

WP/03/18

# IMF Working Paper

---

## Anticipating Arrears to the IMF: Early Warning Systems

*Chikako Oka*

**IMF Working Paper**

Treasurer's Department

**Anticipating Arrears to the IMF:  
Early Warning Systems**

Prepared by Chikako Oka<sup>1</sup>

Authorized for distribution by Jianhai Lin

January 2003

**Abstract**

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

This paper attempts to predict the incidence of arrears to the International Monetary Fund (IMF) by modifying and applying two of the major early warning systems for currency crises: the “signals” approach proposed by Kaminsky, Lizondo, and Reinhart (1997) and the probit-based alternative developed by Berg and Pattillo (1998). The results, based on both in-sample and out-of-sample tests, appear encouraging. While the unique nature of IMF arrears poses some challenges, the models could be useful tools for identifying countries at high risk of incurring arrears to the IMF.

JEL Classification Numbers: F33, F34, F47

Keywords: Arrears, credit risk, default, early warning

Author's E-Mail Address: [coka@imf.org](mailto:coka@imf.org)

---

<sup>1</sup> The author would like to thank, without implication, Birgir Arnason, Lynn Aylward, Pierre Beynet, Eduard Brau, Benedicte Vibe Christensen, Louis Erasmus, Jiro Honda, Magda Kandil, Robin Kibuka, Michael G. Kuhn, Jianhai Lin, Catherine Pattillo, Helga Treichel, Iqbal Zaidi, and seminar participants at the IMF for their helpful comments. All remaining errors are mine.

Contents	Page
I. Introduction .....	3
II. Overview of the Literature.....	4
III. Selection of Explanatory Variables... ..	6
IV. Kaminsky-Lizondo-Reinhart (1997) Signals Approach.....	9
A. Methodology .....	9
B. Implementation.....	13
V. Berg and Pattillo (1998) Probit-Based Alternative .....	20
A. Methodology .....	20
B. Implementation.....	21
VI. Out-of-Sample Performance.....	24
A. Zimbabwe (2001- present).....	25
B. Case B (1999).....	28
C. Comments.....	29
VII. Further Work .....	30
VIII. Conclusions.....	32
References.....	34
<b>Boxes</b>	
1. Different Measures of Efficiency.....	14
2. Composite Indicator and Conditional Probability .....	17
<b>Figure</b>	
1. Out-of-Sample Performance .....	26
<b>Tables</b>	
1. Performance of Indicators: Loss Function and Noise-to-Signal Methods.....	16
2. In-Sample Performance of Signals Approach.....	19
3. Estimation Output of Probit Model .....	22
4. In-Sample Goodness of Fit: Probit Regression and Signals Approach .....	23
5. Performance of Indicators: Out-of-Sample Cases .....	27

## I. INTRODUCTION

In recent years, a number of studies have developed early warning systems to predict currency and banking crises in emerging market economies. However, there have been relatively fewer studies that attempt to systematically predict sovereign debt default.<sup>2</sup> In particular, there is a dearth of studies of payment difficulties encountered by countries with limited or no access to international capital markets. These countries mainly rely on official creditors, including the International Monetary Fund (IMF), for external financing. As part of the effort to safeguard the resources of the IMF, there is a need to monitor payment risks of borrowing countries.

Payment arrears became a matter of serious concern to the IMF in the mid-1980s, a few years after the onset of the debt crisis. The number of countries that failed to repay the IMF on time peaked at 63 in 1986, and the number of countries in protracted arrears to the IMF reached 13 during the period 1988-90.<sup>3</sup> Since the mid-1990s, both the number of countries in arrears to the IMF and the level of overdue payments have declined steadily. In early 2001, Zimbabwe fell into arrears to the IMF and subsequently became the first country that has incurred protracted arrears to the IMF since 1995.<sup>4</sup>

The issue of payment risks to the IMF has gained importance in recent years owing to the increasing concentration of IMF lending to a few countries. The size of IMF loans has increased dramatically since the mid-1990s, mainly due to a sharp increase in the number and the magnitude of currency crises. Consequently, the exposure of the IMF to a handful of crisis-hit countries has increased significantly, which, in turn, raises the issue of safeguarding resources of the IMF.

There has been—to the author's knowledge—only one empirical study of countries' repayment performance vis-à-vis the IMF. Aylward and Thorne (1998) (hereafter AT) use a logit model, with arrears to the IMF as the binary dependent variable, and show that IMF-specific financial ratios and a few macroeconomic indicators are statistically significant in determining the likelihood of a country incurring arrears to the IMF. While AT identify useful indicators and examine the characteristics of arrears to the IMF, they do not seek to predict the timing of such arrears.

This paper attempts to predict arrears to the IMF by modifying and applying two of the major early warning systems developed for currency crises: the "signals" approach developed by

---

<sup>2</sup> Recent studies in this area include Detragiache and Spilimbergo (2001).

<sup>3</sup> The IMF considers arrears protracted when their duration exceeds six months.

<sup>4</sup> As of end-November 2002, six countries were in protracted arrears to the IMF: The Islamic State of Afghanistan, Iraq, Liberia, Somalia, Sudan, and Zimbabwe.

Kaminsky, Lizondo, and Reinhart (1997) (hereafter KLR) and the probit-based alternative by Berg and Pattillo (1998) (hereafter BP). The paper is organized as follows. The next section reviews the study by AT, and briefly discusses the work by KLR and BP. Section III discusses explanatory variables used for this study. Sections IV and V present the methodologies and in-sample results of the KLR signals approach and the BP probit model, respectively. Section VI examines how these two models perform in the out-of-sample test. Section VII discusses possible further work, and Section VIII presents the conclusions of the paper.

## II. OVERVIEW OF THE LITERATURE

### Aylward and Thorne (1998)

In their pioneering study of arrears to the IMF, AT use a logit model to examine which IMF-specific indicators and macroeconomic variables are statistically significant in determining the likelihood of a country incurring arrears to the IMF. AT cover 138 developing countries that had debt service obligations to the IMF during the period 1976–93. Their dependent variable is a binary variable that takes the value one if a country was in arrears to the IMF in year  $t$  and zero otherwise.<sup>5</sup> All the independent variables are lagged one year to reflect the assumed direction of causality. They find the following indicators to be statistically significant: debt outstanding to the IMF relative to quota, the lagged dependent variable, per capita income, economic growth, inflation, and, to a lesser extent, the reserve-imports ratio, and the stock of arrears to non-IMF creditors in relation to total external debt.

The lagged dependent variable, which is an indicator of repayment history vis-à-vis the IMF, is found to have been the overriding explanatory factor, suggesting that arrears tend to persist over a relatively long period. When the lagged dependent variable is excluded from the model, the following ratios in addition to those mentioned above are found to be significant: debt service to the IMF in relation to exports, and the ratios of exports, imports, and government revenue and expenditure to GDP.

AT's findings provide support for the state dependency of repayment performance. They show that factors that contribute to the likelihood of a country incurring arrears to the IMF are different, depending on whether a country has a track record of timely repayments or whether it has incurred late payments to the IMF in the past. Under the entry model, which excludes the data points for the years during which a country had overdue obligations at the time, the following indicators are found to be significant: debt outstanding to the IMF relative to quota, per capita income, economic growth, inflation, and the ratio of reserves to imports. Under the exit model, using the subset of observations where the country was

---

<sup>5</sup> AT count arrears when a country is in continuous arrears for three months or longer, including at least part of year  $t$ .

already overdue to the Fund, only IMF debt outstanding relative to quota and the stock of arrears in relation to total external debt are found to be significant.

In comparing the determinants of IMF arrears versus arrears to creditors other than the IMF, AT find that there is an important difference in terms of which variables have a greater explanatory power. The results indicate that the macroeconomic correlations of IMF arrears reflect a country's overall economic circumstances, whereas those of non-IMF arrears are more closely related to the availability of financial resources. More specifically, IMF arrears are closely associated with per capita income, economic growth, and inflation. On the other hand, arrears to creditors other than the IMF are more closely correlated with the reserve-imports ratio and the ratios of exports, government expenditure, and revenue to GDP. AT suggest that this result might be an indication of the IMF's preferred creditor status: a country facing financial difficulty might slow or cease to repay other creditors, but would still repay the IMF. As macroeconomic problems become insurmountable, however, reflected in low per capita income, slow economic growth, and high inflation, a country eventually incurs arrears to the IMF.

While AT have made a valuable contribution to the analysis of the factors determining countries' repayment performance vis-à-vis the IMF, they do not deal with the timing or probability of arrears. This paper seeks to fill this gap by applying the techniques developed by KLR and by BP to predict currency crises.

#### **Kaminsky-Lizondo-Reinhart (1997)**

KLR propose a nonparametric approach, the signals approach, to predict currency crises. The approach is designed to monitor a number of variables that tend to behave differently prior to currency crises compared to their mean during noncrisis periods, and that way issue warning signals whenever they cross certain thresholds. The thresholds, expressed in percentiles of the distribution of observations, are determined by minimizing the ratio of noise to signal.

The authors claim some success in identifying a group of indicators that are useful in predicting currency crises. They assert in subsequent work that the signals approach could have successfully predicted the Asian crisis in 1997.<sup>6</sup>

#### **Berg and Pattillo (1998)**

BP extend the work by KLR through embedding the signals approach in a multivariate probit framework. Unlike the signals approach, which treats each explanatory variable independently, BP estimate the combined effect of all variables. BP find that the best-performing model is parsimonious as only the following five indicators are found to be

---

<sup>6</sup> Goldstein, Kaminsky, Reinhart (2000).

significant: short-term debt relative to reserves, the real effective exchange rate, the ratio of current account to GDP, change in reserves, and export growth.

Furthermore, BP show that the probit model generally performs better than the KLR signals approach, suggesting that the simple assumption that the probability of crisis goes up in proportion to the changes in variables is better than the step-function implied in the KLR signals approach.

### III. SELECTION OF EXPLANATORY VARIABLES

The selection of indicators used in this study is influenced by AT, as well as the related credit risk literature.<sup>7</sup> In addition, data availability is an important constraint, which precludes the inclusion of certain variables. The variables of interest are grouped into six different categories: debt profile, credit history, availability of resources, policy factors, economic conditions, and political risk.

#### **Debt profile**

This study uses debt outstanding to the IMF relative to quota and debt service to the IMF relative to exports as the key elements of a country's debt profile. The ratio of IMF debt outstanding to quota measures a country's total use of IMF credit relative to its IMF quota. AT find that this is the most significant determinant except for the lagged dependent variable. As AT suggest, it is likely that this IMF-specific indicator represents various aspects of a country's balance of payments problems since a country requests IMF-supported programs only when it faces balance of payments difficulties.

Debt service to the IMF relative to exports is often used to assess a country's capacity to repay the IMF. Debt service to the IMF includes all principal and interest obligations falling due from a country to the IMF. AT find that the inclusion of this variable renders the general debt service ratio (debt service to all external creditors/exports) statistically insignificant. One possible explanation offered by AT is that, given the IMF's preferred creditor status, the likelihood that the IMF will not be repaid is less sensitive to traditional key indicators of creditworthiness.

#### **Credit history**

Countries' repayment history has been found to be one of the strongest determinants of their future repayment performance, as confirmed by AT and many other studies.<sup>8</sup> AT use the ratio of arrears to total external debt and the lags of two dependent variables (arrears to the IMF

---

<sup>7</sup> See AT for a comprehensive review of the literature on external debt repayment problems.

<sup>8</sup> See, for instance, Hajivassiliou (1989); McFadden and others (1985).

and total external arrears) as credit history indicators. However, as this study employs an entry model, which excludes the data points during the years of arrears, it cannot test how the state of being in arrears to the IMF at period  $t$  affects the probability of staying in or exiting from arrears in the next period ( $t+1$ ). A dummy variable for external arrears is not included as an explanatory variable because it is not compatible with the signals approach.

This study includes the stock of arrears to other external creditors relative to total external debt as a credit history indicator. Given the IMF's preferred creditor status, countries are most likely to incur arrears to the IMF only after they become overdue with commercial as well as other official creditors. Hence, an increase in arrears to other external creditors should serve as a warning signal for possible future arrears to the IMF.

### **Availability of resources**

Reserves in relation to imports, change in the level of reserves, and export growth are included in this study as indicators of resource availability. The ratio of reserves to imports is a useful measure of reserve needs for countries with limited access to international capital markets. The weaker a country's reserve position is in relation to its imports, the higher the likelihood that it would fail to repay its external debt obligations on time. AT consistently find this ratio statistically significant across different model specifications.

Recent studies have shown that the ratio of reserves to short-term external debt is a more useful indicator of external vulnerability than the import-based measure of reserve adequacy for countries with access to international capital markets.<sup>9</sup> Nevertheless, the reserve-import ratio is highly relevant to the exercise of anticipating arrears to the IMF because most countries that were in arrears in the past had limited or no access to capital markets. Furthermore, the uniform definition of the reserve-import ratio and its availability across a range of countries make it suitable for this study.

While export growth and the change in the level of reserves are not tested by AT, these variables are the best candidates for signaling the change in resource availability. The availability of reserves for financing balance of payments needs affects a country's repayment performance, particularly when a country with limited access to external financing experiences sharp swings in exports and imports.

### **Policy factors**

While the above indicators provide information about debt sustainability and resource availability, they do not provide information about policy factors, which could contribute to unsustainable external borrowing. To consider this, AT test two policy factors which are of paramount importance in IMF-supported adjustment programs: the fiscal position and

---

<sup>9</sup> See, for instance, Berg and others (1999); IMF (2000).



inflation, both of which are found to be significant. Failure to control government deficits and inflation could put pressures on balance of payments and weaken a country's external position, and therefore raising the likelihood of arrears to the IMF. This study includes the ratio of general government balance to GDP and the rate of inflation as explanatory variables.

### **Economic conditions**

This paper uses the two variables related to domestic economic conditions that AT find significant: per capita income and its growth rate.<sup>10</sup> Many studies have found that low real income growth is positively correlated with a country's deteriorating creditworthiness.<sup>11</sup> To attain external debt sustainability in the long run, the real economic growth rate must exceed the real rate of interest.

AT find that per capita income is an important variable for assessing the likelihood of incurring arrears to the IMF. They explain that, given the IMF's preferred creditor status, protracted arrears to the IMF tend to occur when poorer countries with the fewest options for raising resources or cutting expenses are faced with external imbalances.

### **Political risk**

A number of studies have found that, in addition to economic variables, political factors affect countries' creditworthiness and vulnerability.<sup>12</sup> Indeed, the majority of protracted arrears cases have been associated with difficult political conditions. For instance, all of the six countries currently in protracted arrears to the IMF have been plagued with one form of political strife or other: civil conflicts (Liberia, Somalia, and Sudan), adverse domestic political developments (Zimbabwe), and war (The Islamic State of Afghanistan and Iraq). AT present a study on political factors in their paper and conclude that, despite the apparent strong link between arrears and political risk, it is difficult to establish precise associations between political risk and the timing of IMF arrears.

AT examine the relationship between political factors and arrears to the IMF using a political risk indicator and event study analysis. They find that the average political risk indicator of the arrears countries is not significantly higher than that of the nonarrears countries. This counter-intuitive finding might have resulted from the method being used as

---

<sup>10</sup> The per capita income used here is the log of purchasing-power-parity-based per capita income series (in U.S. dollars) from the IMF's *World Economic Outlook* (WEO).

<sup>11</sup> See, for instance, Feder and Uy (1985); Li (1992).

<sup>12</sup> See, for instance, Berg and Sachs (1988); Bussières and Mulder (1999); Li (1992); McFadden and others (1985).

they sought to distinguish arrears countries from nonarrears countries by the level of political risk. Alternatively, one should compare observations during which countries were in arrears with observations during which countries were not in arrears. The results could be further improved if the observations were limited to the years prior to the emergence of arrears (entry model) because the entry model investigates whether or not a rise in political risk is associated with the emergence of arrears regardless of the persistence of arrears.

This paper revisits the issue of political risk and attempts to incorporate such risk in the model. Different approaches and specifications might find political risk to be a significant determinant of arrears to the IMF. As in AT, the political risk indicator published by a private firm specializing in country risk analysis is used for this study.<sup>13</sup> The main drawback of this indicator is the limited data availability; for example, historical data begin only in 1984 and do not cover all of the countries in the sample.

#### **IV. KAMINSKY-LIZONDO-REINHART (1997) SIGNALS APPROACH**

##### **A. Methodology**

The signals approach developed by KLR is designed to monitor a number of variables that tend to exhibit unusual behavior in the periods preceding a currency crisis and to send a warning signal when a variable reaches an “extreme” level beyond a certain “threshold.” A crisis is defined as a period in which a crisis index is significantly different from its mean.<sup>14</sup>

Observations are classified into four categories: good signal, false signal, missed signal, and good silence (for instance, a “good” signal is any value above a threshold that is followed by a crisis within 24 months). “Optimal” thresholds are determined for each indicator by minimizing the noise-to-signal ratio, i.e., the ratio of false signals to good signals. KLR define thresholds in terms of percentiles of the distribution of observations by country. While

---

<sup>13</sup> “International Country Risk Guide” published by the PRS Group. The political rankings range from 0 (highest risk) to 100 (lowest risk) and consists of 12 components: government stability, socioeconomic conditions, investment profile, internal conflict, external conflict, corruption, military in politics, religion in politics, law and order, ethnic tensions, democratic accountability, and quality of bureaucracy.

<sup>14</sup> The crisis index is the weighted average of the monthly percentage change in the exchange rate and (the negative of) the monthly percentage change in the growth of international reserves. The weights are chosen so that the two components have the same conditional variance. Periods in which the index is more than three standard deviations above its mean are defined as crises.

the percentile thresholds are uniform across countries, the corresponding thresholds in absolute terms are country specific.<sup>15</sup>

The KLR signals approach is a bivariate model in that each indicator is analyzed independent of other variables. KLR include 15 indicators in their model. Subsequently, Kaminsky (1998) calculates a composite index, aggregating signals of different indicators. With the composite index, conditional probabilities of crisis for ranges of composite values can be calculated based on the past values of the indicators.

### **Modifications**

This paper has made the following modifications to the original signals approach in order to reflect the different nature of incidents this study seeks to analyze.

**Arrears to the IMF:** This study aims to predict protracted arrears to the IMF. Therefore, the dependent variable in this study is a binary variable, which takes the value of one if a country is in continuous arrears to the IMF for at least six months, including at least three months of year  $t$ , and zero otherwise.

**Data sample:** The data sample in this study covers 118 developing countries that had outstanding debt to the IMF in any year during the period 1980–97.<sup>16</sup> The data are drawn from the *International Financial Statistics* (IFS), the *World Economic Outlook* (WEO), and the Treasurer’s Department’s records of members’ financial obligations and payments to the IMF. The years during which a country had no debt outstanding to the IMF are excluded from the sample. It should be noted that the vast majority of the countries in sample have never incurred protracted arrears to the IMF.<sup>17</sup>

**“Entry” model:** This paper employs an “entry” model, which excludes from the data sample those years in which the country concerned had protracted arrears to the IMF. Because many of the protracted arrears cases tend to persist, unlike currency crises, there is an issue of state

---

<sup>15</sup> For instance, if the threshold for the ratio of short-term debt to GDP was calculated as 85 percent, this means that all observations beyond the 85<sup>th</sup> percentile of the frequency distribution for each country are categorized as signals for this indicator. The 85<sup>th</sup> percentile could be as small as 1 percent of GDP for small capital importers while it might be 10 percent of GDP for countries with large capital inflows.

<sup>16</sup> Countries classified by the WEO as “advanced economies” are excluded from the sample given their minute likelihood of incurring arrears to the IMF. The inclusion of these countries would only increase “good silence,” or no signal followed by no arrears.

<sup>17</sup> KLR include in their sample only those emerging market countries that had experienced currency crises according to their crisis definition. However, this study does not limit its sample to arrears countries because such a sample would be too limited.

dependency.<sup>18</sup> Another reason for choosing the entry model is that we are interested in predicting arrears that are yet to occur and not in calling when the country exits from arrears.

**Indicators:** As explained in Section III, the following indicators are considered in this study: (1) IMF debt outstanding relative to quota, (2) ratio of IMF debt service to exports, (3) stock of arrears to external creditors relative to total external debt, (4) export growth (in U.S. dollars), (5) change in reserves (in U.S. dollars), (6) reserve-imports ratio, (7) general government balance as a percent of GDP, (8) rate of inflation, (9) PPP-based per capita income (in U.S. dollars), (10) growth rate of per capita income, and (11) political risk indicator.

This paper uses annual data given the lack of availability of monthly or quarterly data for the variables and countries covered in this study. KLR use monthly data as their focus on emerging market economies enables them to use high-frequency data. Nevertheless, the use of annual data in this study is unlikely to limit the effectiveness of the model; AT find that the signs of repayment difficulty emerge several years prior to the incidence of arrears.

**Signaling horizon:** The signaling horizon is a period within which warning signals are supposed to flash prior to the incidence of an event. If an indicator sends a signal that is followed by the incidence within a pre-determined signaling horizon, it is considered a good signal. KLR define this horizon a priori as two years for currency crises. In the case of protracted arrears, however, the signs of fundamental deterioration tend to emerge even earlier. Countries turn to the IMF when they have balance of payments difficulties, and arrears tend to occur when outstanding debt owed to the IMF reaches a very high level. Therefore, the signs of deterioration are evident several years before the emergence of arrears. The signaling horizon in this study is defined as three years, which is confirmed by sensitivity analysis.

**Thresholds:** While KLR calculate thresholds in percentile, this method is not appropriate for this study given the fact that our sample covers 118 countries, most of which have not experienced protracted arrears to the IMF. Percentile thresholds assume that all countries in the sample experience incidents whenever the values of indicators deviate from the normal level relative to their past mean. This procedure would lead to excessive noise when applied to our sample, because most of the countries have not experienced arrears.

Thresholds expressed in absolute terms better suit the nature of protracted arrears. Unlike currency crises where the deviation from the normal indicator level relative to the country mean signals a crisis, protracted arrears tend to occur when a country reaches a certain

---

<sup>18</sup> AT find that the entry and exit models are statistically distinct, meaning that variables that affect the likelihood of falling into (or staying in) arrears differ depending on whether the country is already in arrears or not.

“desperation level,” which is more an absolute concept than a relative one.<sup>19</sup> Percentile thresholds tend to overstate the indicator values of better-off countries that are reasonable, and fail to capture the absolute bleakness of worse-off countries. Therefore, this study defines thresholds in absolute terms rather than in percentile terms. This modification improves performance and makes interpretation easier.

**Performance criteria:** As described in KLR, the performance of indicators can be examined in terms of the matrix below. The cell A represents the number of years in which the indicator issued a good signal, meaning the indicator values beyond the threshold were followed by arrears within three years. B is the number of years in which the indicator issued a false signal, or “noise.” C is the number of years in which the indicator failed to signal impending arrears, and D is the number of years in which the indicator did not issue a signal that would have been a bad signal.

	Arrears within 3 years	No arrears within 3 years
Signal was issued	A (good signal)	B (false signal)
No signal was issued	C (missed signal)	D (good silence)

**Noise-to-signal ratio:** Thresholds are obtained by striking a balance between the risk of missing arrears (type I error) and the risk of issuing false signals (type II error). KLR define the “optimal” threshold as a point which minimizes the noise-to-signal ratio; that is  $[B/(B+D)]/[A/(A+C)]$  in the matrix above. This is equivalent to minimizing the ratio of type II errors to one minus type I errors.<sup>20</sup> A grid search is performed to find the “optimal” value among a range of possible threshold values. As discussed in Box 1, however, minimizing the noise-to-signal ratio often leads to extreme threshold values.

<sup>19</sup> Take the example of a middle-income country with an excellent track record of external debt repayment and macroeconomic management. If the country’s past indicator values lie within a reasonable range, it is highly unlikely that the country would incur arrears to the IMF even when its indicators deviate from the predetermined percentile threshold. On the other hand, a heavily indebted country with a troubled credit history and mismanaged economy is likely to fall into arrears even when its indicator values may not significantly differ from their past means and thus fall short of crossing the percentile threshold.

<sup>20</sup> As discussed in Hali (2000), the noise-to-signal ratio could be interpreted using type I and type II errors. Let the null hypothesis ( $H_0$ ) be when arrears occur and the alternative hypothesis ( $H_A$ ) be when they do not occur. From the matrix above,  $H_0$  is  $A+C$  and  $H_A$  is  $B+D$ . As the size of type I errors is the probability of rejecting  $H_0$  when  $H_0$  is true, type I error is defined as  $C/(A+C)$ . Similarly, the size of type II errors is the probability of not rejecting  $H_0$  when  $H_0$  is false. Thus, type II error is defined as  $B/(B+D)$ . The noise-to-signal ratio,  $[B/(B+D)]/[1-C/(A+C)]$ , is defined as the ratio of type II errors to one minus type I errors.

**Loss function:** Alternatively, thresholds can be chosen to minimize the weighted sum of the share of false signals and missed signals, where weights reflect the preference of decision makers.<sup>21</sup> This is equivalent to minimizing the sum of type I errors [ $C/(A+C)$ ] and type II errors [ $B/(B+D)$ ] with given weights associated with each type of error. This loss function method enables us to calculate more reasonable threshold levels because it is minimizing the sum of two types of errors rather than their ratio. For comparison, this paper implements both the noise-to-signal and loss function methods to determine the thresholds with equal weights attached to each type of error.

## B. Implementation

### Performance of Indicators

Table 1 summarizes the in-sample performance of the indicators using the noise-to-signal ratio and loss function. The first column shows thresholds for each indicator in absolute terms. For example, a warning signal is sent whenever a country's IMF debt service to exports ratio exceeds 8 percent under the noise-to-signal method, and 3.1 percent under the loss function method. For most indicators, thresholds using the noise-to-signal tend to issue signals to fewer countries than those obtained under the loss function method as discussed in Box 1.

The second and third columns show the size of type I errors,  $C/(A+C)$ , and type II errors,  $B/(B+D)$ , respectively. There is a clear trade-off between the two types of errors: the noise-to-signal method yields more restrictive threshold values, therefore issuing fewer false signals (type II error), but missing more arrears (type I error) than the loss function method, and vice versa. The fourth column shows the loss function, the sum of type I and type II errors. The fifth column shows the noise-to-signal ratio, type I errors over one minus type II errors. The sixth column is the share of arrears correctly called,  $A/(A+C)$ . The last column shows the conditional probability of arrears,  $A/(A+B)$  – the probability that arrears occur within 3 years given that there is a signal.

In evaluating the performance of indicators, the loss function method awards indicators with the smallest loss function values. In a similar manner, the noise-to-signal method rewards indicators that have smaller noise-to-signal ratios. While these two methods give us somewhat different performance rankings among indicators, both find a similar set of indicators important in determining arrears to the IMF. Both methods find that the political risk indicator, IMF-specific debt ratios (IMF debt outstanding/quota and IMF debt service/exports), and the reserve-imports ratio are particularly informative indicators. The other indicators are also useful, although the order of importance differs from one method to the other. These indicators are aggregated in a composite indicator as described in Box 2.

---

<sup>21</sup> Demirguc-Kunt and Detragiache (1999). See Box 1 for details.

### Box 1. Different Measures of Efficiency

Determining thresholds for indicators (as in the KLR signals approach) or cut-off points for predicted probabilities (as in probit or logit regressions) requires some evaluation criteria, or efficiency measures.

One of the widely used measures is the noise-to-signal ratio proposed by Kaminsky, Lizondo, and Reinhart (1998). While the noise-to-signal ratio, which is defined as false signals over good signals, is appealing for its simplicity, it is not without problems. As Mulder, Perrelli, and Rocha (2002) correctly point out, the ratio only measures the *proportion* of bad signals to good signals and does not differentiate the cut-off points with different *absolute* number of good signals or bad signals as long as they have the same ratio. In other words, the noise-to-signal ratio of 10 percent could mean that only 10 incidents were called, of which only one was false, or 100 incidents were called, of which 10 were false, out of a large number of incidents. Therefore, minimizing the noise-to-signal ratio often leads to extreme thresholds where hardly any incidents are called since the false signals approach zero. This problem is compounded in this study because the majority of countries in the sample have never experienced arrears, leading to a greater number of false signals with a smaller number of good signals. Because the noise-to-signal ratio becomes smaller either as the noise becomes smaller or good signals become fewer, and given the fact that good signals change less frequently than the false signals, in this sample, the ratio continues to become smaller, leading to extreme thresholds with both few false and good signals.

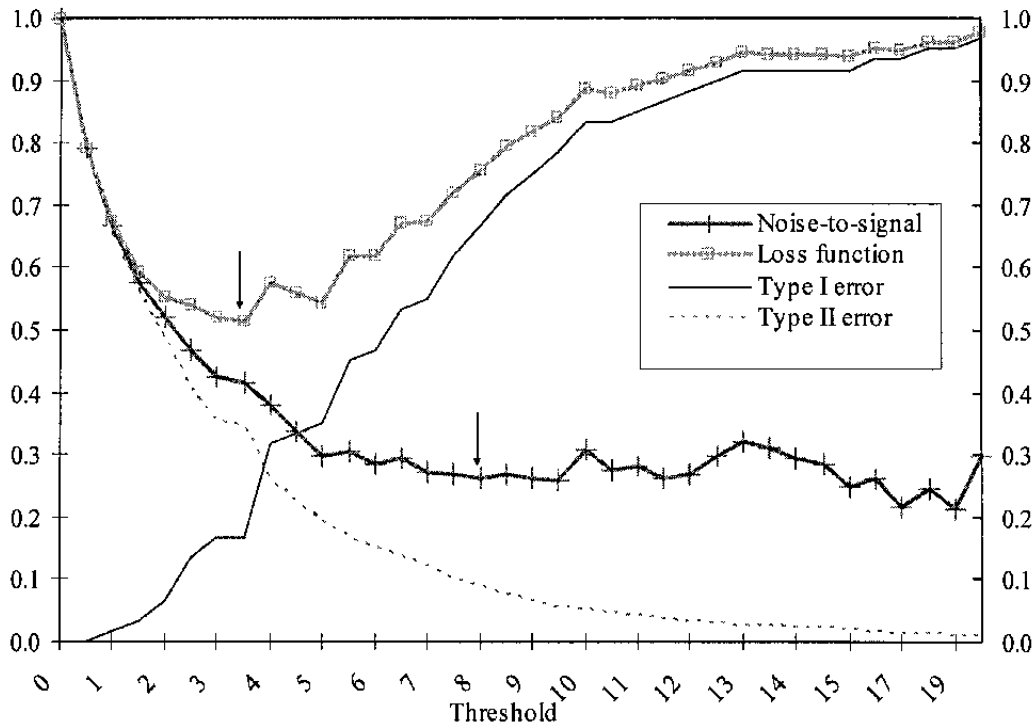
As an alternative to the noise-to-signal ratio, Mulder, Perrelli, and Rocha (2002) propose “signal-to-noise balance” as a criterion for selecting cut-off points for crisis probabilities. The signal-to-noise balance is defined as “the difference between the percentage of pre-crisis periods called correctly and the percentage of false alarms, where both are expressed in terms of the number of (pre-)crisis periods.” This ratio could be interpreted as  $A/(A+C) - B/(A+C)$  in terms of the signal matrix. They claim one of the merits of this measure is that the relative number of classified incidents (i.e. crises, alarms, or tranquil periods) does not affect the ratio. Moreover, this ratio is easily interpretable in the sense that it can reach a maximum value of 100 when all pre-crisis periods are called correctly and no false alarms are issued. In addition, the balance is negative when models issue more false alarms than good alarms per pre-observed crisis periods. Although this measure may be useful for currency crises, it usually leads to a negative value in this study because the number of false signals (B) exceeds the number of arrears incidents (A+C). Consequently, the point that maximizes the value of signal-to-noise balance barely calls any incidence of arrears.

Another method is a loss function approach employed by Demirguc-Kunt and Detragiache (1999) to determine a threshold for banking crisis probabilities in a logit framework. They show that the optimal threshold can be obtained by minimizing the expected loss function, the sum of type I and type II errors with weights associated with each type of error. Weights depend on the costs that decision makers attach to each type of error. For instance, if the cost of missing arrears is greater than the cost of issuing a false alarm, a heavier weight should be attached to type I errors relative to type II errors. Using the above matrix, the loss function could be expressed as follows:  $L(T) = [C/(A+C)]c_1 + [B/(B+D)]c_2$ , where T is the threshold,  $C/(A+C)$  is the size of type I errors,  $c_1$  is the cost attached to it,  $B/(B+D)$  is the size of type II errors, and  $c_2$  is the cost associated with it. Minimizing the loss function yields thresholds that are different from those obtained by the noise-to-signal ratio.

### Box 1. Different Measures of Efficiency (Concluded)

Box Figure 1. Minimizing the Noise-to-Signal Ratio and Loss Function

Debt service to the IMF  
(In percent of exports)



The above figure illustrates how the noise-to-signal ratio and loss function evolve through a range of threshold values, which are, in this example, the ratio of debt service to the IMF to exports. As the threshold is set at a higher level of debt-service ratio, fewer false alarms will be issued, and hence, a smaller size of type II errors. As a trade-off, more arrears would be uncaught, leading to a larger size of type I errors. The figure shows that the loss function, or the sum of type I and type II errors, reaches its minimum point around 3 percent, reflecting this tradeoff.

In contrast, the noise-to-signal ratio does not have a clear minimum point; the ratio remains at a low level beyond about 8 percent. To be precise, the debt service ratio reaches its minimum at 19 percent, at which point only two incidents of past arrears are called. This is because the noise-to-signal ratio measures the ratio, and false alarms approach zero at these extreme values. To avoid using such extreme thresholds, this paper excludes minimum points that call less than 10 percent of total arrears incidents. Moreover, if the noise-to-signal ratio is the same across different threshold values, a less strict threshold value is chosen as the minimum point. In the above example, therefore, the threshold under the noise-to-signal method is set at 8 percent.

The phenomenon that the loss function leads to clear and reasonable minimum values whereas the noise-to-signal ratio yields extreme values is common to most indicators. This result provides support for using the loss function method rather than the noise-to-signal method in this study.



Table 1. Performance of Indicators: Loss Function and Noise-to-Signal Methods

Indicator 1/	Method 2/	Threshold 3/	Type I error	Type II error	Loss Function	Noise-to-Signal Ratio	Share of Arrears Called	Conditional Probability
Political risk indicator	N-S	[37]	0.63	0.07	0.70	0.18	37%	17%
	LF	[53]	0.07	0.39	0.47	0.42	93%	8%
IMF debt service/exports	N-S	8.0%	0.67	0.09	0.75	0.26	33%	15%
	LF	3.1%	0.17	0.35	0.51	0.42	83%	10%
Reserves/imports	N-S	[3.3%]	0.70	0.07	0.77	0.23	30%	18%
	LF	[11.7%]	0.25	0.31	0.56	0.41	75%	10%
IMF debt outstanding/quota	N-S	265%	0.71	0.07	0.78	0.23	29%	17%
	LF	150%	0.30	0.28	0.58	0.40	70%	11%
General government balance/GDP	N-S	[-7.8%]	0.50	0.20	0.70	0.42	51%	10%
	LF	[-6.2%]	0.31	0.31	0.62	0.45	70%	9%
Arrears/total external debt	N-S	4.9%	0.42	0.25	0.67	0.43	58%	10%
	LF	1.0%	0.29	0.35	0.64	0.50	71%	8%
Rate of inflation	N-S & LF	17.9%	0.45	0.32	0.77	0.58	55%	7%
Export growth	N-S	[-25%]	0.88	0.03	0.91	0.27	7%	10%
	LF	[0.5%]	0.43	0.35	0.78	0.61	57%	7%
Percent change in reserves	N-S	[-58%]	0.81	0.04	0.85	0.21	19%	19%
	LF	[-25%]	0.63	0.17	0.80	0.46	37%	10%
Per capita income growth	N-S & LF	[0.7%]	0.60	0.21	0.81	0.52	40%	8%

1/ Indicators are ranked in ascending order of loss function values under the loss function method.

2/ "N-S" indicates the noise-to-signal method while "LF" indicate the loss function method. "N-S & LF" means that both methods obtain the same results.

3/ Threshold values in square brackets indicate that signals are issued when indicator values fall short of the threshold values. Thresholds without brackets issue alarms when the indicator values go beyond the threshold values.

### Box 2. Composite Indicator and Conditional Probability

To take into account different forecasting ability of indicators and aggregate information, composite indicators have been developed. Kaminsky (1998) calculates several composite measures, of which the weighted-sum composite is most commonly used. This composite uses the inverses of the noise-to-signal ratios as weights, thereby giving heavier weights to indicators with smaller noise-to-signal ratios. Kaminsky defines the weighted-sum composite as follows:

$$I_t = \sum_{j=1}^n S_t^j / \omega^j$$

where  $S_t^j = 1$  if variable  $j$  crosses the threshold in period  $t$  and 0 otherwise, and  $\omega^j$  is the noise-to-signal ratio of variable  $j$ . As this study implements both the noise-to-signal and loss function methods, the above equation needs to be altered for the loss function method. Specifically,  $\omega^j$  is redefined as the loss function of variable  $j$ , thereby weighing the signal by the inverse of its loss function instead of its noise-to-signal ratio under the loss function method.

Following Kaminsky, one can infer the likelihood of countries incurring IMF arrears by calculating the sample-based conditional probabilities using the following formula:

$$P(A_{t,t+h} | I_i < I_t < I_j) = \frac{\text{Years with } I_i < I_t < I_j \text{ given arrears occur within h years}}{\text{Years with } I_i < I_t < I_j}$$

where  $P$  denotes probability,  $A_{t,t+h}$  is the occurrence of arrears in the interval  $[t, t+h]$ ,  $h$  is the signaling window (three years),  $I$  is the weighted indicator and the subscripts  $i$  and  $j$  denote upper and lower intervals for the composite indicator. The  $P(A_{t,t+h} | I_i < I_t < I_j)$  denotes the probability that arrears will occur within  $h$  years at time  $t$ , given that the composite indicator  $I_t$  falls within the range  $I_i$  and  $I_j$ .

### In-Sample Goodness of Fit

The assessment of model performance requires selecting the cut-off point, above which data points are categorized as alarms. This paper has so far produced two types of composite indicators: one using the noise-to-signal method and the other using the loss function method. In evaluating the performance of each method, the cut-off point for a composite indicator needs to be selected, which again calls for efficiency measures. The noise-to-signal ratio can not be used to determine the cut-off point of a composite indicator because the ratio is continuously decreasing, leaving no minimum point. Thus, this study determines the cut-off point of a composite indicator by minimizing the loss function (the sum of type I and type II errors) of a composite indicator, irrespective of how the composite indicator is derived.<sup>22</sup>

<sup>22</sup> Berg and Pattillo (1998) initially set the cut-off probability somewhat arbitrarily at 25 percent to assess the goodness of fit. Subsequently, they started using the loss function method to determine the cut-off point.

This study finds that the cut-off point of a composite indicator, where its loss function value is minimized, is 6.1 under the noise-to-signal method with the conditional probability of 7 percent. The cut-off point under the loss function method is 7.5 with the conditional probability of 6 percent. These low conditional probabilities result from the rare occurrences of arrears to the IMF.

The first and second column of Table 2 show the in-sample result of the signals approach under the noise-to-signal method and the loss function method, respectively. At the cut-off composite value of 6.1, the noise-to-signal method correctly calls 76 percent of observations. The model calls 84 percent of arrears periods correctly, capturing all the past incidents of protracted arrears except one case, where only three out of ten indicators are available. The drawback is a large number of false alarms: 86 percent of total alarms are false. This high share of false alarms to total alarms mainly results from the infrequency of IMF arrears in this sample; the total number of arrears periods is only 63 out of 1375 observations.

As for the loss function method, with the cut-off composite value of 7.5, it calls 84 percent of observations correctly, most of which are good silence, no alarm followed by no arrears. The loss function method captures 75 percent of the past arrears periods. In terms of the number of incidence, the model calls 17 protracted arrears cases out of 21. Yet, 81 percent of the alarms are not followed by arrears in the subsequent three years. Nevertheless, considering the fact that the unconditional probability of arrears is as low as 4.6 percent in the sample, the probability of arrears conditional on alarms of 19 percent is an improvement.

It should be noted that there are certain shortcomings in strictly applying the definition of arrears and evaluating the goodness of fit, such as counting the number of false alarms. For instance, some of the signals classified as false alarms were in fact alarms that correctly called short-term arrears, but which did not last for six months, and thus did not meet the definition of the dependent variable. In addition, some alarms were signaling arrears more than three years prior to the incidence. Moreover, countries emerging from arrears still face difficulty and tend to have high composite values (meaning many indicator values are above their thresholds), which are often categorized as false positives. When these issues are taken into account, namely, if signals calling short-term arrears lasting between two and six months and signals flashing four years prior to the emergence of arrears are classified as good signals (A), and if signals one year after the countries emerge from arrears are excluded, both models perform better.

The last two columns of Table 2 show the performance after these adjustments have been made. Under the noise-to-signal method, the ratio of false alarms to total alarms falls from 86 percent to 78 percent, while this ratio declines from 81 percent to 70 percent under the loss function method. The probability of arrears conditional on an alarm goes up from 14 percent to 22 percent under the noise-to-signal method, while it rises from 19 percent to 30 percent with the loss function method.

Table 2. In-sample Performance of Signals Approach

	Strict Definition of Arrears (protracted arrears only)				Broad Definition of Arrears (protracted and short-term arrears)			
	Noise-to-signal		Loss function		Noise-to-signal		Loss function	
Cut-off probability (value of composite) 1/	7	(6.1)	6	(7.5)	7	(6.1)	6	(7.5)
Value of loss function 2/	41		41		33		32	
Type I error 3/	16		25		12		19	
Type II error 4/	25		16		22		13	
Percent of observations correctly called 5/	76		84		79		87	
Percent of arrears in three years correctly called 6/	84		75		88		81	
Percent of tranquil in three years correctly called 7/	75		84		78		87	
False alarms as percent of total alarms 8/	86		81		78		70	
Probability of arrears given signal 9/	14		19		22		30	
Probability of arrears given no signal 10/	1.0		1.4		1.0		1.4	
Number of observations included	1375		1375		1344		1353	
Number of arrears periods covered	63		63		85		84	

	Actual		Actual		Actual		Actual	
	arrears	Tranquil	arrears	Tranquil	arrears	Tranquil	arrears	Tranquil
Alarm issued	53	324	47	205	75	271	68	162
No alarm issued	10	988	16	1107	10	988	16	1107

1/ The cutoff probability (value of composite) above which a forecast is deemed to indicate arrears.

2/ The sum of type I and Type II errors.

3/ Missed arrears as a share of total pre-arrears periods =  $C/(A+C)$ .

4/ False alarms as a share of total tranquil periods =  $B/(B+D)$ .

5/ The number of correct observations as a share of total observations =  $(A+D)/(A+B+C+D)$ .

6/ The number of pre-arrears periods correctly called as a share of total pre-arrears periods =  $A/(A+C)$ .

7/ The number of tranquil periods correctly called as a share of total tranquil periods =  $D/(B+D)$ .

8/ A false alarm is an observation with an estimated probability of crisis above the cutoff not followed by arrears within three years.  $B/(A+B)$ .

9/ The number of pre-arrears periods correctly called as a share of total predicted pre-arrears periods =  $A/(A+B)$ .

10/ The number of periods where tranquility is predicted and arrears actually ensue as a share of total predicted tranquil periods =  $C/(C+D)$ .

Both the noise-to-signal method and the loss function method attain approximately the same loss function values (the sum of type I and type II errors). The important difference, however, lies in the type of error each method minimizes. Table 2 shows that the noise-to-signal method misses fewer arrears (smaller size of type I errors) and issues more false alarms (larger size of type II errors) than the loss function method does.<sup>23</sup> In fact, the ratio of type I errors to type II errors under the loss function method is close to the inverse of the ratio under the noise-to-signal method.

## V. BERG AND PATTILLO (1998) PROBIT-BASED ALTERNATIVE

### A. Methodology

BP build on KLR's pioneering work and embed the KLR signals approach in a multivariate probit framework. Instead of calculating separate thresholds for each indicator, BP estimate the combined impact of the indicators to produce the probability of currency crises. The advantage of the multivariate probit model is that the correlations among different variables are taken into account, and that it is easy to test for statistical significance.

All data across countries and periods are pooled and probit models are estimated using maximum likelihood. The dependent variable in BP is a dummy variable that takes the value one when a crisis occurs in the next 24 months, and zero otherwise. BP's core equation takes the following form:

$$\Pr(C = 1)_{it} = \sum_{k=1}^K \beta_k p(x_{kt})_i + \varepsilon_{it} \quad (1)$$

where  $C = 1$  if there is a crisis in the next 24 months for country  $i$  in month  $t$ . There are  $K$  explanatory variables, all of which are transformed into percentiles by the function  $p(-)$ .

---

<sup>23</sup> This result may seem paradoxical considering the restrictive threshold values that the noise-to-signal method obtained for the indicators as shown in Table 1. This apparent paradox results from the fact that the loss function value (the sum of type I and type II errors) reaches its minimum at a lower composite value under the noise-to-signal method. Because the noise-to-signal method yields more restrictive threshold values, the number of false alarms, or type II errors, is relatively small even at a lower composite value, thus minimizing its loss function at a lower level of composite value. The lower cut-off point for the composite indicator, in turn, leads to a larger number of signals.

## Modifications

As the KLR signals approach is embedded in the BP probit model, all modifications described in Section IV, except for those specific to the signals approach, apply to the probit model. Namely, the binary dependent variable takes the value one if there is an incidence of arrears to the IMF lasting six months or longer in the next three years, and zero otherwise. Independent variables enter in absolute terms rather than in percentiles of the distribution of observations. Reflecting these changes, equation (1) is modified as follows:

$$\Pr(A = 1)_{it} = \sum_{k=1}^K \beta_k x_{ikt} + \varepsilon_{it} \quad (2)$$

where  $\Pr(A = 1)$  is the probability of IMF arrears in the next three years for country  $i$  in year  $t$ .

## B. Implementation

### Estimation Output

Table 3 shows the estimation of probit regressions.<sup>24</sup> Model 1 is the estimate using all explanatory variables, and model 2 is its simplified form, sequentially eliminating variables that are not significant at the 10 percent level.<sup>25</sup> Model 2 indicates that the most significant variable is IMF debt outstanding relative to quota, followed by debt service to the IMF as a percentage of exports, the ratio of non-IMF arrears to total external debt, reserves in relation to imports, and export growth.

The above findings are broadly in line with the results obtained by AT. Their entry model shows that IMF debt outstanding over quota is the most significant determinant of the probability of IMF arrears, with per capita income, economic growth, reserves, and inflation also influencing the likelihood. The introduction of new variables and the use of different model specifications and sample size may account for some of the differences.

---

<sup>24</sup> The political risk indicator is not included in this regression exercise because the significantly smaller number of observations available would reduce the sample size for the estimation and forecasting by almost half. Nevertheless, when included in the test, the political risk indicator is found to be highly significant.

<sup>25</sup> It is worth noting that the probit methodology is less tolerant than the KLR signals approach in terms of the treatment of missing data. Under the probit model, one missing data of an indicator for country  $X$  in year  $Y$  would eliminate all the data for the country in that year. Under the KLR signals approach, however, the composite indicator for country  $X$  in year  $Y$  is included in the sample provided that at least one of the indicators is available.

Table 3. Estimation Output of Probit Model 1/

Variable	Model 1		Model 2	
	Coefficient	Z-statistic	Coefficient	Z-statistic
IMF debt outstanding/quota	0.0039	5.564 ***	0.004	5.562 ***
IMF debt service/exports	0.0455	3.398 ***	0.045	4.072 ***
Stock of non-IMF arrears/external debt	0.0124	2.873 ***	0.014	3.712 ***
Reserves/imports	-0.0231	-3.107 ***	-0.027	-3.625 ***
Export growth	-0.0092	-1.669 *	-0.013	-2.528 **
Change in reserves	-0.0008	-1.155		
General government balance/GDP	-0.0120	-1.259		
Rate of inflation	-0.0006	-0.368		
Economic growth	-0.0169	-1.047		
Per capita income	-0.0482	-0.412		
Constant	-1.8240	-1.957 *	-2.155	-12.685 ***
Sample size	1139		1300	
Observations with dep =0	1088		1245	
Observations with dep =1	51		55	
McFadden R-squared	0.253		0.249	

1/ Asterisks underneath z-statistics indicate significance: \*, \*\*, and \*\*\* denote 10%, 5%, and 1 % significance levels, respectively.

Table 4. In-sample Goodness of Fit: Probit Regression and Signals Approach

	Broad Definition		Strict Definition of Arrears			
	Probit regression	Probit regression	Signals approach			
			Noise-to-signal		Loss function	
Cut-off probability (value of composite) 1/	4	4	7	(6.1)	6	(7.5)
Value of loss function 2/	29	36	41		41	
Type I error 3/	9	13	16		25	
Type II error 4/	20	23	25		16	
Percent of observations correctly called 5/	81	77	76		84	
Percent of arrears in three years correctly called 6/	91	87	84		75	
Percent of tranquil in three years correctly called 7/	80	77	75		84	
False alarms as percent of total alarms 8/	77	86	86		81	
Probability of arrears given signal 9/	23	14	14		19	
Probability of arrears given no signal 10/	0.7	0.7	1.0		1.4	
Number of observations included	1270	1300	1375		1375	
Number of arrears periods covered	78	55	63		63	

	Actual		Actual		Actual		Actual	
	arrears	Tranquil	arrears	Tranquil	arrears	Tranquil	arrears	Tranquil
Alarm issued	71	239	48	292	53	324	47	205
No alarm issued	7	953	7	953	10	988	16	1107

1/ The cutoff probability above which a forecast is deemed to indicate arrears.

2/ The sum of type I and Type II errors.

3/ Missed arrears as a share of total pre-arrears periods =  $C/(A+C)$ .

4/ False alarms as a share of total tranquil periods =  $B/(B+D)$ .

5/ The number of correct observations as a share of total observations =  $(A+D)/(A+B+C+D)$ .

6/ The number of pre-arrears periods correctly called as a share of total pre-arrears periods =  $A/(A+C)$ .

7/ The number of tranquil periods correctly called as a share of total tranquil periods =  $D/(B+D)$ .

8/ A false alarm is an observation with an estimated probability of crisis above the cutoff not followed by arrears within three years.  $B/(A+B)$ .

9/ The number of pre-arrears periods correctly called as a share of total predicted pre-arrears periods =  $A/(A+B)$ .

10/ The number of periods where tranquility is predicted and arrears actually ensue as a share of total predicted tranquil periods =  $C/(C+D)$ .



### **In-sample Goodness of Fit**

Table 4 presents the in-sample goodness of fit of the probit model as well as the signals approach, under the noise-to-signal and loss function methods.<sup>26</sup> The first column shows the result of the probit model using the broader definition of arrears, which is to be compared with the next column on probit using the strict definition, counting protracted arrears only. With the cut-off probability of 4 percent, the probit has the loss function value of 36, lower than that of the signals approach. The size of type I errors under the probit model is markedly smaller than that of the signals approach; the probit calls 87 percent of the arrears periods correctly, capturing 19 incidents of protracted arrears out of 20 covered by the model. The trade-off is a larger size of type II errors: 86 percent of the total alarms are false although the false alarm ratio goes down to 77 percent when the more relaxed definition of arrears is applied.

The in-sample results show that all three models are successful in calling the majority of arrears incidence in the past. Both the probit model and the noise-to-signal method miss only one incidence of arrears while the loss function method misses 4 out of 21. The trade-off is a large number of false alarms: over 80 percent of alarms are false when the strict definition of arrears is used.

In comparing the in-sample performance, the model that attains the lowest loss function value is regarded as the best performer since it is the most successful model in minimizing both types of errors. The probit model obtains the lowest loss function value among the three models regardless of the definitions of arrears, and therefore performs the best in sample.

## **VI. OUT-OF-SAMPLE PERFORMANCE**

It is more interesting to examine how these models perform in forecasting arrears rather than identifying past arrears. Models are estimated using the data sample up to 1997 to allow for an out-of-sample testing for the period after 1998. During the period 1998–2002, there has been only one protracted arrears case (Zimbabwe) and one short-term arrears lasting nearly three months (Case B).<sup>27</sup> Although Zimbabwe is the only protracted arrears case and thus meets the strict definition of dependent variable in this study, Case B should also be

---

<sup>26</sup> It should be noted that the cut-off points used here might not necessarily be the best cut-offs. Depending on the preferences of policymakers, different weights could be attached to type I and type II errors. For instance, if the number of false alarms should be smaller, the weight for type II errors could be increased. In the case of anticipating arrears to the IMF, however, the cost of missing arrears is more likely to be higher than that of issuing false alarms. Therefore, low cut-off points are retained to capture more potential arrears cases.

<sup>27</sup> Information on arrears to the IMF for less than three months is not published under current IMF policy. Therefore, the country in question is addressed in this paper as “Case B.”

considered as a failure to repay the IMF on time and thus be treated separately from the other unrealized predicted arrears.

While this limited incidence of actual arrears out-of-sample makes it difficult to assess the out-of sample performance of the models, we could evaluate the model performance based on the following three criteria: (i) models that place Zimbabwe and Case B well above their respective cut-offs are better than those that do not; (ii) the best performing model should assign the highest probability to actual arrears cases among others prior to the emergence of arrears; (iii) models that issue fewer false alarms while calling the arrears cases correctly should be regarded as better than those that call many arrears that did not occur.

#### **A. Zimbabwe (2001- present)**

The most recent case of protracted arrears is Zimbabwe, which has been in arrears to the IMF since February 2001. It is an important, and in a strict sense, the only out-of-sample test to examine how the models would have performed in anticipating Zimbabwe's arrears had they been available prior to the emergence of arrears.

The deterioration in Zimbabwe's economy originated in unsustainable macroeconomic policies, notably the monetization of large fiscal deficit and gross currency overvaluation, which was further exacerbated by deteriorating governance. The launching of the fast-track land reform program and growing disrespect for the rule of law have led to Zimbabwe's international isolation.

Although all three models predict Zimbabwe's arrears one to three years prior to the emergence of arrears in 2001, the loss function method outperforms the others in capturing Zimbabwe's deterioration and in issuing fewer false alarms.

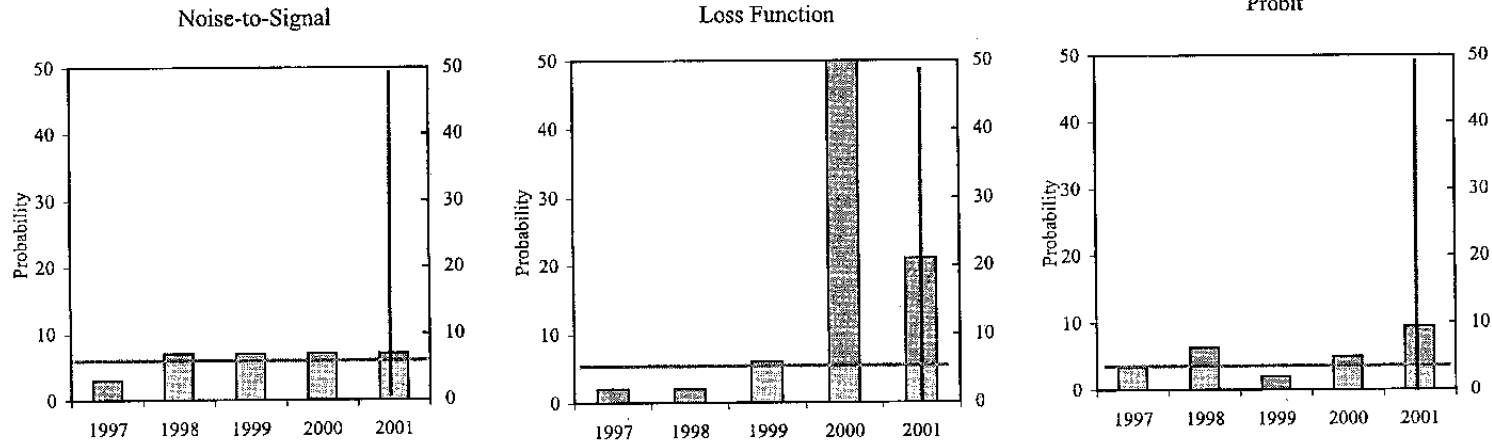
#### **Signals approach**

Figure 1 presents the probability of IMF arrears generated by the signals approach, under the noise-to-signal and loss function methods, and the probit model. Under the noise-to-signal method, the probability narrowly crosses its cut-off in 1998 and stays constant prior to 2001. The probability under the loss function method goes beyond its cut-off in 1999 and jumps to a 50 percent level in 2000.

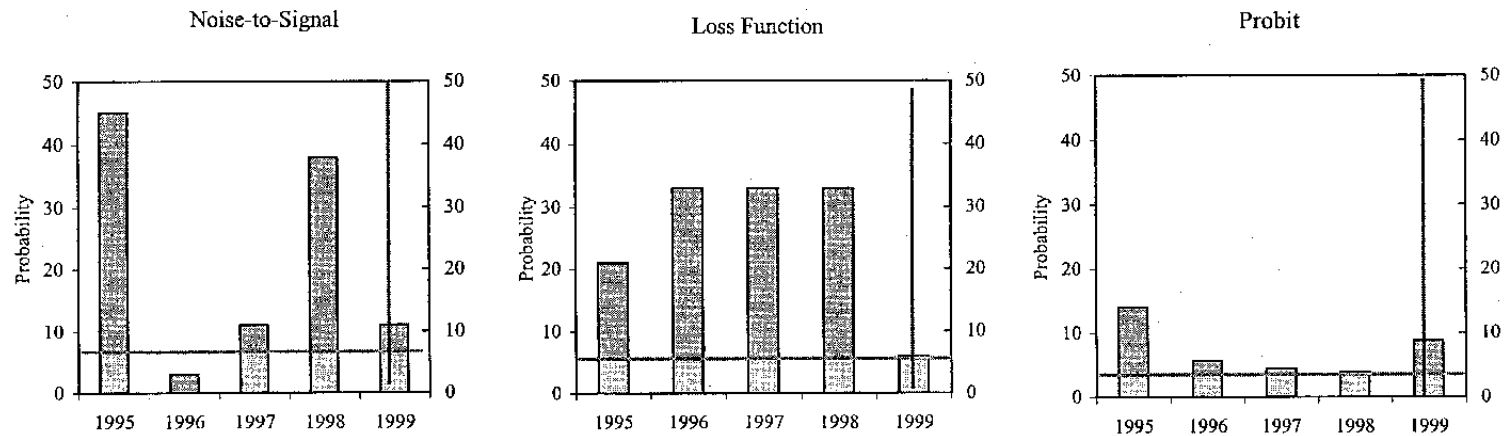
Table 5 shows the indicators that issued signals one to three years prior to the emergence of arrears and presents the evolution of composite indicators under the noise-to-signal and loss function methods. One of the stark differences between the two results is that the number of indicators signaling is greater under the loss function method than under the noise-to-signal method, which is reasonable given more relaxed threshold values of the former method. Four indicators are sending signals in 1999 and 2000 under the noise-to-signal method, whereas six and eight indicators, respectively, are beyond their thresholds in 1999 and 2000 under the loss function method.

Figure 1. Out-of-Sample Performance<sup>1/</sup>

Zimbabwe



Case B



<sup>1/</sup> Vertical lines represent the emergence of arrears while horizontal lines indicate the threshold cut-off, above which alarms are issued.

Table 5. Performance of Indicators: Out-of-Sample Cases 1/

Zimbabwe

Indicators	Noise-to-Signal Method				Loss Function Method			
	1998	1999	2000	2001	1998	1999	2000	2001
IMF debt outstanding/quota								
IMF debt service/exports							X	
Non-IMF arrears/external debt			X			X	X	
Reserves/imports	X	X			X	X		
Change in reserves							X	
Export growth					X	X	X	
General government balance/GDP		X	X			X	X	
Inflation	X	X	X		X	X	X	
Economic growth	X	X	X		X	X	X	
Political risk							X	
Composite value	6.1	8.5	8.4		5.6	8.8	12.3	

Case B

Indicators	Noise-to-Signal Method				Loss Function Method			
	1996	1997	1998	1999	1996	1997	1998	1999
IMF debt outstanding/quota					X	X	X	
IMF debt service/exports					X			
Non-IMF arrears/external debt	X	X	X		X	X	X	
Reserves/imports		X				X		
Change in reserves								
Export growth		X			X	X	X	
General government balance/GDP			X			X	X	
Inflation	X		X		X		X	
Economic growth		X	X			X	X	
Political risk			X		X	X	X	
Composite value	4.1	10.4	13.9		10.0	11.4	10.9	

1/ "X" indicates that the indicator value goes beyond the threshold value.

Up to two years prior to the incidence of arrears, Zimbabwe consistently ranks among the top five countries with the highest probability evaluated by either of the methods. The noise-to-signal method ranks Zimbabwe third in 1999 and fourth in 2000 while the loss function method ranks Zimbabwe fourth in 1999 and first in 2000.

In terms of false alarms, the loss function method performs better than the noise-to-signal method. In 1999, the former calls seven countries while the latter calls ten countries; in 2000, the former predicts two cases and the latter predicts six cases.

### **Probit model**

The probit model narrowly calls Zimbabwe's arrears. In 1998, the probability goes beyond the cut-off to 6 percent while it falls below the cut-off in the following year. In 2000, the probability rises to 5 percent, which is slightly above the cut-off. In 2001, the probability rises to 9 percent, but the data in 2001 would not have been available prior to the emergence of arrears in early 2001.

In terms of false alarms, performance is mixed. The probit calls 19 countries in 1998, of which Zimbabwe ranks 11<sup>th</sup>; in 1999, the probit calls eight countries while Zimbabwe is not called; in 2000, only three countries are predicted, of which Zimbabwe ranks second, after Case B.

### **B. Case B (1999)**

Case B incurred short-term arrears to the IMF for 79 days in late 1999. This occurred after the country emerged from protracted arrears to the IMF in the mid-1990s. Case B went through a civil war during the 1990s, which severely disrupted its economy. The short-term arrears in late 1999 came about as the government endeavored to regain economic stability. This incidence fell short of turning into a protracted arrears case owing to the government's effort to stabilize the economy and donor support that helped clear the arrears.

While all three models call Case B one to four years prior to the incidence, the noise-to-signal and loss function methods appear to indicate the deterioration more clearly than the probit model does.<sup>28</sup> Evaluated both in terms of calling the incidence and in issuing fewer false alarms, the loss function method performs the best.

### **Signals approach**

As shown in Figure 1, both the noise-to-signal and loss function methods perform well in capturing Case B's arrears. The noise-to-signal method calls Case B in 1995, 1997 and 1998,

---

<sup>28</sup> Strictly speaking, only the performance after 1998 serves as an out-of-sample test because the model is estimated with the sample covering up to 1997.

of which the probability in 1995 is the highest.<sup>29</sup> The probability under the loss function method consistently crosses its cut-off between 1995 and 1998.

Table 5 shows that a larger number of indicators are signaling under the loss function method than under the noise-to-signal method. In 1998, five indicators are signaling under the noise-to-signal method, and seven are issuing alarms under the loss function method.

In terms of the number of false alarms, the loss function method fares better than the noise-to-signal method. In 1997, the noise-to-signal method classifies nine countries as potential arrears cases, of which Case B ranks second. In 1998, it calls 19 countries, of which Case B ranks second. On the other hand, in 1997 the loss function method predicts five cases, of which Case B is assigned the highest probability. Also in 1998, the loss function calls six potential cases and gives Case B the highest probability.

### **Probit model**

The probit model has more mixed results in calling Case B's short-term arrears. While the probability between 1995 and 1997 is above the cut-off point, the probability declines as 1999 approaches. In 1998, the probability falls to 3.9 percent, slightly below the cut-off point. However, in 1999, the year of arrears emergence, the probability goes up to 8.7 percent, well above the cut-off point.

Unlike the signals approach, the probit model does not give high probabilities to Case B prior to 1999, primarily because the probit model covers fewer indicators and those indicators with greater weights, mainly IMF-specific debt ratios, did not deteriorate markedly until 1999. In terms of false alarms, the probit model calls 19 countries and it does not predict Case B in 1998. In 1999, it calls nine countries and assigns the second highest probability to Case B.

### **C. Comments**

After examining out-of-sample cases, there remains a question of the different predictions produced by the models. This issue concerns both the nature of the models and the particular incidence of arrears to the IMF.

First, there is an intrinsic difference between the signals approach and probit regressions. As Berg and Pattillo (1998) point out, the signals approach assumes that the probability of crisis, arrears in this case, in the subsequent period is a step function of the value of the indicator. Namely, the probability of event is zero when the variable is below the threshold and one, above the threshold. The drawback of this assumption is that it leaves out all the information

---

<sup>29</sup> Often times, countries emerging from protracted arrears face mounting economic and political problems as they strive to stabilize the economy. Case B, which had been in protracted arrears until the mid-1990s, is a case in point.

when the value of the indicator falls short of the threshold. In addition, once an indicator value crosses the threshold, the signals approach does not take into account how far the value goes beyond the threshold. Under the probit model, the probability of an event increases in proportion to the values of indicators.

Second, the number of indicators covered is different: the signals approach covers 10 indicators while the probit covers only five. This is because the probit model takes into account the correlations among the indicators, whereas the signals approach treats each indicator independent of the others. In general, the former is more rigorous and parsimonious, but it might leave out useful information provided by the other variables available under the signals approach.<sup>30</sup>

In the case of Zimbabwe, the loss function method of the signals approach performs the best in calling Zimbabwe's payment difficulties because the characteristics of the case better fit the specification of the approach. Zimbabwe experienced steady deterioration in various aspects of its economy and governance in the years leading up to the emergence of arrears: negative growth, fiscal imbalances, high inflation, dwindling reserves, and rising political risk, most of which are not covered by the probit model. On the other hand, the two Fund-specific debt ratios, which take up the largest weight under the probit model, were not at an alarming level per se. Conversely, if a sharp movement in the indicators were to set off the incidence of arrears, the probit model would be in a better position to capture it.

While the loss function method of the signals approach achieved the best out-of-sample result in this study, it would be premature to judge general out-of-sample performance based on such limited cases. Particularly, if we are concerned about missing potential arrears, we should monitor the output of all three models. According to the in-sample results, the loss function method issues the least number of false alarms, but it misses the largest number of arrears among the three models.

## VII. FURTHER WORK

Before concluding, it is useful to discuss some of the challenges that this study has faced and to highlight issues that could be further explored in subsequent studies. Perhaps the most difficult part of modeling IMF arrears is the limited sample, both in terms of time and scope, which might conflict with the two assumptions embedded in panel-data estimation: temporal stability and country heterogeneity.

---

<sup>30</sup> For instance, the political risk indicator is not included in the probit model because its strict treatment of missing data would severely limit the total number of observations available. See footnote 25 for details.

### **Temporal stability**

The overwhelming majority of arrears to the IMF occurred in the aftermath of the debt crisis of 1980s. The emergence of arrears has dwindled since the mid-1990s. In fact, there has been only one protracted arrears case since the mid-1990s. There is, then, an issue of temporal stability, whether the information obtained during the debt crisis years is still applicable to the current period. AT test a debt crisis dummy in their logit regressions and find it significant.

The changing nature of IMF financial assistance also affects the temporal stability of the model. The introduction of the Heavily Indebted Poor Country (HIPC) Initiative has helped prevent the emergence of IMF arrears among HIPCs.<sup>31</sup> Therefore, countries that might have incurred arrears to the IMF without HIPC assistance have managed to avoid falling into arrears.

A significant change in IMF lending during the late 1990s and early 2000s has been the sharp increase in the size of loans to a very small group of countries. Starting from the Mexican crisis of 1994-95, the IMF has considerably increased its exposure to countries faced with financial crises. Most of these unprecedented levels of loans were approved under the Supplemental Reserve Facility (SRF) to counteract destabilizing pressures from capital markets.<sup>32</sup> Since most of the protracted arrears in the past occurred with much smaller IMF loan packages, estimating models with that sample and then applying them to the countries with sizeable loans would inevitably raise their arrears probability to a very high level regardless of other indicator values. This leads to another specification issue: country heterogeneity.

### **Country heterogeneity**

Country heterogeneity accounts for the differences where one country may default when its indicator values reach certain levels while another country may continue to service its debt even when it experiences similar levels of strains.<sup>33</sup> Persistence of country heterogeneity with

---

<sup>31</sup> In 1996, the IMF and the World Bank jointly adopted the HIPC Initiative, a framework which envisages comprehensive debt relief by the international community, including multilateral institutions, to achieve debt sustainability of heavily indebted poor countries. In order to receive HIPC assistance, a country needs to build a track record of strong policy performance.

<sup>32</sup> The SRF was established in 1997 to provide financial assistance for members facing exceptional balance of payment difficulties due to a sudden and disruptive loss of confidence reflected in its depleting foreign reserves.

<sup>33</sup> See, for instance, Hajivassiliou (1998); McFadden and others (1985) for discussions on country heterogeneity.



panel data violates assumptions about the randomness of the error term. While it is difficult to take into account country-specific effects in a probit framework, one could experiment with fixed-effects logit models. Alternatively, one could improve the pooled estimation by categorizing countries into groups according to their characteristics, for instance their rankings of indicator values, and estimate different models for each group.

In this study, the sample of protracted arrears cases is limited to a small group of fairly poor countries: nearly half of the protracted arrears countries are now classified as HIPCs. Given this limited sample, applying the estimated models to over 100 developing countries might not yield very meaningful results, especially if the country in question is not representative of the past arrears cases. To address this issue, the coverage of countries could be more targeted, for instance, limiting the sample to PRGF-eligible countries, or the definition of arrears could be broadened, for instance, including short-term arrears and arrears to other official creditors.

## VIII. CONCLUSIONS

The purpose of this paper has been to test whether it is possible to anticipate arrears to the IMF in a systematic manner. Specifically, two of the leading early warning system models for currency crises, the KLR signals approach and the BP probit model, have been modified and applied to the incidence of protracted arrears to the IMF. The results, evaluated both in-sample and out-of-sample, are encouraging, albeit with some limitations.

Both the signals and probit models find that countries tend to incur arrears to the IMF when IMF debt outstanding is high relative to quota, the ratio of IMF debt service to exports is high, reserves relative to imports are low, export growth is sluggish, and arrears to other external creditors are on the rise. Also, the signals approach finds the following indicators to be useful: political risk, the government fiscal balance as a percentage of GDP, the rate of inflation, per capita income growth, and change in reserves.

In terms of in-sample performance, approximately 80 percent of arrears periods are called by any of the three models examined in this paper. While they issue a large number of false alarms, if our purpose is to identify a list of countries that require close monitoring, false alarms are less of a problem than missed arrears. In terms of in-sample performance, the probit model outperforms either method of the signals approach. While an out-of-sample test is limited to Zimbabwe and Case B, all models, in particular the loss function method of the signals approach, capture their payment difficulties one to four years in advance.

The models examined in this paper have important differences, both in terms of the embedded assumptions and data requirements. The probit model takes into account the correlations of the variables, uses fewer variables, and captures the magnitude of the indicator values. The signals approach, on the other hand, covers a wide range of variables irrespective of their correlations and evaluates whether the indicator values cross certain threshold values. Therefore, the signals approach, the loss function method in particular, is suited to predicting arrears cases where various aspects of the economy as well as

governance deteriorate, perhaps not dramatically but steadily, as in the case of Zimbabwe. The probit model is better at capturing a spike, or a dramatic movement in the indicator values.

The implementation of early warning models for IMF arrears would pose some challenges that are common to these types of exercises as well as those that are unique to IMF arrears. Temporal stability and country homogeneity that are assumed under probit estimation using panel data might be even more problematic than usual because IMF arrears are very rare events that mostly occurred during the 1980s and early 1990s. While models estimated from the past statistics are able to capture potential arrears that are characteristic of the past cases, such as Zimbabwe, they are not as well-equipped to predict cases where countries have large access to private capital and problems arise in sharp swings in capital flows. Despite these limitations, the models are useful tools for systematically identifying countries that experience serious external difficulties, and therefore, pose a high risk of incurring arrears to the IMF.

## REFERENCES

- Aylward, Lynn, Rupert Thorne, 1998, "An Econometric Analysis of Countries' Repayment Performance to the International Monetary Fund," IMF Working Paper 98/32 (Washington: International Monetary Fund).
- Berg, Andrew, Catherine Pattillo, 1998, "Are Currency Crises Predictable? A Test," IMF Working Paper 98/154 (Washington: International Monetary Fund).
- \_\_\_\_\_, 1999, "Predicting Currency Crises: The Indicator Approach and an Alternative," *Journal of International Money and Finance*, Vol. 18 (August), pp. 561-86.
- Berg, Andrew, and others, 1999, *Anticipating Balance of Payments Crises: The Role of Early Warning Systems*, IMF Occasional Paper No. 186 (Washington: International Monetary Fund).
- Berg, Andrew, and Jeffrey D. Sachs, 1998, "The Debt Crisis: Structural Explanations of Country Performance," *Journal of Development Economics*, Vol. 29, pp. 271-306.
- Bussi eres, Matthieu, and Christian Mulder, 1999, "Political Instability and Economic Vulnerability," IMF Working Paper 99/46 (Washington: International Monetary Fund).
- Demirguc-Kunt, Asli, and Enrica Detragiache, 1999, "Monitoring Banking Sector Fragility: A Multivariate Logit Approach," IMF Working Paper 99/147 (Washington: International Monetary Fund).
- Detragiache, Enrica, and Antonio Spilimbergo, 2001, "Crises and Liquidity: Evidence and Interpretation," IMF Working Paper 01/2 (Washington: International Monetary Fund).
- Feder, Gershon, and Lily V. Uy, 1985, "The Determinants of International Creditworthiness and Their Policy Implications," *Journal of Policy Modeling*, Vol. 7, pp. 133-56.
- Frankel, Jeffrey A., and Andrew K. Rose, 1996, "Currency Crashes in Emerging Markets: An Empirical Treatment," *Journal of International Economics*, Vol. 41, pp. 351-66.
- Goldstein, Morris, Graciela L. Kaminsky, and Carmen M. Reinhart, 2000, *Assessing Financial Vulnerability: an Early Warning System for Emerging Markets* (Washington: Institute for International Economics).
- Hajivassiliou, Vassilis, 1989, "Do the Secondary Markets Believe in Life After Debt?," World Bank Policy Planning and Research Working Paper No. 252 (Washington: World Bank).

- Hali, Edison J., 2000, "Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System," International Financial Discussion Paper No. 675 (Washington: Board of Governors of the Federal Reserve System).
- International Monetary Fund, 2000, "Debt- and Reserve-Related Indicators of External Vulnerability," *Staff Paper for Board Seminar* (Washington: International Monetary Fund). Available via the Internet at <http://www.imf.org/external/np/pdr/debtres/index.htm>.
- Kaminsky, Graciela L., 1998, "Currency and Banking Crisis: a Composite Leading Indicator," International Financial Discussion Paper No. 629 (Washington: Board of Governors of the Federal Reserve System).
- Kaminsky, Graciela, Saul Lizondo, and Carmen M. Reinhart, 1997, "Leading Indicators of Currency Crises," IMF Working Paper 97/79 (Washington: International Monetary Fund).
- Kaminsky, Graciela, Saul Lizondo, and Carmen M. Reinhart, 1998, "Leading Indicators of Currency Crises," *Staff Papers*, International Monetary Fund, Vol. 45 (March), pp. 1-48.
- Li, Carmen., 1992, "Debt Arrears in Latin America: Do Political Variables Matter?" *Journal of Development Studies*, Vol.28, pp. 668-88.
- McFadden, Daniel, and others, 1985, "Is there a Life after Debt? An Econometric Analysis of the Creditworthiness of Developing Countries," in *International Debt and the Developing Countries*, ed. by Gordon Smith and John Cuddington (Washington: World Bank).
- Mulder, Christian, Roberto Perrilli, and Manuel Rocha, 2002, "The Role of Corporate, Legal, and Macroeconomic Balance Sheet Indicators in Crisis Detection and Prevention," IMF Working Paper 02/59 (Washington: International Monetary Fund).
- PRS Group, International Country Risk Guide, various years. Available to subscribers via the internet at <http://www.icrgonline.com>.