

IMF Working Paper

© 1999 International Monetary Fund

This is a *Working Paper* and the author(s) would welcome any comments on the present text. Citations should refer to a *Working Paper of the International Monetary Fund*. The views expressed are those of the author(s) and do not necessarily represent those of the Fund.

WP/99/147

INTERNATIONAL MONETARY FUND

Research Department

Monitoring Banking Sector Fragility: A Multivariate Logit Approach

Prepared by Aslı Demirgüç-Kunt and Enrica Detragiache¹

Authorized for distribution by Peter Wickham

October 1999

Abstract

This paper explores how a multivariate logit empirical model of banking crisis probabilities can be used to monitor banking sector fragility. The proposed approach relies on readily available data, and the fragility assessment has a clear interpretation based on in-sample statistics. The model has better in-sample performance than currently available alternatives, and the monitoring system can be tailored to fit the preferences of the decision maker regarding type I and type II errors. The framework can be useful as a preliminary screen to economize on precautionary costs.

JEL Classification Numbers: E44, E47, G21

Keywords: Banking crises, bank fragility, monitoring.

Author's E-Mail Address: ADemirguckunt@worldbank.org, Edetrageiache@imf.org

¹Aslı Demirgüç-Kunt is a Principal Economist in the Development Research Group of the World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, the International Monetary Fund, their Executive Directors, or the countries they represent. The authors wish to thank Anqing Shi for capable research assistance.

Contents	Page
I. Introduction	3
II. The Literature	4
III. Estimating In-Sample Banking Crisis Probabilities in a Multivariate Logit Framework .	6
A. Sources of Fragility—The 1994 Mexican Crisis According to the Empirical Model	8
B. Out-of-Sample Probability Forecasts	10
IV. Building An Early Warning System Using Estimated Crisis Probabilities	11
A. Loss Function for the Decision-Maker	12
B. Choosing the Optimal Threshold	15
V. Using Estimated Crisis Probabilities to Construct a Rating System for Bank Fragility .	18
VI. An Application to the Banking Crises of 1996-97	20
VII. Conclusions	23
Text Tables	
1. Banking Crises and Estimated Crisis Probabilities	7
2. Banking Crisis Probabilities—Logit Regression Results	9
3. Decomposition of Estimated Banking Crisis Probability Mexico, 1992-93	10
4. A Rating System for Banking Sector Fragility	19
Figures	
1. Crisis Threshold and In-Sample Classifications Accuracy	14
2. Loss Function for Various Values of the Cost Parameters	16
3. Optimal Probability Threshold	17
4. Out-of-Sample Crisis Probabilities Using Forecasted and Actual Data	22
References	26

I. INTRODUCTION

The last two decades have seen a proliferation of systemic banking crises, as documented, among others, by the comprehensive studies of Lindgren, Garcia, and Saal (1996) and Caprio and Klingebiel (1996). Most recently, the economic crises experienced by five East Asian countries (Indonesia, Malaysia, South Korea, the Philippines, and Thailand) were accompanied by deep financial sector problems. While in some cases the troubles were foreseen, in others most observers (including the Fund, the Bank, and the major credit rating agencies) were caught by surprise. Similarly, three years earlier, the Mexican devaluation and associated banking crisis had caught many observers and market participants by surprise.

The spread of banking sector problems and the difficulty of anticipating their outbreak have raised the issue of improving monitoring capabilities both at the national and supra-national level, and, particularly, of using statistical studies of past banking crises to develop a set of indicators of the likelihood of future problems. In our previous work (Demirgüç-Kunt and Detragiache, 1998a and 1999), we developed an empirical model of the determinants of systemic banking crises for a large panel of countries. Using a multivariate logit framework, we estimated the probability of a banking crisis as a function of various explanatory variables. That research showed that there is a group of variables, including macroeconomic variables, characteristics of the banking sector, and structural characteristics of the country, that are robustly correlated with the emergence of banking sector crises. This paper explores how the information contained in that empirical relationship can be utilized to monitor banking sector fragility.²

The basic idea is to estimate a specification of the multivariate logit model used in our previous work that relies mainly on explanatory variables whose future values are routinely forecasted by professional forecasters, the Fund, or the Bank. Out-of-sample banking crisis probabilities are then computed using the estimated coefficients and forecasted values of the explanatory variables. Using the information provided by in-sample estimation results, these forecasted probabilities are used to make a quantitative assessment of fragility. More specifically, we examine two different monitoring frameworks: in the first, the monitor wants to know whether forecasted probabilities are high enough to trigger a response or not. The response is defined to be a costly action of some sort, for instance gathering new specific information, scheduling on-site bank inspections, taking preventive regulatory measures, or others. Each possible threshold for taking action has a cost in terms of type I error (failure to identify a crisis) and type II error (false alarm), a cost that can be quantified on the basis of in-sample classification accuracy. Naturally, the choice of the criterion depends on the cost of either type of error to the monitor. For instance, if the monitoring system is used as a preliminary screen to determine which cases warrant further analysis, then a system that

²Other studies using limited dependent variable econometric models to estimate banking crises probabilities are Eichengreen and Rose (1998) and Hardy and Pazarbaşıoğlu (1998). These studies do not address issues of forecasting.

tolerates a fair amount of type II errors but incurs few type I errors will be preferable to one that is likely to miss a lot of crises. Conversely, if the “warning system” is used to put pressure on country authorities to take drastic policy actions to prevent an impending disaster, then a more conservative criterion is desirable. The framework developed here will be sufficiently flexible to accommodate alternative preferences for the decision-maker, and it will make explicit the costs and benefits of alternative criteria.

In the second monitoring framework examined, the monitor is simply interested in rating the fragility of the banking system. Depending on the rating, various courses of action may follow, but these are not explicitly modeled. In this case, it is desirable for a rating to have a clear interpretation in terms of crisis probability, so that different ratings can be compared. We examine one such example. As an illustration of the monitoring procedures developed in the first part of the paper, in the second part of the paper we conduct a limited out-of-sample forecasting exercise by constructing forecast probabilities for the six banking crises that occurred in 1996-97, namely Jamaica in 1996 and the five East Asian crises in 1997.

The paper is organized as follows: the next section will briefly review existing literature on banking system fragility indicators; Section III presents an adapted version of our empirical model of banking crises. Section IV discusses how out-of-sample probability forecasts obtained from the model can be used to obtain an early warning system. Section V contains an application to the crises of 1996-97, while Section VI concludes.

II. THE LITERATURE

An extensive literature has reviewed episodes of banking crises around the world, examining the developments leading up to the crisis as well as the policy response. This work, while it does not directly address the issue of leading indicators of banking sector problems, points to a number of variables that display “anomalous” behavior in the period preceding the crises. For instance, Gavin and Hausman (1996) and Sachs, Tornell, and Velasco (1996) suggest that credit growth be used as an indicator of impending troubles, as crises tend to be preceded by lending booms. Mishkin (1994) highlights equity price declines, while, in his analysis of Mexico’s 1995 crisis, Calvo (1996) suggests that monitoring the ratio of broad money to foreign exchange reserves may be useful in evaluating banking sector vulnerability to a currency crisis.

Honohan (1997) performs a more systematic evaluation of alternative indicators: he uses a sample of 18 countries that experienced banking crises and six that did not. The crisis countries are then divided into three groups (of equal size) according to the type of crisis (macroeconomic, microeconomic, or related to the behavior of the government). The average value of seven alternative indicators for the crisis countries is then compared with the average for the control group of countries. This exercise shows that banking crises associated with macroeconomic problems were characterized by a higher loan-to-deposit ratio, a higher

foreign borrowing-to-deposit ratio, and higher growth rate of credit. Also, a high level of lending to the government and of central bank lending to the banking system were associated with crises related to government intervention. On the other hand, banking crises deemed to be of microeconomic origin did not appear to be associated with abnormal behavior on the part of the indicators examined in the study.

Rojas-Suarez (1998) proposes an approach based on bank level indicators, similar in spirit to the CAMEL system used by U.S. regulators to identify problem banks. The author argues that in emerging markets (particularly in Latin America), CAMEL indicators are not good signals of bank strength, and that more information can be obtained by monitoring the deposit interest rate, the spread between the lending and deposit rate, the rate of credit growth, and the growth of interbank debt. Because these variables are to be measured against the banking system average, however, this approach appears more adequate for identifying weaknesses specific to individual banks rather than a situation of systemic fragility. Also, the approach requires bank level information, which is often not readily available outside of developed countries.

The most comprehensive effort to date to develop a set of early warning indicators for banking crises (and for currency crises) is that of Kaminsky and Reinhart (1999), further refined in Kaminsky (1998). These studies examine the behavior of 15 macroeconomic indicators for a sample of 20 countries which experienced banking crises during 1970-95.³ The behavior of each indicator in the 24 months prior to the crisis is contrasted with the behavior during “tranquil” times. A variable is deemed to signal a crisis if at any time it crosses a particular threshold. If the signal is followed by a crisis within the next 24 months, then it is considered correct; otherwise it is considered noise. The threshold for each variable is chosen to minimize the in-sample noise-to-signal ratio. The authors then compare the performance of alternative indicators based on the associated type I and type II errors, on the noise-to-signal ratio, and on the probability of a crisis occurring conditional on a signal being issued.⁴ The indicator with the lowest noise-to-signal ratio and the highest probability of crisis conditional on the signal is the real exchange rate, followed by equity prices and the money multiplier. These three indicators, however, have a large incidence of type I error, as they fail to issue a signal in 73-79 percent of the observations during the 24 months preceding a crisis. The incidence of type II error, on the other hand, is much lower, ranging between 8 percent and 9 percent. The variable with the lowest type I error is the real interest rate, which signals in 30 percent of the pre-crisis observations. The high incidence of type I error relative to type II error may not be a desirable feature of a warning system if the costs of false alarms are small relatively to the costs of missing a crisis. Since, presumably, the likelihood of a crisis is greater when several indicators are signaling at the same time, Kaminsky (1998) develops

³For a study of early warning indicators of currency crises, see also IMF (1998).

⁴Actually, the authors use an “adjusted” version of the noise-to-signal ratio, computed as the ratio of the probability of type II error to one minus the probability of a type I error.

“composite” indexes, such as the number of indicators that cross the threshold at any given time, or a weighted variant of that index, where each indicator is weighted by its signal-to-noise ratio, so that more informative indicators receive more weight. The best composite indicator outperforms the real exchange rate in predicting crises in sample, but is worse at predicting noncrisis observations.⁵

The approach developed in the following sections will allow the policy-maker to choose a warning system that reflects the relative cost of type I and type II error, and it will offer a natural way of combining the effect of various economic forces on banking sector vulnerability. By making better use of all available information, the system will deliver lower overall in-sample forecasting errors than those associated with individual indicators. Also, we will examine a problem that is not addressed by Kaminsky and Reinhart, namely that of a monitor who wishes to use information contained in the statistical analysis of past crisis episodes not just to “call” or “not call” a crisis, but to obtain a more nuanced assessment of banking sector fragility.

III. ESTIMATING IN-SAMPLE BANKING CRISIS PROBABILITIES IN A MULTIVARIATE LOGIT FRAMEWORK

The starting point of our analysis is an econometric model of the probability of a systemic banking crisis. In previous work (Demirgüç-Kunt and Detragiache, 1998a and 1999), we have estimated various alternative specifications of a logit regression for a large sample of developing and developed countries, including both countries that experienced banking crises and countries that did not. Details on sample selection, the construction of the banking crisis variable, and the choice of explanatory variables can be found in our previous papers. To form the basis of an easy-to-use monitoring system, we have estimated a specification of our empirical model that includes only variables available from the International Financial Statistics or other publicly available data bases, and that are routinely forecasted by the Fund in its biannual World Economic Outlook (WEO) exercise or by professional forecasters. As it turns out, this is not the specification that fits the data the best. The regression is estimated using a panel of 766 observations for 65 countries during 1980-95.⁶ In this panel, 36 systemic banking crises were identified, so that crisis observations make up 4.7 percent of the sample. Table 1 lists the crisis episodes. The set of explanatory variables capturing macroeconomic conditions includes the rate of growth of real GDP, the change in the terms of trade, the rate of depreciation of the exchange rate (relative to the

⁵Kaminsky (1998) finds that a crisis probability computed taking into account the number of indicators signaling increased substantially before the 1997 crisis in the Philippines, Malaysia, and Thailand, but not in Indonesia. Korea was not in the sample.

⁶Due to lack of data or breaks in the series, for some countries part of the sample period may be excluded. Also, years in which banking crises are ongoing are excluded from the sample.

Table 1. Banking Crises and Estimated Crisis Probabilities

	Crisis year	Estimated Probability
Chile	1981	.231
Colombia	1982	.066
Ecuador	1995	.439
Finland	1991	.066
Guyana	1993	.007
Indonesia	1992	.107
India	1991	.069
Israel	1983	.999
Italy	1990	.015
Jordan	1989	.334
Japan	1992	.037
Kenya	1993	.361
Sri Lanka	1989	.036
Mexico	1982	.527
	1994	.099
Mali	1987	.035
Malaysia	1985	.067
Nigeria	1991	.011
Norway	1987	.036
Nepal	1988	.018
Panama	1988	.539
Peru	1983	.244
Philippines	1981	.035
Papua New Guinea	1989	.121
Portugal	1986	.064
El Salvador	1989	.055
Sweden	1990	.036
Swaziland	1995	.633
Thailand	1983	.027
Turkey	1991	.158
	1994	.482
Tanzania	1988	.035
Uruguay	1981	.329
United States	1980	.238
Venezuela	1993	.494
South Africa	1985	.196

U.S. dollar), the rate of inflation and the fiscal surplus as a share of GDP. The explanatory variables capturing characteristics of the financial sector are the ratio of broad money to foreign exchange reserves and the rate of growth of bank credit lagged by two periods. Finally, GDP per capita is used as a proxy for the structural characteristics of the economy.

The estimated coefficients of the logit regression are reported in Table 2. Low GDP growth, a high real interest rate, high inflation, strong growth of bank credit in the past, and a large ratio of broad money to reserves are all associated with a high probability of a banking crisis. Exchange rate depreciation, the terms of trade variable and the fiscal surplus, on the other hand, are not significant. Table 1 shows the estimated crisis probabilities for the 36 episodes included in the sample. The probabilities range from a low of 1.1 percent for Nigeria to a high of 99.9 percent for Israel. About 70 percent of the episodes have an estimated probability of 4 percent or above, while only 17 percent have an estimated probability of over 50 percent.

A. Sources of Fragility—The 1994 Mexican Crisis According to the Empirical Model

One of the advantages of the multivariate logit model is that the sources of fragility can be easily identified by calculating the contribution of each explanatory variable to a change in the estimated crisis probability. As an illustration, we analyze the factors that contributed to the sharp increase in the estimated crisis probability in Mexico in 1993, a prelude to the crisis beginning in 1994. Table 3 reports the results. The last two rows of the table contain estimated crisis probability in 1992 and 1993. The first column gives the percent change in each explanatory variable between 1992 and 1993. The next two columns report the “weights” given to each factor in 1993 and 1992, respectively. These weights are obtained by multiplying the estimated regression coefficient of each variable with the corresponding value of the variable. Negative weights indicate that the variable in question tended to decrease the estimated crisis probability. In 1993, high past credit growth, high real interest rates, and high inflation were the main underlying reasons why the crisis probability was high in Mexico. The table also reports change in factor weights between 1992 and 1993, and the corresponding change in crisis probability. Because logit is nonlinear the sum of the contribution of each variable does not always add up to the total change in probability. Looking at macro factors, one sees that Mexico had a negative growth shock which increased the crisis probability significantly. There was also a significant increase in real interest rates and a minor terms of trade shock. At the same time, appreciation of the exchange rate, a lower rate of inflation and a lower budget surplus helped offset some of this increase. Financial sector variables played a less important role in explaining the overall increase in probability, slightly offsetting the impact of the macro factors. Vulnerability of the financial system to capital outflows—measured by M2/reserves ratio—decreased slightly, leading to a 1 percent decrease in crisis probability. Credit growth slowed down, leading to a 2 percent lower crisis

Table 2. Banking Crisis Probabilities—Logit Regression Results 1/

Explanatory Variables	Estimated Coefficients
Growth	-.172*** (.034)
Terms of trade change	-.021 (.018)
Depreciation	.007 (.006)
Real interest rate	.065*** (.016)
Inflation	.020** (.010)
Fiscal surplus/GDP	.066 (.036)
M2/Reserves	.013*** (.005)
Credit growth _{t-2}	.015** (.008)
GDP per capita	-.039 (.033)
No. of Crises	36
No. of Obs.	766
model χ^2	61.46***
AIC	249

1/ Standard errors in parentheses, *, ** and *** indicate significance levels of 10 percent, 5 percent and 1 percent respectively.

Table 3. Decomposition of Estimated Banking Crisis Probability
Mexico, 1992-93

Explanatory variable	Percentage Change in Variable, 1992-93	Weight in 1993	Weight in 1992	Change in Weight, 1992-93	Contribution to Percentage Change in Probability
GDP growth	-125	0.154	-0.624	0.778	105
Terms of Trade	-16	-0.034	-0.041	0.007	1
Depreciation	-119	-0.002	0.010	-0.012	-1
Real interest rate	386	0.327	0.067	0.259	28
Inflation	-31	0.202	0.295	-0.093	-8
Fiscal surplus/GDP	-79	0.022	0.102	-0.080	-7
M2/reserves	-16	0.057	0.068	-0.011	-1
Lagged credit growth	-4	0.498	0.517	-0.019	-2
GDP per capita	-1	-0.070	-0.070	0	0
ESTIMATED CRISIS PROBABILITY:					
1992	0.054				
1993	0.116				

probability. Finally, GDP per capita—which we use as a proxy of institutional development—did not change significantly in this period. Thus, decomposing the crisis probability helps understanding which factors played a role in bringing about the crisis, at least according to the empirical model.

B. Out-of-Sample Probability Forecasts

Because the purpose of monitoring is to obtain an assessment of future fragility, the next step is to obtain forecasts of banking crisis probabilities. These can be easily obtained as follows: let β be a $1 \times N$ vector containing the N estimated coefficients of the logit regression reported in Table 1, and let z_t be a $N \times 1$ vector of out-of-sample values of the explanatory variables for country i at date t . Of course, these values can be true forecasts, estimates of past values, data for countries/time periods not included in the sample, or ranges of values to construct alternative scenarios. Then, the out-of-sample probability of a banking crisis for country i at date t is

$$p_{it} = \frac{\exp[\beta z_{it}]}{1 + \exp[\beta z_{it}]}$$

Once out-of-sample probabilities are computed, the question arises of how to interpret them: is a 10 percent crisis probability high or low? Should a policy-maker undertake preventive actions when faced with such a probability? Should a surveillance agency issue a warning? The next section will address the issue of how to use the forecasted probabilities to monitor banking sector fragility.

IV. BUILDING AN EARLY WARNING SYSTEM USING ESTIMATED CRISIS PROBABILITIES

The first monitoring framework considered is one in which, after forecast probabilities are obtained as described in the preceding section, the decision-maker has to choose whether the probability is large enough to issue a warning. This is the framework implicit in Kaminsky and Reinhart (1999). Issuing a warning will lead to some sort of preventive action: for instance, the decision-maker may invest in further information gathering, such as the acquisition of bank-level balance sheet data, or discussions with senior bank managers, bank supervisory agencies in the country, or other market participants. Alternatively, the monitoring system may be used to decide whether to take preventive policy measures, such as tightening of prudential capital or liquidity requirements for banks, or a reduction in interest rates to ease pressures on bank balance sheets. For a warning system to be useful it must be the case that preventive measures can substantially reduce the costs of a crisis. We will assume that this is the case. Also, preventive measures are usually costly: tighter prudential requirements may cause banks to cut credit, perhaps leading to a credit crunch; looser monetary policy may lead to higher inflation, and so on. Thus, a useful warning system should minimize “false alarms,” namely situations in which preventive measures are taken while no real crisis is pending.

The choice of the threshold for issuing a warning will generally depend on three aspects: first, the probability of type I and type II errors associated with the threshold, which, assuming that the sample of past crises is representative of future crises, can be assessed on the basis of the in-sample frequency of the two errors. Clearly, the higher the threshold that forecasted probabilities must cross before a warning is issued, the higher will be the probability of a type I error and the lower will be the probability of a type II error, and vice versa. The second parameter on which the choice of the threshold depends is the unconditional probability of a banking crisis, which can also be assessed based on the in-sample frequency of crisis observations: if crises tend to be rare events, then the overall likelihood of making a type I error is relatively small, and vice versa. Finally, the third aspect that affects the choice of a warning threshold is the cost to the decision-maker of taking preventive action relative to the cost of an unanticipated banking crisis. In general, these costs are themselves forecasts of the true costs, and making a good decision requires having good

forecasts of the costs. A policy-maker that tends to underestimate the cost of a crisis or to overestimate the cost of taking preventive policy action, will be too conservative in the choice of a warning threshold, and vice versa.⁷

A. Loss Function for the Decision-Maker

Based on the above considerations, a more formal analysis of the decision process behind the choice of a warning system may be stated as follows. Let T be the threshold chosen by the decision-maker, so that if the forecasted probability of a crisis for country I at time t exceeds T , then the system will issue a warning. Let $p(T)$ denote the probability that the system will issue a warning, and let $e(T)$ be the joint probability that a crisis will occur **and** the system issues no warning. Further, let c_1 be the cost of taking preventive actions as a result of having received a warning signal, and let c_2 be the additional cost of a banking crisis if it is not anticipated (if anticipating a crisis can prevent it altogether, then c_2 is the entire cost of the crisis). Presumably, c_1 is substantially smaller than c_2 if further information gathering can be relied upon to provide useful information, and if the knowledge that a crisis is impending allows policy-makers to take effective preventive measures. Then, a simple linear expected loss function for the decision-maker may be defined as follows:

$$L(T) \equiv p(T)c_1 + e(T)c_2.$$

This expression can be rewritten using the notions of type I and type II error. Let $a(T)$ be the type I error associated with threshold T (the probability of not receiving any warning conditional on a crisis occurring), and let $b(T)$ be the probability of a type II error (the probability of receiving a warning conditional on no crisis taking place). Also, let w denote the (unconditional) probability of a crisis. Then the loss function of the decision-maker can be rewritten as:

$$L(T) = c_1[(1-a(T))w + b(T)(1-w)] + c_2a(T)w = wc_1\left[1 + \left(\frac{c_2 - c_1}{c_1}\right)a(T) + b(T)\left(\frac{1-w}{w}\right)\right].$$

The second part of the equality above shows that the higher is the cost of missing a crisis relative to the cost of taking preventive action (the larger is c_2 relative to c_1), the more concerned the decision-maker will be about type I error relative to type II error, and vice versa. Also, the higher is the unconditional probability of a banking crises (measured by the parameter w), the more weight the decision-maker will place on type II errors, as the

⁷For estimates of the fiscal costs of recent banking crises, see Caprio and Klingebiel (1996).

frequency of false alarms is greater when crises tend to be rare events.⁸ Notice also that minimizing the noise-to-signal ratio (in our notation, $b(T)/(1 - a(T))$)—the criterion chosen by Kaminsky and Reinhart to construct and rank alternative signals—does not generally lead to minimizing the expected loss function specified above.

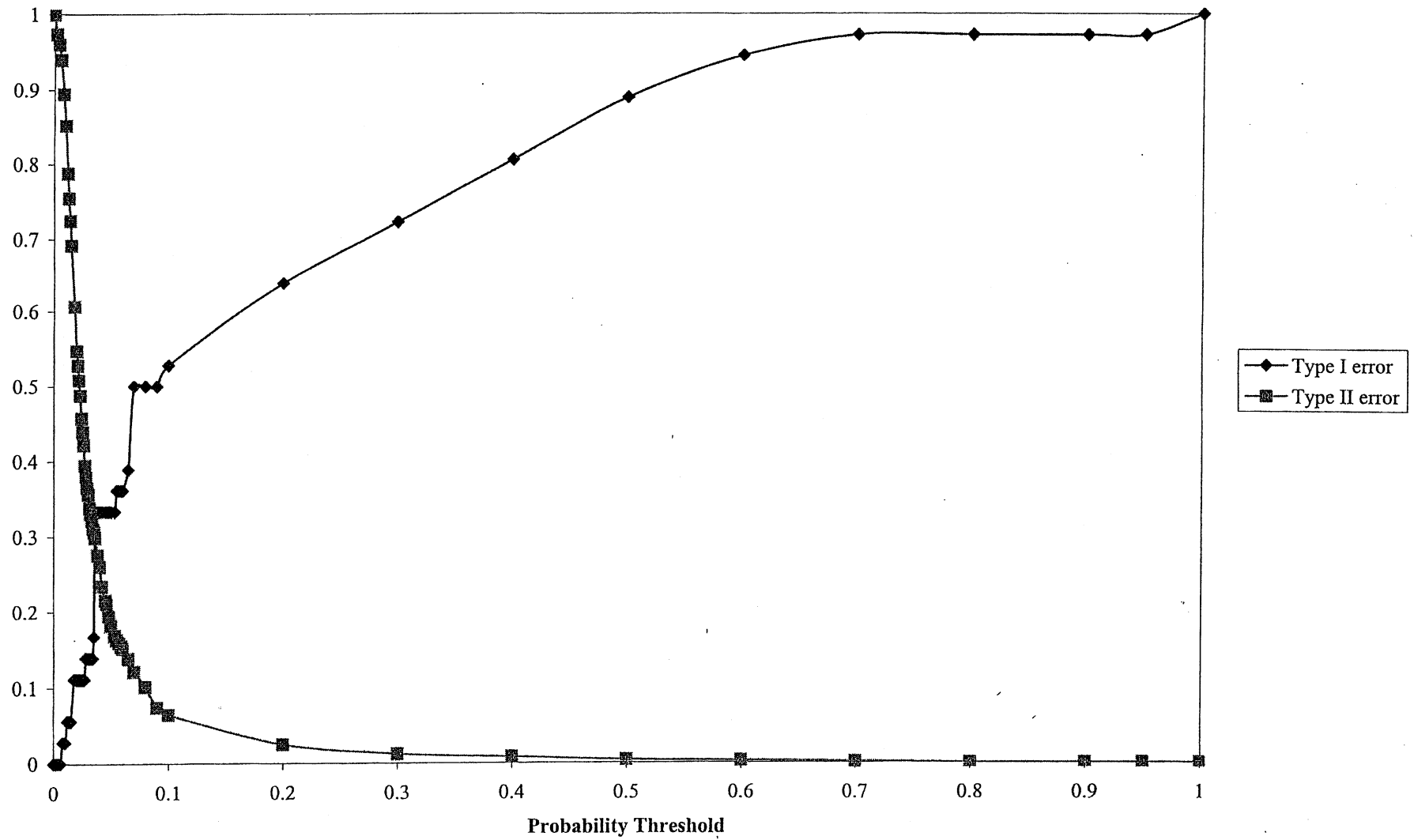
Using in-sample frequencies as estimates of the true parameters, the parameter w should be equal to the frequency of banking crises in the sample, namely 0.047. The functions $a(T)$ and $b(T)$, that trace how error probabilities change with the threshold for issuing warnings, can be obtained from the in-sample estimation results as follows: given a threshold of—say— $T = 0.05$, we can obtain $a(0.05)$, i.e. the associated probability of type I error, as the percentage of banking crises in the sample with an estimated crisis probability below 0.05. Similarly, $b(0.05)$, the probability of issuing a warning when no crisis occurs, is the percentage of noncrisis observations with an estimated probability of crisis above 0.05. Figure 1 shows the functions $a(T)$ and $b(T)$ for $T \in [0, 1]$ computed from the estimation results of Section III above. Of course, $a(T)$ is increasing, as the probability of not issuing a warning when a crisis occurs increases as the threshold rises, while $b(T)$ is decreasing. The two functions cross at $T = 0.036$, where the probabilities of either type of error is about 30 percent.

Figure 1 also shows that crisis probabilities estimated through our multivariate logit framework can provide a more accurate basis for an early warning system than the indicators developed by Kaminsky and Reinhart (1999): as discussed in Section II above, the indicator of banking crises associated with the lowest type I error in the Kaminsky-Reinhart framework is the real interest rate, with a type I error of 70 percent and a type II error of 19 percent. With our model, as shown in Figure 1, a threshold for type I error of slightly over 70 percent (72 percent, to be precise) comes at the cost of a type II error of only 1.2 percent. Similarly, the best indicator of banking crises according to Kaminsky and Reinhart is the real exchange rate, with a type I error of 73 percent and a type II error of 8 percent (resulting in an adjusted noise-to-signal ratio of 0.30). With our model, a type II error of 7.4 percent can be obtained by choosing a probability threshold of 0.09, and it is associated with a type I error of only 53 percent, resulting in an adjusted noise-to-signal ratio of 0.25. We conjecture that the better performance of the multivariate logit model stems from its ability to combine into one number (the estimated crisis probability) all the information provided by the various economic variables monitored.⁹

⁸A risk-averse decision-maker would place greater weight on minimizing type I errors relative to type II errors, since type I errors are more costly. We are indebted to a referee for suggesting this point to us.

⁹It should be pointed out that the logit parameters are estimated using maximum likelihood, and the likelihood function does not take into account the different costs of type I and type II errors. Another avenue to improve the warning system could be to choose parameters to minimize the decision-maker's loss functions.

Figure 1. Crisis Threshold and In-Sample Classification Accuracy



B. Choosing the Optimal Threshold

By way of illustration, we have computed loss functions for three alternative configurations of the cost parameters of the decision-maker. The cost of taking further action as a result of a warning c_1 is normalized to one in all three scenarios, while the cost of suffering an unanticipated crisis c_2 takes the values 20, 10, and 5 respectively. The three resulting loss functions are plotted in Figure 2.¹⁰ The values of the warning threshold that minimize the loss functions are, respectively, $T=0.034$, $T=0.09$, and $T=0.20$. In other words, a decision-maker whose cost of missing a crisis is 10 times the cost of taking precautionary measures would issue an alarm every time the forecasted probability of crisis exceeds 9 percent, and similarly for the other cases. Thus, as expected, as the cost of missing a crisis increases relatively to the cost of taking preventive action, the optimal threshold of the warning system falls, resulting in a warning system with fewer type I errors and more type II errors.

Figure 3 shows the optimal probability threshold for a broad range of values of the parameter c_2 , namely $c_2 \in [2, 40]$, while c_1 is kept constant at 1. For values of c_2 between 40 and 15 the optimal probability threshold for issuing a warning is $T = 0.034$. With this criterion, the probability of not issuing a warning when a crisis occurs is about 14 percent, while the probability of mistakenly issuing a warning is 31 percent. As c_2 declines below 15, the threshold increases to 0.09 (type I error of 50 percent, and type II error of 7.4 percent), and remains there until c_2 reaches 8. At this point, the threshold jumps to 0.20, as the decision-makers is very concerned about false alarms. Finally, if the cost of missing a crisis is as low as 2-3 times that of issuing a false warning, then the optimal threshold is 0.30, corresponding to a type I error as high as 72.2 percent and a type II error as low as 1.2 percent.

To fully appreciate the nature of the warning system, it is worth pointing out that the probability of a type I error is *not* the probability of missing a crisis. To obtain the probability of missing a crisis, the probability of a type I error must be multiplied by the unconditional probability of a crisis, which in our sample is 0.047. Similarly, the probability of issuing a wrong warning is the size of the type II error times the frequency of noncrisis observations. With a threshold of $T=0.09$, the probability of missing a crisis is, therefore, only 2.3 percent, since crises occur rarely. In contrast, the probability of receiving a false alarm is 7.1 percent, because noncrisis observations tend to be the majority.

So, based on our framework for forecasting crisis probabilities, warning systems associated with a relatively low incidence of type I error (below 15 percent) give rise to a fairly large amount of false alarms, in part because crises tend to be infrequent events. If the system is used as a preliminary screen, and further information gathering can provide an

¹⁰ To keep the image sufficiently clear in the relevant range, we have omitted values of the loss functions for $T > 0.30$. The functions continue to increase in the omitted range.

Figure 2. Loss Function for Various Values of the Cost Parameters

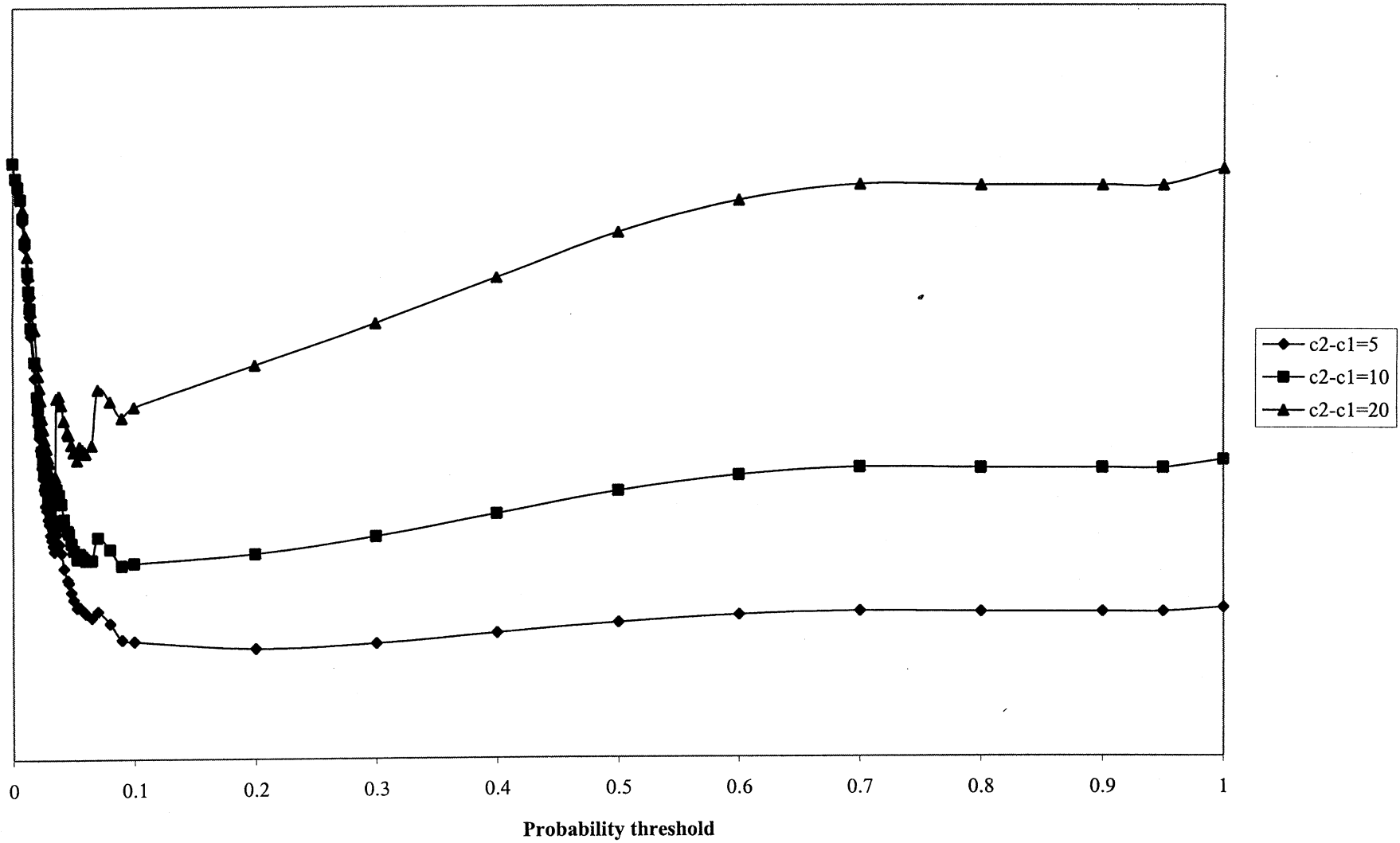
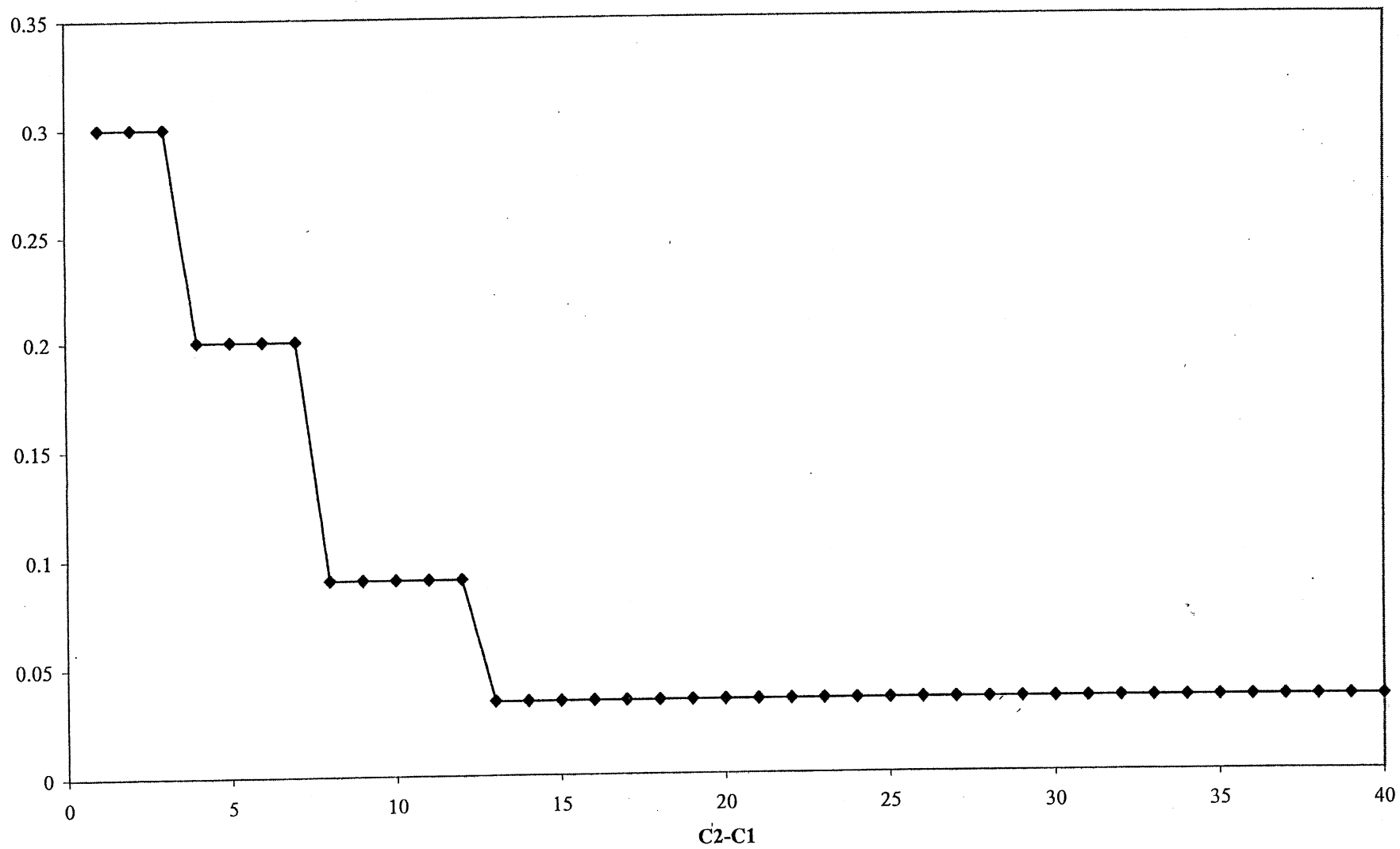


Figure 3. Optimal Probability Threshold



effective way to sort out cases in which the banking system is sufficiently sound, then the decision-maker would be willing to accept the high incidence of type II error. It should also be pointed out that, in some cases, what is considered a false alarm by the model may actually be a useful signal. To illustrate this point, we have examined the “false alarms” generated in-sample by a threshold of 0.047. As it turns out, in 21 cases the “false positives” were observations in the two years immediately preceding a crisis, suggesting that the conditions that eventually led to a full-fledged crisis were in place (and were detectable) a few years in advance. In other cases, the “false alarms” may have corresponded to episodes of fragility that were not sufficiently severe to be classified as full-fledged crises in our empirical study, or where a crisis was prevented by a prompt policy response. Thus, an assessment of the accuracy of the warning system based on in-sample classification accuracy may exaggerate the incidence of type II errors. On the other hand, as usual, out-of-sample predictions are subject to additional sources of error relative to in-sample prediction: the forecasted values of the explanatory variables include forecast errors, and there may be structural breaks in the relationship between banking sector fragility and the explanatory variables which make predictions based on past behavior inadequate. Also, despite the large size of our panel, the number of systemic banking crises in the sample (36) is still relatively small, so that small sample problems may affect the estimation results. Obviously, as more data become available and the size of the panel is extended, this problem should become less severe.

V. USING ESTIMATED CRISIS PROBABILITIES TO CONSTRUCT A RATING SYSTEM FOR BANK FRAGILITY

In this section, we consider the problem of a monitor whose task is to rate the fragility of a given banking system. The rating will then be used by other agents to decide on a possible policy response, but the monitor is not necessarily aware of the costs and benefits of such policy actions. Another rationale for using fragility classes instead of a critical threshold as a monitoring device is that small changes in the critical threshold may lead to substantial differences in type I and type II errors, as evident from Figure 1. To construct fragility classes, it seems desirable for the classification criterion to have a clear interpretation in terms of type I and type II error. This has two advantages: first, agents who learn the rating can do their own cost/benefit calculations when they decide whether or not to take action; second, the fragility of two systems that are assigned two different ratings can be compared based on a clear metric.

The starting point for constructing the rating system is once again the set of forecasted crisis probabilities obtained using the coefficients estimated in the multivariate logit regression of Section III above. Clearly, a country with a forecasted probability of x should be deemed more fragile than one with an estimated probability of $y < x$. To establish fragility “classes,” one can partition the interval $[0, 1]$, which is the set of possible forecasted crisis probabilities, into a number of subintervals, and assign a rating to all estimated probabilities within a given class. Obviously, there are no objective criteria for choosing one particular partition, but a number of considerations help narrowing down partitions that may be useful. First, because the

frequency of crises in the sample is small, choosing fine partitions would give rise to misleading results, because there would be many classes with no observed crises in them. For instance, as shown by the flat section of the type I error curve in Figure 1, in our sample there are no crises episodes with an estimated crisis probability between 4 percent and 5 percent. On the other hand, there are episodes with an estimated probability between 3 percent and 4 percent. If we choose one of the classes to be the interval [0.04-0.05] and the other the interval [0.03, 0.04], then it would appear that fragility decreases with the estimated crisis probability, an obviously misleading conclusion. Thus, due to the small number of crises relative to sample observations, only fairly coarse partitions will give rise to sensible results. Another caveat is that the empirical distribution of the estimated probabilities is strongly skewed towards zero: only 8.5 percent of the observations have probabilities larger than 10 percent, and over 45 percent are in the 0-2 percent range. Thus, partitioning the unit interval by subsets of the same size would result in a very uneven number of observations belonging to each class, with very few observations in the highest probability intervals.

Based on these considerations, we have constructed an example of a rating system with four fragility classes (Table 4). The system uses “intuitive” thresholds of type I error to determine the upper bound of each class. More specifically, the upper bounds of each of the four classes have been chosen so that the type I error associated with the bounds are 10 percent, 30 percent, 50 percent, and 100 percent respectively. According to this criterion, observations with forecasted crisis probability below 1.8 percent belong to the lowest fragility class. Observations with probability between 1.8 percent and 3.6 percent are in the second lowest class; the third group has forecasted probabilities up to 7 percent, while observations with forecasted probabilities above 7 percent are classified in the highest fragility group. Table 4 also reports the values of the type II error associated with the upper bound of each class. These values are (about) 60 percent, 30 percent, 12 percent, and zero respectively.

Table 4. A Rating System for Banking Sector Fragility

Class	Probability Interval	Type I error	Type II error	No. of observations	Crises/ observations
I	0-0.018	0-0.10	1-0.60	291	0.01
II	0.018-0.036	0.10-0.30	0.60-0.30	232	0.03
III	0.036-0.07	0.30-0.50	0.30-0.12	136	0.05
IV	0.07-1	0.50-1	0.12-0	107	0.17

To illustrate the meaning of the fragility groupings, consider that, if all observations with forecasted probability in classes higher than the first (i.e., observations with probability above 1.8 percent) were treated as crises, then the likelihood of missing a crisis (given that one takes place) would be below 10 percent. On the other hand, the probability of falsely calling a crisis would be over 60 percent. Another way to put it is that, in sample, 90 percent of the crisis observations have a probability higher than the probabilities in the lowest fragility class. Similarly, if one were to classify as crises only observations with forecasted probability in the highest two fragility classes, then the probability of missing a crisis would be 30 percent and the probability of a false alarm would fall to 30 percent as well.

As an additional measure of the degree of fragility associated with each class, we have computed the fraction of sample observations in each class that corresponds to an actual banking crisis. This measure goes from 1.5 percent for the lowest fragility class to 16.8 percent for the highest. Thus, the likelihood that an observation in the highest fragility class is a crisis is 16.8 percent; this may seem quite low, but it should be compared to the unconditional probability of crisis of only 4.7 percent (the sample frequency of crises). To put it another way, finding a crisis probability in the highest fragility class tells the analyst that that observation is three and a half times more likely to correspond to a crisis than the average observation. Clearly, these rating systems are just examples of many possible alternatives, and depending on the purposes of the monitor one alternative may be preferable to the other. What is important is that the meaning of the fragility score and the criteria used in rating be made clear to potential users.

VI. AN APPLICATION TO THE BANKING CRISES OF 1996-97

As an illustration of the performance of the monitoring mechanisms developed in the preceding sections, we consider how the system would have fared in relation to the six banking crises that took place in 1996-97, that is after the end of the sample period used in the estimation exercise of Section III above. The six banking crises took place in Jamaica in 1996, and Indonesia, Korea, Malaysia, the Philippines, and Thailand in 1997. Early accounts and analyses of the events surrounding the five Asian crises can be found, for instance, in IMF (1997), Radelet and Sachs (1998), and Goldstein and Hawkins (1998).

To compute out-of-sample banking crisis probabilities for the six countries in 1996 and 1997 we use two alternative sets of values for the explanatory variables. The first set consists of actual realizations of the variables. The out-of-sample probabilities obtained in this way are not true forecasts, of course. In particular, for the five Asian countries these figures reflect the large exchange rate depreciations that took place in the second half of 1997 and their immediate consequences. Since these events were largely unanticipated by observers, it is of interest to try to assess whether signs of increasing banking sector fragility would have been apparent before the depreciations took place. To this end, and, more generally, to assess the performance of the monitoring system when true forecasts are used, we also compute out-of-sample crisis probabilities using forecasts of the explanatory variables as of April-

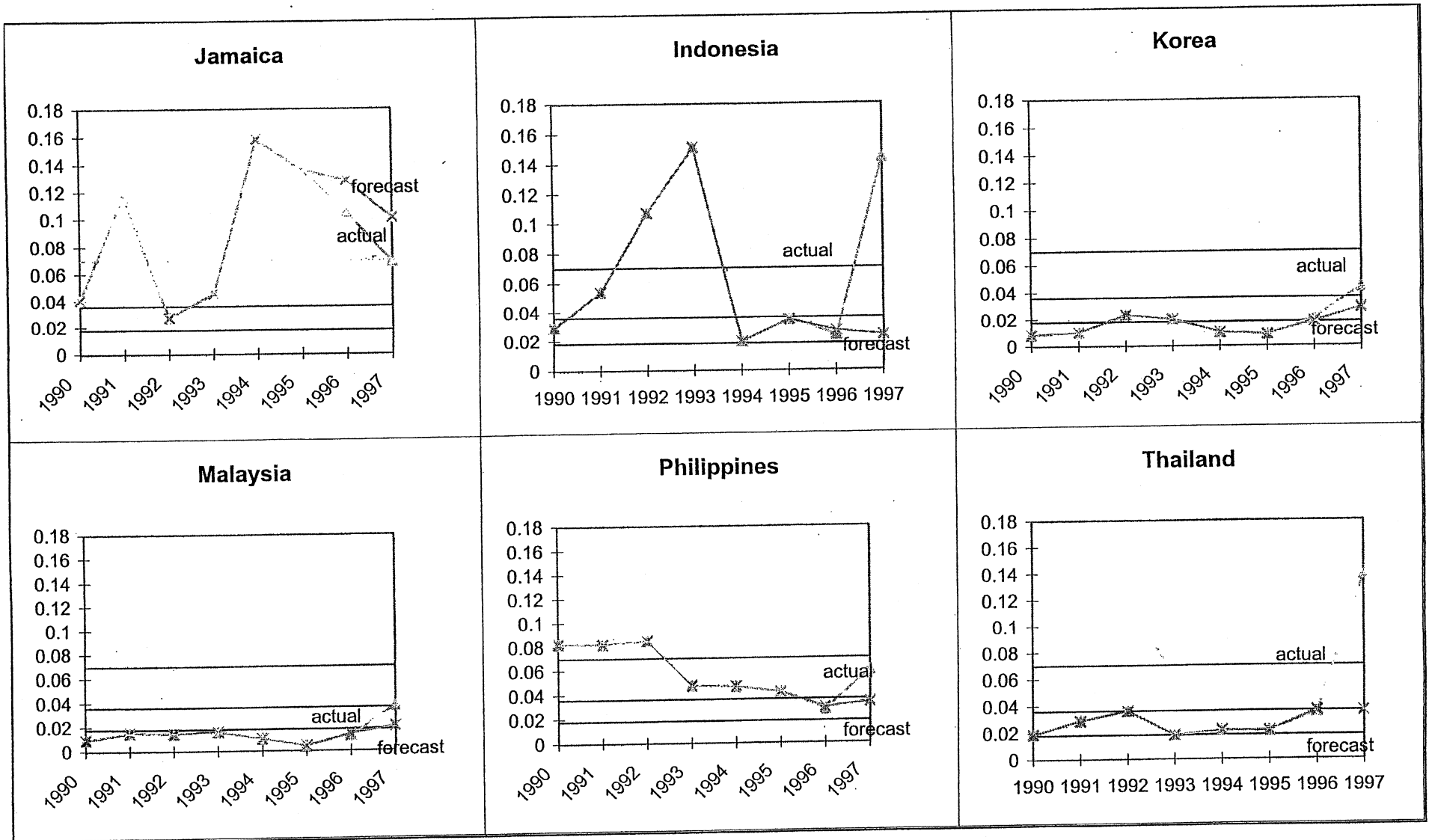
May 1997. Comparison between the two forecasts will show to what extent errors in forecasting explanatory variables would have clouded the fragility assessment based on our model. The forecasted values of the explanatory variables are taken, where available, from the FT Currency Forecaster and from Consensus Forecasts. These publications survey several prominent private sector forecasters and publish the means of the professional forecasts. For the five Asian countries, the growth rate of real GDP, inflation, exchange rate depreciation and the real interest rate are from the FT Currency Forecaster; broad money is from Consensus Forecasts, while the remaining values (and all of the values for Jamaica) are from the May 1997 round of the IMF's semiannual World Economic Outlook (WEO) exercise.¹¹ To compute out-of-sample crisis probabilities using realized values of the explanatory variables, we have used IFS numbers when available, and WEO February 1998 numbers otherwise.

Figure 4 shows estimated crisis probabilities for the six countries in 1990-95, as well as probability forecasts for 1996-97. The two lines for 1996 and 1997 correspond to the two alternative sets of explanatory variables, April-May 1997 forecasts and actual realizations.¹² To give a fragility assessment based on these probabilities, we have chosen the rating system of Table 4 (see Section V above). The horizontal lines in the figures mark the boundaries between each of the four fragility classes, corresponding to fragility rating (type I error) of 10 percent, 30 percent, and 50 percent respectively. Based on forecasts as of April-May 1997, estimated crisis probabilities were relatively low for the five Asian countries, while Jamaica was well into the highest fragility zone as early as 1995. This is not surprising, since all the Asian countries had a very good macroeconomic performance in the years up to 1996, a performance that, by-and-large, was expected to continue. In Jamaica, the forecasted crisis probability was 14 percent in 1995 and 12.80 percent in 1996. Analysis of the factors contributing to the increase in crisis probability indicates that in both years the two main factors were high real interest rates and high inflation. Strong past credit growth and a favorable fiscal position also contributed to fragility in 1995, but not in 1996. The two most fragile Asian countries were Thailand and the Philippines, with a forecasted crisis probability of about 3.5 percent in 1997. This would have placed the two countries on the borderline between the second and third fragility zone based on our rating system. In Thailand, the main factor contributing to bank fragility both in 1996 and in 1997 was the high level of the real interest rate, while strong past credit growth was also a factor. In contrast with Jamaica, however, where GDP growth was lackluster, in Thailand the large predicted rate of growth of GDP worked as an offsetting factor, keeping the overall crisis probability relatively small.

¹¹There are three exceptions to the above-mentioned criteria: for the Philippines, broad money comes from the WEO. For Korea, no forecast of reserves was available, so we arbitrarily assumed reserves to return to their 1995 value in 1997.

¹²The differences between "forecasted" and "actual" figures for 1996 are due to revisions of 1996 data in the February 1998 WEO.

Figure 4. Out-of-Sample Crisis Probabilities Using Forecasted and Actual Data



In the Philippines, the predicted probability increased over twenty percent between 1996 and 1997, mainly due to the high rate of growth of credit two years earlier. The real interest rate was lower than in Thailand, but so was GDP growth. Indonesia, Korea, and Malaysia, all had forecasted crisis probabilities below 3 percent both in 1996 and in 1997 and would have been placed in the second fragility class (actually, Malaysia would have even received the lowest fragility rating in 1996). As in the other two Asian countries, the expectation of continued stability of the exchange rate and, especially, of continued strong GDP growth more than offset the elements of fragility coming from high real interest rates (not, however, in Korea) and strong past credit expansion. In Indonesia, the relatively high rate of inflation also tended to increase bank fragility.

Not surprisingly, the picture obtained by estimating crises probabilities using the latest available data would have been quite different for the five Asian countries, while for Jamaica no striking dissimilarities emerge. Estimated crises probabilities are in the highest fragility class for Indonesia and Thailand, and in the second highest for the other three countries. Malaysia, with a probability of 3.7 percent, appears to be the least fragile.¹³ The decomposition of the probability tells some interesting stories: first, of course, the exchange rate depreciation had an important direct effect on fragility in all five countries. On the other hand, in 1997 inflation was not much higher than forecasted, so it was not among the main factors contributing to increased banking system vulnerability according to our model. In all five countries except Korea lower-than-forecasted GDP growth was one of the main contributing factors, and so was the higher-than-expected real interest rate (except in Thailand).

To summarize, an analysis of banking system fragility using the methods developed in this paper would have clearly indicated an impending banking crisis in Jamaica, while signs of fragility were present in Thailand and the Philippines, the overall image of the five Asian economies would have been a rather reassuring one, as expectations of continued strong economic growth and stable exchange rates would have offset the negative impact of relatively high real interest rates and strong past credit expansion.

VII. CONCLUSIONS

The econometric study of systemic banking crises is a relatively new field of study, and the development and evaluation of monitoring and forecasting tools based on the results of those studies is at an embryonic stage at best. The purpose of this paper has been not so much to propose one or more "ready-to-use" procedure for decision-makers, but rather to highlight what elements need to be evaluated in developing such a procedure, and to explore some possible avenues. The multivariate logit econometric model used here to estimate banking crises probabilities relies solely on aggregate variables which are readily available for a large

¹³Of the five Asian countries, Malaysia is the only one without an IMF program so far.

number of countries, and whose future values are routinely forecasted by professional forecasters or by the Fund, so forecasts of crisis probabilities can be produced at very low cost. Using these forecast probabilities, we have developed two monitoring tools. The first is an “early warning system” that issues a signal in case the forecasted crisis probability exceeds a certain threshold. The appropriate threshold for issuing a warning can be chosen based on the costs of missing a crisis and the benefits of avoiding false alarms. The second monitoring tool is a rating system for bank fragility. In this case, forecasted crisis probabilities are used to classify a particular banking system in one of a few fragility classes. Each fragility class is constructed so that it has a clear interpretation in terms of the likelihood of a crisis, and the fragility level of different classes can be compared based on a well-defined metric. Both monitoring tools can be used to economize on precautionary costs, by pointing to high fragility cases for which more in-depth monitoring efforts are warranted.

The evaluation of banking sector fragility performed along these lines is subject to several potential errors common to all exercises based on forecasts: first, the regression coefficients used to compute crisis probability forecasts are only estimates of the true parameters. Second, new crises may be of a different nature than those experienced in the past, so that the coefficients derived from in-sample estimation may be of limited use out of sample. The latter problem may be particularly severe since banking crises tend to be rare events, and, even though the panel used for in-sample estimation is quite large (766 observations), crisis episodes only number 36. The third source of errors is that forecasts of the explanatory variables are likely to incorporate forecast errors, as vividly illustrated by the example of the five recent Asian crises. Large forecast errors, in turn, may severely distort the fragility assessment resulting from our procedures.¹⁴ One way to reduce the impact of forecast errors is to develop alternative scenarios for the explanatory variables, and to examine banking sector fragility in the context of such scenarios. This would seem particularly useful, because in many cases banking crises are triggered by “extreme” behavior in one or more explanatory variables (a currency collapse, a bout of inflation, a drastic deterioration in the terms of trade) in a context in which other elements also contribute to overall fragility. Routine forecasts of economic variables usually do not encompass extreme events of this sort,

¹⁴Another direction in which this work can be extended is to explore alternative model specifications, and compare them from the point of view of their usefulness for forecasting (see, for instance, Diebold (1997)). Here we have used a specification developed in our previous work after eliminating explanatory variables for which forecasts were not readily available. It could be that an even more parsimonious specification is more suitable for forecasting purposes. We leave this issue to future extensions.

which, instead, tend to be discussed as “risk elements” of the overall picture.¹⁵ The framework used here would lend itself quite easily to the evaluation of fragility in alternative scenarios, as the contribution of each individual explanatory variable to the forecasted crisis probability can be clearly isolated.

Another important *caveat* is that, while aggregate variables can convey information about the general economic conditions that tend to be associated with banking sector fragility, they are silent about the situation at individual banks or in specific segments of the banking sector, so crises that may develop from specific weaknesses in some market segments and spread through contagion would not be detected. Also, informed observers who are familiar with a particular country are likely to be in a better position to detect signs of incoming trouble, so the information generated by a quantitative approach such as ours should not replace but rather complement other sources of information.

A final message from this exercise is that, to be useful, a monitoring system must be designed to fit the preferences of the decision-maker, and so the development of a system must be the outcome of an interactive process that involves both econometricians and policy-makers.

¹⁵This is certainly true of Fund forecasts, which often tend to be excessively optimistic (Mussa and Savastano (1999)). In the case of the Asian countries, we have computed crisis probabilities using the most pessimistic forecasts in the Consensus Forecasts group, but this did not lead to a substantial increase in forecasted crisis probabilities.

References

- Calvo, Guillermo A., 1996, "Capital Flows and Macroeconomic Management: Tequila Lessons," *International Journal of Finance & Economics*, No. 1, pp. 207-24.
- Caprio, Gerry, and D.Klingebl, 1996, "Dealing with Bank Insolvencies: Cross Country Experience," (Washington: The World Bank).
- Demirgüç-Kunt, Asli, and E. Detragiache, 1998a, "The Determinants of Banking Crises in Developing and Developed Countries," *IMF Staff Papers*, March, (Washington: International Monetary Fund).
- , 1999, Financial Liberalization and Financial Fragility, in B. Pleskovic and J.E. Stiglitz (Eds.) *Proceedings of the 1998 World Bank Conference on Development Economics*, (Washington: The World Bank).
- Diebold, Francis, *Elements of Forecasting*, (Cincinnati, Ohio: South-Western College Publishing).
- Eichengreen, Barry, and Andrew K. Rose, 1998, "Staying Afloat While the Wind Shifts: External Factors and Emerging-Market Banking Crises," NBER Working Paper No. 6370 (Cambridge, Massachusetts: MIT Press).
- Gavin, Michael, and Ricardo Hausman, 1996, "The Roots of Banking Crises: the Macroeconomic Context," in Hausman and Rojas-Suarez (Eds.), *Volatile Capital Flows: Taming their Impact on Latin America*, (Washington: Inter-American Development Bank).
- Goldstein, Morris, and John Hawkins, 1998, "The Origins of the Asian Financial Turmoil," Reserve Bank of Australia.
- Hardy, Daniel, and Ceyla Pazarbaşıoğlu, 1998, "Leading Indicators of Banking Crises: Was Asia Different?," IMF Working Paper No. 91, (Washington: International Monetary Fund).
- Honohan, Patrick, 1997, "Banking System Failures in Developing and Transition Countries: Diagnosis and Predictions," BIS Working Paper No. 39, (Basle, Switzerland: Bank of Economic Settlements).
- International Monetary Fund, 1997, *World Economic Outlook*, Interim Assessment, December, (Washington: International Monetary Fund).
- , 1998, *World Economic Outlook*, (Washington: International Monetary Fund).

Kaminsky, Graciela, and C.M. Reinhart, 1999, "The Twin Crises: The Causes of Banking and Balance of Payments Problems," *American Economic Review*, Vol. 3 No. 89, pp. 473-500.

Lindgren, C.J., G. Garcia, and M. Saal, 1996, *Bank Soundness and Macroeconomic Policy*, (Washington: International Monetary Fund).

Mussa, Michael, and Miguel Savastano, 1999, "The IMF Approach to Stabilization," mimeo, (Washington: International Monetary Fund).

Radelet, Steven, and J. Sachs, 1998, "The Onset of the Asian Financial Crises," forthcoming in P. Krugman (Ed.), *Currency Crises*, (forthcoming, Chicago: University of Chicago Press).

Rojas-Suarez, Liliana, 1998, "Early Warning Indicators of Banking Crises: What Works for Developing Countries?," unpublished manuscript, February.