ELSEVIER

Journal of International Money and Finance
18 (1999) 561–586

Journal of International Money and Finance

www.elsevier.com/locate/jimonfin

Predicting currency crises: The indicators approach and an alternative

Andrew Berg, Catherine Pattillo*

Research Department, International Monetary Fund, 700 19th St. N.W., Washington D.C., 20431, USA

Abstract

In recent years, a number of researchers have claimed success in systematically predicting which countries are more likely to suffer currency crises, most notably Kaminsky, Lizondo and Reinhart (1998). This paper evaluates the KLR approach to anticipating currency crises and develops and tests an alternative. First, we try to answer the question: if we had been using the KLR model in late 1996, how well armed would we have been to predict the Asia crisis? Second, we analyze a more general probit-based model of predicting currency crises. In the process, we test several basic assumptions underlying the indicators approach. © 1999 Elsevier Science Ltd. All rights reserved.

JEL classification: F31, F47.

Keywords: Currency crises; Vulnerability indicators; Asian crisis; Balance of payments crises; Crisis prediction

1. Introduction

In recent years, a number of researchers have claimed success in systematically predicting which countries are more likely to suffer currency crises. Perhaps the most prominent model proposed before 1997 for predicting currency crises is the indicators approach of Kaminsky et al. (1998) (hereafter KLR), who monitor a large set of monthly indicators that signal a crisis whenever they cross a certain threshold. The

* Corresponding author. E-mail: cpattillo@imf.org

PII: S0261-5606(99)00024-8

authors claim some success in developing a set of indicators that reliably predict the likelihood of crisis.[1]

It may seem unlikely that currency crises should be systematically predictable. Early theoretical models of currency crises suggested that crises may, however, be predictable even with fully rational speculators (Krugman, 1979; Blanco and Garber, 1986). In 'second generation' models, 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 the relative probabilities of crises in response to a shock such as a global downturn in confidence in emerging markets.[2]

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. 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.[3] 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 focused 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. 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' crises 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.

Kaminsky (1998) asserted that this method can be applied successfully to the 1997

[1] KLR was originally issued as a working paper in 1996. 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. For example, IMF (1998); Kaminsky (1998); Radelet and Sachs (1998); Sachs (1997); Corsetti et al. (1998) and Tornell (1998). Other important pre-1997 models include Sachs et al. (1996) and Frankel and Rose (1995). The out-of-sample performance of these two, as well as KLR, is analyzed in Berg and Pattillo (1998).

[2] See Flood and Marion (1998) for a survey of this literature.

[3] Initially successful early warning systems might thus cease to work following publication.

crises.[4] However, while she presents out-of-sample estimates of the probability of currency crisis, she does not provide tests of whether these forecasts are better than, for example, guesswork. Furman and Stiglitz (1998), on the other hand, apply the KLR methodology to predicting the Asia crisis and conclude that it does not work well.

This paper evaluates the KLR approach to anticipating currency crises and develops and tests an alternative. First, we try to answer the question: if we had been using the KLR model in late 1996, how well armed would we have been to predict the Asia crisis? Second, we analyze a more general probit-based model of predicting currency crises. In the process, we can test several basic assumptions underlying the indicators approach.

The paper is organized as follows. Section 2 implements the KLR model over the pre-1997 period. We duplicate the original results as closely as possible, using where possible the original data. Next, we re-estimate the model 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. In Section 3 we apply a probit regression technique to the same data and crisis definition as in KLR. Section 4 uses both models to forecast the probability of crisis for 1997. We generate a ranking of countries according to predicted probability of crisis in 1997 for each model, then compare the predicted and actual rankings. Section 5 concludes.

2. Kaminsky–Lizondo–Reinhart (1997) signals approach

2.1. 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.[5] KLR chose 15 indicator variables based on theoretical priors and on the availability of monthly data.[6] An indicator issues a signal whenever it moves beyond a given threshold level.

[4] See also Goldstein (1998).

[5] Means, standard deviations and weights are country-specific. Weights 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%. Note that lack of data precluded the inclusion of domestic interest rates in the crisis definition.

[6] Indicators are: (1) international reserves (in $US); (2) imports (in $US); (3) exports (in $US); (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) and index of equity prices (measured in $US). The indicator is defined

We can consider the performance of each indicator in terms of the matrix below. 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. For each indicator, KLR find the 'optimal' threshold, defined as that threshold which minimizes the noise-to-signal ratio B/A.[7]

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

The thresholds are calculated in terms of the percentiles of each country's distribution for the variable in question. An optimal threshold for a given predictor, such as domestic credit growth, might be 80, for example, meaning that a signal is considered to be issued whenever domestic credit growth in a given country is in the highest 20% of observations for that country. The optimal threshold is constrained to be the same across countries. Thus, minimizing the noise-to-signal ratio for the sample of countries yields an optimal threshold percentile for each indicator that is the same for all countries. The corresponding country-specific threshold value of the underlying variable associated with that percentile will differ across countries, however.

The KLR approach is bivariate, in that each indicator is analyzed, and optimal thresholds calculated, 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 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.[8]

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).

[7] Note that KLR attach no cost to missing crises (observations of type C). KLR actually minimize what they call the 'adjusted' noise-to-signal ratio, defined as (B/(B+D)/A/(A+C)). This amounts to minimizing B/A, however, since (A+C)/(B+D) is a function of the frequency of crises in the data and does not depend on the threshold (we thank Robert Hodrick for this observation).

[8] The conditional probabilities are generated as follows:

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

where k is the sum of the weighted indicators signaling. $\mathrm{Prob}(C^{i}_{t,t+24}|k_t = j)$ 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.

2.2. Implementation

2.2.1. Reproduction of KLR results

We first attempted to reproduce the KLR results using the same 20-country, 1970–95 sample they use.[9] Following KLR, we first examined the effectiveness of the approach by determining the extent to which each individual indicator is useful in predicting crises.

Table 1 presents information on the performance of individual indicators from our reproduction. The first column 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, (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.[10]

Our results are broadly similar to those of KLR, 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 found a noise-to-signal ratio substantially below unity while we found a ratio above unity. Thus, although KLR found 12 informative indicators, that is those with noise-to-signal ratios below unity, we found only eight of these to be informative.[11]

2.2.2. Modifications

Having reproduced as nearly as we could the KLR results, we changed the sample, and tried two other indicators. We modified the sample in two ways. First, we estimated 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. Second, we changed the sample of countries: we omitted the five European countries from the sample and added other emerging market economies. This sample is more appropriate for our concern with crises in ‘emerging markets’ and also serves as an informal

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

[10] Note that the KLR approach does not lend itself to hypothesis testing; their technique gives no indication of when results are statistically significant.

[11] 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. Some of this discrepancy may come from differences in the raw data. We have found that seemingly small differences due to revisions in International Financial Statistics (IFS) data can strongly influence the results, and furthermore they and we separately ‘cleaned’ the data of errors.

Table 1
Performance of indicators

Indicator[a]	KLR sample rerun				23-Country sample,1970–95:4			
	Noise/signal (adjusted)[b] (1)	P(crisis/signal) $-P$(crisis)[c] (2)	Threshold percentile (3)	Number of crises with data (4)	Noise/signal (adjusted)[b] (5)	P(crisis/signal) $-P$(crisis)[c] (6)	Threshold percentile (7)	Number of crises with data (8)
Real exchange rate[d]	0.24	32	90	70	0.25	29	90	70
M2/reserves	0.46	16	87	69	0.39	17	87	68
Exports	0.49	15	90	69	0.60	9	90	68
International reserves	0.53	13	89	69	0.44	15	90	68
Excess M1 balances[e]	0.69	7	90	68	0.60	9	90	67
Domestic credit/GDP	0.71	6	85	66	0.78	4	84	65
Real interest rate[f]	0.74	5	82	42	0.76	4	88	38
M2 multiplier	0.80	4	84	69	1.14	−2	81	68
Imports	1.19	−3	90	69	1.16	−2	86	68
Industrial production	1.23	−3	85	58	1.14	−2	85	54
Terms of trade	1.42	−5	80	63	0.93	1	90	55
Lending rate/deposit rate[f]	1.44	−5	89	34	1.04	−1	89	29
Bank deposit	1.53	−6	90	69	1.63	−6	90	68
Stock price index	1.81	−9	87	47	1.59	−6	80	44
Real interest differential[f]	1.97	−9	82	42	1.34	−4	82	38
Current account/GDP[f]					0.45	14	90	51
M2/reserves (level)[f]					0.42	16	90	70

[a] Measured as 12-month growth rate except as indicated.
[b] 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)]).
[c] P(crisis/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 + B)] in terms of the matrix in the text). P(crisis) is the unconditional probability of a crises, (A + C)/(A + B + C + D).
[d] Deviation from deterministic trend.
[e] Residual from regression of real M1 on real GDP, inflation, and a deterministic trend.
[f] Measured in levels.

test of robustness of the KLR approach.[12] The last four columns of Table 1 show that indicator performance over the larger sample is broadly similar to results using the KLR sample. 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). In what follows, we focus on the 23-country sample estimated through April 1995.

We tried 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 focused on the level of this variable. KLR did not use the current account. We found 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 0.45 and 0.39 respectively. The current account/GDP is also highly informative, with a noise-to-signal ratio of 0.42.[13]

So far we have looked at each indicator separately. We can also, following Kaminsky (1998) calculate the weighted-sum based probabilities of crisis.[14] This produces a series of estimated probabilities of crisis for each country. These should be interpreted as the predicted probability of crisis within the next 24 months, based on the (weighted) number of indicators signaling in a given month.[15]

How good are these forecasts? For zero/one dependent variables, it is natural to ask what fraction of the observations are correctly called. A cut-off level for the predicted probability of crisis is defined such that a crisis is predicted if the estimated probability is above this threshold. The resulting goodness-of-fit data are shown in the first two columns of Table 2 for two cut-offs: 50 and 25%.[16]

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). These measures evaluate the accuracy of probability forecasts. In addition, the global squared bias (GSB) measures forecast calibration. The QPS ranges from 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.[17]

---

[12] 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.

[13] 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.

[14] Two issues regarding the treatment of missing data in the KLR framework deserve mention. A key variable is *c24*, which is defined to equal one if there is a crisis in the next 24 months. This variable is defined as long as one observation is available (either a crisis or non-crisis month) in the relevant 24 month period. Secondly, the weighted sum of indicators signaling is calculated provided that data on at least one of the indicators is available. The weighted-sum based probabilities are calculated using the same principle.

[15] Unlike Kaminsky (1998) we used only the good indicators, i.e. those with noise-to-signal ratio less than one.

[16] See Table 2 footnotes for precise definitions of ‘correctly called’ and related terms.

[17] For 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, $t + 24$ 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

What can we conclude? The first column of Table 2 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% 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%) 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% of alarms (that is, observations where the predicted probability of crisis is above 50%) are false, in that no crisis in fact ensues within 24 months. As the second column of Table 2 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%. With this lower cut-off, 41% 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% cut-off, the probability of a crisis within 24 months is 37% if there is an alarm, much higher than the unconditional probability of crisis of 16% in this sample. Now, however, 63% of alarms are false.

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.[18] A Pesaran-Timmermann[19] test rejects, at the 1% signifi-

squared error for probability forecasts is the QPS: $\text{QPS} = 1/T\sum_{t=1}^{T} 2(P_t - R_t)^2$. The analogy is rough, however, because P_t is not the forecast of the event (which is a zero/one variable) but the probability of the event. Large errors are penalized more heavily under the LPS, given by: $\text{LPS} = 1/T\sum_{t=1}^{T}[(1 - R_t)\ln(1 - P_t) + R_t\ln(P_t)]$. Overall forecast calibration is measured by the global squared bias $\text{GSB} = 2(\bar{P} - \bar{R})^2$, where $\bar{P} = 1/T\sum_{t=1}^{T} P_t$, $\bar{R} = 1/T\sum_{t=1}^{T} R_t$. Calibration compares the mean forecasted probability to the observed relative frequencies. See Diebold and Lopez (1996) for more discussion.

[18] The uninformative benchmark consists of the predictions of someone who called a crisis randomly 20% of the time, the same as the frequency of called crises in the KLR predictions using the 25% cut-off. She would have called correctly 70% of total, 20% of pre-crisis and 80% of tranquil periods. 82% of her alarms would have been false. Note that this is not the same as guessing based on the unconditional frequency of crisis. The latter is a poor benchmark because one could do better by simply always predicting the event that is most common in the sample (in our case, tranquility).

[19] Pesaran and Timmermann (1992) develop a statistic for a non-parametric test of the statistical independence of two zero/one variables (in our case, the actual and predicted pre-crisis variable). In the null hypothesis, the two variables are distributed independently but have the same mean (that is, the frequency of 'ones' is the same in the two series), as in footnote 18. Under the null hypothesis, the test statistic has a normal distribution in large samples.

The difference between the results in Table 2 and the benchmark (detailed in [18]) may seem small (for example, 77 vs 70% of observations correctly called). Note, though, that with 6680 observations, such an increase in the fraction correctly called is highly unlikely to happen by chance. One way to see this cance level, the hypothesis that the original KLR model does no better at calling observations than guesses based on the unconditional probability of crisis, using the 25% cut-off.[20]

2.2.3. 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, p. 23). Our analysis of the in-sample success of the KLR-type models suggests that 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.

As to the assessment of which variables are potentially important leading indicators, those we find useful are also so classified 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. We find fewer potentially useful indicators, though, not finding information in the M2 multiplier growth rate, growth of imports, growth of industrial production, ratio of lending to deposit rates, bank deposit growth rate, stock price index growth and the real interest differential.

3. A probit-based alternative model

3.1. Methodology

In this section, 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. 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

is to note that while even the naive benchmark gets most of the tranquil periods right, the increase in performance in the relatively rare pre-crisis observations (41% correct vs 20% in the benchmark) is much more striking.

[20] With the 50% cut-off, the null hypothesis is also rejected at the 1% level.

Table 2
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 of 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 of observations correctly called	82	83	85	84	85
Percent of pre-crisis periods correctly called[a]	9	9	16	7	19
Percent of tranquil periods correctly called[b]	98	99	99	100	98
False alarms as percent of total alarms[c]	44	30	29	11	34
Goodness-of-fit (cut-off probability of 25%)					
Percent of observations correctly called	77	75	81	78	80
Percent of pre-crisis periods correctly called[a]	41	46	44	48	47
Percent of tranquil periods correctly called[b]	85	81	89	84	87
False alarms as percent total alarms[c]	63	65	57	63	59

[a] 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.
[b] 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.
[c] A false alarm is an observation with an estimated probability of crisis above the cut-off (an alarm) not folowed by a crisis within 24 months.

test for the statistical significance of individual variables and the constancy of coefficients across time and countries.[21]

Fig. 1 presents various possible relationships between the probability of crisis (on the vertical axis) and the value of a variable *P*, measured, as in KLR, in percentiles (on the horizontal axis).[22] The KLR assumption, in terms of Fig. 1, is that α_1 and α_3 are zero while α_2 is equal to 1. That is, they 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 one at or above the threshold. 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.

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. For each indicator we estimate equations of the form:

$$\text{Prob}(c = 1) = \text{f}(0 + \alpha_1 p(x) + \alpha_2 I + \alpha_3 I(p(x) - T)), \tag{1}$$

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.[23] Thus, α_1, α_2, and α_3 in Eq. (1) correspond to the α's in Fig. 1. We use the

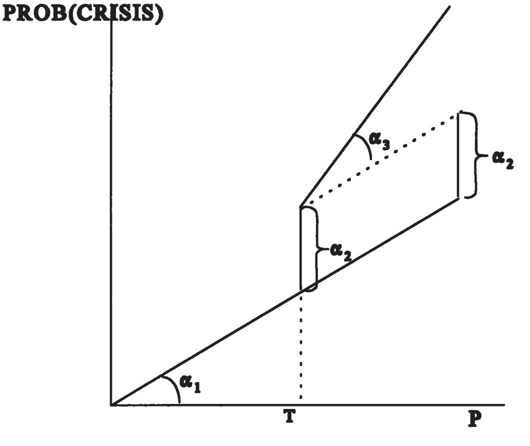

Fig. 1. Relationship between predictive variable and probability of crisis.

[21] We ignore the potential correlations in the errors, both across countries at a point in time as well as the serial correlation 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.

[22] *P* is measured in percentiles in that the observations on the underlying predictive variable for a given country, for example, 12 month percent changes in real domestic credit, are expressed in terms of percentiles of the distribution of that variable for the country in question.

[23] The probit models are estimated over the 1970:1–1995:4 period.

thresholds T calculated from the KLR algorithm, since we are interested primarily in testing their approach against a more general alternative.[24]

Estimates of Eq. (1) for deviations of the real exchange rate from trend, an important predictive variable, indicate that its relationship with the probability of crisis is of the general form shown in Fig. 1, linear with a jump at the threshold and a higher slope thereafter. The choppy line in Fig. 2 represents the fraction of times the observation of a given percentile for RER deviations is followed by a crisis within 24 months in the pooled data. The other line represents the estimated relationship 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.

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.[25]

Multivariate probits are the natural extension to the bivariate probits discussed so far. Table 3 presents estimates of three probit models that explain whether a crisis occurs in the next 24 months (hereafter designated BP models).[26] 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.[27] Model 3 is the result of a simplification starting with the most general piecewise-linear specification for all the vari-

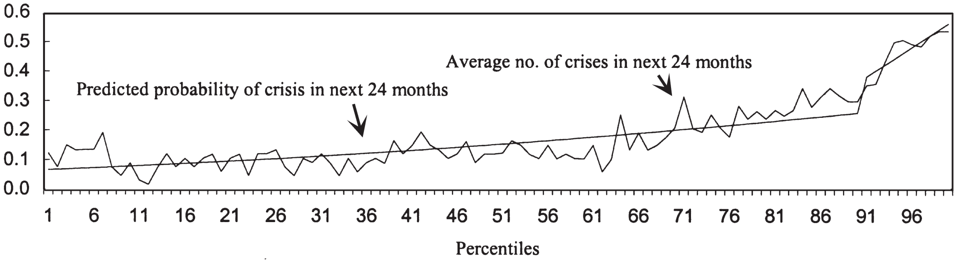

Fig. 2. Average no. of crises in next 24 months by percentile of variable real exchange rate deviations.

[24] 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.

[25] The pure 'indicators' hypothesis ($\alpha_1 = \alpha_3 = 0$) is generally rejected at well below the 0.01% significance level.

[26] We 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. The probit methodology is less forgiving of missing observations for explanatory variables than the KLR methodology (see footnote 14).

[27] For 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% level.

Table 3
Multivariate probit models[a]

	Model 1. Indicator		Model 2. Linear		Model 3. Piecewise-linear		
Variable[b]	Coefficient	*T*-statistic	Coefficient	*T*-statistic	Variable[c]	Coefficient	*T*-statistic
Real exchange rate deviations	0.31564	15.89	0.00232	13.50	Real exchange rate deviations (*I*)	0.12057	3.35
Current account	0.10461	5.63	0.00178	9.50	Real exchange rate deviations (*P*)	0.00114	5.58
Reserve Growth			0.00128	6.20	Real exchange rate deviations ($I^*(P-T)$)	0.01078	2.42
Export Growth	0.05686	3.24	0.00064	3.65	Current account (*I*)	0.06328	1.77
M2/reserves	0.08561	4.37	0.00053	2.80	Current account (*P*)	0.00189	8.67
M2/reserves growth	0.12770	7.37			Current account ($I^*(P-T)$)	−0.01313	−2.38
					Reserve growth (*P*)	0.00107	4.74
					Export growth (*P*)	0.00070	3.94
					M2/reserves (*I*)	0.06372	3.47
					M2/reserves growth (*I*)	0.03960	2.33
Sample size	4928		5025		Sample size	4928	
Log-likelihood	−1958		−1970		Log-likelihood	−1858	
Pseudo-R^2	0.099		0.115		Pseudo-R^2	0.145	

[a] 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. Model 3 started with the most general piecewise-linear specification for all the variables (allowing the estimation, for each variable, of the slope below the threshold (*P*), the jump at the threshold (*I*), and the slope above the threshold ($I^*(P-T)$) and simplified to the most parsimonious representation of the data.

[b] All coefficients should be positive since variables such as reserve growth, export growth, and real exchange rate deviations from trend have been multiplied by −1.

[c] α_1 is the coefficient of *P*, α_2 is the coefficient on *I*, the zero/one indicates variable, and α_3 is the coefficient on $I^*(P-T)$, and represents additional effect of changes in the variable above the threshold.

ables. 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 threshold, we used a general-to-specific procedure to simplify to the most parsimonious representation of the data.[28]

Model 1 of Table 3 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.[29] 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.

3.2. In-sample performance

How well do the different models perform? The results in Tables 2 and 3 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 generally 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.[30]

[28] We 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. On the general-to-specific approach see Ericsson et al. (1991) and Pagan (1987).

[29] Note 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.

[30] Pesaran-Timmermann tests reject the null of noninformative forecasts at well below the 1% level of significance for all the probit models. A Davidson and MacKinnon encompassing test of the non-nested linear and indicator probits shows that neither encompasses the other.

4. Predicting 1997

4.1. 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 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.

We have examined the performance of each individual indicator in 1996 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.[31] 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 4 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% prior to actual crises. The goodness-of-fit rows show that only 4% of the time was the predicted probability of crisis above 50% in cases when there was a crisis within the next 24 months, during the 1995:5 to 1997:12 period.[32] As before, we may be interested in using a lower cut-off probability to define a crisis. Table 4 shows that the Kaminsky (1998) probability estimates are above 25% in 25% of the pre-crisis observations. As we observed in-sample, most alarms are false at the 25% 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% 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% cut-off) is 40%, which is somewhat higher than the unconditional probability

[31] Tables are available upon request.

[32] At the time of collection, April 1998, data were not available for December 1997 for all countries. More recent analyses have confirmed that the results reported here both for the KLR and BP models are not significantly effected by the addition of complete data for 1997 (some partial exceptions are noted below).

Table 4
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 of observations correctly called	74	73	78	78	78
Percent of pre-crisis periods correctly called[a]	4	0	2	0	5
Percent of tranquil periods correctly called[b]	100	100	99	100	98
False alarms as percent of total alarms[c]	17	No crisis called	50	No crisis called	56
Goodness-of-fit: (cut-off probability of 25%)					
Percent of observations correctly called	69	69	76	79	76
Percent of pre-crisis periods correctly called[a]	25	32	16	80	48
Percent of tranquil periods correctly called[b]	85	83	93	79	84
False alarms as percent of total alarms[c]	63	60	61	49	54

[a] 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.
[b] A tranquil period is correctly called when the estimated probability of crisis is below the cut-off probability and no crisis ensues within24 months.
[c] 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.

of 27%. A Pesaran-Timmermann test rejects the hypothesis that the forecasts are uninformative at the 1% level of significance.[33]

So far we have examined the ability of the models to predict the approximate *timing* of crises for each country.[34] 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 5 shows countries' actual crisis index and predicted probability of crisis in 1997 for the various different forecasting methods.[35] 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.[36]

The KLR-based forecasts are 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% of the variance is explained. The addition of the current account and the level of M2/reserves brings the R^2 up to 36%.

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. The overall goodness-of-fit for the out-of-sample predictions illustrates 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 models.

4.2. A 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 3, then extend the explanatory vari-

[33] The test statistic cannot reject the null that the forecast is uninformative for the 50% cut-off, as no crises are called. See footnote 19 for a discussion of the test.

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

[35] The predicted crisis probability is the average of the probabilities during 1996:1–12, using the out-of-sample estimates. Averaging over, for example, 1996:7 to 1996:12 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.

[36] 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 5
Correlation of actual and predicted rankings based on KLR approach

	Actual		KLR weighted-sum of indicators[b]				BP probit models[d]					
	Crisis index[a]	Rank	Original specification		Augmented[c]		Indicator		Linear		Piecewise-linear	
Country			Probability	Rank	Probability	Rank	Probability	Rank	Probability	Rank	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

Continued

Table 5
Continued

	Actual		KLR weighted-sum of indicators[b]				BP probit models[d]					
	Crisis index[a]	Rank	Original specification		Augmented[c]		Indicator		Linear		Piecewise-linear	
Country			Probability	Rank	Probability	Rank	Probability	Rank	Probability	Rank	Probability	Rank
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	0.2	48
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[e]				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

[a]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 eviation. Values above three are defined as a crisis and are shown in bold.
[b]Based on average of noise-to-signal weighted probabilities from during 1996:1–12, using out-of-sample estimates.
[c]Augmented with the inclusion of the current account and M2/reserves in levels.
[d]All probit models probabilities are average predicted probabilities for 1996:1–12, where model was estimated up to 1995:4.
[e]Spearman Rank Correlation of the fitted values and the actual crisis index and its *P*-value. The R^2 is from a regression of fitted values on actual values.

ables to generate predictions for the period 1995:5–1997:12.[37] The estimated probabilities can be evaluated using the probability scores and goodness-of-fit measures discussed above.

Table 4 shows that on all the scoring measures, the probits perform better than the probabilities based on the weighted-sum of indicators signaling.[38] 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% cut-off. Using the looser standard whereby a probability of crisis above 25% 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% in 80% of the periods that precede a crisis. 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% 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% cut-off) is 51%, much higher than the unconditional probability of 22%.[39]

The linear model performs much better out-of-sample than the more general piecewise-linear model that includes 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 5 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 model that had the best goodness-of-fit has the lowest, though still significant, correlation with the actual ranking.[40]

[37] The probit-based probabilities are derived from the models in Table 3. Again, we did not have complete data for December 1997. The results are not significantly changed by the inclusion of more recent data, except as noted below.

[38] An exception is that the indicator probit has a higher GSB than the KLR-based probabilities. As described in [17], the scores measure the total size of the errors, similar to the mean squared error in ordinary least squares. Lower scores are better.

[39] These predictions are also statistically significantly better than uninformed guesses at the 1% level.

[40] The contrast between the results of the rankings and goodness-of-fit comparisons is somewhat surprising but not inexplicable. 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

We can flesh out these results by examining the performance of the linear probit in predicting crises for a sub-sample of four crisis and one non-crisis country in 1997 (Table 6).[41] 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% for several months in 1996, and in Malaysia the probabilities were above 30%. The probabilities are also reasonably high for Indonesia, ranging from 25 to 28%, while the model is somewhat less successful for Korea, where the estimated probability of crisis was between 20 and 33%.[42] Turning to Brazil, a non-crisis country during this period, the probabilities ranged from 25 to 37%.[43]

4.3. Summary out-of-sample assessment

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, among the probits, the linear specification performs best in terms of the probability scores and goodness-of-fit. This suggests that the superior in-sample performance of the piecewise-linear specification may have reflected 'overfitting.' Third, we can compare the BP probit-based alternatives to the KLR probabilities. The KLR forecasts perform better than some of the probits on some of the measures, so this comparison is not unambiguous. Overall, though, the probits seem to work better. In particular the linear specification has much better scores and goodness-of-fit.

5. Conclusion

This paper has examined the extent to which the KLR indicators model, originally formulated and estimated prior to 1997, would have helped predict the 1997 currency crises. We have also compared the predictions of this model with a probit-based alternative, which we dub the BP model.

---

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%. The addition of December 1997 data (not available as of April 1998 when the data for these results were collected) reverses the order of the ranking correlations, with the linear BP model performing somewhat better than KLR.

[41] 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.

[42] Probits that excluded the current account largely failed to predict a crisis in Indonesia.

[43] Berg and Pattillo, 1998, also present detailed results for non-crisis countries Mexico and Argentina. Both have crisis probabilities below 30%, with Argentina's well below 20% for most of 1996 and 1997.

Table 6
Summary measures for selected countries: Asian countries

Date	Indonesia				Korea				Malaysia				Thailand				Brazil			
	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]
		Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]	
1995:01	1(9)	11	16	14	2(10)	42	36	16	2(9)	30	26	28	0(10)	11	10	15	3(10)	55	36	26
1995:02	1(9)	11	16	15	1(10)	30	26	16	2(9)	30	26	28	0(10)	11	10	17	3(10)	55	36	27
1995:03	1(9)	11	16	15	1(10)	30	26	18	2(9)	30	26	31	0(10)	11	10	20	3(10)	55	36	28
1995:04	1(9)	11	16	17	1(10)	30	26	18	2(9)	30	26	34	0(10)	11	10	23	3(10)	55	36	37
1995:05	1(9)	11	16	16	1(10)	30	26	19	2(9)	30	26	36	0(10)	11	10	21	2(10)	42	36	35
1995:06	1(9)	11	16	17	1(10)	30	26	19	2(9)	30	26	37	0(10)	11	10	22	2(10)	42	36	36
1995:07	0(9)	11	10	23	1(10)	30	26	17	2(9)	30	26	36	0(10)	11	10	26	1(10)	30	26	33
1995:08	1(9)	11	16	24	1(10)	30	26	18	2(9)	30	26	35	0(10)	11	10	24	1(10)	30	26	31
1995:09	1(9)	11	16	23	1(10)	30	26	18	2(9)	30	26	35	0(10)	11	10	27	1(10)	30	26	31
1995:10	1(9)	11	16	27	1(10)	30	26	19	2(9)	30	26	36	0(10)	11	10	30	2(10)	42	36	28
1995:11	1(9)	11	16	28	1(10)	30	26	19	2(9)	30	26	35	0(10)	11	10	28	2(10)	42	36	28
1995:12	1(9)	11	16	26	2(10)	42	36	23	1(9)	21	16	34	0(10)	11	10	30	2(10)	42	36	28
1996:01	1(10)	11	16	28	1(10)	30	26	19	2(9)	30	26	38	1(10)	11	16	30	2(10)	42	36	25
1996:02	1(10)	16	16	27	1(10)	30	26	20	1(9)	21	16	40	1(10)	11	16	30	2(10)	42	36	25
1996:03	1(10)	11	16	26	1(10)	30	26	22	1(9)	21	16	33	2(10)	16	28	32	2(10)	42	36	26
1996:04	1(10)	11	16	24	1(10)	30	26	23	1(9)	21	16	41	1(10)	11	16	34	2(9)	42	36	NA
1996:05	1(10)	11	16	25	1(10)	30	26	22	1(8)	21	16	NA	1(10)	11	16	36	2(9)	42	36	NA
1996:06	1(10)	11	16	26	1(10)	30	26	24	1(8)	21	16	NA	2(9)	16	28	40	1(9)	30	26	NA
1996:07	1(9)	11	16	NA	2(10)	42	36	28	1(8)	21	16	NA	1(9)	11	16	40	1(9)	30	26	NA
1996:08	1(9)	11	16	NA	1(10)	16	16	29	0(8)	11	10	NA	1(9)	11	16	41	1(9)	30	26	NA
1996:09	1(9)	11	16	NA	2(10)	21	16	33	0(8)	11	10	NA	2(9)	16	28	42	1(9)	30	26	NA

Table 6
Continued

Date	Indonesia				Korea				Malaysia				Thailand				Brazil			
	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]	No. of good indicat. signal-ing[a]	Weighted-sum prob.		Probit based prob.[d]
		Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]			Original spec.[b]	Extra vars.[c]	
1996:10	1(9)	11	16	NA	1(10)	16	16	30	0(8)	11	10	NA	1(9)	11	16	41	1(9)	30	26	NA
1996:11	1(9)	11	16	NA	1(10)	16	16	33	0(8)	11	10	NA	2(9)	16	28	42	2(9)	42	36	NA
1996:12	0(9)	11	10	NA	1(10)	16	16	27	0(4)	11	10	NA	1(9)	11	16	43	2(9)	42	36	NA
1997:01	0(9)	11	10	NA	2(10)	21	16	29	0(4)	11	10	NA	1(9)	11	16	40	2(9)	42	36	NA
1997:02	0(9)	11	10	NA	2(10)	21	16	33	0(3)	11	10	NA	2(9)	16	28	44	2(9)	42	36	NA
1997:03	0(9)	11	10	NA	2(10)	21	16	28	0(3)	11	10	NA	1(9)	11	16	43	2(9)	42	36	NA
1997:04	0(9)	11	10	NA	1(10)	16	16	27	0(3)	11	10	NA	0(8))	11	10	NA	2(8)	42	36	NA
1997:05	0(9)	11	10	NA	1(10)	16	16	NA	0(3)	11	10	NA	2(8)	30	26	NA	2(8)	42	36	NA
1997:06	0(9)	11	10	NA	1(9)	16	16	NA	0(3)	11	10	NA	2(7)	30	26	NA	2(8)	42	36	NA
1997:07	0(9)	11	10	NA	1(9)	16	16	NA	1(3)	21	16	NA	2(7)	32	28	NA	3(9)	42	36	NA
1997:08	1(9)	16	16	NA	1(9)	16	16	NA	1(3)	21	16	NA	3(7)	42	36	NA	3(9)	42	36	NA
1997:09	1(7)	16	16	NA	0(9)	11	10	NA	1(3)	21	16	NA	2(6)	32	28	NA	3(9)	55	36	NA
1997:10	1(7)	16	16	NA	1(9)	16	16	NA	1(3)	21	16	NA	2(6)	32	28	NA	4(9)	55	36	NA
1997:11	1(7)	16	16	NA	2(9)	32	28	NA	1(2)	21	16	NA	1(1)	21	16	NA	1(5)	30	26	NA
1997:12	0(2)	11	10	NA	1(4)	21	16	NA	0(0)	NA	NA	NA	1(1)	21	16	NA	1(4)	30	26	NA

[a]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.
[b]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.
[c]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.
[d]Predicted probabilities of crisis from a probit regression of impending crisis on the indicator variables measured linearly in percentiles.

The KLR-based probabilities of crisis have some predictive value out-of-sample. When this model issued an alarm during the 1995:5 to 1996:12 period, a crisis would actually have followed in 1997 37% of the time.[44] This compares to a 27% 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%. The addition of two variables to the KLR model, the level of the current account and M2/reserves, improves performance somewhat.[45]

We also estimated a set of alternative models (BP probit-based models) using the data and crisis 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, though not unambiguously, 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% of the time.

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

- The alternative method reproduces most of the KLR conclusions regarding which variables are important predictors of crisis. In particular, both approaches demonstrate that the probability of a currency crisis increases when the bilateral real exchange rate is overvalued relative to trend, reserve growth and export growth are low, and the growth of M2/reserves is high. Our analysis suggests, in addition to KLR, that a large current account deficit and a high ratio of M2 to reserves are important risk factors.
- With regard to the 1997 crises, it is noteworthy that both models make significant out-of-sample predictions despite the omission of some heavily emphasized phenomena such as poor banking supervision and weak corporate governance.

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

---

[44] An alarm here is defined as a predicted probability above 25%. These alarms are significant predictors of crises at the 1% level.

[45] Furman and Stiglitz (1998) apply the KLR methodology to predicting the Asia crisis and conclude that it does not work well, noting some success but also many false positives. They dismiss what success they do observe largely on the argument that the method of measuring predictive variables in terms of percentiles is biased in favor of predicting crises in countries that have little volatility in predictive variables. For example, even a relatively small real exchange rate appreciation results in a large percentile deviation in relatively tranquil countries, such as the Asia crisis countries. The KLR model does not, however, tend to systematically overpredict crises in-sample in tranquil countries or in the Asian crisis countries, contrary to their hypothesis.

cal 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. While a more complicated piecewise-linear specification does better in-sample, its poorer out-of-sample performance suggests that this may reflect 'overfitting.' There is, however, some evidence for nonlinearities of the sort assumed in KLR.

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.[46] These models are clearly not the last word. A variety of specification issues appear worth exploring, particularly in the context of probit-based models estimated on panel data. A variety of alternative predictive variables could also be analyzed, the most obvious being fiscal policy and short-term external debt. Other interesting possibilities include political variables and the degree of openness of the capital account.[47]

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 help predict future crises.

Acknowledgements

We would like to thank, without implication, Graciela Kaminsky and Carmen Reinhart for help reproducing and interpreting their results, Brooks Calvo, and Nada Mora for superb research assistance, Eduardo Borensztein, Hali Edison, Robert Hodrick, Steve Kamin, Hashem Pesaran, many IMF colleagues, and participants and discussants in the Journal of International Money and Finance/Fordham University conference on the Asia Crisis for useful comments.

References

Berg, A., Pattillo, C., 1998. Are currency crises predictable? A test. IMF Working Paper 98/154. International Monetary Fund.

Blanco, H., Garber, P., 1986. Recurrent devaluation and speculative attacks on the Mexican peso. Journal of Political Economy 94, 148–166.

Bussiere, M., 1998. Political instability and economic vulnerability. International Monetary Fund, unpublished.

Corsetti, G., Pesenti, P., Roubini, N., 1998. Paper tigers? A preliminary assessment of the Asian crisis. NBER-Bank of Portugal International Seminar, unpublished.

[46] We have also not addressed the issue of data revisions and other measurement problems. These could be important issues in practice, as suggested by misleading or incorrect estimates of Korean reserves and Indonesian short-term external debt prior to their 1997 crises.

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

Diebold, F.X., Lopez, J.A., 1996. Forecast evaluation and combination. Technical Working Paper No. 192, National Bureau of Economic Research, 1996.
Ericsson, N.R., Campos, J., Tran, H., 1991. PC-GIVE and David Hendry's Econometric Methodology. International Finance Discussion Paper 406, Board of the Governors of the Federal Reserve System.
Flood, R., Marion, N., 1998. Perspectives on the recent currency crisis literature. Working Paper No. 6380, National Bureau of Economic Research, 1998.
Frankel, J., Rose, A., 1995. Currency crashes in emerging markets: an empirical treatment. Journal of International Economics 41, 351–366.
Furman, J., Stiglitz, J.E., 1998. Economic crises: evidence and insights from East Asia. World Bank, unpublished.
Goldstein, M., 1998. Early warning indicators and the Asian financial crisis. Institute for International Economics, unpublished.
Kaminsky, G., 1998. Currency and banking crises: a composite leading indicator. Board of Governors of the Federal Reserve System, unpublished.
Kaminsky, G., Reinhart, C., 1996. The twin crises: the causes of banking and balance-of-payments problems. International Finance Discussion Paper 544, Board of the Governors of the Federal Reserve System.
Kaminsky, G., Lizondo, S., Reinhart, C., 1998. Leading indicators of currency crises. International Monetary Fund Staff Papers 45, 1–48.
Krugman, P., 1979. A model of balance-of-payments crises. Journal of Money, Credit and Banking 11, 311–325.
Pagan, A., 1987. Three econometric methodologies: a critical appraisal. Journal of Economic Surveys 1, 3–24.
Pesaran, M.H., Timmermann, A., 1992. A simple nonparametric test of predictive performance. Journal of Business and Economic Statistics 10, 461–465.
Radelet, S., Sachs, J., 1998. The East-Asian financial crisis: diagnosis, remedies, prospects. Harvard Institute for International Development, unpublished.
Sachs, J., 1997. What investors should learn from the crisis that has forced Thailand to seek an IMF loan. Financial Times (London), July 30.
Sachs, J., Tornell, A., Velasco, A., 1996. Financial crises in emerging markets: the lessons from 1995. Brookings Papers on Economic Activity 1, 147–215.
Tornell, A., 1998. Common fundamentals in the tequila and Asian crises. Harvard University, unpublished.
IMF, 1998. World Economic Outlook, 1998. A survey by the staff of the International Monetary Fund, financial surveys, International Monetary Fund, May 1998.