

IMF Working Paper

© 1998 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/98/154

INTERNATIONAL MONETARY FUND

Research Department

Are Currency Crises Predictable? A Test

Prepared by Andrew Berg and Catherine Pattillo¹

Authorized for distribution by Eduardo Borensztein

November 1998

Abstract

This paper evaluates three models for predicting currency crises that were proposed before 1997. The idea is to answer the question: if we had been using these models in late 1996, how well armed would we have been to predict the Asian crisis? The results are mixed but somewhat encouraging. One model, and our modifications to it, provide useful forecasts, at least compared with a naive benchmark. The head-to-head comparison also sheds light on the economics of currency crises, the nature of the Asian crisis, and issues in the empirical modeling of currency crises.

JEL Classification Numbers: F31, F47

Keywords: Currency crises, vulnerability indicators, Asian crisis, balance of payments crises, crisis prediction

Authors E-Mail Address: aberg@imf.org; cpattillo@imf.org

¹We would like to thank, without implication, Graciela Kaminsky, Andy Rose and Aaron Tornell for help reproducing and interpreting their results, Brooks Calvo, Maria Costa, Manzoor Gill, and Nada Mora for superb research assistance, and Eduardo Borensztein, Steve Kamin, Hashem Pesaran, and many IMF colleagues for useful comments.

	Page
Summary	4
I. Introduction	5
II. Kaminsky-Lizondo-Reinhart (1997) Signals Approach	8
A. Methodology	8
B. Implementation	9
C. Predicting 1997	25
III. Frankel and Rose (1996) Probit Model Using Multi-Country Sample	36
A. Methodology	36
B. Implementation	37
C. Predicting 1997	42
IV. Sachs, Tornell, and Velasco (1996) Cross-Country Regressions	43
A. Methodology	43
B. Implementation	46
C. Predicting 1997	50
V. Discussion	53
VI. Conclusion	55
Tables	
1. Performance of Indicators	10
2. Proportion of Indicators Signaling a Crisis	13
3. Comparing Predictive Power of Alternative Composite Indicators—In Sample	17
4. Testing Indicators Against More General Piecewise-Linear Specifications in Bivariate Probit Models	20
5. Multivariate Probit Models	23
6. Comparing Predictive Power of Alternative Composite Indicators—Out-of-Sample	26
7. Correlation of Actual and Predicted Rankings Based on KLR Approach	28
8a. Summary Measures for Selected Countries: Asian Countries	33
8b. Summary Measures for Selected Countries: Latin American Countries	34
9. Frankel and Rose: Probit Estimates of Probability of a Currency Crash, 1970–92 ..	38
10. Frankel and Rose: Probit Estimates of Probability of a Currency Crash, 1970–96 ..	41
11a. Currency Crash Probabilities Based on FR Probit Model	44
11b. Currency Crash Probabilities Based on Revised FR Models	45
12. STV: 1994/5 Regressions	48
13. STV: Predicted and Actual Crises in 1997	51
14. Correlation of Actual and Predicted Rankings Based on KLR, BP, FR, and STV ..	54

Figures:

1.	KLR Weighted-Sum Crisis Probabilities for Selected Countries	15
2.	Relationship Between Predictive Variable and Probability of Crisis	18
3.	Average No. Of Crises in Next 24 Months by Percentile of Variable	21
4a..	KLR Weighted-Sum Crisis Probabilities for Selected Countries	30
4b.	Crisis Probabilities Based on Linear Probit Model for Selected Countries	35
References		60

Summary

This paper evaluates three different pre-1997 econometric models for predicting currency crises. The question is, If we had been using these models in late 1996, would we have been better able to predict the Asian crisis? We chose the three papers: Kaminsky, Lizondo, and Reinhart (1997) (hereafter KLR); Frankel and Rose (1996) (FR); and Sachs, Tornell, and Velasco (1996) (STV).

The answer to the question posed in the title of this paper is, "Yes, but not very well." The answer is "yes" since the KLR forecasts are clearly better than a naive benchmark of pure guesswork. We say "not very well" because none of these models reliably predicts the timing of crises, that is, *whether* there would be crises in 1997. False alarms almost always outnumber appropriate warnings. Moreover, the statistically significant results imply that some of the models clearly outperform pure guesswork, not that they do better than the analysis of informed observers.

These models may give insight into the nature and causes of crises independently of their value as predictors. They demonstrate that the probability of a currency crisis increases when domestic credit growth is high, the bilateral real exchange rate is overvalued relative to trend, and reserves are low when measured as a ratio to broad money. All but STV also suggest that a large current account deficit tends to precede crises. For the 1997 crises, some of the models make significant out-of-sample predictions despite the omission of some heavily emphasized phenomena such as weak banking supervision and poor corporate governance. Apparently, the Thai crisis was relatively easy to predict; the Indonesian crisis was perhaps less foreseeable. This exercise has also shed some light on empirical modeling issues.

These models are not the last word. We also estimate an alternative model that provides generally better forecasts than the best tested alternative.

I. INTRODUCTION

In recent years, a number of researchers have claimed success in systematically predicting which countries are more likely to suffer currency crises. The Asia crisis has stimulated further work in this area, with several papers already claiming to be able to “predict” the incidence of this crisis using pre-crisis data.²

It may seem unlikely that currency crises should be systematically predictable. In practice, they usually seem to come as a surprise. Since the exchange rate is an asset price, it is reasonable to doubt that sharp and predictable movements are consistent with the actions of forward-looking speculators. Early theoretical models of currency crises suggested that crises may, however, be predictable even with fully rational speculators. They emphasized an inconsistency between the maintenance of a currency peg and other economic policies. The signs of this inconsistency, such as large government deficits or declining reserves, should help predict crises. A central insight of these models is that even if the crisis and its timing are fully predictable (for example, because excessive money creation is leading to a steady loss of reserves), speculators will wait until reserves are below some critical level before they attack.³

Later analyses have extended this picture in several ways, partly inspired by the apparent absence of weakening fundamentals prior to the successful attacks on various EMU currencies in the early 1990s. These so-called “second generation” models have made the point that currency crises could represent not the result of a deteriorating underlying situation but instead a “jump” from one equilibrium, the pegged regime, to another, the devalued or floating regime. As with a bank in the absence of deposit insurance, two equilibria are possible: one with default (here devaluation) and one without. In this view, a country may be in a situation in which an attack, while not inevitable, might succeed if it were to take place. The exact timing of crises would be essentially unpredictable. Even here, though, it may be possible to identify whether a country is in a zone of vulnerability, that is whether fundamentals are sufficiently weak that a shift in expectations could cause a crisis. In this case, the relative vulnerability of different countries might predict relative severity of crises in response to a shock such as a global downturn in confidence in emerging markets.⁴

It is one thing to say that currency crises may be predictable in general, however, and another that econometric models that are estimated using historical data on a panel or cross-section of countries can foretell crises with any degree of accuracy. Here the question is whether crises are sufficiently similar across countries and over time to allow generalizations from past experience, and whether adequate data on the signs of crisis are available. Each crisis episode presents unique features, and many factors that may indicate a higher probability

²IMF (1998), Kaminsky (1998), Sachs and Radelet (1998b), Sachs (1997), Corsetti et al. (1998), Tornell (1998).

³Krugman (1979), Blanco and Garber (1986).

⁴See Flood and Marion (1998) for a survey of this literature.

of crisis, such as inadequate banking supervision or a vulnerable political situation, are not easily quantified.

The possible endogeneity of policy to the risk of crisis may also limit the predictability of crises. For example, authorities within a country, or their creditors, might react to signals so as to avoid crises.⁵ On the other hand, a focus by market participants on a particular variable could result in its precipitating a crisis where one might not otherwise have occurred.

Ultimately, the question of whether crises are predictable can only be settled in practice. The recent work claiming success in predicting crises has focussed almost exclusively on in-sample prediction, that is on formulating and estimating a model using data on a set of crises, then judging success by the plausibility of the estimated parameters and the size of the prediction errors for this set of crises.⁶ The key test is not, however, the ability to fit a set of observations after the fact, but the prediction of future crises. Can the model predict the crises that are not in the sample used in its estimation? Given the relatively small number of crises in the historical data, the danger is acute that specification searches through the large number of potential predictive variables may yield spurious success in “explaining” crisis within the sample. The possibility that the determinants of crises may vary importantly through time also suggests the importance of testing the models out-of-sample.

The flurry of work between the 1994 and 1997 crises and the large number of crises observed in 1997 provides an excellent opportunity to test existing state-of-the-art “early warning systems” out-of-sample. This paper evaluates three different models proposed before 1997 for predicting currency crises. The idea is to try to answer the question: if we had been using these models in late 1996, how well armed would we have been to predict the Asia crisis?

We chose the following three approaches based on their promise as early warning systems and their success within sample:

- Kaminsky, Lizondo, and Reinhart (1997) (hereafter KLR) monitor a large set of monthly indicators that signal a crisis whenever they cross a certain threshold. This approach has the potential attraction that it produces thresholds beyond which a crisis is more likely. This accords with the common practice of establishing certain warning zones, such as current account deficits beyond 5 percent of GDP or reserves less than three months of imports. The authors claim some success in developing a set of indicators that reliably predict the likelihood of crisis. Moreover, Kaminsky (1998) and Goldstein (1998) have asserted that this method can be applied successfully to the 1997 crises.

⁵Initially successful early warning systems might thus cease to work following publication.

⁶Exceptions are Tornell (1998), discussed below, and Kaminsky (1998), which, while it presents out-of-sample estimates of the probability of currency crisis, does not provide tests of whether these forecasts are better than, for example, guesswork.

- Frankel and Rose (1996) (FR) develop a probit model of currency crashes in a large sample of developing countries. Their use of annual data permits them to look at variables, such as the composition of external debt, that are available only at that frequency.
- Sachs, Tornell, and Velasco (1996) (STV) restrict their attention to a cross-section of countries in 1995, analyzing the incidence of the “tequila effect” following the Mexico crisis. They concentrate on a more structured hypothesis about the cause of this particular episode, emphasizing interactions among weak banking systems, overvalued real exchange rates, and low reserves. They claim to explain most of the cross-country pattern of currency crisis in emerging markets in 1994–1995. Their approach has also been applied to analyzing the Asia crisis.⁷

The paper is organized as follows. Sections 2 through 4 implement each model in turn. For each method, we:

- Briefly describe the methodology.
- Duplicate the original results as closely as possible, using where possible the original data. We also re-estimate over the same sample, fixing any errors in the original estimates and using currently available and hence revised data.
- Reestimate the models using data through 1996 in order to forecast for 1997, as would a researcher who at the end of 1996 aimed to predict crises the following year. We use two samples of countries: the same as the original paper, and another common sample for purposes of comparing across the three methods.
- Make a few plausible extensions or improvements. These changes are in some cases inspired by events in 1997, but again we estimate using data only through 1996.
- Use the models to forecast the probability or severity of crisis for 1997. We generate a ranking of countries according to predicted probability or severity of crisis in 1997 for each model, then compare the predicted and actual rankings.

Section 5 summarizes and discusses the results. A conclusion follows in Section 6.

⁷Tornell (1998), Sachs, and Radelet (1998b) and Corsetti et al. (1998) estimate variants of STV for 1997. IMF (1998) constructs a composite indicator of crisis based on the STV approach and argues that it accords well with the pattern of country experience in the Asia crisis.

II. KAMINSKY-LIZONDO-REINHART (1997) SIGNALS APPROACH

A. Methodology

KLR propose the monitoring of several indicators that tend to exhibit unusual behavior prior to a crisis. A currency crisis is defined to occur when a weighted average of monthly percentage depreciations in the exchange rate and monthly percentage declines in reserves exceeds its mean by more than three standard deviations.⁸ KLR choose 15 indicators based on theoretical priors and on the availability of monthly data.⁹ An indicator issues a signal whenever it moves beyond a given threshold level. A “good” signal is one that is followed by a crisis within 24 months. An “optimal” set of thresholds is calculated, defined as a set that minimizes the noise-to-signal ratio: i.e., the ratio of false signals to good signals.

Thresholds are defined relative to the percentiles of the distribution of the indicator by country. For example, the threshold for real exchange rate deviations might be the 85th percentile, so that any value of the real exchange rate deviation above this percentile would constitute a signal. The percentiles are calculated relative to each country's empirical distribution of the variable in question. To continue the example, the threshold value of the real exchange rate deviation for each country is the 85th percentile of *that country's* distribution of real exchange rate deviations. Thus, minimizing the noise-to-signal ratio for the sample of countries yields a percentile for each indicator that is uniform across countries, but the corresponding country-specific thresholds associated with that percentile will differ across countries.

⁸Weights are calculated so that the variance of the two components of the index are equal. Weights and the mean and standard deviation of the exchange rate component of the index are calculated separately for low and high inflation periods, where the latter are defined as the collection of months for which inflation in the previous six months was greater than 150 percent. Note that lack of data precluded the inclusion of domestic interest rates in the crisis definition. Eichengreen, Rose, and Wyplosz (1996), who analyze currency crises in developed countries, include domestic interest rates.

⁹Indicators are: (1) international reserves (in U.S. dollars); (2) imports (in U.S. dollars); (3) exports (in U.S. dollars); (4) terms of trade; (5) deviations of the real exchange rate from a deterministic time trend (in percentage terms); (6) the differential between foreign and domestic real interest rates on deposits; (7) “excess” real M1 balances, where excess is defined as the residuals from a regression of real M1 balances on real GDP, inflation, and a deterministic time trend; (8) the money multiplier of M2; (9) the ratio of domestic credit to GDP; (10) the real interest rate on deposits; (11) the ratio of (nominal) lending to deposit rates; (12) the stock of commercial bank deposits; (13) the ratio of broad money to gross international reserves; (14) an index of output; and (15) an index of equity prices (measured in U.S. dollars). The indicator is defined as the annual percentage change in the level of the variable (except for the deviation of the real exchange rate from trend, “excess” real M1 balances, and the three interest rate variables).

Some notation may help with this last point. Let x_{it} be a variable that may help predict crises, such as the 12-month growth in exports for country I in period t. The percentile is then $p(x_{it})$, the number between zero and 100 representing where x_{it} fits in the distribution of x_i . $I(p(x_{it}))$ is the indicator, taking a value of 1 when $p(x_{it})$ is above the threshold percentile for that indicator.

The KLR approach is bivariate, in that each indicator is analyzed, and optimal thresholds calculated, separately. Kaminsky (1998) aggregates the separate indicator series into a single crisis index by computing a weighted average of the indicators $I(p(x_{it}))$, with the weights based on the noise-to-signal ratios of each indicator. She then calculates a probability of crisis for each value of the aggregate index by observing how often within the sample a given value of the aggregate index is followed by a crisis within 24 months.

B. Implementation

1. Reproduction of KLR Results

We first attempt to reproduce the KLR results using the same 20-country, 1970–95 sample they use.¹⁰ Following KLR, we first examine the effectiveness of the approach by determining the extent to which each individual indicator is useful in predicting crises.

Table 1 presents information of the performance of individual indicators, from KLR and from our reproduction. Consider the performance of each indicator in terms of the matrix below.¹¹

	Crisis within 24 months	No crisis within 24 months
Signal was issued	A	B
No signal was issued	C	D

The cell A represents the number of months in which the indicator issued a good signal, B is the number of months in which the indicator issued a bad signal or “noise,” C is the number of months in which the indicator failed to issue a signal which would have been a good signal, and D is the number of months in which the indicator did not issue a signal that would have been a bad signal.

The first column in Table 1 shows the noise-to-signal ratio estimated for each indicator. This is defined as the number of bad signals as a share of possible bad signals

¹⁰Argentina, Bolivia, Brazil, Chile, Colombia, Denmark, Finland, Indonesia, Israel, Malaysia, Mexico, Norway, Peru, Philippines, Spain, Sweden, Thailand, Turkey, Uruguay, and Venezuela.

¹¹All tables follow the References section of the paper.

Table 1. Performance of Indicators

Indicator	KLR Sample Rerun				Original KLR Estimates				23-Country Sample, 1970-95:4			
	Noise/Signal (adjusted) 1/ (1)	P(Crisis/Signal) -P(Crisis) 2/ (2)	Threshold Percentile (3)	Number of Crises with data (4)	Noise/Signal (adjusted) 1/ (5)	P(Crisis/Signal) -P(Crisis) 2/ (6)	Number of Crises with data (7)	Noise/Signal (adjusted) 1/ (8)	P(Crisis/Signal) -P(Crisis) 2/ (9)	Threshold Percentile (10)	Number of Crises with data (11)	
Real exchange rate 3/	0.24	32	90	70	0.19	39	72	0.25	29	90	70	
M2/reserves growth rate	0.46	16	87	69	0.48	17	70	0.39	17	87	68	
Export growth rate	0.49	15	90	69	0.42	20	70	0.60	9	90	68	
International reserves growth rate	0.53	13	89	69	0.55	16	72	0.44	15	90	68	
Excess M1 balances 4/	0.69	7	90	68	0.52	15	66	0.60	9	90	67	
Domestic credit/GDP growth rate	0.71	6	85	66	0.62	11	62	0.78	4	84	65	
Real interest rate	0.74	5	82	42	0.77	6	44	0.76	4	88	38	
M2 multiplier growth rate	0.80	4	84	69	0.61	11	70	1.14	-2	81	68	
Import growth rate	1.19	-3	90	69	1.16	-3	71	1.16	-2	86	68	
Industrial production growth rate	1.23	-3	85	58	0.52	16	57	1.14	-2	85	54	
Terms of trade growth rate	1.42	-5	80	63	0.77	6	58	0.93	1	90	55	
Lending rate/deposit rate	1.44	-5	89	34	1.69	-9	33	1.04	-1	89	29	
Bank deposit growth rate	1.53	-6	90	69	1.20	-4	69	1.63	-6	90	68	
Stock price index growth rate	1.81	-9	87	47	0.47	18	53	1.59	-6	80	44	
Real interest differential	1.97	-9	82	42	0.77	6	44	1.34	-4	82	38	
Current account/GDP								0.42	16	90	70	
M2/reserves (level)								0.45	14	90	51	

1/ Ratio of false signals (measured as a proportion of months in which false signals could have been issued $[B/(B+D)]$) to good signals (measured as a proportion of months in which good signals could have been issued $[A/(A+C)]$).

2/ $P(\text{Crisis}/\text{Signal})$ is the percentage of the signals issued by the indicator that were followed by at least one crisis within the subsequent 24 months $[A/(A+C)]$ in terms of the matrix in the text. $P(\text{crisis})$ is the unconditional probability of a crisis, $(A+C)/(A+B+C+D)$.

3/ Deviation from deterministic trend.

4/ Residual from regression of real M1 on real GDP, inflation, and a deterministic trend.

$(B/(B+D))$ divided by the number of good signals as a share of possible good signals $(A/(A+C))$. The threshold percentile, chosen to minimize this ratio, is shown in column 3. Column 2 shows how much higher is the probability of a crisis within 24 months when the indicator emits a signal than when it does not. When the noise-to-signal ratio is less than 1, this number is positive, implying that crises are more likely when the indicator signals than when it does not. Indicators with noise-to-signal ratios equal to or above unity are not useful in anticipating crises.

Our results are broadly similar to those of KLR, though we are not able to match exactly the KLR results, as columns 5 through 7 of Table 1 show.¹² The patterns are quite similar, though column 1 shows slightly weaker performance than reported by KLR for most of the indicators. Differences are starker for four indicators, for which KLR find a noise-to-signal ratio substantially below unity while we find a ratio above unity. Thus, although KLR find 12 informative indicators, that is those with noise-to-signal ratios below unity, we find only eight of these to be informative.¹³

There are a number of possible reasons for the differences in results. We have found that our implementation of the KLR definition of crisis results in a set of crisis dates that do not fully match the KLR crisis dates as reported in Kaminsky and Reinhart (1996). Specifically, we fail to match 14 out of 76 KLR crises.¹⁴ Some of this discrepancy may come from differences in the raw data.¹⁵ We have found that seemingly small differences due to revisions in IFS data can strongly influence the results, and furthermore they and we separately “cleaned” the data of errors.¹⁶

¹²KLR do not report their threshold percentiles.

¹³These are: deviations of the real exchange rate from trend, the growth in M2 as a fraction of reserves, export growth, change in international reserves, “excess” M1 balances, growth in domestic credit as a share of GDP, the real interest rate, and the growth in the money multiplier of M2 (these indicators are also all informative in the KLR analysis).

¹⁴In Kaminsky and Reinhart (1996) there are 76 crises. We find 72 over the same sample. The match is not as good as this suggests, however, as we find some crises that KLR do not and vice versa. There are 14 KLR crises that we do not find and cannot account for (due to small differences in procedure regarding windowing and other identifiable factors). In addition, Kaminsky and Reinhart add a crisis for Chile that is not produced by their definition (personal communication with the author).

¹⁵We have not had access to their data. Our stock price indices are from a different source than KLR, possibly accounting for the difference with regard to this variable.

¹⁶Other possible sources include the fact that we may not have exactly matched the KLR procedure, all the details of which are not fully specified in the paper.

As a first step toward considering the extent to which a group of indicators is useful in predicting crises, Table 2 shows the proportion of good indicators signaling a crisis (good indicators here are those with a noise-to-signal ratio less than 1). In more than one-half of the crises, at least 60 percent of the good indicators were signaling, while this was the case in slightly more than one-third of the tranquil periods. Indicators often emit false signals of crisis, however. Indeed, 98 percent of the times that at least 60 percent of the good indicators were signaling, there was *no* crisis within 24 months.

Having reproduced as nearly as we could the KLR results, we carry out three sorts of modifications. First we change the sample and try two other indicators. In the following subsection, we modify the basic methodology. Specifically, we depart from the entire “indicators” methodology that looks for discrete thresholds and calculates noise-to-signal ratios. Instead, we apply a probit regression technique to the same data and crisis definition as in KLR. In the process we test some of the basic assumptions of the KLR approach.

We modify the sample in two ways. First, we estimate only through April 1995. This reflects the information available to the analyst just before the Thai crisis of July 1997, since the evaluation of an observation requires knowing whether there will be a crisis within 24 months.¹⁷ Second, we change the sample of countries. This will allow the KLR results to be more comparable to those of the other two papers under consideration, as well as serving as an informal test of robustness of the KLR approach. We omit the five European countries from the sample and add other emerging market economies.¹⁸ Our 23-country sample is the union of the emerging market economies in the KLR set and the countries in the STV sample.¹⁹ The last four columns of Table 1 show that indicator performance over the larger sample is broadly similar to results using the KLR sample. At least for the informative indicators, the thresholds appear fairly similar. The average noise-to-signal ratio falls a little for the informative indicators in the 23-country sample (as well as for the entire set of indicators). The most important changes in the noise-to-signal ratios are that the growth of the M2 multiplier is no longer informative while the change in terms of trade is, though only marginally, with a ratio above 0.9. In what follows, we focus on the 23-country sample estimated through April 1995.

¹⁷KLR estimate through 1995 but are not explicit about the final month. We have assumed they use data through December.

¹⁸ We add the following countries to the 15 KLR emerging market economies: India, Jordan, Korea, Pakistan, South Africa, Sri Lanka, Taiwan Province of China, and Zimbabwe.

¹⁹ The complete sample of countries comprises Argentina, Bolivia, Brazil, Chile, Colombia, India, Indonesia, Israel, Jordan, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, South Africa, Spain, Sri Lanka, Taiwan Province of China, Thailand, Turkey, Uruguay, Venezuela, and Zimbabwe.

Table 2. Proportion of Indicators Signaling a Crisis

Proportion of Indicators Signaling a Crisis	Crisis Months		Tranquil Months	
	Number	Percent	Number	Percent
80-100 percent	22	32	788	12
60-79 percent	24	35	1627	26
40-59 percent	13	19	1755	28
20- 39 percent	8	12	1577	25
Less than 20 percent	2	3	610	10

We try two more candidate indicators: the level of M2 to reserves and the ratio of the current account to GDP. KLR used the rate of growth of M2/reserves, but most discussions of crisis vulnerability have focussed on the level of this variable. KLR did not use the current account. We find that the level of M2/reserves is informative, as Table 1 shows. It has about the same noise-to-signal ratio as the rate of change, at 0.42 and 0.39 respectively. The current account/GDP is also highly informative, with a noise-to-signal ratio of 0.45.²⁰

So far we have looked at each indicator separately. Kaminsky (1998) calculates a single composite indicator of crisis as a weighted-sum of the indicators, where each indicator is weighted by the inverse of its noise-to-signal ratio. She then calculates time-series

²⁰ The current account is measured as a moving average of the previous four quarters. We use our interpolated monthly GDP series to form the ratio of the current account to the moving average of GDP over the same period.

probabilities of crisis for each country, based on the sample distribution of this composite indicator.²¹ Figure 1 displays these probabilities and shows some increase in the probability of crisis preceding particular crises for Korea, Indonesia, Malaysia, Philippines and Thailand, as well as for Argentina, Brazil and Mexico.²²

As with other aspects of the KLR methodology, it is somewhat difficult to assess the success of these estimates of the probability of crisis. Figure 1 itself does not tell a clear story. The KLR approach does not lend itself to hypothesis testing; their technique gives no indication of when results are statistically significant.²³

There are, nonetheless, several ways to systematically evaluate the KLR models, as shown in the first two columns of Table 3. For zero/one dependent variables, it is natural to ask what fraction of the observations are “correctly called,” where, for example, a crisis period is correctly called when the estimated probability of crisis is above a given cut-off level and a crisis ensues within 24 months. Such “goodness-of-fit” data are shown in Table 3 for two cut-offs: 50 percent and 25 percent. The in-sample probability forecasts can also be evaluated with analogs of a mean squared error measure, the quadratic probability score (QPS) and log probability score (LPS), that evaluate the accuracy of probability forecasts. In addition, the global square bias (GSB) measures forecast calibration. The QPS ranges from

²¹ Following Kaminsky (1998) the conditional probabilities are generated as follows:

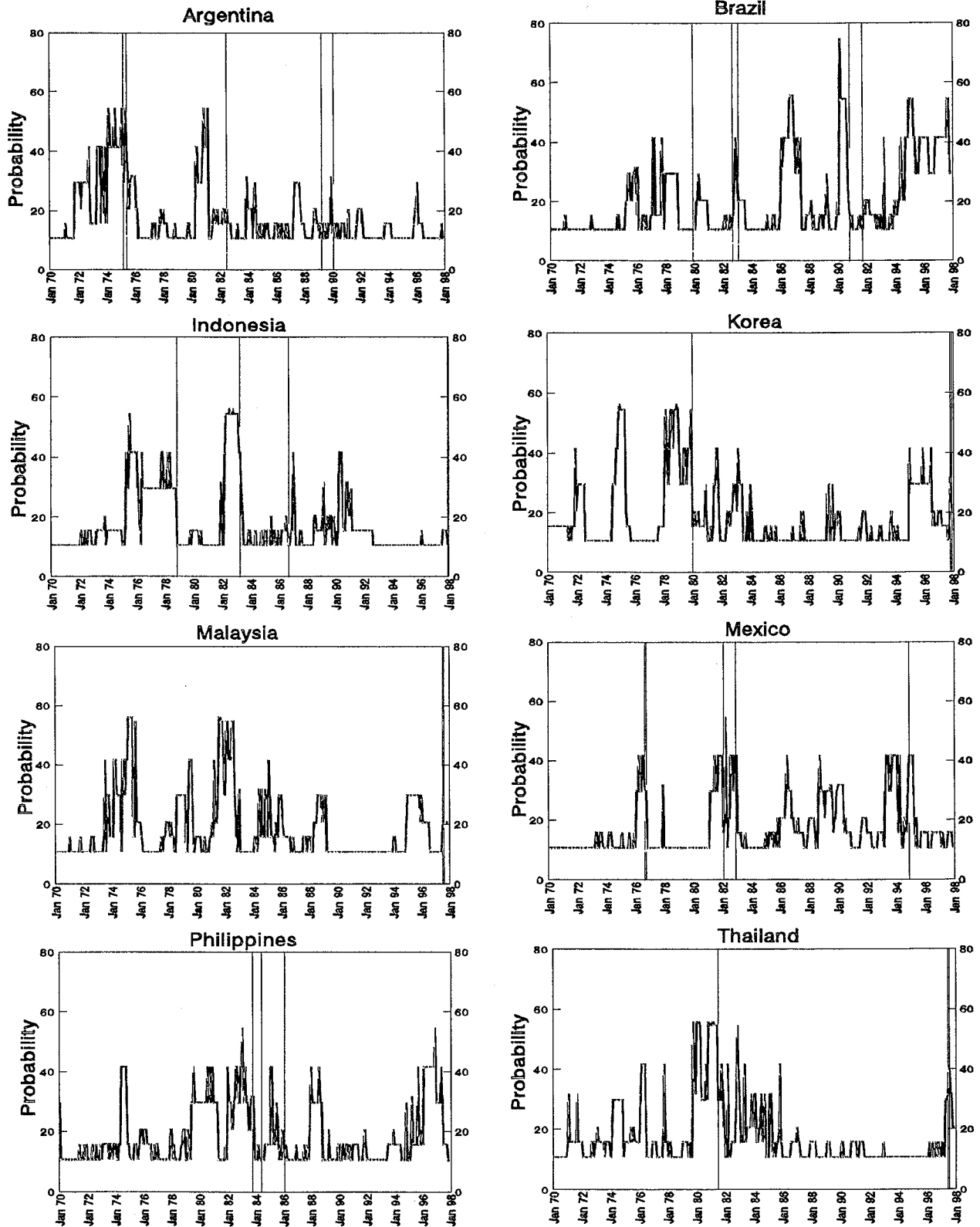
$$Prob(C_{t,t+24}^i | k_t = j) = \frac{\text{Months with } k = j \text{ and a crisis within 24 months}}{\text{Months with } k = j}$$

where k is the sum of the weighted indicators signaling. $Prob(C_{t,t+24}^i | k_t = j)$ is the probability of a crisis for country I in the time interval $\{t, t+24 \text{ months}\}$ given that the weighted-sum of the indicators signaling at time t is equal to j . Unlike Kaminsky (1998), we use only the good indicators, i.e. those with noise-to-signal ratio less than one.

²² The probabilities for 1995:5–1997:12 are out-of-sample estimates. Data through 1995:4 was used to calculate the thresholds for the indicators, and the probabilities as in footnote 20. The probability time series was extended by applying the probabilities to the realizations of weighted indicators signaling in 1995:5–1997:12. The vertical lines represent crisis dates. The pictures are “choppy” because the generation of probabilities in this manner produces time series that alternate between a limited number of values. The continuous variable k is categorized into nine ranges. The probability of crisis can thus take one of only nine values.

²³ Confidence intervals could presumably be generated by Monte Carlo methods, but we do not pursue that here.

Figure 1: KLR Weighted-Sum Crisis Probabilities for Selected Countries 1/



1/ Vertical line represents crisis dates.

zero to 2, and the LPS ranges from zero to infinity, with a score of zero corresponding to perfect accuracy for both. The GSB also ranges from zero to 2, where zero corresponds to perfect global calibration.²⁴

What can we conclude? The first column of Table 3 displays the scores and goodness-of-fit measures for our reproduction of the KLR weighted-sum-based probabilities, excluding our additional variables. The model correctly calls most observations at the 50 percent cut-off, almost entirely through correct prediction of tranquil periods (that is, those that are not followed by crises within 24 months). Almost all (91 percent) of the crisis months (that is, observations followed by a crisis within 24 months) are missed. Even with so few crisis observations correctly called, 44 percent of alarms (that is, observations where the predicted probability of crisis is above 50 percent) are false, in that no crisis in fact ensues within 24 months. As the second column of Table 3 shows, the addition of the current account and M2/reserves in levels only modestly improves the performance of the KLR-based probabilities.

If we are more interested in predicting crises than predicting tranquil periods and are not so worried about calling too many crises, we may want to consider an alarm to be issued when the estimated probability of crisis is above 25 percent. With this lower cut-off, 41 percent of crisis observations are correctly called by the original KLR model. Alternatively, we may ask how often an alarm is actually followed by a crisis within 24 months. With the 25 percent cut-off, the probability of a crisis within 24 months is 37 percent if there is an alarm, much higher than the unconditional probability of crisis in this sample of 16 percent. Now, however, 63 percent of alarms are false.

Still, these predictions are better than guesses. It is true that since most observations are tranquil, even an uninformative model can, by almost always calling for no crisis, predict correctly most of the time. But the model does significantly better than this uninformative benchmark. A Pesaran-Timmermann test rejects, at the 1 percent significance level, the hypothesis that the original KLR model does no better at calling crises than guesses based on the unconditional probability of crisis, using the 25 percent cut-off.²⁵

²⁴For each of the methods we can generate T probability forecasts where $P_t = \text{Prob}(C_{t, t+24})$ is the probability of crisis in the period $[t, t+24 \text{ months}]$. R_t is the actual times series of observations on $C_{t, t+24}$; $R_t = 1$ if a crisis occurs between t and $t+24$ and equals zero otherwise. The analog to mean squared error for probability forecasts is the QPS: $QPS = 1/T \sum_{t=1}^T 2(P_t - R_t)^2$. The analogy is rough, however, because P_t is not the forecast of the event (which is a 0/1 variable) but the probability of the event. Large errors are penalized more heavily under the LPS, given by: $LPS = 1/T \sum [(1 - R_t) \ln(1 - P_t) + R_t \ln(P_t)]$. Overall forecast calibration is measured by the global squared bias $GSB = 2(\bar{P} - \bar{R})^2$, where $\bar{P} = 1/T \sum P_t$, $\bar{R} = 1/T \sum R_t$. Calibration compares the mean forecasted probability to the observed relative frequencies. See Diebold and Lopez (1996) for more discussion.

²⁵With the 50 percent cut-off, the hypothesis is also rejected at the 1 percent level.

Table 3. Comparing Predictive Power of Alternative Composite Indicators--In Sample

	KLR-Based Weighted-Sum Probabilities		Alternative Probit Models		
	Original Specification	Augmented with Current Account and level of M2/reserves	Indicator	Linear	Piecewise- linear
Accuracy and Calibration Scores:					
Quadratic Probability Score	0.270	0.267	0.237	0.236	0.226
Log Probability Score	0.436	0.432	0.391	0.386	0.371
Global Squared Bias	0.00002	0.00002	0.00046	0.00040	0.00046
Goodness-of-Fit: (cut-off probability of 50 percent)					
Percent of observations correctly called	82	83	85	84	85
Percent of pre-crisis periods correctly called 1/	9	9	16	7	19
Percent of tranquil periods correctly called 2/	98	99	99	100	98
False alarms as percent of total alarms 3/	44	30	29	11	34
Goodness-of-Fit: (cut-off probability of 25 percent)					
Percent of observations correctly called	77	75	81	78	80
Percent of pre-crisis periods correctly called 1/	41	46	44	48	47
Percent of tranquil periods correctly called 2/	85	81	89	84	87
False alarms as percent of total alarms 3/	63	65	57	63	59

1/ A pre-crisis period is correctly called when the estimated probability of crisis is above the cut-off probability and a crisis ensues within 24 months.

2/ A tranquil period is correctly called when the estimated probability of crisis is below the cut-off probability and no crisis ensues within 24 months.

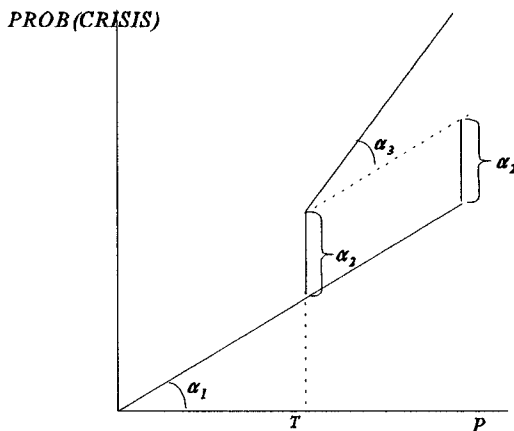
3/ A false alarm is an observation with an estimated probability of crisis above the cut-off (an alarm) not followed by a crisis within 24 months.

2. A Probit-Based Alternative Model

In this section we deviate fairly substantially from the KLR methodology. Specifically, we embed the KLR approach in a multivariate probit framework in which the independent variable takes a value of one if there is a crisis in the subsequent 24 months and zero otherwise. This has three advantages: we can test the usefulness of the threshold concept; we can aggregate predictive variables more satisfactorily into a composite index, taking account of correlations among different variables; and we can easily test for the statistical significance of individual variables and the constancy of coefficients across time and countries.²⁶

KLR assume that the probability of crisis in the subsequent 24 months is a step function of the value of the indicator, equal to zero when the indicator variable is below the threshold and 1 at or above the threshold. They assume, for example, that when the real exchange rate continues to appreciate after it is already above the threshold, this does not increase the probability of crisis. In general, the relationship between a given indicator variable and the probability of crisis could take many more forms than a simple step function. Figure 2 presents various possible relationships between the probability of crisis (on the vertical axis) and the value of a variable $P(x)$, measured as in KLR in percentiles (on the horizontal axis). The KLR assumption, in terms of Figure 2, is that α_1 and α_3 are zero while α_2 is equal to 1. Other possibilities are also plausible. For example if α_1 is non-zero and equal to α_3 , while α_2 is equal to zero, then there is a linear relationship between the indicator measured in percentiles and the probability of a crisis. That is, to continue the example, increases in the degree of overvaluation increase the risk no matter how overvalued the exchange rate already is.

Figure 2. Relationship Between Predictive Variable and Probability of Crisis



²⁶We ignore the potential serial correlation in the errors that may be introduced by the fact that the left-hand-side variable (which takes a 1 if there is a crisis sometime in the next 24 months) is serially correlated.

We propose to let the data resolve the question of whether a step-function is in fact a reasonable description of the relationship between indicator variables and the probability of a crisis. To this end, we run bivariate probit regressions on the pooled panel in which the dependent variable is the KLR variable that takes a value of 1 if there is a crisis in the subsequent 24 months and zero otherwise. For each indicator we estimate equations of the form:

$$\text{prob}(c24=1)=f(\alpha_0+\alpha_1p(x)+\alpha_2I+\alpha_3I(p(x)-T)) \quad (1)$$

where $c24 = 1$ if there is a crisis in the next 24 months, $p(x)$ = the percentile of the variable x , and $I = 1$ if the percentile is above some threshold T and zero otherwise.²⁷ Thus, α_1 , α_2 , and α_3 in equation 1 correspond to the α 's in Figure 2. We use the thresholds T calculated from the KLR algorithm, since we are interested primarily in testing their approach against a more general alternative.²⁸

Table 4 presents estimates of equation 1 for three important predictive variables: deviations of the real exchange rate from trend, the current account deficit as a share of GDP, and the growth rate of the ratio of M2/reserves. Consider first the real exchange rate. Column 1 of Table 4 indicates that α_1 , α_2 and α_3 are all significant. The data cannot reject that the relationship between RER deviations and the probability of crisis is of the general form shown in Figure 2, linear with a jump at the threshold and a higher slope thereafter. The first panel of Figure 3 gives a richer view of the relationship between overvaluation and the probability of crisis. The choppy line in this figure presents the fraction of times the observation of a given percentile for RER deviations is followed within 24 months by a crisis in the pooled data. The other line represents the estimated relationship shown in the first column of Table 4 and discussed above. The message of this figure is that while the jump at the threshold is significant, it does not capture an important part of the variation in the probability of crisis as a function of RER deviations.

²⁷ The probit models are estimated over the 1970:1–1995:4 period.

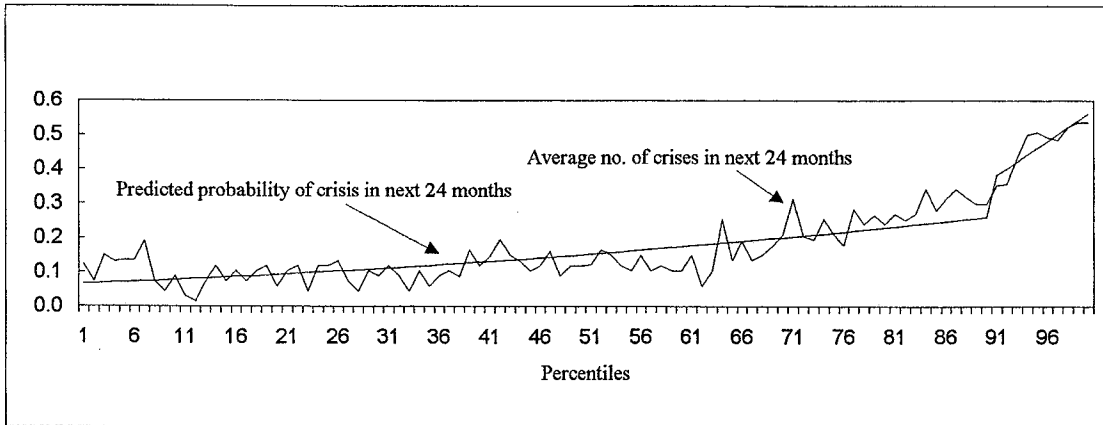
²⁸ This procedure is biased in favor of finding significant jump coefficients. Since we use the data itself to identify the biggest jump (through the KLR method), the subsequent tests will tend to find that the jumps we have found are unusually large. The tests we perform thus overestimate the statistical significance of the jump coefficient α_2 . We have also achieved some success with general functional forms that do not require prior knowledge of the break point.

Table 4. Testing Indicators Against More General Piecewise-Linear Specifications in Bivariate Probit Models

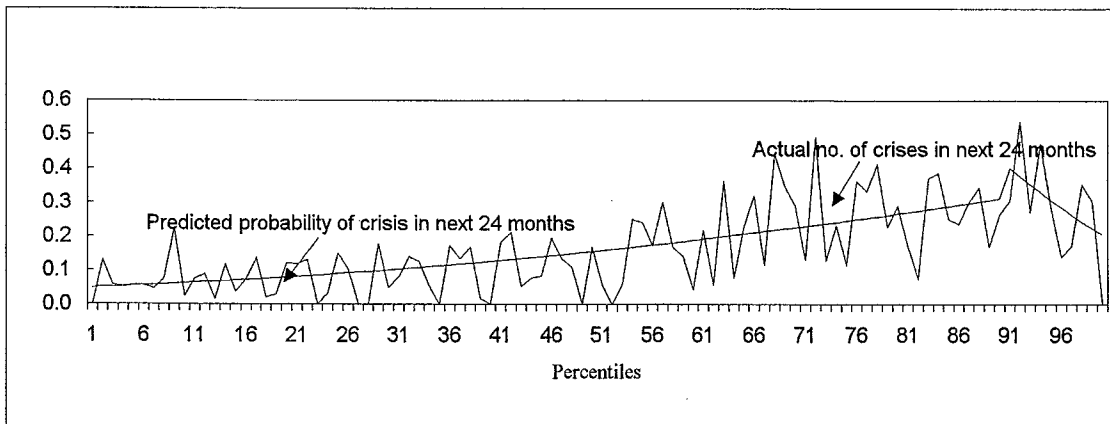
Variable	Real Exchange Rate		Current Account		M2/Reserve Growth	
	(1) Unrestricted		(2) Unrestricted		(3) Unrestricted	
	Coefficient	T-statistic	Coefficient	T-statistic	Coefficient	T-statistic
Percentile (α_1)	0.0024	12.55	0.0031	14.91	0.0017	7.96
Indicator (α_2)	0.0818	2.66	0.0824	2.31	0.0999	3.33
Indicator *(percentile - Threshold)(α_3)	0.0101	2.41	-0.0182	-3.27	-0.0010	-0.31
Sample size	6824		5424		6072	
Log Likelihood	-2959		-2298		-2677	
Pseudo R ²	0.08		0.06		0.04	
					0.0017	7.96
					0.0916	5.15
					-2677	
					0.04	
					6072	
					-2677	
					0.04	
					0.0017	7.96
					0.0916	5.15
					-2677	
					0.04	
					6072	
					-2677	
					0.04	

Figure 3. Average No. of Crises in Next 24 Months by Percentile of Variable

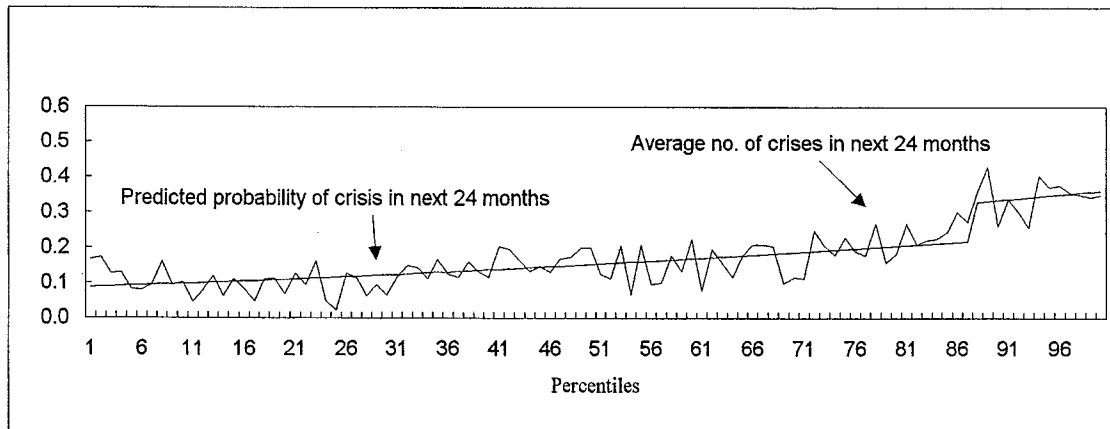
Real Exchange Rate Deviations



Current Account as Percent of GDP



M2/Reserves Growth Rate



Turning to the current account, we again find all three coefficients statistically significant. The second panel of Figure 3 shows that the jump, while statistically significant, appears not to be economically important compared to the strong linear effect below the threshold.²⁹

For the M2/reserves growth variable, we cannot reject that $\alpha_3=0$, as shown in column 3. The data reject the further restriction of $\alpha_2=0$, which would result in a model that is linear in percentiles. The simplification supported by the data is a linear model with a jump at the threshold, as shown in the third panel of Figure 3.

While the outcome of this analysis varies somewhat across indicators, the general lesson is that although the jump in probability of crisis at the threshold is often statistically significant, the underlying percentile variable is usually also important in explaining the variation in crisis probability.

Multivariate probits are the natural extension to the bivariate probits discussed so far. First, they easily accommodate broader functional forms, and we have seen that the bivariate probits cast some doubt on the zero/one indicator approach of KLR. A further advantage is that the estimation of a multivariate version of equation (1) is a natural way to combine the information from the various indicator variables into a single estimate of the probability of crisis. The composite indicators proposed by Kaminsky (1998), based on a weighted-sum of indicators, ignore possible correlations among different indicators, unlike the multivariate probits. Finally, the probits allow the calculation of standard errors and other measures of statistical significance.

Table 5 presents estimates of three probit models that explain whether a crisis occurs in the next 24 months (hereafter designated BP models).³⁰ Model 1 uses the indicator form of the variables, where the indicator equals 1 above the threshold and zero otherwise. In model 2 the variables enter linearly, expressed as percentiles of the country-specific distribution of observations.³¹ Model 3 is the result of a simplification starting with the most general piecewise-linear specification for all the variables. From a starting point that allowed the estimation, for each variable, of the slope below the threshold, the jump at the threshold, and the slope above the

²⁹We have no explanation for the decline in probability of crisis for values of the current account that exceed the threshold.

³⁰We omit the real interest rate, terms-of-trade growth, industrial production growth, stock price growth and real interest differential variables from the probit models because the significantly smaller number of observations available would greatly change the sample.

³¹For models 1 and 2, we simplify the general regression by first eliminating variables with negative coefficients, and then retaining all variables significant at the 10 percent level.

Table 5. Multivariate Probit Models

Variable 1/	Model 1		Model 2		Model 3	
	Indicator	T-statistic	Linear	T-statistic	Piecewise Linear	T-statistic
	Coefficient		Coefficient		Coefficient	
Real exchange rate deviation	0.31564	15.89	0.00232	13.50	0.12057	3.35
Current Account	0.10461	5.63	0.00178	9.50	0.00114	5.58
Reserve Growth			0.00128	6.20	0.01078	2.42
Export Growth	0.05686	3.24	0.00064	3.65	0.06328	1.77
M2/Reserves	0.08561	4.37	0.00053	2.80	0.00189	8.67
M2/Reserves Growth	0.12770	7.37			-0.01313	-2.38
					0.00107	4.74
					0.00070	3.94
					0.06372	3.47
					0.03960	2.33
Sample Size	4928		5025		4928	
Log-Likelihood	-1958		-1970		-1858	
Pseudo-R ²	0.099		0.115		0.145	

threshold, we used a general-to-specific procedure to simplify to the most parsimonious representation of the data.³²

Model 1 of Table 5 shows that the probability of crisis is increased when the following variables exceed their thresholds: real exchange rate deviations, the current account, reserve growth, export growth, and both the level and growth rate of M2/Reserves.³³ These variables also increase the probability of crisis when entered linearly in model 2, except for the growth rate of M2/reserves, while reserve growth itself is now significant. In the simplified piecewise-linear model 3, two variables (real exchange rate deviations and current account) enter with a significant slope below the threshold, a jump at the threshold, and a steeper slope above the threshold; two variables (reserve and export growth) enter linearly; and for two variables (M2/reserves and M2/reserves growth) only the jump at the threshold is significant.

How well do the different models perform? The results in Tables 3 (on page 16) and 5 allow us to draw two main conclusions. First, the probits tend to slightly outperform the KLR-based probabilities. The most direct comparison involves the indicator probit which uses as predictive variables the zero/one signals from the KLR indicators; here the only difference with KLR is the use of the probits to derive probabilities of crisis from the individual indicators. This model outperforms the KLR-based probabilities in terms of scores and goodness-of-fit. Second, the ranking among the various probit models is ambiguous. The piecewise-linear has the best pseudo- R^2 and lowest scores, as is not surprising given that it is a generalization of the other two models (none of these measures give any weight to parsimony). It does not outperform in goodness-of-fit, however. The indicator probit and the linear probit perform similarly: the linear model has better scores but generally worse goodness-of-fit.³⁴

3. Summary In-Sample Assessment

Given the non-statistical nature of most of the KLR analysis, it is somewhat difficult to evaluate the success of this approach. KLR conclude that “the signals approach can be useful as the basis for an early warning system of currency crises” (KLR, page 23). Their grounds are largely that most of the indicators have low noise-to-signal ratios, most indicators signal ahead of most crises, and most crises are preceded by multiple signals. We find similar though somewhat weaker results in our larger sample. Our analysis of the in-sample success of the

³²We do not investigate the undoubted path dependency of this procedure. We simplify the general regression by first sorting the variables in ascending order of the significance (measured by an F-test of the significance of all three terms for each predictive variable), then attempting for each variable to set first α_3 , then α_1 , then α_2 equal to zero.

³³Note that here as elsewhere variables such as reserve growth, export growth and real exchange rate deviations from trend have been multiplied by -1 and thresholds defined accordingly, so that an increase in an variable should increase the probability of a crisis.

³⁴A Davidson and MacKinnon encompassing test of the non-nested linear and indicator probits shows that neither encompasses the other.

KLR-type models suggests the approach can indeed be useful and the model does significantly better than guesses based on the unconditional probability of crisis. Nonetheless, most crises are still missed and most alarms are false. In evaluating the KLR indicator approach against our modifications, we find that the probit models generally perform slightly better. The in-sample performance of the linear, indicator and piecewise-linear models is broadly similar.

As to the assessment of which variables are potentially important leading indicators, although we find fewer potentially useful indicators, ours are also classified as useful indicators by KLR (except for those we have added). These are deviations of the real exchange rate from trend, growth of exports, change in international reserves, "excess" M1 balances, growth in domestic credit as a share of GDP, the real interest rate, terms of trade growth, the level and growth of M2/reserves, and the current account.

C. Predicting 1997

1. Original KLR Model

The KLR approach has generated a variety of different ways to forecast 1997 outcomes. First, we can see which indicators were signaling prior to the 1997 crises. We have already calculated the optimal thresholds and resulting noise-to-signal ratios for the different indicators. To forecast for the post-April 1995 period, we apply these thresholds to the values of the predictive variables after this date, determining whether they are issuing signals or not.³⁵ We have examined the performance of each individual indicator in 1996 for each of the eight Asian and Latin American countries discussed above.³⁶ To summarize this large amount of information, no particular indicators flashed in all of the crisis countries. The only indicators to signal in more than one country were the growth rate of exports, which flashed in both Thailand and Korea, the growth of M2/reserves, which signaled in both Thailand and Malaysia, and reserve growth, which flashed in Korea, Malaysia and Thailand.

More interesting for purposes of forecasting crisis than looking at each individual indicator is combining the information from the different variables into a summary measure of crisis probabilities. The first column of Table 6 shows the performance of the Kaminsky (1998) composite measures of the probability of crisis based on the weighted-sum of indicators signaling. A natural question is whether the estimated probability of crisis is above 50 percent prior to actual crises. The goodness-of-fit rows show that only 4 percent of the time was the predicted probability of crisis above 50 percent in cases when there was a crisis within the next 24 months, during the 1995:5 to 1997:12 period. As before, we may be interested in using a lower cut-off probability to define a crisis. Table 6 shows that the Kaminsky (1998) probability estimates are above 25 percent in 25 percent of the pre-crisis

³⁵Note that an observer in April 1997, for example, would have been able to observe the signals emitted in prior dates but would not yet know whether these signals were good or false, as he would not yet have observed whether there was a crisis in the subsequent 24 months.

³⁶Tables are available upon request.

Table 6. Comparing Predictive Power of Alternative Composite Indicators--Out-of-Sample

	KLR-Based Weighted-Sum Probabilities		Alternative Probit Models		
	Original Specification	Augmented with Current Account and level of M2/reserves	Indicator	Linear	Piecewise- linear
Accuracy and Calibration Scores:					
Quadratic Probability Score	0.402	0.398	0.325	0.281	0.299
Log Probability Score	0.606	0.596	0.501	0.433	0.452
Global Squared Bias	0.01774	0.01946	0.02987	0.00581	0.01256
Goodness-of-Fit: (cut-off probability of 50 percent)					
Percent of observations correctly called	74	73	78	78	78
Percent of pre-crisis periods correctly called 1/	4	0	2	0	5
Percent of tranquil periods correctly called 2/	100	100	99	100	98
False alarms as percent of total alarms 3/	17	No crisis called	50	No crisis called	56
Goodness-of-Fit: (cut-off probability of 25 percent)					
Percent of observations correctly called	69	69	76	79	76
Percent of pre-crisis periods correctly called 1/	25	32	16	80	48
Percent of tranquil periods correctly called 2/	85	83	93	79	84
False alarms as percent of total alarms 3/	63	60	61	49	54

1/ A pre-crisis period is correctly called when the estimated probability of crisis is above the cut-off probability and a crisis ensues within 24 months.

2/ A tranquil period is correctly called when the estimated probability of crisis is below the cut-off probability and no crisis ensues within 24 months.

3/ A false alarm is an observation with an estimated probability of crisis above the cut-off (an alarm) not followed by a crisis within 24 months.

observations. As we observed in-sample, most alarms are false at the 25 percent cut-off. The addition of the current account and level of M2/reserves variables improves out-of-sample performance slightly, as shown in the second column. In particular, 32 percent of the pre-crisis observations are called correctly.

This may sound like poor performance. It is worth noting, though, that these forecasts are significantly better than random guesses, both economically and statistically. The forecasts from the augmented KLR model in column (2), for example, suggest that the probability of a crisis within 24 months conditional on an alarm (using the 25 percent cut-off) is 40 percent, which is somewhat higher than the unconditional probability of 27 percent. And a Pesaran-Timmermann test rejects the hypothesis that the forecasts are no better than guesses based on the unconditional probability of crisis in the sample at the 1 percent level of significance.

So far we have examined the ability of the models to predict the approximate *timing* of crises for each country.³⁷ We can also evaluate the cross-sectional success of the models' predictions in identifying which countries are vulnerable in a period of global financial turmoil such as 1997. The question here is whether the models assign higher predicted probabilities of crisis to those countries that had the biggest crises. Forecasting performance can be evaluated in this manner by comparing rankings of countries based on the predicted and actual crisis indices. Table 7 shows countries' actual crisis index and predicted probability of crisis in 1997 for the various different forecasting methods.³⁸ The table also shows the Spearman correlation between the actual and predicted rankings and its associated p-value, as well as the R^2 from a bivariate regression of the actual rankings on the predictions.³⁹

The KLR-based forecasts are clearly somewhat successful at ranking countries by severity of crisis. The actual rankings of countries in 1997 by their crisis index are significantly correlated with forecasts from the weighted-sum of indicators-based probabilities. With the original KLR variables, 28 percent of the variance is explained. The addition of the current account and the level of M2/reserves brings the R^2 up to 36 percent.

³⁷We say approximate because the models only attempt to place the crisis within a 24 month window.

³⁸The predicted crisis probability is the average of the probabilities during 1996:1–12, using the out-of-sample estimates. Averaging over for example 1996:1 to 1996:6 gives somewhat different results. The actual crisis index used to rank the countries for 1997 is the maximum value of the monthly crisis index for each country during 1997.

³⁹The Spearman correlation, like the commonly used Pearson correlation coefficient, varies from -1 to 1. It is more appropriate for measuring correlation in rankings. One important feature is that it is less sensitive to extreme values. The p-value is the probability of observing a correlation of that absolute value or higher under the null hypotheses that the two rankings are uncorrelated.

Table 7. Correlation of Actual and Predicted Rankings based on KLR Approach

Country	Actual		KLR Weighted-Sum of Indicators 2/		BP Probit Models 4/							
	Crisis Index 1/	Rank	Original Specification Probability	Rank	Indicator Probability	Rank	Linear Probability	Rank	Piecewise Linear Probability	Rank		
Thailand	10.19	1	12.42	16	20.31	7	0.20	4	0.38	2	0.25	6
Korea	9.52	2	25.27	4	22.76	5	0.28	3	0.26	9	0.31	4
Indonesia	4.48	3	11.18	18	15.77	11	0.15	8	0.26	8	0.24	7
Malaysia	4.42	4	17.27	8	14.58	13	0.20	5	0.38	1	0.32	3
Zimbabwe	4.40	5	32.25	3	25.95	3						
Taiwan Province of China	3.37	6	22.69	5	23.92	4	0.19	6	0.30	5	0.19	12
Colombia	3.01	7	16.91	9	15.46	12	0.17	7	0.36	4	0.35	2
Philippines	2.68	8	40.58	1	34.52	1	0.29	2	0.22	12	0.21	10
Brazil	0.82	9	36.67	2	32.08	2	0.38	1	0.25	10	0.46	1
Turkey	0.65	10	17.37	7	15.91	10	0.10	13	0.14	18	0.16	13
Venezuela	0.62	11	14.49	14	13.89	16	0.10	13	0.09	20	0.08	20
Pakistan	0.57	12	15.49	10	16.56	9	0.14	11	0.28	6	0.24	9
South Africa	0.52	13	21.78	6	18.98	8	0.12	12	0.23	11	0.21	11
Jordan	0.45	14	14.08	15	13.83	18	0.10	13	0.21	13	0.15	15
India	0.39	15	10.77	20	10.12	21	0.10	13	0.14	19	0.15	14
Sri Lanka	0.36	16	12.01	17	11.51	19	0.10	13	0.19	14	0.14	16
Chile	0.24	17	11.18	18	10.58	20	0.10	13	0.18	15	0.12	17
Bolivia	0.18	18	10.77	20	10.12	21	0.10	13	0.07	21	0.05	22
Argentina	0.15	19	14.82	12	13.84	17	0.10	13	0.15	17	0.10	19
Mexico	0.15	20	14.49	13	14.30	14	0.10	13	0.06	22	0.05	21
Peru	0.12	21	14.90	11	20.41	6	0.14	10	0.27	7	0.24	8
Uruguay	-0.02	22	10.77	20	10.12	21	0.10	13	0.18	16	0.11	18
Israel	-0.11	23	10.77	20	14.27	15	0.15	9	0.37	3	0.29	5
Correlation 5/			0.543		0.600		0.666		0.474		0.566	
P-value			0.007		0.003		0.001		0.026		0.006	
R ²			0.284		0.359		0.475		0.233		0.327	

1/ The KLR crisis index (a weighted average of percentage changes in the exchange rate and reserves) is standardized by subtracting the mean and dividing by the standard deviation. Values above three are defined as a crisis and are shown in bold.

2/ Based on average of noise-to-signal weighted probabilities from during 1996:1-12, using out-of-sample estimates.

3/ Augmented with the inclusion of the current account and M2/reserves in levels.

4/ All probit models probabilities are average predicted probabilities for 1996:1-12, where model was estimated up to 1995:4.

5/ Spearman Rank Correlation of the fitted values and the actual crisis index and its p-value. The R² is from a regression of fitted values on actual values.

To get a richer sense of how useful this general approach would have been, we now examine more closely the predictions of the KLR-based model for four Asian crisis countries (where crisis is identified according to the KLR definition): Korea, Indonesia, Malaysia, and Thailand, and one Asian and three Latin American non-crisis countries: Philippines, Argentina, Brazil and Mexico. Figure 4a presents the KLR composite measure of estimated probability of crisis, with vertical lines at crisis dates.⁴⁰

The weighted-sum based probability measure does not paint a clear picture of substantial risks in crisis compared to non-crisis countries. Two non-crisis countries, Brazil and the Philippines, consistently present risks of crisis above 30 percent during 1996. One crisis country, Korea, also presents risks above 30 percent, while Malaysia is generally above 20 percent. Estimated crisis risks remain below 17 percent in 1996 for both the crisis and non-crisis countries Argentina, Mexico, Indonesia and Thailand.

In sum, the KLR approach shows some promise. In particular, the fitted probabilities from the weighted-sum of indicators are significant predictors of crisis probability in 1997. This suggests the model may be useful in identifying which countries are vulnerable in a period following a global financial shock. Still, the overall explanatory power is fairly low, as demonstrated by the low R^2 statistic in the regression of the actual on the predicted crisis rankings. Both the overall goodness-of-fit for the out-of-sample predictions and the analysis of the eight cases illustrate the low predictive power of the weighted-sum based probabilities in predicting the timing of crisis. We have already seen that within sample, our probit-based alternatives to the KLR model perform slightly better. We now turn to an examination of the out-of-sample performance of the BP probit model.

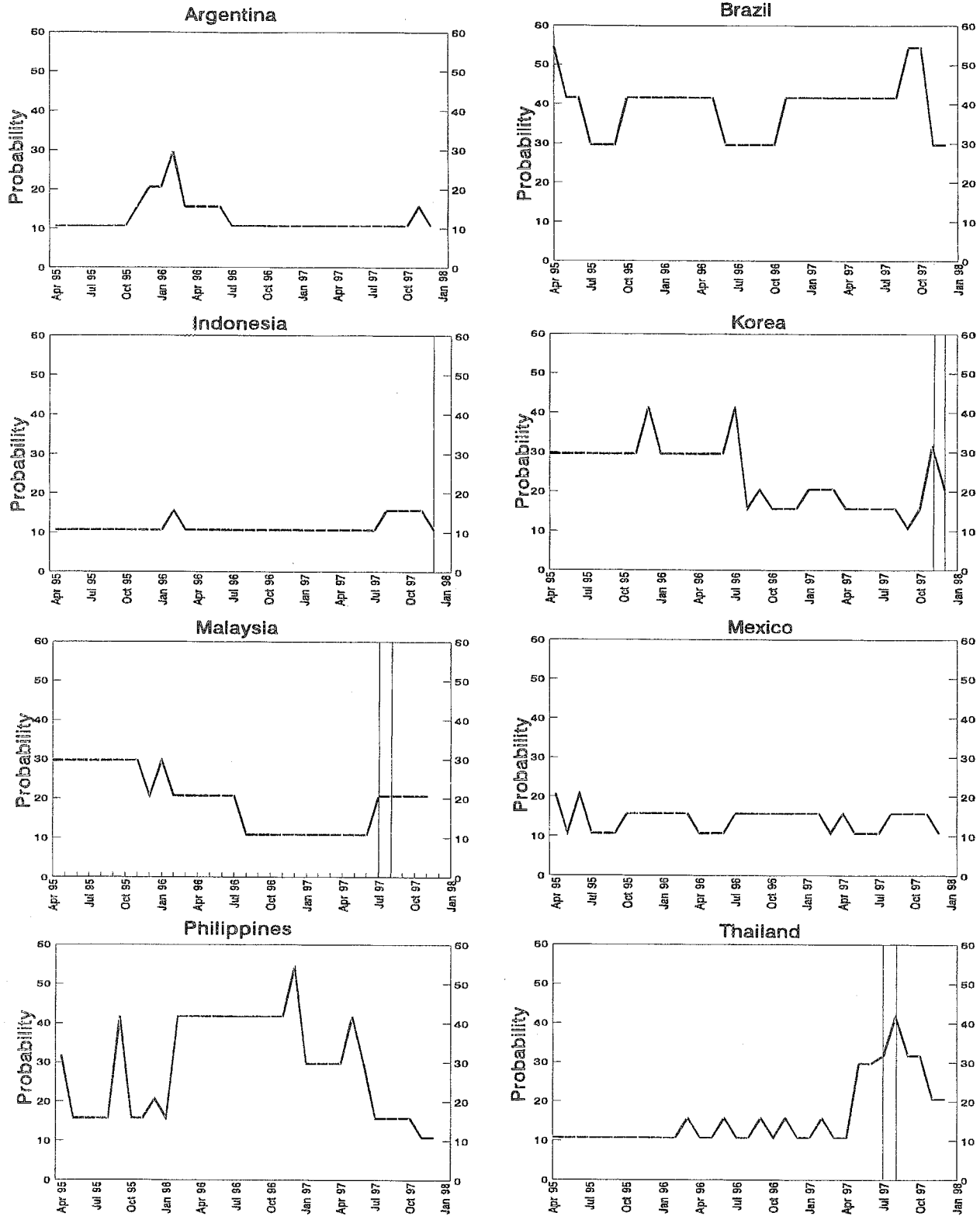
2. BP Probit-Based Alternative

To test the various probit models out-of-sample, we use data through 1995:4 to estimate the regression coefficients, as in Table 5, then extend the explanatory variables to generate predictions for the period 1995:5–1997:12.⁴¹ The estimated probabilities can be evaluated using the probability scores and goodness-of-fit measures discussed above.

⁴⁰Figure 4a is the original KLR specification, without the current account and M2/reserves in levels. Table 8 presents the same data, along with estimates based on the KLR model augmented with additional variables as well as the alternative probit models, to be discussed below.

⁴¹The weighted-sum based probabilities are based on the nine indicators with noise-to-signal ratio less than 1 in the 1970–1995:4 sample. The probit-based probabilities are derived from the models in Table 4, which represent simplifications to the most parsimonious representation of the data.

Figure 4a: KLR Weighted-Sum Crisis Probabilities for Selected Countries 1/



1/ Vertical line represents crisis dates.

Table 6 shows that on all the scoring measures, the probits perform better than the probabilities based on the weighted-sum of indicators signaling.⁴² The linear model has the best scores, though the piecewise-linear model is close behind. None of the models correctly calls many crises observations at the 50 percent cutoff, where a correct call is an observation that results in an estimated probability of crisis higher than the cutoff with an actual crisis ensuing within 24 months. Using the looser standard whereby a probability of crisis above 25 percent is considered an alarm, the linear and piecewise-linear probits perform well, much better than the weighted-sum based probabilities. The linear probit generates a probability of crisis above 25 percent in 80 percent of the periods that precede a crisis.⁴³ Reflecting their greater prediction success, the probit models have a lower share of false alarms (crisis calls not followed by a crisis as a share of total crisis calls), as low as 49 percent for the linear model. Putting it slightly differently, for this model the probability of crisis within 24 months conditional on an alarm (using the 25 percent cutoff) is 51 percent, much higher than the unconditional probability of 22 percent.⁴⁴

The linear model performs much better out-of-sample than the more general piecewise-linear model that include a role for discrete jumps in the risk of crisis at the KLR thresholds. This suggests that the threshold and indicator concept add little to the explanatory power of the simple linear model in predicting crisis timing, at least for 1997. The worse out-of-sample performance of the indicator and piecewise-linear models (and similar or better in-sample performance) is consistent with the greater risk of data-mining in the indicator and piecewise-linear approaches.

As with the KLR models, we can also evaluate the performance of the probit models in predicting the cross-country incidence of crisis in 1997. Table 7 shows that country rankings based on all the probit forecasts are significantly correlated with actual crisis rankings in 1997. Forecasts based on the indicator probit rank countries more accurately than the weighted-sum of indicators-based forecasts, with an R^2 close to one half. This superior performance is consistent with previous results that the KLR weighted-sum-of-indicators forecasts are outperformed by the analogous probit model. Somewhat anomalously, the other two probit models perform worse than the indicator probit. In particular, the ranking based on the linear

⁴²An exception is that the indicator probit has a higher GSB than the KLR-based probabilities. As described in footnote 24 on page 16, the scores measure the total size of the errors, similar to the mean squared error in ordinary least squares. Lower scores are better.

⁴³Of course the accuracy of correctly calling tranquil periods falls, from 100 percent to 79 percent.

⁴⁴These predictions are also statistically significantly better than guesses based on the unconditional probability at the 1 percent level.

model that had the best goodness-of-fit has the lowest, though still significant, correlation with the actual ranking.⁴⁵

We can flesh out these results by examining the performance of the linear probit in predicting crisis for our sub-sample of four crisis and four non-crisis countries in 1997 (Table 8a, 8b and Figure 4b).⁴⁶ The linear probit present a fairly clear picture of the prospects of crisis for most of these countries. Consider first the crisis countries. In Thailand estimated probabilities of crisis were above 40 percent for several months in 1996, and in Malaysia the probabilities were above 30 percent. The probabilities are also reasonably high for Indonesia, ranging from 25 to 28 percent, while the model is somewhat less successful for Korea, where the estimated probability of crisis was between 20 and 33 percent. Turning to the non-crisis countries, in the Philippines probabilities ranged from 20 to 23 percent. None of the Latin American countries yielded crisis probabilities above 30 percent in 1996, and only Brazil was above 20 percent for any length of time.

We have examined model performance in predicting, out-of-sample, crisis timing and cross-sectional severity of crisis during 1997. Several conclusions emerge. First, all the models examined perform significantly better than chance would imply, both at predicting whether or not a crisis will occur as measured by goodness-of-fit and at predicting the cross-country severity of crisis. Second, we can compare the BP probit-based alternatives to the KLR probabilities based on the weighted-sum of indicators signaling. The KLR forecasts perform better than some of the probits on a few of the measures, so this comparison is not unambiguous. Overall, though, the probits seem to work better. Moreover, in contrasting the BP probit methodology with the KLR probabilities, the most direct comparison involves the indicator probit, as it also uses indicator predictive variables. Here in particular the probit generally outperforms. Third, among the probits, the linear specification performs best in terms of the probability scores, goodness-of-fit and the eight cases examined more closely.

⁴⁵The contrast between the results of the rankings and goodness-of-fit comparisons is somewhat surprising but not inexplicable. The two measures are somewhat different and they need not correspond. The goodness-of-fit measure examines only whether crisis calls are correct or not and ignores the size of errors. The rankings comparison considers whether the highest probabilities of crisis are associated with the *largest* crises; the magnitude of the crisis, however, as distinct from whether or not there is a crisis, is not a factor in any of the models. These results are sensitive to the exact sample of countries involved in the ranking comparison. For example, eliminating Israel (one of the largest outliers) from the sample increases the R^2 of the rankings predictions of the percentile probit model from 23 to 42 percent.

⁴⁶ It is not always possible to calculate the probit probabilities for the entire out-of-sample period because data on the current account and GDP were not available as of the date at which this data was collected (April 1998). We discuss the problem of data lags in the conclusion.

Table 8a. Summary Measures for Selected Countries: Asian Countries

Date	Indonesia				Korea				Malaysia				Philippines				Thailand			
	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Prob. Based Prob.	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Prob. Based Prob.	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Prob. Based Prob.	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Prob. Based Prob.	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Prob. Based Prob.
1995:01	1(9)	11	16	14	2(10)	42	36	16	2(9)	30	26	28	0(9)	11	10	24	0(10)	11	10	15
1995:02	1(9)	11	16	15	1(10)	30	26	16	2(9)	30	26	28	0(9)	11	10	19	0(10)	11	10	17
1995:03	1(9)	11	16	15	1(10)	30	26	18	2(9)	30	26	31	0(9)	16	16	22	0(10)	11	10	20
1995:04	1(9)	11	16	17	1(10)	30	26	18	2(9)	30	26	34	2(9)	32	28	23	0(10)	11	10	23
1995:05	1(9)	11	16	16	1(10)	30	26	19	2(9)	30	26	36	1(9)	16	16	23	0(10)	11	10	21
1995:06	1(9)	11	16	17	1(10)	30	26	19	2(9)	30	26	37	1(9)	16	16	26	0(10)	11	10	22
1995:07	0(9)	11	10	23	1(10)	30	26	17	2(9)	30	26	36	1(9)	16	16	22	0(10)	11	10	26
1995:08	1(9)	11	16	24	1(10)	30	26	18	2(9)	30	26	35	1(9)	16	16	24	0(10)	11	10	24
1995:09	1(9)	11	16	23	1(10)	30	26	18	2(9)	30	26	35	2(9)	42	36	26	0(10)	11	10	27
1995:10	1(9)	11	16	27	1(10)	30	26	19	2(9)	30	26	36	1(9)	16	16	26	0(10)	11	10	30
1995:11	1(9)	11	16	28	1(10)	30	26	19	2(9)	30	26	35	1(9)	16	16	25	0(10)	11	10	28
1995:12	1(9)	11	16	26	2(10)	42	36	23	1(9)	21	16	34	2(9)	21	16	23	0(10)	11	10	30
1996:01	1(10)	11	16	28	1(10)	30	26	19	2(9)	30	26	38	1(9)	16	16	23	0(10)	11	10	30
1996:02	1(10)	16	16	27	1(10)	30	26	20	1(9)	21	16	40	2(9)	42	36	23	1(10)	11	16	30
1996:03	1(10)	11	16	26	1(10)	30	26	22	1(9)	21	16	33	2(9)	42	36	20	2(10)	16	28	32
1996:04	1(10)	11	16	24	1(10)	30	26	23	1(9)	21	16	41	2(8)	42	36	NA	1(10)	11	16	34
1996:05	1(10)	11	16	25	1(10)	30	26	22	1(8)	21	16	NA	2(8)	42	36	NA	1(10)	11	16	36
1996:06	1(10)	11	16	26	1(10)	30	26	24	1(8)	21	16	NA	2(8)	42	36	NA	2(9)	16	28	40
1996:07	1(9)	11	16	NA	2(10)	42	36	28	1(8)	21	16	NA	2(8)	42	36	NA	1(9)	11	16	40
1996:08	1(9)	11	16	NA	1(10)	16	16	29	0(8)	11	10	NA	2(8)	42	36	NA	1(9)	11	16	41
1996:09	1(9)	11	16	NA	2(10)	21	16	33	0(8)	11	10	NA	2(8)	42	36	NA	2(9)	16	28	42
1996:10	1(9)	11	16	NA	1(10)	16	16	30	0(8)	11	10	NA	2(8)	42	36	NA	1(9)	11	16	41
1996:11	1(9)	11	16	NA	1(10)	16	16	33	0(8)	11	10	NA	2(8)	42	36	NA	2(9)	16	28	42
1996:12	0(9)	11	10	NA	1(10)	16	16	27	0(4)	11	10	NA	3(8)	55	36	NA	1(9)	11	16	43
1997:01	0(9)	11	10	NA	2(10)	21	16	29	0(4)	11	10	NA	1(8)	30	26	NA	1(9)	11	16	44
1997:02	0(9)	11	10	NA	2(10)	21	16	33	0(3)	11	10	NA	1(8)	30	26	NA	2(9)	16	28	44
1997:03	0(9)	11	10	NA	2(10)	21	16	28	0(3)	11	10	NA	1(8)	30	26	NA	1(9)	11	16	43
1997:04	0(9)	11	10	NA	1(10)	16	16	27	0(3)	11	10	NA	1(8)	30	26	NA	0(8)	11	10	NA
1997:05	0(9)	11	10	NA	1(10)	16	16	27	0(3)	11	10	NA	2(8)	42	36	NA	2(8)	30	26	NA
1997:06	0(9)	11	10	NA	1(9)	16	16	NA	0(3)	11	10	NA	1(8)	30	26	NA	2(7)	30	26	NA
1997:07	0(9)	11	10	NA	1(9)	16	16	NA	1(3)	21	16	NA	1(8)	16	16	NA	2(7)	32	28	NA
1997:08	1(9)	16	16	NA	1(9)	16	16	NA	1(3)	21	16	NA	1(8)	16	16	NA	3(7)	42	36	NA
1997:09	1(7)	16	16	NA	0(9)	11	10	NA	1(3)	21	16	NA	1(8)	16	16	NA	2(6)	32	28	NA
1997:10	1(7)	16	16	NA	1(9)	16	16	NA	1(3)	21	16	NA	1(8)	16	16	NA	2(6)	32	28	NA
1997:11	1(7)	16	16	NA	2(9)	32	28	NA	1(2)	21	16	NA	0(3)	11	10	NA	1(1)	21	16	NA
1997:12	0(2)	11	10	NA	1(4)	21	16	NA	0(0)	NA	NA	NA	0(1)	11	10	NA	1(1)	21	16	NA

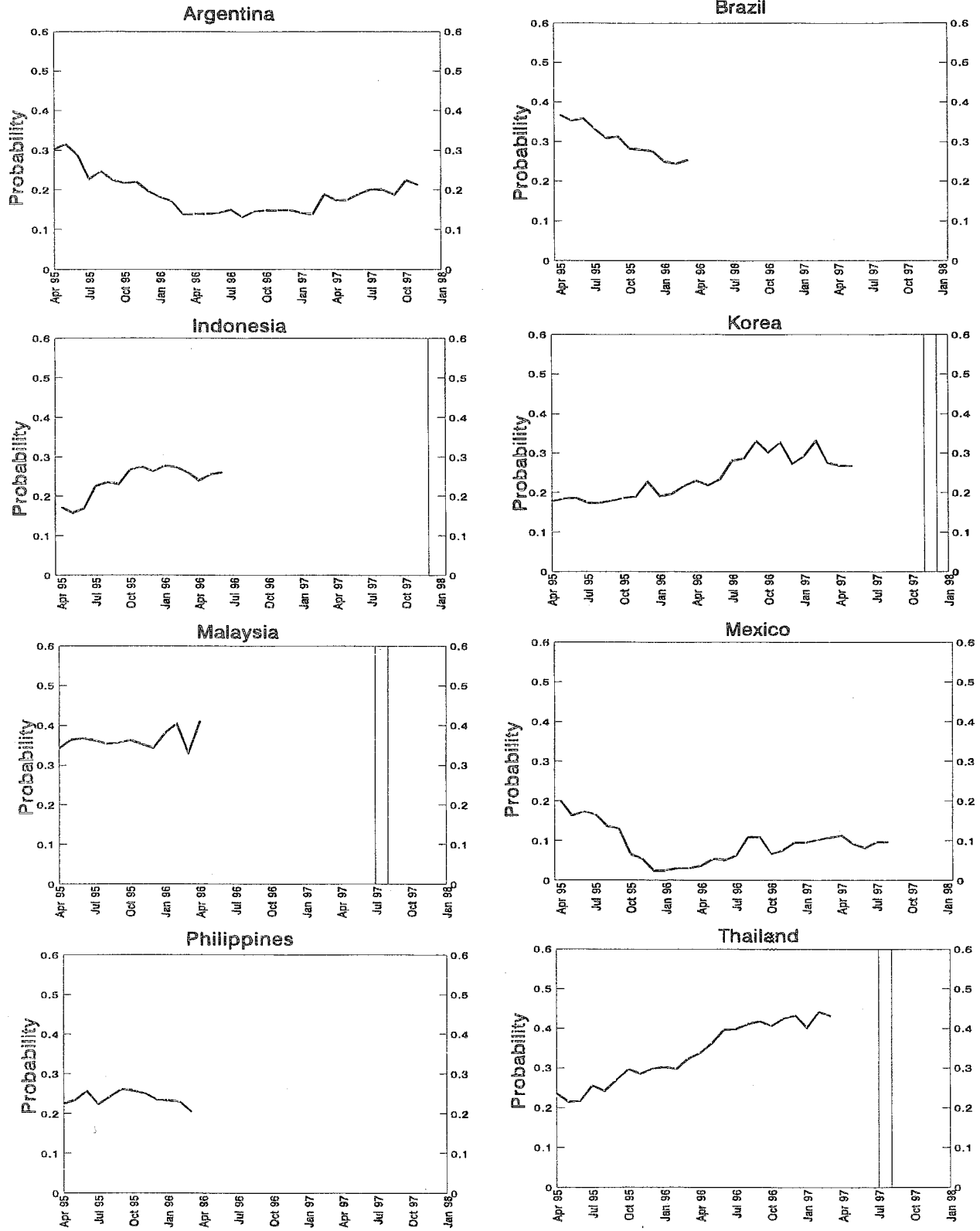
1/ Number of good indicators (with noise-to signal ratio less than unity) that are signaling, with the number for which data are available in parenthesis. There are ten good indicators.
 2/ Predicted probabilities based on weighted sum of the good indicators, where each indicator is weighted by the inverse of its adjusted noise-to-signal ratio, with original KLR variables.
 3/ Predicted probabilities based on weighted sum of the good indicators, where each indicator is weighted by the inverse of its adjusted noise-to-signal ratio, with original KLR variables, augmented with the inclusion of the current account and M2/reserves in levels.
 4/ Predicted probabilities of crisis from a probit regression of impending crisis on the indicator variables measured linearly in percentiles.

Table 8b. Summary Measures for Selected Countries: Latin American Countries

Date	Argentina				Brazil				Mexico			
	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Probit Based Prob.	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Probit Based Prob.	No. of Good Indicat. Signaling	Weighted-Sum Prob. Original Spec.	Extra vars	Probit Based Prob.
1995:01	0(10)	11	10	33	3(10)	55	36	26	3(10)	55	36	26
1995:02	0(10)	11	10	33	3(10)	55	36	27	3(10)	55	36	27
1995:03	0(10)	11	10	34	3(10)	55	36	28	3(10)	55	36	28
1995:04	0(10)	11	10	30	3(10)	55	36	37	3(10)	55	36	37
1995:05	0(10)	11	10	32	2(10)	42	36	35	2(10)	42	36	35
1995:06	0(10)	11	10	29	2(10)	42	36	36	2(10)	42	36	36
1995:07	0(10)	11	10	23	1(10)	30	26	33	1(10)	30	26	33
1995:08	0(10)	11	10	25	1(10)	30	26	31	1(10)	30	26	31
1995:09	0(10)	11	10	22	1(10)	30	26	31	1(10)	30	26	31
1995:10	0(10)	11	10	22	2(10)	42	36	28	2(10)	42	36	28
1995:11	0(10)	16	16	22	2(10)	42	36	28	2(10)	42	36	28
1995:12	0(10)	21	16	20	2(10)	42	36	28	2(10)	42	36	28
1996:01	2(10)	21	16	18	2(10)	42	36	25	2(10)	42	36	25
1996:02	3(10)	30	26	17	2(10)	42	36	25	2(10)	42	36	25
1996:03	1(10)	16	16	14	2(10)	42	36	26	2(10)	42	36	26
1996:04	1(10)	16	16	14	2(9)	42	36	NA	2(9)	42	36	NA
1996:05	1(10)	16	16	14	2(9)	42	36	NA	2(9)	42	36	NA
1996:06	1(10)	16	16	14	1(9)	30	26	NA	1(9)	30	26	NA
1996:07	0(10)	11	10	15	1(9)	30	26	NA	1(9)	30	26	NA
1996:08	0(10)	11	10	13	1(9)	30	26	NA	1(9)	30	26	NA
1996:09	0(10)	11	10	15	1(9)	30	26	NA	1(9)	30	26	NA
1996:10	0(10)	11	10	15	1(9)	30	26	NA	1(9)	30	26	NA
1996:11	0(10)	11	10	15	2(9)	42	36	NA	2(9)	42	36	NA
1996:12	0(10)	11	10	15	2(9)	42	36	NA	2(9)	42	36	NA
1997:01	0(10)	11	10	14	2(9)	42	36	NA	2(9)	42	36	NA
1997:02	0(10)	11	10	14	2(9)	42	36	NA	2(9)	42	36	NA
1997:03	0(10)	11	10	19	2(9)	42	36	NA	2(9)	42	36	NA
1997:04	0(10)	11	10	17	2(8)	42	36	NA	2(8)	42	36	NA
1997:05	0(10)	11	10	18	2(8)	42	36	NA	2(8)	42	36	NA
1997:06	0(10)	11	10	19	2(8)	42	36	NA	2(8)	42	36	NA
1997:07	0(1)	11	10	20	3(9)	42	36	NA	3(9)	42	36	NA
1997:08	0(10)	11	10	20	3(9)	42	36	NA	3(9)	42	36	NA
1997:09	0(10)	11	10	19	3(9)	55	36	NA	3(9)	55	36	NA
1997:10	0(10)	11	10	22	4(9)	55	36	NA	4(9)	55	36	NA
1997:11	1(10)	16	16	21	1(5)	30	26	NA	1(5)	30	26	NA
1997:12	0(4)	11	10	NA	1(4)	30	26	NA	1(4)	30	26	NA

1/ Number of good indicators (with noise-to signal ratio less than unity) that are signaling, with the number for which data are available in parenthesis. There are ten good indicators.
 2/ Predicted probabilities based on weighted sum of the good indicators, where each indicator is weighted by the inverse of its adjusted noise-to-signal ratio, with original KLR variables.
 3/ Predicted probabilities based on weighted sum of the good indicators, where each indicator is weighted by the inverse of its adjusted noise-to-signal ratio, with original KLR variables, augmented with the inclusion of the current account and M2/reserves in levels.
 4/ Predicted probabilities of crisis from a probit regression of impending crisis on the indicator variables measured linearly in percentiles.

Figure 4b: Crisis Probabilities based on Linear Probit Model for Selected Countries 1/



1/ Vertical line represents crisis dates.

III. FRANKEL AND ROSE (1996) PROBIT MODEL USING MULTI-COUNTRY SAMPLE

A. Methodology

FR estimate the probability of a currency crash using annual data for more than 100 developing countries from 1971–1992, a much broader sample of countries than the other two papers. The use of annual data may restrict the applicability of the approach as an early warning system, but it permits the analysis of variables such as the composition of external debt for which higher frequency data are rarely available. FR test the hypothesis that certain characteristics of capital inflows are positively associated with the occurrence of currency crashes: low shares of FDI; low shares of concessional debt or debt from multilateral development banks; and high shares of public sector, variable rate, short-term and commercial bank debt.⁴⁷

FR define a currency crash as a nominal exchange rate depreciation of at least 25 percent that also exceeds the previous year's change in the exchange rate by at least 10 percent. Thus, the type of currency crisis considered does not include speculative attacks successfully warded off by the authorities through reserve sales or interest rate increases. FR argue that it is more difficult to identify successful defenses, since reserve movements are noisy measures of exchange market intervention and interest rates were controlled for long periods in most of the countries in the sample.

⁴⁷The complete list of variables is as follows. Domestic macroeconomic variables: (1) the rate of growth of domestic credit, (2) the government budget as percent of GDP, (3) and the growth rate of real GNP. Measures of vulnerability to external shocks include: (1) the ratio of total debt to GNP, (2) the ratio of reserves to imports, (3) the current account as a percentage of GDP, and (4) the degree of overvaluation, defined as the deviation from the average bilateral real exchange over the period. Foreign variables are represented by (1) the percentage growth rate of real OECD output (in U.S. dollars at 1990 exchange rates and prices), and (2) a "foreign interest rate" constructed as the weighted average of short-term interest rates for the United States, Germany, Japan, France, the United Kingdom and Switzerland, with weights proportional to the fractions of debt denominated in the relevant currencies. Characteristics of the composition of capital inflows are expressed as a percentage of the total stock of external debt and include (1) amount of debt lent by commercial banks, (2) amount which is concessional, (3) amount which is variable rate, (4) amount which is public sector, (5) amount which is short-term, (6) amount lent by multilateral development banks (includes the World Bank and regional development banks but not the International Monetary Fund), and (7) the flow of FDI as a percentage of the debt stock.

B. Implementation

Table 9 (column 1) presents our reproduction of the FR benchmark probit regression.⁴⁸ The coefficients reflect the effect of one-unit changes in regressors on the probability of a currency crash (expressed in percentage points) evaluated at the mean of the data.⁴⁹ Significant results are starred. FR conclude from this and a variety of similar regressions that the probability of a crisis increases when output growth is low, domestic credit growth is high, foreign interest rates are high, and FDI as a proportion of total debt is low. They also found support for the prediction that crashes tend to occur when reserves are low and the real exchange rate is overvalued.⁵⁰

We made several revisions to the FR benchmark regression before updating it to 1996. As with the other papers, we used currently available, and hence revised, data from the same World Bank source as FR.⁵¹ This changed not only some of the data but also the sample, because some of the data that had previously been available, largely from the early 1970s, are now considered to be of unacceptable quality, while other formerly unavailable observations now had data. The net effect is to increase the number of observations from 780 in FR to 881, though the overlap of common data points is only 729 observations. In addition, we corrected an error in the original FR calculation of the overvaluation variable.⁵²

⁴⁸The reproduction of this result, with data and programs provided by the authors, was uneventful, in some contrast with the other two papers under consideration.

⁴⁹Thus, an increase in the share of short-term debt in total debt by 1 percentage point would increase the estimated probability of crisis by 0.23 percentage points.

⁵⁰Although the authors highlight the importance of low reserves and overvaluation in their conclusion, their results show significant effects were not robust and were found in fewer than half of the specifications they tested. The result that faster domestic growth reduces the probability of crisis is also not robust, as illustrated by the benchmark regression itself.

⁵¹Most of the data come from the World Development Indicators and Global Development Finance databases of the World Bank.

⁵²We also made two other technical modifications. First, we used percent changes instead of log differences in comparing the devaluations with the 25 percent crisis threshold. Second, we changed the implementation of the "windowing" procedure to more closely match the FR intent of ensuring that only the first of a sequence of crises was counted in the sample. See Milesi-Ferretti and Razin (1998) who recommended these two modifications.

Table 9. Frankel and Rose: Probit Estimates of Probability of a Currency Crash, 1970-92

	Model 1 FR Benchmark		Model 2 Revised	
	dF/dx	z ^{1/}	dF/dx	z ^{1/}
Commercial bank share of total debt	0.03	0.2	-0.07	-0.6
Concessional share	-0.14	-2.1 **	-0.12	-1.6
Variable rate share	-0.03	0.2	0.20	1.5
Short-term share	0.23	2.0 **	0.28	2.3 **
FDI/debt	-0.31	-2.5 **	-0.53	-3.7 ***
Public sector share	0.19	2.2 **	0.18	2.0 **
Multilateral share	-0.06	-0.8	0.08	1.0
Debt/GNP	-0.04	-1.7 *	-0.02	-1.5
Reserves/imports	-0.01	-3.4 ***	-0.01	-3.3 ***
Current account/GDP	0.02	0.2	-0.03	-0.3
Over-valuation 2/	0.08	2.5 **	0.15	4.0 **
Government budget surplus/GDP	0.16	1.1	0.10	0.7
Domestic credit growth	0.10	3.2 ***	0.08	3.8 ***
GDP growth rate	-0.16	-1.3	-0.07	-1.5
Foreign interest rate	0.80	2.6 ***	0.48	1.4
Northern (OECD) growth	-0.85	-1.5	-1.17	-1.9 *
Sample size	780		881	
Pseudo R ²	0.17		0.17	

Goodness of Fit

	Model 1			Model 2		
	Tranq.	Crash	Total	Tranq.	Crash	Total
Cut-off Probability of 50 percent ^{3/}						
Predicted tranquility	707	64	771	777	88	865
Predicted cash	4	5	9	7	9	16
Total	711	69	780	784	97	881
Cut-off Probability of 25 percent ^{4/}						
Predicted tranquility	678	52	730	743	62	805
Predicted cash	33	17	50	41	35	76
Total	711	69	780	784	97	881

1/ *, **, and *** denote significance at the 10, 5 and 1 percent levels respectively.

2/ Defined as the deviation from the average real exchange rate over the period.

3/ A crisis is correctly called when the estimated probability of crisis is above 50 percent if a crisis ensues within 24 months. A tranquil period is correctly called when the estimated probability of crisis is below 50 percent and there is no crisis within 24 months.

4/ A crisis is correctly called when the estimated probability of crisis is above 25 percent if a crisis ensues within 24 months. A tranquil period is correctly called when the estimated probability of crisis is below 25 percent and there is no crisis within 24 months.

The net effect of all these changes is shown in the second regression of Table 9. Overall, the model performs somewhat better than the original FR regression. The corrected overvaluation variable now has a much stronger and more significant effect. Higher northern (OECD) growth now significantly decreases the risk of crisis, and the effect of foreign interest rates is smaller and insignificant.⁵³

We now estimate the model through 1996 for purposes of generating predictions for 1997. As the first regression in Table 10 shows, the results are similar to the 1970 to 1992 regressions. A large share of debt which is concessional now reduces the risk of crisis.⁵⁴

The sample of countries used in these regressions is substantially different from those in the KLR and STV regressions. In particular, a large number of least-developed countries (such as the CMEA zone countries) and small island economies (for example, São Tomé, Cape Verde, and Vanuatu) are included. Because of concerns that crises in these countries may have different determinants and to maximize comparability with the other papers, we have rerun the FR regression over a smaller sample of 41 countries made up of all developing countries with per capita incomes above \$1,000 and population above 1 million for which there are data.⁵⁵

The results are broadly similar, as regression 2 of Table 10 shows. The most notable changes are that the ratio of reserves to imports is no longer significant whereas the current account and the fiscal balance now are.

Finally, as with the other models under consideration we try various plausible potential improvements to the original FR specification, in addition to changing the sample. We concentrate here on trying alternative explanatory variables.⁵⁶

⁵³For the overvaluation variable itself, the correction is the source of the improvement. For the other variables, the changes in sample resulting from the data revision are more important than the data revisions themselves, the changes in the windowing procedure and definition of crisis, or the correction of the overvaluation variable in driving these changes in results.

⁵⁴For purposes of predicting 1997 outcomes, we also estimate this regression with the government budget as a share of GDP excluded from this regression, because this variable is not available for 1996 as would be required for forecasting 1997. This omission makes little difference.

⁵⁵Milesi-Ferretti and Razin(1998) raise these sample issues and extract this smaller sample, for which they get improved results compared with FR.

⁵⁶This is in violation of the out-of-sample spirit of this paper, as clearly the selection of new variables is influenced by recent experience.

- We have seen that the ratio of reserves to imports does not seem to matter. Measuring **reserves as a ratio to short-term external debt and to broad money (M2)** have both been suggested as alternative ways of measuring the adequacy of reserves.⁵⁷ We find that both reserves/short-term external debt and reserves/M2 are separately significant predictors of crisis. When all three reserve ratios are included (Table 10 regression 3), reserves/M2 is significant at the 1 percent level, while reserves/short-term external debt can be rejected at the 10 percent significance level. The ratio of reserves to imports is insignificant and wrongly signed.⁵⁸
- The degree of **openness** of the economy may indicate the flexibility of the adjustment mechanism in the country and hence the probability of crisis. We found that more open economies, as measured by the share of exports and imports in GDP, were significantly less likely to suffer a crisis.⁵⁹
- Changes in the **terms-of-trade** had no apparent impact on the likelihood of crisis, while measuring the **debt composition variables as a share of GDP** rather than total debt also had no effect. **Interacting short-term external debt with credit growth**, in the spirit of STV, also did not help predict crises.

Regression 4 of Table 10 includes reserves/M2 and the degree of openness of the economy, as a result of this specification search. This model suggests that the probability of a crash increases when concessional debt and FDI are small and public sector debt large as a share of total external debt, the ratio of reserves/M2 is low, the current account deficit is large, the real exchange rate is overvalued, domestic credit growth is high, foreign interest rates are high, and the country is not open to trade.

The diagnostic statistics show that the models rarely generate a predicted probability of crash above 50 percent. Using this threshold, model 1 estimated through 1996 correctly predicts 89 percent of the observations and model 4 correctly predicts 90 percent, but the majority of the correct predictions are for tranquil periods. Model 1 correctly predicts only eight out of the 105 crashes; model 4 does better, predicting one-third of the crashes in the sample.

⁵⁷See for example Calvo and Mendoza (1996) on Mexico for an emphasis on the ratio of M2 to reserves and Sachs and Radelet (1998a) on the Asia crises for a focus on short-term external debt/reserves.

⁵⁸There is some non-robustness in the results for reserve ratios that calls for further analysis. For example, the inverse ratios M2/reserves and short-term external debt/reserves are not always significant predictors of crisis.

⁵⁹Milesi-Ferretti and Razin (1998) make this argument and include this variable in a similar regression with some success.

Table 10. Frankel and Rose: Probit Estimates of Probability of A Currency Crash, 1970-96

	Large Sample		Small Sample		Model 3		Model 4	
	Model 1		Model 2		Modified (Alt.)		Modified	
	dF/dx	z 1/	dF/dx	z 1/	dF/dx	z 1/	dF/dx	z
Commercial bank share of total debt	0.028	-0.3	0.022	0.1	0.053	1.0	0.121	1.0
Concessional share	-0.133	-1.9 *	-0.296	-2.3 **	-0.144	-3.2 ***	-0.305	-3.3 ***
Variable rate share	0.128	1.1	0.020	0.1	-0.062	-0.9	-0.089	-0.6
Short-term share	0.270	2.2 **	0.106	0.6	0.005	0.1	0.161	1.1
FDI/debt	-0.458	-3.3 ***	-0.795	-2.5 **	-0.251	-1.7 *	-0.576	-1.8 *
Public sector share	0.164	1.8 *	0.212	1.7 *	0.096	2.0 *	0.247	2.4 **
Multilateral share	0.064	0.8	0.021	0.1	0.018	0.3	0.045	0.4
Debt/GNP	-0.018	-1.4	-0.025	-0.5	0.001	0.1 *	0.021	0.6
Reserves/imports	-0.015	-3.9 ***	-0.007	-1.3	0.000	0.6		
Reserves/short-term debt					0.000	-1.6		
Reserves/M2					-0.092	-2.1 **	-0.206	-3.5 ***
Current account/GDP	-0.025	-0.2	-0.697	-2.4 **	-0.278	-2.4 **	-0.679	-2.8 ***
Over-valuation 2/	0.120	3.3 ***	0.172	2.9 ***	0.048	2.4 **	0.107	2.4 **
Government budget surplus/GDP	0.114	0.8	0.280	2.6 ***	0.280	2.6 ***	0.595	2.6 ***
Domestic credit growth	0.080	3.7 ***	0.182	4.4 ***	0.053	2.9 ***	0.119	3.1 ***
GDP growth rate	-0.080	-1.6	-0.058	-0.9	-0.006	-0.3	-0.017	-0.3
Foreign interest rate	0.331	1.0	1.007	2.0 **	0.052	2.3 **	0.909	2.3 **
Northern (OECD) growth	-1.522	-2.5 **	0.414	0.4	-0.102	-0.2	0.033	0.0
Open					-0.060	-3.5 ***	-0.239	-4.1 ***
Sample size	940		464		438		448	
Pseudo R ²	0.15		0.20		0.32		0.32	
Goodness of Fit								
	Model 1		Model 2		Model 3		Model 4	
	Tranq.	Crash	Tranq.	Crash	Tranq.	Crash	Tranq.	Crash
	Total		Total		Total		Total	
Cut-off Probability of 50 percent 3/								
Predicted tranquility	830	97	398	50	448	40	381	421
Predicted crash	5	8	6	10	16	7	7	27
Total	835	105	404	60	464	60	388	448
Cut-off Probability of 25 percent 4/								
Predicted tranquility	792	72	373	34	407	22	342	369
Predicted crash	43	33	31	26	57	41	41	79
Total	835	105	404	60	464	60	383	448

1/ *, **, and *** denote significance at the 10, 5 and 1 percent levels respectively.

2/ Defined as the deviation from the average real exchange rate over the period.

3/ A crisis is correctly called when the estimated probability of crisis is above 50 percent if a crisis ensues within 24 months. A tranquil period is correctly called when the estimated probability of crisis is below 50 percent and there is no crisis within 24 months.

4/ A crisis is correctly called when the estimated probability of crisis is above 25 percent if a crisis ensues within 24 months. A tranquil period is correctly called when the estimated probability of crisis is below 25 percent and there is no crisis within 24 months.

When an estimated probability of above 25 percent followed by a crash is considered success, the results look better. Model 4, for example, generates a probability above 25 percent before 63 percent of crises. About half of warnings defined this way (41 out of 79) were not followed by a crash.⁶⁰

The FR models thus show some promise for predicting crises based on this in-sample assessment. There is a fair amount of parameter stability across samples, and many sensible variables are significant predictors of crisis. The overall explanatory power is fairly low, though our modifications lead to some improvement here.

C. Predicting 1997

The FR models estimated through 1996 can easily generate out-of-sample predictions for 1997. Forecasting with these models presents one complication not faced until now, however: some pre-crisis explanatory variables are still unavailable, even as of mid-1998.⁶¹ In particular, data on the 1996 government budget deficit is only available for 13 countries in the larger FR sample. We have “filled in” this variable for 1996 from other sources.⁶² This makes some difference for the out-of-sample forecasts, compared with the alternative of re-estimating the models without the deficit. Indeed, for models 2 through 4, the 1997 forecasts are slightly better when the models are reestimated and predictions generated without the deficit variable. In practice, the implementation of any of the models in the paper would involve filling in recent values of many of the explanatory variables from alternative data sources. For this reason, and in the interest of avoiding the use of the out-of-sample information to aid in specifying the models, we keep the deficit variables in the reported forecasts.

⁶⁰A Pesaran-Timmermann test, rejects, at the 1 percent level, the hypothesis that the predictions using both thresholds are uncorrelated with the actual incidence of crisis.

⁶¹Some of the predictive variables used in the KLR model were also still not available for late 1996 and early 1997 as of mid-1998, as shown in Table 8. The forecasts from the probit regressions were not produced for these observations. The weighted-sum of indicators signaling and associated probabilities of crisis can still be calculated even when some indicators are not available.

⁶²We used IFS data where available, and IMF WEO estimates otherwise. Where the measures overlap, the correlation between the World Bank deficit measure and the IFS measure is generally around 0.95 (depending on the sample). The correlation between the WEO deficit measure and the World Bank measure is somewhat lower at around 0.85.

Table 11a shows predicted probabilities of crisis and actual values of the nominal exchange rate depreciation for 1997 for 44 countries for which data are available, based on the updated FR benchmark regression for this larger sample, model 1 of Table 10.⁶³ Table 11b reports actual exchange rate depreciation and predicted probabilities of crisis, with associated country rankings, for the models based on the smaller sample, models 2 and 4 of Table 10.

Overall, the forecasts are at best moderately successful, with correlations ranging from 18 to 24 percent. The fraction of the variance of the rankings accounted for (measured by the R-squared statistic) is always below 7 percent, and the predictions are not significant, with the lowest p-values at 11 percent.⁶⁴ A closer examination of the eight countries discussed in detail above illustrates this weakness. In the large sample (Table 11a), Thailand (the only crisis country of the eight for which data are available) has a 10 percent probability of crisis, while Brazil, Mexico and Argentina have probabilities of 9 percent, 18 percent, and 8 percent respectively. In the smaller sample models in Table 11b, Indonesia and Thailand have lower estimated probabilities of crisis than Brazil, Mexico, and Argentina.

In sum, the FR model and extensions fail to provide much useful guidance on crisis probabilities in 1997.

IV. SACHS, TORNELL, AND VELASCO (1996) CROSS-COUNTRY REGRESSIONS

A. Methodology

Both KLR and FR examine the predictive power of a large number of variables on a panel of countries. This creates two problems. First, the use of a panel relies on the assumption that all crises can be explained in the same way. Second, the analysis of a large number of possible explanatory variables means that, even with a large panel of crises, it is not possible to consider all the nonlinearities and interaction effects that may be important. An alternative is to concentrate on a smaller number of episodes that can reasonably be considered similar, while focussing on a small number of variables deemed critical based on a priori reasoning. This is the path taken by STV.

STV analyze the impact of Mexico's financial crisis of December 1994 on other emerging markets in 1995 (the so-called Tequila effect). They examine the determinants of the magnitude of the currency crisis in a cross-section of countries in 1995. This approach cannot hope to shed light on the timing of crises. Rather, it may answer the question of which countries are most likely to suffer serious attacks in the event of a change in the global environment. This approach is attractive, even for our purposes, for a number of reasons.

⁶³The use of annual frequency does not work well for the crisis variable in 1997; because the devaluations happened toward the end of a year following some within-the-year appreciation, none of the Asian countries are identified as crisis countries in 1997.

⁶⁴Testing over the common sample of models 2 and 4 does not change these results much.

Table 11a. Currency Crash Probabilities Based on FR Probit Model

Country Name	Actual 1997	Outcome	Probabilities of Crisis in 1997	
	Crisis Index 1/ Rank	Rank	Table 10 Model 1 Probability	Rank
Romania	132.5	1	1.2	37
Turkey	86.4	2	17.2	4
Ecuador	25.4	3	14.2	6
Madagascar	25.4	4	4.6	25
Ghana	25.2	5	0.0	43
Thailand	23.9	6	9.9	11
Hungary	22.5	7	9.3	13
Uruguay	18.5	8	25.0	1
Venezuela	17.2	9	2.6	31
Mauritius	14.8	10	12.3	10
Equatorial Guinea	14.2	11	6.7	19
Cote d'Ivoire	14.2	11	5.3	23
Congo	14.2	11	6.7	20
Cameroon	14.2	11	8.4	15
Pakistan	13.9	15	7.4	18
Nicaragua	12.0	16	0.0	44
Colombia	10.0	17	7.9	17
Morocco	9.3	18	9.5	12
Peru	8.4	19	13.3	8
Guinea	8.3	20	5.7	22
Brazil	7.5	21	8.8	14
Swaziland	7.2	22	0.1	42
Sri Lanka	6.7	23	1.8	35
Haiti	6.4	24	13.6	7
Bangladesh	5.0	25	2.2	32
Ethiopia	5.0	26	0.4	39
Mexico	4.1	27	17.5	3
Dominican Republic	3.8	28	15.4	5
Bolivia	3.5	29	0.8	38
Uganda	3.2	30	2.8	30
Kenya	2.8	31	3.0	29
Fiji	2.6	32	4.7	24
India	2.4	33	2.1	33
Chile	1.6	34	3.2	28
Seychelles	1.2	35	2.0	34
Guatemala	0.3	36	12.5	9
Maldives	0.3	37	4.2	26
Nigeria	0.1	38	6.5	21
Argentina	0.0	39	8.4	16
Jordan	0.0	40	3.5	27
Panama	0.0	40	23.7	2
Egypt	0.0	42	0.1	41
China	-0.2	43	0.2	40
Rwanda	-1.6	44	1.2	36
Correlation 2/				0.24
p-value				0.11
R ²				0.06

1/ Values in bold indicate crises according to the FR definition. Note that Turkey and Ecuador do not have crises even though the index is above 25 because the index did not exceed the previous year's value by 10 points.

2/ Spearman Rank Correlation of the fitted values and the actual crisis index and its p-value. The R² is from a regression of fitted values on actual values.

Table 11b. Currency Crash Probabilities Based on Revised FR Models

Country Name	Actual 1997		Probabilities of Crisis in 1997			
	Crisis Index 1/	Rank	Table 10 Model 2		Table 10 Model 4	
			Probability	Rank	Probability	Rank
Romania	132.5	1	0.0	23
Turkey	86.4	2	15.8	3	20.1	2
Ecuador	25.4	3	9.7	6	6.2	7
Indonesia	24.7	4	2.4	11
Thailand	23.9	5	6.3	10	0.4	16
Hungary	22.5	6	8.1	7
Uruguay	18.5	7	21.0	2	26.2	1
Venezuela	17.2	8	1.2	18	0.0	21
Mauritius	14.8	9	3.9	15	0.0	20
Cote d'Ivoire	14.2	10	1.4	17	1.0	12
Pakistan	13.9	11	0.6	19	0.8	14
Philippines	12.5	12	0.9	13
Colombia	10.0	13	6.0	12	5.4	9
Morocco	9.3	14	4.5	14	5.0	10
Peru	8.4	15	29.0	1	15.7	4
Brazil	7.5	16	6.5	9	15.6	5
Sri Lanka	6.7	17	0.2	22	0.0	22
Paraguay	6.2	18	0.2	18
Mexico	4.1	19	10.7	4
Dominican Republic	3.8	20	5.4	13	0.3	17
Bolivia	3.5	21	0.5	20
India	2.4	22	0.4	21
Chile	1.6	23	3.4	16	0.4	15
Guatemala	0.3	24	6.2	11	7.8	6
El Salvador	0.1	25	5.6	8
Argentina	0.0	26	7.8	8	16.5	3
Panama	0.0	27	10.2	5	0.0	23
Egypt	0.0	28	0.0	24	0.1	19
Correlation 2/				0.18		0.22
p-value				0.41		0.31
R ²				0.03		0.05
Number of observations				24		23

1/ Values in bold indicate crises according to the FR definition. Note that Turkey and Ecuador do not have crises even though the index is above 25 because the index did not exceed the previous year's value by 10 points.

2/ Spearman Rank Correlation of the fitted values and the actual crisis index and its p-value.

The R² is from a regression of fitted values on actual values.

- The timing may be much harder to predict than the incidence of a crisis across countries.
- The determinants of crisis episodes may have varied importantly over time.
- STV can impose more economic structure on their analysis by focussing on a particular set of crises (those occurring at one time). STV argue that a key feature of the 1995 crises was that the attacks hit hard only at already vulnerable countries. In a rational panic, investors identify a country as being likely to suffer from a large devaluation in the face of an outflow, and validate their own concerns by fleeing the country. Thus, countries with overvalued exchange rates and weak banking systems were subject to more severe attacks, but only if they had low reserves relative to monetary liabilities (so that they could not easily accommodate the capital outflow) and weak fundamentals (so that fighting the attack with higher interest rates would be too costly).

For our purposes, it is important that the crises that affected mostly Asian countries in 1997 be broadly similar to the 1995 crises. Although there are certainly important differences in the two episodes, the standard for similarity need not be very high. The other two papers under consideration assume all crises of the last several decades are identical. If the Tequila and Asian crises are such that a model formulated in 1995 has no explanatory power, then doubt must be cast on efforts that make a stronger demands on parameter constancy.⁶⁵

B. Implementation

STV examine data on a cross-section of 20 emerging markets. They define a crisis index [IND] as the weighted-sum of the percent decrease in reserves and the percent depreciation of the exchange rate, from November 1994 to April 1995. The central argument, and result, is that while the occurrence and timing of the crises was clearly a product of contagion, the variation in the crisis index across countries is largely explicable. They find that countries had more severe attacks when their banking systems were weak (proxied by a lending boom variable [LB] measuring growth in loans to the private sector from 1990 through 1994) and when the exchange rate was overvalued (measured as the degree of depreciation from 1986–89 to 1990–94 [RER]). Moreover, they find that these factors only

⁶⁵The IMF (1998) argues that the STV results apply to the Asian crisis and constructs a composite indicator of crises on that basis. Sachs (1997) argues that Thailand's 1997 crisis "has the same hallmarks [as the 1995 crises]: overvaluation of the real exchange rate, coupled with booming bank lending, heavily directed at real estate." Sachs and Radelet (1998a) argue that the 1997 and 1995 crises shared important characteristics, though their interpretation of post-Thailand Asian crises relies more heavily on contagion effects. Sachs and Radelet (1998b), Tornell (1998) and Corsetti et al. (1998) apply models in the STV spirit to both sets of crises.

matter for countries with low reserves [DLR], measured as having a Reserves/M2 ratio in the lowest quartile, *and* “weak fundamentals” [DWF], which means having RER in the lowest three quartiles or LB in the highest three quartiles.

Thus, they estimate an equation of the form :

$$IND = \beta_1 + \beta_2 RER + \beta_3 LB + \beta_4 RER \cdot DLR + \beta_5 LB \cdot DLR + \beta_6 RER \cdot DWF + \beta_7 LB \cdot DWF$$

where their hypotheses are:

(1) Countries with a depreciated real exchange rate suffer a less severe crisis, but this only matters for countries with low reserves and weak fundamentals, so that

$$\beta_2 = 0, \beta_2 + \beta_4 = 0, \beta_2 + \beta_4 + \beta_6 < 0$$

(2) Lending booms increase the severity of crisis, but only for countries with low reserves and weak fundamentals, so that

$$\beta_3 = 0, \beta_3 + \beta_5 = 0, \beta_3 + \beta_5 + \beta_7 > 0$$

Regression 1 of Table 12 reproduces the original STV benchmark regression, using their data.⁶⁶ The results emphasized by STV are, first, that the effect of RER is significantly negative for countries with low reserves and weak fundamentals (the sum of estimates of $\beta_2 + \beta_4 + \beta_6$ is negative), and the effect of LB is significantly positive for these same countries (the

⁶⁶Regression 1 differs slightly from the published STV benchmark, mainly because we have corrected an error in the calculation of RER for Taiwan Province of China, in STV. The resulting differences are statistically, numerically, and economically small. In addition, the data used both in the STV benchmark and regression 1 differ slightly from that described and published in STV. First, the data published in STV (but not that used in their regressions) contain several typographical errors, which we have corrected with the help of the authors. Second, here and in the STV regression the lending boom variable was calculated differently for Peru than for the other countries and as defined in the appendix of STV. Specifically, LB is defined as the growth from 1990 through 1994 in the ratio of domestic credit to the private sector to GDP. For Peru, however, the base year actually used is apparently 1991. This is presumably because the hyperinflation and stabilization of 1989/1990 led to a tiny base of credit/GDP and would have resulted in a large outlier for Peru if calculated as defined in STV. Third, the measure of reserves for South Africa apparently includes gold reserves, as is standard for that country but contrary to the description in the appendix of STV.

Table 12. STV: 1994/5 Regressions

Reg. No.	Regression	No. of Countries	R ²	Rbar ²	Results 1/, 2/							
					$\beta_2 = 0$	$\beta_2 + \beta_4 = 0$	$\beta_2 + \beta_4 + \beta_6 < 0$	$\beta_3 = 0$	$\beta_2 + \beta_3 = 0$	$\beta_2 + \beta_3 + \beta_7 > 0$		
With quartiles												
1	STV benchmark, rerun fixing Taiwan Province of China	20	0.69	0.55	6.61 (2.311)	0.28 (0.539)	-2.16 (0.994)	1.86 (0.810)	-5.49 (2.648)	3.88 (1.463)		
2	STV with revised data	20	0.66	0.50	6.46 (1.860)	-0.06 (0.228)	-1.59 (1.133)	1.74 (0.654)	-16.87 (7.292)	3.75 (1.405)		
3	23-country sample	23	0.44	0.22	4.18 (1.811)	-0.34 (0.205)	-3.48 (1.615)	0.94 (0.617)	-8.11 (6.578)	1.89 (1.354)		
4	Alt. RER definition(1) RER 1994/1990	23	0.41	0.19	-4.18 (2.329)	0.78 (0.875)	-4.29 (1.784)	-0.42 (0.157)	0.22 (0.771)	2.15 (1.582)		
5	Alt. RER definition(2) RER 1994/average (1986 to 1989)	23	0.43	0.21	2.22 (2.566)	-0.30 (0.134)	-2.69 (1.141)	0.46 (0.710)	-5.72 (5.779)	1.80 (1.351)		
With halves												
6	STV benchmark from STV	20	0.68	0.54	-0.52 (1.545)	0.48	-6.58	0.26 (0.547)	-1.80	3.32		
7	STV with revised data	20	0.66	0.50	-0.20 (0.952)	0.15 (0.881)	2.51 (4.657)	0.17 (0.399)	2.06 (3.832)	4.94 (2.209)		
8	23-country sample	23	0.64	0.51	-0.26 (1.054)	1.50 (1.216)	-14.45 (3.992)	0.93 (0.724)	5.50 (9.222)	1.93 (0.731)		

1/ Coefficients in bold are significant at the 5-percent level. Underlined coefficients are significantly inconsistent with the STV hypothesis. Figures in parenthesis are standard errors.

2/ The β 's are coefficients from the regression $IND = \beta_2 RER + \beta_3 LB + \beta_4 RER * DLR + \beta_5 RER * DLR + \beta_6 RER * DWF + \beta_7 LB * DWF$, where RER is the degree of real depreciation, LB is a measure of the lending boom, DLR is a dummy variable for countries with low reserves, and DWF is a dummy for countries with weak fundamentals (see text for explanations).

sum of estimates of $\beta_3 + \beta_5 + \beta_7$ is positive). Moreover, the interaction of these two variables with low reserves alone is insignificant.⁶⁷ The high R-squared of the regression (0.69) is taken to indicate that the model explains fairly well the pattern of contagion.

Before applying this model to the 1997 Asia crisis, we carry out some additional tests over the STV sample. First, the data used in STV has been revised. Because we will be using the currently available, and hence revised, data when we apply this approach to 1997, we first apply the revised data to the original STV regression for the 1994-95 crises. Thus, regression 2 of Table 12 represents the result of applying the STV model to the STV sample, using the data now available for that period. The revisions to the data appear small, but the cumulative effect is to substantially change some of the estimates.⁶⁸ Most notably, the effect of RER with low reserves and weak fundamentals ($\beta_2 + \beta_4 + \beta_6$) is now insignificantly different from zero, while the coefficient on LB with low reserves ($\beta_3 + \beta_5$) increases significantly.

For purposes of comparing forecasts with the other approaches discussed in this paper, line 3 of Table 12 presents the result of running the benchmark regression over the same sample of 23 countries to which we apply the KLR approach. There are important differences with regression 2. Most notably, the effect of a lending boom with weak fundamentals and low reserves ($\beta_3 + \beta_5 + \beta_7$) is much smaller and is no longer significant.⁶⁹

We have also tried estimating variants of the STV regressions for the 1994–1995 sample based on different definitions of the real exchange rate variable. The STV definition in terms of the average level of the real exchange rate in the 1990 through 1994 period divided by the average level during 1986 through 1989 clearly has an arbitrary element, and STV themselves calculate various alternative definitions. Given the changes induced in real exchange rates in some countries by the 1995 crises themselves, alternative definitions might work better. Table 12 regression 4 measures the real exchange rate change as the percent change in the real exchange rate from 1990 to 1994, while regression 5 measures RER as the level of the real exchange rate in 1994 compared with its average over the 1986 to 1989 period. The regressions are quite similar to the benchmark specification in regression 3.

⁶⁷The result that RER *increases* the severity of crisis when not interacted (that is, a more depreciated currency implies a bigger crisis) is noted as anomalous, while the fact that the un-interacted lending boom increases the severity of the crisis is a milder puzzle. Another anomalous result in regression 1, compared with the STV hypotheses, is that the lending boom significantly decreases the expected severity of the crisis when interacted with low reserves alone ($\beta_2 + \beta_3 < 0$). This anomaly appears also in STV, though in STV the sum is slightly smaller and (barely) insignificant.

⁶⁸The R-squared statistics in regressions of the revised series on the STV series (for example, the revised crisis index on the STV crisis index) are all above 0.95.

⁶⁹STV demonstrate that their results are reasonably robust to the exclusion of various countries from their sample.

The definitions of low reserves and weak fundamentals in terms of which quartile of the sample the country finds itself are somewhat arbitrary. For this reason, STV vary the definition of low reserves and weak fundamentals so that countries in different fractions of the sample qualify. For example, regression 6 of Table 12 reproduces the STV results for the case where "low reserves" is defined as having a reserves/M2 ratio in the bottom *half* of the sample, while "weak fundamentals" is having low reserves or a an exchange rate depreciation in the lower *half* of the sample. The main results continue to hold. Regressions 7 and 8 of Table 12 present the re-estimation of regression 5 with revised data and correcting the Taiwan Province of China crisis variable. Unlike with the quartile regressions, this changes the results: most importantly, RER with low reserves and weak fundamentals ($\beta_2 + \beta_4 + \beta_6$) now has the wrong sign, though it is insignificant.⁷⁰

The fragility of the STV results with respect to the data revisions that have taken place since their estimations and to the addition of three countries to the sample casts some doubt on the usefulness of this specification for the Asia crises. We nonetheless generate predictions for 1997 based on these estimates drawn from the Tequila crisis.

C. Predicting 1997

The application of the STV model to 1997 is not as straightforward as with the other two approaches. Because the model is formulated and estimated over a cross-section of countries, it is not clear how to update for 1997. We attempt two approaches. First, in true out-of-sample fashion, we mechanically update the STV variables and apply the coefficients from the STV regressions for the Tequila crisis to obtain predicted values for the 1997 crises. For the dependent variable that measures the severity of the crisis, we measure percent depreciation of the nominal exchange rate from April 1997 through December 1997. For the explanatory variables, we move all the definitions forward two years. We then calculate forecasts of devaluation using the coefficient estimates from the STV benchmark specification estimated for the Tequila crisis.

Column 1 of Table 13 shows the country rankings based on the actual value of the crisis index for 1997, defined, analogously to STV, as the change in the nominal exchange rate between April and December 1997. Column 2 present country rankings based on applying the exact coefficients from the published STV benchmark regression to the updated LB and RER variables and associated dummy variables. Columns 3 through 7 present alternative forecasts based on regressions 1 through 5 of Table 12, our reestimations of the 1994–1995 STV regressions.

None of these forecasts performs very well. The most successful specification, based on Table 12 regression 4, employs one of the alternative definitions of RER. Its forecast

⁷⁰In this case, part of the reason for the difference is that, even using the (typo-corrected) STV data, we were not able to reproduce regression 5.

Table 13. STV: Predicted and Actual Crises in 1997

Country Name	Actual Crisis Rankings Apr.-Dec 1997 (1)	Orig. STV Benchmark (2)	Predicted Crisis Rankings					In-Sample Forecasts 1/	
			Out-of-Sample Forecasts (Table 12)		Model 3			Model 1	Model 2
			Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2
Indonesia	1	14	14	14	14	9	12	11	13
Thailand	2	5	5	6	7	5	6	6	4
Korea	3	13	13	13	12	11	11	10	11
Malaysia	4	4	4	5	6	6	5	8	3
Zimbabwe	5	21	19	19	23	12	21	1	10
Philippines	6	1	1	1	1	1	1	2	1
Turkey	7	8	7	7	9	21	9	3	2
Colombia	8	19	20	20	18	4	20	21	19
Taiwan Province of China	9	12	12	12	11	22	14	12	8
Pakistan	10	18	18	17	17	20	15	14	18
Uruguay	11	6	6	8	3	3	4	17	14
South Africa	12	15	15	15	15	16	22	9	22
India	13	11	11	2	5	19	7	16	20
Brazil	14	7	8	9	4	2	3	15	15
Sri Lanka	15	17	17	18	16	17	16	13	16
Chile	16	20	21	21	19	14	18	19	17
Jordan	17	16	16	16	20	15	17	4	9
Mexico	18	23	23	23	21	18	23	20	23
Israel	19	10	10	11	10	8	8	7	7
Peru	20	3	3	4	8	23	13	18	21
Venezuela	21	22	22	22	22	13	19	22	12
Bolivia	22	9	9	10	13	10	10	5	5
Argentina	23	2	2	3	2	7	2	23	6
Correlation		0.08	0.10	0.11	0.11	0.23	0.12	0.46	0.25
P-value		0.73	0.66	0.61	0.61	0.29	0.58	0.03	0.26
R ²		0.01	0.01	0.01	0.01	0.05	0.02	0.21	0.06

1/ See text for explanation of models.

rankings of crisis severity are insignificant predictors of the actual rankings and explain only 5 percent of the variance of the actual country rankings. Among the countries we have been examining in more detail, this model yields large expected crises for Malaysia and Thailand but also for Brazil and to a lesser extent Argentina. It misses Indonesia and Korea. The other specifications, including the benchmark, perform somewhat worse.

In light of this predictive failure, we have also considered a much less ambitious test of the STV model, justified by the idea that we may reasonably expect some constancy of the general model of crisis episodes even if parameter constancy fails to hold. In this spirit, our second approach is to re-estimate the regression using 1996 and 1997 data. This application of the STV model to the 1997 crisis meets with little success. The results vary strongly depending on the exact specification, but the fit is always relatively poor.⁷¹ Compared with its application to the 1994 crisis, the coefficients are economically and statistically different, and the explanatory power of the regressions is much lower. STV argue based on the Tequila crisis that lending booms combined with weak fundamentals and low reserves are a recipe for crisis. This hypothesis receives some support, in that the relevant coefficients are significant and consistent with that view. No evidence is found, however, for the importance of the real exchange rate.

We can observe more directly how much the re-estimation of the STV model using 1997 data improves the predictions for 1997 compared with the out-of-sample forecasts, which as we have already seen perform poorly. The in-sample forecasts are substantially better, as is to be expected, but are still not very useful. Columns 8 and 9 of Table 13 present the predicted crisis index and rankings based on two regressions estimated over 1996/1997 data. The first represents a specification that follows the benchmark STV specification while the second mirrors the specification in regression 4 of Table 12. The first forecast is somewhat useful, with a significant correlation of 0.46 with the actual rankings and an R^2 statistic of 0.21. It is remarkable, though, that the STV regression re-estimated with 1997 data performs somewhat worse than the KLR out-of-sample forecasts and much worse than the modified KLR forecasts.

A recent paper (Tornell (1998)) may seem to contradict the results in this paper. Tornell estimates a model very similar to STV, stacking observations from the 1994/1995 crisis and the 1997 crisis. He finds that his new model: (1) fits fairly well, with significant coefficients plausibly signed; (2) has coefficients that appear stable between the two sets of crises; and (3) when fitted with the 1994 observations only and forecasting for 1997, produces good predictions, much better than the STV forecasts examined here and comparable to the KLR weighted-sum of indicators-based probabilities.⁷²

⁷¹An appendix is available from the authors on request.

⁷²The Tornell (1998) forecasts are significant at the 1 percent level in a regression of the actuals on the forecasts, with an R-squared statistic of 0.24.

Rather than providing a counter example to the results presented here, this effort illustrates the importance of testing models out of the sample used to formulate them, as we do here. A variety of apparently small modifications characterizes the difference between the specification in STV and Tornell (1998), and yet these re-specifications apparently make the difference between success and failure in predicting the incidence of the 1997 crises “out-of-sample.”⁷³

This suggests that specification uncertainty can be as important as parameter uncertainty across crisis episodes, at least for techniques such as STV that rely on a small number of observations and relatively complex models. Only the application of models to episodes that postdate the design of the model provides an appropriately tough test. Unfortunately for our purposes, the apparent need for a separate specification search for the new set of crises casts some doubt on the usefulness of this sort of approach for predicting future crises.

V. DISCUSSION

This paper has assessed how well the various forecasts work in several ways. Summarizing some of these results, Table 14 directly contrasts the performance of each of the three methods in ranking countries by probability (KLR and FR) or severity (STV) of crisis in 1997, comparing the predicted and actual rankings. The only successful pure out-of-sample forecasts in terms of country rankings are those based on the KLR noise-to-signal weighted-sum of indicators (column 1). In addition, the KLR model augmented with additional variables (column 2) and the BP probit model (columns 3 and 4) also provide useful forecasts. None of the STV and FR-based predictions are helpful.

We have throughout paid particular attention to eight key Asian and Latin American countries to illustrate these results. None of the three methods we began with tells a very clear story about these countries:

- The original weighted-sum of indicators specification correctly predicts relative tranquility for Argentina and Mexico and predicts a fairly high severity of crisis in Thailand and Malaysia but issues much stronger warnings for Brazil and the Philippines while missing Indonesia and Thailand.
- The STV regressions based on the 1994-1995 observations do not seem to provide very useful forecasts. They predict a fairly high severity of crisis in Thailand and Malaysia but miss Indonesia and Korea. They correctly predict tranquility for Mexico and predict an intermediate outcome for Brazil, but assign high risk to Argentina and the Philippines.

⁷³Examples of subtle specification issues include the exact definition of RER, LB, weak fundamentals and low reserves, as well as details such as the need in STV to adjust the measurement LB for Peru to avoid an unwanted outlier (see footnote 64). Tornell (1998) like STV does contain a substantial number of robustness checks (such as dropping individual countries and varying the definition of low reserves).

Table 14. Correlation of Actual and Predicted Rankings based on KLR, BP, FR, and STV

	KLR		BP		FR		STV				
	Actual 1997 Crises Index	Predicted Probabilities of Crisis in 1997 Noise-to-signal Weighted Sum of Indicators 1/ Original 2/ Augmented 3/	Predicted Probabilities of Crisis in 1997		Actual 1997 Crisis Index	Predicted Probabilities of Crisis in 1997 4/		Actual Crisis Index April-Dec 1997	Predicted Probabilities of Crisis in 1997		
			Indicators Table 5	Linear Model 1		Table 5	Model 2		Table 10	Model 3	
Thailand	1	16	7	4	2	3	7	7	2	7	5
Korea	2	4	5	3	9	3	7	11	3	12	11
Indonesia	3	18	11	8	8	2	7	7	1	14	9
Malaysia	4	8	13	5	1				4	6	6
Zimbabwe	5	3	3						5	23	12
Taiwan Province of China	6	5	4	6	5				9	11	22
Colombia	7	9	12	7	4	8	8	6	8	18	4
Philippines	8	1	1	2	12	7	7	8	6	1	1
Brazil	9	2	2	1	10	10	6	5	14	4	2
Turkey	10	7	10	13	18	1	3	2	7	9	21
Venezuela	11	14	16	13	20	5	10	12	21	22	13
Pakistan	12	10	9	11	6	6	11	9	10	17	20
South Africa	13	6	8	12	11				12	15	16
Jordan	14	15	18	13	13				17	20	15
India	15	20	21	13	19	14	13	13	13	5	19
Sri Lanka	16	17	19	13	14	11	14	13	15	16	17
Chile	17	18	20	13	15	15	9	10	16	19	14
Bolivia	18	20	21	13	21	13	12	13	22	13	10
Argentina	19	12	17	13	17	16	5	3	23	2	7
Mexico	20	13	14	13	22	12	4	4	18	21	18
Peru	21	11	6	10	7	9	1	4	20	8	23
Uruguay	22	20	21	13	16	4	2	1	11	3	3
Israel	23	20	15	9	3				19	10	8
Correlation 5/		0.54	0.60	0.67	0.48		0.33	0.12		0.11	0.23
P-value		0.007	0.003	0.001	0.026		0.253	0.694		0.612	0.295
R ²		0.28	0.36	0.47	0.23		0.11	0.02		0.01	0.05

1/ Based on average of weighted-sum probabilities during 1996:1-12, using out-of-sample estimates.

2/ Original KLR variables.

3/ Addition of current account and M2/reserves in levels to original variables

4/ Average predicted probabilities for 1996:1-12, where model was estimated upto 1995:4.

5/ Spearman Rank Correlation of the fitted values and the actual crisis index and its p-value. The R² is from a regression of fitted values on actual values.

- The FR predictions also perform without distinction. The estimated probability of crisis is reasonably high for Thailand in some versions, but is higher for Brazil, Mexico, and Argentina, countries which did not suffer crises (there is insufficient data to forecast the other countries under consideration). An important problem with the FR model is the definition of crisis; it fails in 1997 to identify most of the countries that are commonly accepted to have experienced a crisis.

The BP probit model's somewhat better performance overall is mirrored in the story it tells for these eight countries, particularly in its linear form. It ranks the crisis countries Thailand, Korea, Malaysia and Indonesia in the top ten countries at risk and correctly predicts tranquility for Argentina and Mexico. It attaches somewhat high risks to the Philippines and Brazil, which is plausible *ex post* given the difficulties both countries have faced in 1997.⁷⁴

VI. CONCLUSION

This paper has examined the extent to which models formulated and estimated prior to 1997 would have helped predict the 1997 currency crises. For each model, we have tried to reproduce as closely as possible the original specification, and we have also tried reasonable modifications, the most substantial of which we dub the BP model. We have evaluated the results in several ways. For the models with a time dimension (KLR, FR and BP) we have reported the out-of-sample goodness of fit statistics, which measure whether models successfully call future crises. We have also looked at how each model would have ranked countries in 1997 in terms of probability of crisis and compared this result with the actual cross-country ranking of crisis severity. Finally, we have examined a few country cases in more detail.

The results are mixed. The most successful "pure" pre-1997 forecasts we study are the KLR-based probabilities of crisis derived from the weighted-sum of signaling indicators. When this model issued an alarm during the 1995:5 to 1996:12 period, a crisis would actually have followed in 1997 37 percent of the time.⁷⁵ This compares to a 27 percent unconditional probability of crisis in 1997. Moreover, its forecasted cross-country ranking of severity of crisis is a significant predictor of the actual ranking, with an R^2 of 28 percent. Its ranking of the eight key countries we have been examining in detail does not, however, shed light on what was to come in 1997. The other two methods examined, FR and STV, provide forecasts that would have been of little use. These forecasts are insignificant predictors of actual outcomes and, although positively correlated with actual outcomes, explain very little of the variance of actual countries' experience.

⁷⁴Probits that excluded the current account, reported in a previous version of this paper, largely failed to predict a crisis in Indonesia. Note that we employ the KLR definition of crisis here.

⁷⁵An alarm here is defined as a predicted probability above 25 percent. These alarms are significant predictors of crises at the 1 percent level.

We also modified the models in various ways. The addition of two variables to the KLR model, the level of the current account and M2/reserves, improves performance somewhat. Various plausible modifications to the STV and FR models did not yield useful forecasts, even some, such as the inclusion of short-term external debt, actually inspired by events in 1997.

We also estimated a set of alternative models (BP probit-based models) using the data and crises definition of the KLR method but with a different approach to generating crisis probabilities from the data. These models did not exist prior to the crises they attempt to predict and to that extent do not generate pure out-of-sample forecasts. However, the methodological innovations were not inspired by events in 1997, nor did we use success or failure in predicting 1997 outcomes to aid in the specification of the alternative models. The BP probit models provide generally better forecasts than the KLR models. The probit in which the predictive variables enter linearly issues alarms in 1995:5 to 1996:12 that are followed by crises 51 percent of the time. We also examine other probit specifications that do not embody the KLR indicators assumption and find that, while the results are not unambiguous, the linear model is the most successful.

The answer to the question posed in the title of this paper is “yes, but not very well.” The answer is “yes” since the KLR forecasts, and *a fortiori* the BP modifications, are clearly better than a naive benchmark of pure guesswork. We say “not very well” because none of the models reliably predicts the timing of crises, that is *whether* there would be crises in 1997. False alarms always outnumber appropriate warnings (except for the BP linear probit). Moreover, the statistically significant results imply that some of the models clearly outperform pure guesswork, not that they do better than the analysis of informed observers.

The head-to-head testing performed here may give insight into the nature and causes of these crises independent of the value of the models as predictors.

- All three approaches demonstrate that the probability of a currency crisis increases when domestic credit growth is high, the bilateral real exchange rate is overvalued relative to trend, and the ratio of M2 to reserves is high. All but STV also suggest that a large current account deficit is an important risk factor. Some evidence is also found for the importance of other variables, such as export growth, the size of the government budget deficit, and the share of FDI in external debt.
- With regard to the 1997 crises, it is noteworthy that some of the models make significant out-of-sample predictions despite the omission of some heavily emphasized phenomena such as poor banking supervision and weak corporate governance. It appears that the Thai crisis was relatively easy to predict, a conclusion that is consistent with Thailand’s role in setting off the crises. The Indonesian crisis was perhaps less foreseeable, lending some credence to the view that weak fundamentals may have played a relatively small role here. The models tend to do well in predicting tranquility in Mexico but have a harder time with Brazil and the Philippines, not unreasonably, but also in some cases with Argentina and Malaysia.

- The models shed some light on the question of how to measure reserve adequacy. The ratio of reserves to short-term debt, reserves to M2 and even the traditional measure of reserves to imports have received attention in the literature. We find in the context of the FR-type regressions that the best measure for predicting currency crises is reserves to M2.

The out-of-sample comparison of different approaches provides some insight into important issues in the empirical modeling of currency crises.

- The data sets used are not without anomalies and old data are frequently revised. This seems to have been an important reason for the difficulties we have encountered in reproducing some results, particularly those of KLR. The small-sample STV approach appears even more sensitive to sample and data revisions.
- Specification uncertainty may be as important as parameter uncertainty, as least for STV-type approaches, which represent a more complex specification fitted to many fewer observations. We have found that re-estimating the relatively simple (in terms of specification) KLR and FR models over different samples of countries and longer time periods has preserved the basic results of the models. In this context, it may not be surprising that the more recent applications to the Asia crisis perform much better than the original STV model itself.
- The data do not clearly support one of the basic ideas of the KLR indicator approach: that it is useful to interpret predictive variables in terms of discrete thresholds, the crossing of which is particularly significant for signaling a crisis. Both direct statistical tests and the generally superior performance of the BP linear model suggest that a better simple assumption is that the probability of crisis goes up linearly with changes in the predictive variables. There is, however, some evidence for nonlinearities of the sort assumed in KLR.⁷⁶

Where do we go from here? Implementation of an early warning system along the lines of the BP probits would pose some challenges that we have avoided here. Most importantly, we have largely ignored the problem that data on predictive variables are in many cases available only with a long lag. To take one egregious example, data on the 1996 government budget deficit are only available from the World Bank data base for 13 of 94 countries in the FR sample, even as of June 1998. Although data may be obtainable from alternative sources, these data may be sufficiently different from the series used to estimate the model that substantial errors could be introduced, as we saw to some extent with the deficit variable in

⁷⁶One apparent inconsistency in our results, comparing the probits and the KLR-based probabilities out-of-sample, is the contrast between the superior performance of the linear probit as measured by goodness of fit and the relative success of the indicator probit and the KLR-based probabilities at ranking the countries in 1997 by severity of crisis. It may be that the non-linear indicator variables are more effective at distinguishing especially severe crises from other observations, despite the overall better performance of the linear model.

the FR regressions. In any early warning system actually implemented, this problem would be much more important, because we have benefited from the year-and-a-half lag between late 1996 and the time when we carried out our forecasts.⁷⁷

These models are clearly not the last word. Recent papers have already shown that it is possible to improve the in-sample fit of models of the Asia crisis using STV-style models. Careful reparameterization, the modification of some of the explanatory variables and inclusion of further variables, changes in the definition of the interaction variables, and other modifications have led to substantially better fits. The STV-type approach presents promise as a way to understand the nature of crises after the fact. The contrast between the failure of the original STV specification to predict the 1997 crises and the much greater success of the post-Asia STV-type models may suggest caution, however, in using these types of models as early warning systems.

A variety of specification issues appear worth exploring, particularly in the context of probit-based models estimated on panel data. We have not evaluated the robustness of the results to alternative definitions of currency crisis.⁷⁸ We have not yet looked at robustness to different time periods for the panel models. We have not explored alternatives to the 24-month-ahead prediction structure of the KLR model. It may be more plausible, for example, to give more weight to more recent observations than more distant ones in calculating the probability of a crisis in a given month. While we found the simple linear formulation to be quite successful, we also found evidence of nonlinearity in the relationship between predictive variables and the probability of crisis, though not of a KLR-type step function. This deserves further exploration in a way that does not depend on prior calculation of the KLR thresholds.⁷⁹ We have maintained the KLR practice of measuring variables in terms of percentiles; this issue could also be explored.⁸⁰

⁷⁷We have also not addressed the issue of data revisions to the extent that we did not use data as reported in May 1997. This could be an important issue in practice, as suggested by incorrect estimates of Korean reserves and Indonesian short-term external debt prior to their 1997 crises.

⁷⁸Milesi-Ferretti and Razin (1998) point out the importance of this consideration for the FR approach.

⁷⁹Recall that we used the KLR method to calculate the break-points for the stepwise linear probit models. We also experimented, somewhat successfully, with more general polynomials that did not rely on knowing the break-points.

⁸⁰As a more general specification issue, we have followed the practice of attempting to predict crises that are defined as a discontinuous function of a continuous crisis index. In a typical probit application, the latent variable is unobserved, but here this is not the case. Thus, the use of probits potentially results in a loss of information and obscures the fact that the prediction of crisis could be considered to be an aspect of the general problem of predicting movements

(continued...)

A variety of alternative predictive variables could be analyzed, the most obvious being short-term external debt. Other interesting possibilities include political variables and the degree of openness of the capital account.⁸¹ We can also consider the interaction of variables, motivated by the STV insight that particular variables may increase the probability of crisis only when other variables are also creating vulnerability. An important feature of currency crises we have not yet explored is contagion, which undoubtedly plays a role at least in the timing of crises.⁸² Finally, a variety of difficult-to-measure structural factors have been central to at least recent crises. It may be possible in the future to incorporate measures of the strength of regulatory frameworks, corporate governance and other such factors.⁸³

We can be confident that future papers will predict past crises. Some of the positive results in this paper suggest that they may also be able to predict future crises.

⁸⁰(...continued)
in the exchange rate and other variables.

⁸¹Bussiere (1998) finds political variables to predict the severity of crisis in a Tornell (1998) type model.

⁸²J.P. Morgan (1998) claim some success in predicting crises with the help of a measure of contagion that depends simply on the number of crises in other countries in recent months. They conclude that closing out positions with a sufficiently high probability of crisis according to their model would have yielded much higher returns over 1992-1997 than always staying invested. They perform some out-of-sample tests as in Tornell (1998); that is they estimate over part of the sample that they used to formulate the model then fit the rest of the sample. This is not an out-of-sample test in our sense. Indeed, the higher yields they obtain derive almost entirely from strong results in 1997 (the model would have lost money most other years), a year when in retrospect contagion was clearly important.

⁸³Johnson et al. (1998) find that measures of corporate governance help explain the severity of crisis in 1997.

References

- Blanco, Herminio and Peter M. Garber, 1986, "Recurrent Devaluation and Speculative Attacks on the Mexican Peso," *Journal of Political Economy*, Vol. 94 (February), pp. 148–66.
- Bussiere, Matthieu, 1998, "Political Instability and Economic Vulnerability" (unpublished; International Monetary Fund).
- Calvo, Guillermo A. and Enrique G. Mendoza, "Mexico's balance-of-payments crisis: a chronicle of a death foretold," *Journal of International Economics*, Vol. 41 (1996), 235–64.
- Corsetti, Giancarlo, Paolo Pesenti and Nouriel Roubini, 1998, "Paper Tigers? A Preliminary Assessment of the Asian Crisis" (unpublished; Lisbon: NBER-Bank of Portugal International Seminar, June).
- Demirgüç-Kunt, Asli and Enrica Detragiache, 1998, "Monitoring Banking Sector Fragility: A Multivariate Logit Approach with an Application to the 1996–97 Banking Crises" (unpublished; Washington: International Monetary Fund).
- Diebold, Francis X. and Jose A. Lopez, 1996, "Forecast Evaluation and Combination," NBER Technical Working Paper No. 192 (Cambridge, Massachusetts, MIT Press).
- Eichengreen, Barry, Andrew K. Rose and Charles Wyplosz, 1995, "Exchange Market Mayhem: The Antecedents and Aftermath of Speculative Attacks," *Economic Policy*, Vol. 21 (October), pp. 249–312.
- Estrella, Arturo and Frederic S. Mishkin, 1998, "Predicting U.S. Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics* Vol. 80, (February), pp 45–61.
- Flood, Robert and Nancy Marion, 1998, "Perspectives on the Recent Currency Crisis Literature" NBER Working Paper No. 6380 (Cambridge, Massachusetts, MIT Press).
- Frankel, Jeffrey and Andrew Rose, 1995, "Currency Crashes in Emerging Markets: An Empirical Treatment," *Journal of International Economics*, Vol. 41 (December), pp. 351–66.
- Goldstein, Morris, 1998, "Early Warning Indicators and The Asian Financial Crisis," (unpublished; Washington: Institute for International Economics).
- International Monetary Fund, 1998, *World Economic Outlook* (Washington: International Monetary Fund, May).

- Johnson, Simon, Peter Boone, Alasdair Breach and Eric Friedman, "Corporate Governance in the Asian Financial Crisis, 1997–1998," (unpublished; Cambridge: MIT Sloan School of Management).
- Kaminsky, Graciela, 1998 "Currency and Banking Crises: A Composite Leading Indicator" (unpublished; Washington: Board of Governors of the Federal Reserve Board, February).
- _____ and Reinhart M. Carmen, 1998 "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems," International Finance Discussion Paper No. 544 (Washington: Board of the Governors of the Federal Reserve System, March).
- _____, 1998, "Financial Crises in Asia and Latin America: Then and Now," *American Economic Review* Vol. 88 (May), pp. 448–48.
- _____, Lizondo Saul and Carmen Reinhart, 1998, "Leading Indicators of Currency Crises," *Staff Papers* (Washington: International Monetary Fund) Vol. 45 (March), pp. 1–48.
- Krugman, Paul, "A Model of Balance-of-Payments Crises," 1979, *Journal of Money, Credit and Banking*, Vol. 11 (August), pp 311–25.
- Masson, Paul, "Contagion: Monsoonal Effects, Spillovers, and Jumps Between Multiple Equilibria," IMF Working Paper (forthcoming).
- Milesi-Ferretti, Gian Maria and Assaf Razin, 1998, "Current Account Reversals and Currency Crises" IMF Working Paper 98/89 (Washington: International Monetary Fund).
- Persaud, Avinash, "Event Risk Indicator Handbook," 1998, Global Foreign Exchange Research: Technical Series (London: J.P. Morgan, January 29, 1998).
- Radelet, Steven and Jeffrey Sachs, 1998a, "The Onset of the East Asian Financial Crisis," (unpublished; Boston: Harvard Institute for International Development).
- _____, 1998b, "The East-Asian Financial Crisis: Diagnosis, Remedies, Prospects" (unpublished; Boston: Harvard Institute for International Development, April).
- Sachs, Jeffrey, Aaron Tornell and Andrés Velasco, 1996, "Financial Crises in Emerging Markets: The Lessons from 1995," *Brookings Papers on Economic Activity: 1*, Brookings Institution, pp. 147–215.
- Sachs, Jeff, 1997, "What Investors Should Learn from the Crisis that Has Forced Thailand to Seek an IMF Loan," *Financial Times* (London), July 30.
- Tornell, Aaron, 1998, "Common Fundamentals in the Tequila and Asian Crises," (unpublished; Boston: Harvard University).