries (see Appendix 1). These include all the major LDCs, in terms of population, GDP, external debt, and debt-servicing problems, and they account for 83% of total external debt of all countries covered by the World Bank's *World Debt Tables* for 1989–90.⁵

3.3. Testing banker judgement (1)

The null hypothesis is that banker judgement (as expressed by the *II* rating) has no predictive power, and it is addressed firstly by the runs test (Rohatgi, 1984, pp. 740–745). The advantage of this over tests based on predictive performance (see below) is that it does not require the specification of a cutoff rating value between arrears and non-arrears cases. Since no prediction can be involved using this test, it is carried out on the training and forecasting periods combined, in order to maximize sample size.

The sample is sorted numerically on the *II* ratings. Predictive power is sought over a one-year horizon: rating values in each year (1979 to 1989) are associated with values of the criterion variable *y* in the following year (1980 to 1990). If the rating system has predictive power (a false null), then this 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 country-year-cases, and the correct procedure is to test each year separately.

Table 1 reveals that the null is rejected at the 5 percent level in only four years out of eleven, 1980, 1986, 1988 and 1989, and in 1986 and 1989 it cannot be rejected at the 1 percent level. The overall result suggests that banker judgement may have limited ability to predict creditworthiness over a one-year horizon, although with rejection of the null in three years out of four in 1986–89, it may be that bankers' ability to predict creditworthiness is improving.

⁵ Of the 79 rated LDCs, 10 are excluded as being highly unrepresentative of the generality of indebted LDCs (Hong Kong, Israel, Singapore, South Africa, and the 6 oil sheikhdoms of the Arabian peninsula); 2 (Iraq and Libya) are excluded because of data shortages; a further 13 (Angola, Barbados, Cyprus, Cuba, Ethiopia, Grenada, Haiti, Lebanon, Mauritius, Seychelles, Sierra Leone, Tanzania, Uganda) are not covered by the EIU data base, from which the data used in Section 4 are drawn. In aggregate, these 13 accounted for only about 1.5% of LDC debt outstanding in 1986.

3.4. Testing banker judgement (2)

The second test produces specific classifications of the sample, which may be compared directly with the predictions of the multivariate models. A one-year forecast horizon is used, so that rating values at date t are used to predict debt-servicing status at date $t + 1$. The rating value of each country-year-case is compared with a cutoff value. If the rating is above the cutoff, the case is classified as ‘non-arrears’; otherwise as ‘arrears’. The cutoff is estimated from the sample (I_{it}, y_{it+1}) , where $t \in [1979, 1986]$. The cutoff is then used with 1987–89 rating values to yield predictions of debt-servicing status for 1988–90.

Two types of error are possible, each with an associated cost. When a country that moves into arrears is misclassified as creditworthy (a type I error), creditors’ cash-flows will suffer, and the value of their assets will fall. A type II error involves misclassifying a creditworthy country into the opposite group. In this case, the cost is an opportunity cost, of missing a profitable lending opportunity. In the literature it is usually assumed, plausibly, that the first type of error is perceived by creditors as having the greater average cost – see Sargen (1977) and Taffler and Abassi (1984). In the present context, what matters is the *relative* magnitude of the average costs of the two types of error, and this will be referred to below as C : the ratio of the average cost of a type I error to that of a type II.

A cutoff is taken to be optimal if it minimizes the expected total cost of misclassification. If prior probabilities of group membership equal sample proportions, then minimizing expected cost is equivalent to minimizing ex post cost. The optimal cutoff depends on the value of the cost ratio C ; Sargen (1977) and Taffler and Abassi (1984) use a value of $C = 3$, while Somerville (1991, Ch. VII.3.2) proposes a value of 3.75. For $C = 3$, the optimal cutoff is determined empirically at an I rating value of 40.7. Moreover, this cutoff value is optimal for all values of C lying in the interval $[2.1, 6.0]$ (see Appendix 2 for further details).

Using the cutoff of 40.7, the rating values for 1987 to 1989 may be used to derive the predictions for 1988–1990 that are summarized in Table 2. All ‘arrears’

Table 1

Runs test statistic: based on the numbers of runs of arrears and non-arrears cases, among the *Institutional Investor* ratings

| | | | |
|-------------|--------------|-------------|--------------|
| 1979 | -1.18 | 1985 | -0.59 |
| 1980 | -3.29 | 1986 | -2.02 |
| 1981 | 0.05 | 1987 | -1.36 |
| 1982 | -0.45 | 1988 | -3.10 |
| 1983 | -0.87 | 1989 | -2.52 |
| 1984 | -1.26 | | |

After ordering sample numerically on the rating, runs relative to y_{it+1} in each year t are counted.

H_0 (randomness) rejected at the 5 percent level if test statistic (based on number of runs) falls outside $[-1.96, 1.96]$; for 1 percent significance level, acceptance region is $[-2.57, 2.57]$.

Table 2

Institutional Investor: Rating values 1987–89 used to predict debt-servicing status 1988–90

| Actual category | | Predictions 1988–1990 | |
|-----------------|-----|--------------------------|-----------------------|
| | | Arrears | Non-arrears |
| Arrears | 99 | 99 (100%) | Type I errors: 0 (0%) |
| Non-arrears | 63 | Type II errors: 39 (62%) | 24 (38%) |
| Total | 162 | 138 | 24 |

(54 countries, 3 years)

Predictions based on cutoff rating value of 40.7. For definition of 'arrears' etc., see Section 2.

cases are correctly forecast, but the type II error rate is very high: 39 of the country-year-cases that yielded no arrears of debt-service in 1988–90, i.e. 62% of that category, are erroneously forecast to fall into the 'arrears' category. This arises because of significant overlap in the distributions of ratings of arrears and non-arrears cases: for example, a cutoff of 35.07 (i.e. 5 points lower) avoids 10 type II error, but yields an additional 6 of type I.⁶

There is clearly scope for improvement in these forecasting results.

4. Multivariate models for assessing country risk

This section explores the power of two typical multivariate models derived from recent data. The research reported here utilizes multivariate statistical models whose derivation is described extensively in Somerville (1991). Descriptions of the methodology, and of the binary dependent variable, are given in Section 2 above. The initial data set includes 70 variables, covering all those that might be considered as relevant on a priori grounds, and including those that have been used by previous researchers. The source of the data is the data base of *The Economist Intelligence Unit*, which includes a very large set of financial and economic data for less-developed countries.

The objective is to find a parsimonious set of predictor variables yielding maximal separation between the two groups. Two statistical techniques, linear discriminant analysis and logit analysis, are used because earlier papers have shown no clear superiority of one over the other (e.g. Hamer, 1983; Saini and Bates, 1984; Lo, 1986; Morgan, 1986). In each case an interactive stepwise procedure governs the selection of the variables that appear in the model: at each step, the 'entering' variable is discarded if it is highly correlated with a variable

⁶ Application of ex post cost-minimizing cutoffs in each year 1987–89 has only a marginal effect on the type II error rate, which remains high. It is conceivable that the cutoff should be adjusted ex ante to allow for shifts since 1979–86 in the distribution of rating values. Such adjustment, whether in terms of mean or median, results in an unacceptably high type-I error rate.

Table 3
Linear discriminant model

| Discriminant score $z_{it} = \beta + \gamma' x_{it}$: concerns group membership at $t + 1$ | | | | |
|---|------------|--------------------|-------------------------------|--|
| Estimated discriminant function (jackknifed coefficients) | | | Partial F (5 & 296 d.f.) | Mosteller–Wallace contribution ^a |
| β, γ | x | | | |
| 6.6715 | (constant) | | | |
| +0.0732 | NARY | (net assets/GDP)% | 164.4 | 53.3% |
| –0.0655 | DCPI | (inflation)% | 47.5 | 21.0% |
| –0.0669 | INPS | (int./debt.serv.)% | 19.2 | 13.9% |
| +0.1215 | DGDP | (GDP growth)% | 8.1 | 6.3% |
| +0.0733 | INVR | (investment/GDP)% | 6.5 | 5.5% |
| | | | | <u>100.0%</u> |

Wilks' $\lambda = 0.41$; $\chi^2(5) = 268.2$. All partial F ratios significant at 5%.

Definition of variables: **NARY**: numerator is end-period foreign exchange reserves, SDRs and reserve position with IMF, less total public and private disbursed short-, medium- and long-term external debt. **DCPI**: annual percentage change of 12-month average of consumer price index. **DGDP**: annual percentage change of real GDP. **INPS**: interest payments/debt service in year, relating to total public and private short-, medium- and long-term external debt. **INVR**: gross fixed capital formation/GDP. Sample period: 1979–1986 for predictors, 1980–1987 for dependent variable y .

^a I.e. to the Mahalanobis distance between the group centroids.

that has already entered, or if they are highly loaded on the same principal component of the data set.

The best-fitting discriminant function includes five variables, and is set out in Table 3. An 'arrears' classification is associated with a negative discriminant score. The variables in the model are listed in order of entry. The partial F ratios (' F -to-remove'), which are all significant at the 5 percent level, reveal the same ranking and indicate the relative importance of the discriminating variables. The χ^2 value associated with Wilks' lambda provides a test of the hypothesis that the model has no discriminating power, and this is rejected at the 5 percent level.

The logit model, which is set out in Table 4, is also statistically significant at the 5 percent level, as is each of its estimated coefficients. In terms of specification, the models are very similar, and all coefficients have the correct sign: the emergence of arrears is associated negatively with net assets/GDP, GDP growth, and investment/GDP (discriminant model only), and positively with inflation, and interest/debt-service. Full definitions of these variables are given at the foot of Table 3.

Saini and Bates (1984) identify parameter instability as a serious problem with models of this type. Somerville (1991, Ch. X) explores this, via intertemporal partitioning of the data set, and also using the bootstrapping technique (Efron, 1982). While there is some evidence of instability, it is not conclusive. The key question is the out-of-sample performance, which is now addressed.

Table 4
Logit model

$$P(y_{it+1} = 1 | \mathbf{x}_{it}) = e^{b+c'\mathbf{x}_{it}} / (1 + e^{b+c'\mathbf{x}_{it}})$$

| Estimated logit function (jackknifed coefficients) | | $\chi^2(1)$ to remove | Coeff/SE (Wald) |
|---|-------------------------|--------------------------|--------------------|
| b, c | x | | |
| -8.9914 | (constant) | 102.6 | -6.773 |
| -0.0833 | NARY (net assets/GDP)% | 122.3 | -6.852 |
| +0.0843 | DCPI (inflation)% | 33.4 | 3.754 |
| -0.2308 | DGDP (GDP growth)% | 19.6 | -4.403 |
| +0.0661 | INPS (int./debt.serv.)% | 13.6 | 3.404 |

Improvement $\chi^2(1) = 13.6$. Goodness-of-fit: Hosmer–Lemeshow $\chi^2(8) = 14.3$. (Hosmer and Lemeshow, 1989.)

Definition of variables and sample: see Table 3.

To enable the models to be used to make predictions, it is necessary to estimate least-cost cutoffs (see Section 3.4 above, and Appendix 2) over the 1979–1986 training sample. The following cutoffs are optimal for the indicated ranges of the cost ratio C , i.e. the relative average cost of a type I error:

Discriminant score: 0.0, for $C \in [1.0, 4.2]$

Logit probability: 31.1%, for $C \in [1.6, 6.0]$

Table 5 shows that these models have a similar predictive ability out of the sample period, and this is confirmed statistically by a standard χ^2 test.⁷ Both models are quite insensitive to the value of the cutoffs used for forecasting.⁸

A more extensive description of these results is contained in Somerville (1991, Tables IX.3 and IX.5), where the forecasts for 1988, 1989 and 1990 are compared. It is found that the type I error rate declines slightly over time, the type II rate rises slightly, and the total rate remains approximately constant. This suggests that the estimated parameters are stable.

⁷ See Conover (1980, pp. 144–151). Computed $\chi^2 = 136.63$, while $\chi^2(0.05; 1) = 3.84$. Thus the null (no significant relationship between the classifications made by the two models) is rejected at the 5 percent level.

⁸ In each case, the cutoff may be represented by a notional cutoff case with a given vector of predictors. This is not directly observable, but it is possible to evaluate $\gamma'x^D$ and cx^L , where γ , c , x^D and x^L are, respectively, vectors of discriminant and logit coefficients, and of predictor values for each function at its cutoff. Scalar multiplication of $\gamma'x^D$ (or $c'x^L$) equates to multiplying the components of x^D (or x^L) by that scalar. Given a multiple of 1.01, the cutoff value of each predictor rises by 1% (absolute value), which weakens the requirement for a 'non-arrears' forecast. Both models make an extra type-I error, and the discriminant model makes one fewer of type II. Alternatively, when the cutoff values of the predictors are reduced in value by 1% (absolute value), there are no changes in the models' forecasts.

Table 5
Multivariate models: Predicted values 1987-89, used to predict debt-servicing status, 1988-90

| Actual category | Discriminant model Predictions 1988-1990 | | Logit model Predictions 1988-1990 | |
|-----------------|---|--------------------------|--------------------------------------|-----------------------|
| | Arrears | Non-arrears | Arrears | Non-arrears |
| Arrears | 99 | Type I errors: 11 (11%) | 91 (92%) | Type I errors: 8 (8%) |
| Non-arrears | 63 | Type II errors: 11 (17%) | Type II errors: 14 (22%) | 49 (78%) |
| Total | 162 | 99 | 105 | 60 |

Discriminant scores and logit probabilities 1987-1989 are evaluated by applying the estimated models to 1987-89 data for the predictors. The above results are obtained using as cutoffs a zero discriminant score, and a logit probability of 31.1%.

Table 6

Institutional Investor ratings and multivariate models: forecast errors out-of-sample

| Forecasts by: | Forecast errors 1988–90 | | |
|---|-------------------------|----------|----------|
| | Type I | Type II | Overall |
| <i>Institutional Investor</i> rating, 1987–1989 | 0 (0%) | 39 (62%) | 39 (24%) |
| Forecast discriminant scores, 1987–1989 | 11 (11%) | 11 (17%) | 22 (14%) |
| Forecast logit probabilities, 1987–1989 | 8 (8%) | 14 (22%) | 22 (14%) |

Debt-servicing status at $t + 1$ classified by rating or model at t ; least-cost cutoffs derived from training sample in each case. Cutoffs: *II* rating 40.7; discriminant score 0; logit probability 31.1%.

In computing error rates, the denominators are, respectively: the number of arrears cases (type I error rate), the number of non-arrears cases (type II error rate), the total number of cases (overall error rate).

5. Comparisons of forecasting performance

The forecasting errors shown in Tables 2 and 5 are summarized in Table 6. The predictions of the *II* ratings are heavily biased towards the ‘arrears’ classification, with no type I errors, but a very high type-II error rate (62%). In contrast, the predictions of the statistical models are more balanced, with type I and type II error rates of, respectively, 11 and 17% (discriminant), and 8 and 22% (logit).⁹ These comparisons may be summarized in terms of the overall error rates, making no distinction between the two error types. The high type-II error rate of the *II* ratings yields an overall error rate of 24%, which is markedly higher than those of the models: 14% (logit) and 14% (discriminant), respectively. The introduction to this paper asked whether formal statistical models can outperform judgementally based methods of country-risk assessment, and the conclusion is now reached that they can, in terms of their overall error rates. Thus, a larger role for formal models in credit allocation to LDCs may be indicated.

However, if enough weight is placed on type I errors, it is possible for the ratings to achieve a lower misclassification cost than the model. In exploring this, we focus on the logit model only, given the similarities between the two models’ predictions.

The errors shown in Table 6 are derived from forecasts based on a least-cost cutoff for the training-sample period. In each case, there is a range of values for the cost ratio (i.e. C , the relative average cost of type I error) for which the cutoff

⁹ Nevertheless, the standard χ^2 test suggests that the predictions of ratings and models are statistically similar: taking the ratings and either model, the null (no relation between the two sets of predictions) is rejected at the 5 percent level. Comparing the *II* forecasts with those of the logit and discriminant models respectively, computed $\chi^2 = 37.4$, and 33.2, while $\chi^2(0.05;1) = 3.84$. See footnote 7.

is in fact least-cost, and these ranges have as their intersection the interval [2.1,6.0]. For values in this interval, total misclassification costs are given by:

Logit model: $8C + 14$

Institutional Investor ratings: 39

It is easy to see that the *II* ratings will yield a lower misclassification cost than the multivariate model if C exceeds 3.1 (i.e. if $8C + 14 > 39$). However, the basis of this performance by the rating system is the achievement of a minimal number of type I errors, at the expense of a very high type II error rate. The rating system is 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%, respectively. The rating, given the cost-minimizing cutoffs, does not go to this extreme. However, it classifies 85% of cases as arrears (using an ex ante cost-minimizing cutoff), compared with a true proportion ex post of 61%. At plausible values of the cost ratio, the rating system outperforms the model by avoiding type I errors, but it achieves this by giving a low rating to 39 LDCs that turn out, ex post, to be creditworthy.

6. Discussion and conclusions

This paper sets out to explore the judgemental accuracy of bankers, in the context of country risk assessment, and to compare it with the performance of formal statistical models in terms of forecasting power over a one-year horizon.

The *II* country credit ratings, which are based on banker judgement, are shown to be biased towards an adverse view of the creditworthiness of LDCs during 1987–1989. This 'overpessimism' contrasts with the findings of Taffler and Abassi (1984) of 'overoptimism' at the beginning of the decade (i.e. 1980–1983) when, moreover, evidence was available that should have suggested a revision of opinion. In both periods, the average view of bankers appears to have been biased, and may possibly be interpreted as evidence of judgemental failings. This may be a consequence of bankers' failure to revise their assessments of prior probabilities of debt-servicing problems in the face of new evidence: in the earlier period, the sequence of events that was publicly signalled by the Mexican moratorium of August 1982; in the later period, the recent improvements in debt indicators (World Bank, 1990–91, p. 4). Thus, bankers appear to be poor Bayesians.¹⁰ In fact, in the earlier period Taffler and Abassi (1984) find that good credit ratings tended not to be influenced even by the event of rescheduling itself. This relative

¹⁰ See Holt and Morrow (1992) for experimental evidence on this point.

stability of the pattern of rating values has persisted, for the 54 countries in the sample: in each year $t = 1980$ to 1989, the rank-order and product-moment correlation coefficients between the rating values and their values at $t - 1$ exceed 0.95, and in most years they equal 0.98 or 0.99.

Taffler and Abassi (1984, p. 558) suggest that bankers may be influenced by 'sentiment' about particular countries, rather than by economic fundamentals. Keynes (1936) and more recently Scharfstein and Stein (1990) argue that financial decision-takers may possess a herd instinct. Moreover, Keynes argues (pp. 157–158) that there are social pressures on an unorthodox individual to conform with herd behaviour in this context: "If he is successful, that will only confirm the general belief in his rashness; and if in the short run he is unsuccessful, which is very likely, he will not receive much mercy. Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally."

An alternative explanation for the poor performance of the *II* ratings could possibly lie in the particular dimension of risk that they may reflect. Portfolio theory distinguishes between systematic risk and specific (unsystematic) risk. The emergence of debt-service arrears reflects total risk. However, to the extent that banks hold well-diversified portfolios, then the respondents to the *II* surveys may have been reporting their views of countries' levels of systematic risk. This suggests a direction for further research.

Despite its better overall predictive success rate, the multivariate modelling approach has been shown to be not necessarily superior, when allowance is made for differential misclassification costs. The threshold value of the cost ratio C , above which the rating system dominates the logit model on a total cost basis, is 3.1. A banker is likely to be more fearful of lending to a subsequent defaulter, than of suffering the opportunity-cost of missing some profitable business. Consequently, it may be that the cost ratio is well above this value.

Possibly this indicates that some form of combination of methods is appropriate: as Bunn and Wright suggest (1991, p. 501): "the issue of the interaction of judgemental and statistical methods is ... identified as a more worthwhile line of inquiry."

This suggests a possible second direction for further research, along the lines of the study by Mascarenhas and Sand (1989), although it is noteworthy that, if the *II* rating is forced into the logit and discriminant models of Section 4 as an additional predictor, then the forecasting power of those models is in fact reduced.

Finally, further research should explore longer forecast horizons, as one year is rather short for bankers' purposes.

Appendix 1

Member countries of the data set

| | | | |
|------------|-------------|-------------|------------|
| Algeria | El Salvador | Malawi | Sudan |
| Argentina | Gabon | Mexico | Syria |
| Bangladesh | Guatemala | Morocco | Taiwan |
| Bolivia | Honduras | Nicaragua | Thailand |
| Brazil | India | Nigeria | Trinidad |
| Cameroon | Indonesia | Pakistan | Tunisia |
| Chile | Iran | Panama | Turkey |
| China | Ivory Coast | Papua N.G. | Uruguay |
| Congo | Jamaica | Paraguay | Venezuela |
| Colombia | Jordan | Peru | Yugoslavia |
| Costa Rica | Kenya | Philippines | Zambia |
| Dom. Rep. | Liberia | Senegal | Zaire |
| Ecuador | Malaysia | S. Korea | Zimbabwe |
| Egypt | | Sri Lanka | |

Appendix 2: Choosing cutoffs

Expected total costs of misclassification are given by (Mensah, 1983, p. 23):

$$\text{Expected cost} = [C \times P \times T1/N1] + [(1 - P) \times T2/N2]$$

where:

- P : prior probability of 'arrears';
- $T1, T2$: numbers of type I and II errors;
- C : average cost per type I error;
- $N1, N2$: numbers of cases of each type.

$T1/N1$ and $T2/N2$ are the conditional probabilities of type I and II errors, respectively. However, if prior probabilities are assumed to equal sample proportions, then minimizing expected cost is equivalent to minimizing the ex post cost: $[C \times T1] + T2$. The cutoffs for the ratings and the two models are all determined by minimizing this quantity, in each case over the domain of all possible cutoffs. For the discriminant model, the intercept is adjusted to put the cost-minimizing discriminant score at zero.

References

- Bunn, D. and G. Wright, 1991, Interaction of judgemental and statistical forecasting methods: issues and analysis, *Management Science* 37, 501–518.

- Burton, F.N. and H. Inoue, 1985, An appraisal of the early warning indicators of sovereign loan default in country risk evaluation systems, *Management International Review* 25, 45–56.
- Calverley, J., 1990, *Country risk analysis* (2nd ed.) (Butterworths, London).
- Clemen, R.T., 1989, Combining forecasts: a review and annotated bibliography, *International Journal of Forecasting* 5, 559–583.
- Conover, W.J., 1980, *Practical nonparametric statistics* (2nd ed.) (Wiley, New York).
- Cooper, J.C.B., 1985, Country creditworthiness: a canonical correlation analysis, Working Paper (Glasgow College of Technology, Glasgow).
- Cosset, J.-C. and J. Roy, 1991, The determinants of country risk ratings, *Journal of International Business Studies* 22, 135–142.
- Dhonte, P., 1975, Describing external debt situations: a roll-over approach, *IMF Staff Papers* 22, 159–186.
- Efron, B., 1982, *The jackknife, the bootstrap, and other resampling plans* (Society for Industrial and Applied Mathematics, Philadelphia).
- Evans, J.St.B.T., 1989, *Bias in human reasoning – causes and consequences* (Lawrence Erlbaum, Hove).
- Feder, G. and R. Just, 1977, A study of debt servicing capacity applying logit analysis, *Journal of Development Economics* 4, 25–38.
- Feder, G. and K. Ross, 1982, Risk assessments and risk premiums in the Eurodollar market, *Journal of Finance* XXXVII, 679–691.
- Feder, G., R. Just and K. Ross, 1981, Projecting debt servicing capacity of developing countries, *Journal of Financial and Quantitative Analysis* XVI, 651–669.
- Frank, C.R. and R. Cline, 1971, Measuring debt servicing capacity: an application of discriminant analysis, *Journal of International Economics* 1, 327–344.
- Hamer, M., 1983, Failure prediction: sensitivity of classification accuracy to alternative statistical methods and variable sets, *Journal of Accounting and Public Policy* 2, 289–307.
- Heffernan, S.A., 1986, *Sovereign risk analysis* (Allen and Unwin, London).
- Holt, D.L. and P.C. Morrow, 1992, Assessment judgements of auditors and bank lenders: a comparative analysis of conformance to Bayes' theorem, *Accounting, Organizations and Society* 17, 549–559.
- Hosmer, D.W. and S. Lemeshow, 1989, *Applied logistic regression* (Wiley, New York).
- IMF, 1986, 1988, 1989, *International Capital Markets – Developments and Prospects* (IMF, Washington).
- Kahneman, D., P. Slovic and A. Tversky, eds., 1982, *Judgement under uncertainty: heuristics and biases* (Cambridge University Press, Cambridge).
- Keynes, J.M., 1936, *The general theory of employment, interest and money* (MacMillan, London).
- Kugler, P., 1984, Economic indicators and debt reschedulings 1977–1981: empirical results from a logit model, *Rivista Internazionale di Scienze Economiche e Commerciali*, part 10–11, 992–1005.
- Lee, S.H., 1991, Ability and willingness to service debt as explanation for commercial and official rescheduling cases, *Journal of Banking and Finance* 15, 5–27.
- Libby, R. and B.L. Lewis, 1982, Human information processing research in accounting: the state of the art in 1982, *Accounting, Organizations and Society* 7, 231–285.
- Lo, A.W., 1986, Logit versus discriminant analysis – a specification test and application to corporate bankruptcies, *Journal of Econometrics* 31, 151–178.
- Lloyd-Ellis, H., G.W. McKenzie and S.H. Thomas, 1989, Using country balance sheet data to predict debt rescheduling, *Economics Letters* 31, 173–177.
- Mascarenhas, B. and O. Sand, 1985, Country risk assessment systems in banks: patterns and performance, *Journal of International Business Studies* 16, 19–35.
- Mascarenhas, B. and O. Sand, 1989, Combination of forecasts in the international context: predicting debt reschedulings, *Journal of International Business Studies* 20, 539–552.
- Mayo, A.L. and A.G. Barrett, 1978, An early-warning model for assessing developing country risk, in: S.H. Goodman, ed., *Financing and risk in developing countries* (Praeger, New York) pp. 81–87.

- Mensah, Y.M., 1983, The differential bankruptcy predictive ability of specific price level adjustments: some empirical evidence, *The Accounting Review* LVIII, 228–246.
- Moghadam, M.R. and H.S. Samavati, 1991, Predicting debt rescheduling by less-developed countries: a probit model approach, *Quarterly Review of Economics and Business* 31, 3–14.
- Moghadam, M.R., H.S. Samavati and L. Haber, 1991, The determinants of debt rescheduling: the case of Latin America, *Southern Economic Journal* 58, 510–517.
- Morgan, J.B., 1986, A new look at debt rescheduling indicators and models, *Journal of International Business Studies* 17, 37–54.
- Nisse, J., 1987, Political risks: the benefits of hindsight, *The Banker* 137, 35–39.
- Oral, M., O. Kettani, J.-C. Cosset and M. Daouas, 1992, An estimation model for country risk rating, *International Journal of Forecasting* 8, 583–593.
- Peterson, C.R. and L.R. Beach, 1967, Man as an intuitive statistician, *Psychological Bulletin* 68, 29–46.
- Rohatgi, V.K., 1984, *Statistical inference* (Wiley, New York).
- Saini, K.G. and P.S. Bates, 1978, Statistical techniques for determining debt-servicing capacity for developing countries: analytical review of the literature and further empirical results, Research Paper no. 7818 (Federal Reserve Bank of New York, New York).
- Saini, K.G. and P.S. Bates, 1984, A survey of the quantitative approaches to country risk analysis, *Journal of Banking and Finance* 8, 341–356.
- Sargen, N., 1977, Economic indicators and country risk appraisal, *Economic Review: Federal Reserve Bank of San Francisco* 4, 19–35.
- Scharfstein, D.S. and J.C. Stein, 1990, Herd behaviour and investment, *American Economic Review* 80, 465–479.
- Schmidt, R., 1984, Early warning of debt rescheduling, *Journal of Banking and Finance* 8, 357–370.
- Shanteau, J., 1989, Cognitive heuristics and biases in behavioural auditing: review, comments and observations, *Accounting, Organizations and Society* 14, 165–177.
- Shapiro, H.D., 1989, 10th anniversary survey – a modest uptick, *Institutional Investor* (September), 129–147.
- Shefrin, H. and M. Statman, 1985, The disposition to sell winners and ride losers too long: theory and evidence, *Journal of Finance* XL, 777–792.
- Slovic, P., 1972, Psychological study of human judgment: implications for investment decision making, *Journal of Finance* XXVII, 779–799.
- Somerville, R.A., 1991, Country risk: multivariate models and human judgement, Unpublished PhD thesis (The City University, London).
- Taffler, R.J. and B. Abassi, 1984, Country risk: a model for predicting debt servicing problems in developing countries, *Journal of The Royal Statistical Society Series A* 147, 541–568.
- Tversky, A. and D. Kahneman, 1982, Judgement under uncertainty: heuristics and biases, in: D. Kahneman, P. Slovic and A. Tversky, eds., *Judgement under uncertainty: heuristics and biases* (Cambridge University Press, Cambridge) 3–20.
- World Bank (1989–90, 1990–91), *World Debt Tables* (The World Bank, Washington).