IMF Working Paper

"Rules of Thumb" for Sovereign Debt Crises

Paolo Manasse and Nouriel Roubini

INTERNATIONAL MONETARY FUND

IMF Working Paper

Fiscal Affairs Department

"Rules of Thumb" for Sovereign Debt Crises

Prepared by Paolo Manasse and Nouriel Roubini[1]

Authorized for distribution by Manmohan S. Kumar

March 2005

Abstract

This Working Paper should not be reported as representing the views of the IMF.
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 contains an empirical investigation of the set of economic and political conditions that are associated with a likely occurrence of a sovereign debt crisis. We use a new statistical approach (Binary Recursive Tree) that allows us to derive a collection of "rules of thumb" that help identify the typical characteristics of defaulters. We find that not all crises are equal: they differ depending on whether the government faces insolvency, illiquidity, or various macroeconomic risks. We also characterize the set of fundamentals that can be associated with a relatively "risk free" zone. This classification is important for discussing appropriate policy options to prevent crises and improve response time and prediction.

---

[1] Paolo Manasse is a Professor of Economics at Università di Bologna and a Technical Assistance Advisor at the International Monetary Fund; Nouriel Roubini is an Associate Professor of Economics at Stern School of Business, New York University. We wish to thank Guido Ascari, James Daniel, Mark De Broeck, Christian Keller, Manmohan S. Kumar, Pier Carlo Padoan, Indira Rajaraman, as well as participants at seminars at Fiscal Affairs Department, International Monetary Fund, and Bocconi University.

Contents   Page

I. INTRODUCTION

Following the debt crises of the 1980s, sovereign debt defaults have become more frequent. Episodes of outright default include Russia, Ecuador, Argentina; in other cases, formal default was avoided via a debt restructuring under a coercive threat of default as in Ukraine, Pakistan, and Uruguay; and in other cases, default was averted through large scale IMF financial support as in Mexico, Brazil, and Turkey.

While there has been a significant amount of research regarding debt crises in general, and about the policy responses to these sovereign defaults,[2] the macroeconomic and structural weaknesses leading to them are still not properly understood; there is little comparative empirical work on the sovereign debt crises of the last decade. Many policymakers and analysts continue to use simple rules of thumb to judge risks and to assess fiscal sustainability (IMF, 2003), as well as the soundness of macroeconomic policies. Too often, these rules are not based on a rigorous quantitative analysis, and may miss some core elements that led to these sovereign debt crises.

Our aim is to provide answers to the following basic questions. What set of economic and political conditions is empirically associated with a likely occurrence of a sovereign debt crisis? Can one derive thresholds for vulnerability indicators that may signal a higher likelihood of a sovereign debt crisis? Part of the motivation for the paper stems from so-called surveillance failures, namely cases where international financial institutions, such as the IMF, as well as rating agencies, private sector agents, and academics failed to correctly assess the likelihood of a sovereign default.

In the paper, we use a new statistical approach and derive a set of "rules of thumb" that help identify the typical characteristics of defaulters. In the process, we identify empirically different typologies of debt crises. We find that not all crises are equal: they differ depending on whether the government faced insolvency, illiquidity, or various macroeconomic weaknesses and risks. This classification is crucial for discussing appropriate policy options for preventing crises and responding to them once they occur. For example, it is often argued that solvent but illiquid countries with large amounts of short-term debt may need IMF support to avoid a liquidity run or "roll-off" crisis. Conversely, highly indebted countries may face a debt crisis, unless there is a strong and credible fiscal consolidation. Also, it is argued that conditionality should set targets indicating that a country's macroeconomic fundamentals are heading towards a relatively "safe" zone. In the paper these concepts of liquidity crisis, insolvency crisis, crises triggered by weak macrofundamentals, and relatively "safe zones" are made precise. Unless the diagnosis is correct, it is hard to get the policy cure right.

---

[2] See for example Roubini and Setser (2004) for a systematic analysis of the crises in emerging market economies in the last decade, and on how they were resolved.

This empirical analysis is based on a dataset containing annual observations for 47 emerging market economies from 1970 to 2002. A country is defined to be in a "debt crisis" if it is classified as being in default by Standard & Poor's, or if it receives a large nonconcessional IMF loan (where "large" means in excess of 100 percent of quota). Standard & Poor's rates sovereign issuers in default when a government fails to meet principal or interest payment on an external obligation on due date (including exchange offers, debt equity swaps, and buy back for cash).

We employ the Binary Recursive Tree methodology (BRT) for classification and prediction.[3] BRT is a computer-intensive data mining technique that selects explanatory variables, their critical values, and their interactions in order to identify "safe" from "crisis-prone" types. The main conclusions of our empirical analysis are as follows.

First, out of 50 candidate variables, 10 predictor variables turn out to be sufficient for classification and prediction: total external debt/GDP ratio; short-term debt reserves ratio; real GDP growth; public external debt/fiscal revenue ratio; CPI inflation; number of years to the next presidential election; U.S. treasury bills rate; external financial requirements (current account balance plus short-term debt as a ratio of foreign reserves); exchange rate overvaluation; and exchange rate volatility.

Second, a relatively "safe" country type is described by a handful of economic prerequisites: low total external debt (below 49.7 percent of GDP); low short-term debt (below 130 percent of reserves); low public external debt (below 214 percent of fiscal revenue); and an exchange rate that is not excessively overappreciated (overvaluation below 48 percent).

Third, three major types of risks are identified: (i) solvency (or debt unsustainability); (ii) illiquidity; and (iii) macroexchange rate risks. The debt unsustainability risk types are characterized by: external debt in excess of 49.7 percent of GDP, and together with monetary or fiscal imbalances, as well as large external financing needs that signal illiquidity as an element of debt unsustainability. Liquidity risk types are identified by moderate debt levels, but with short-term debt in excess of 130 percent of reserves coupled with political uncertainty and tight international capital markets. Macroexchange rate risk types arise from the combination of low growth and relatively fixed exchange rates. Each of these risk types differ in their likelihood of producing a crisis.

The analysis has one important, albeit simple, implication for sustainability analysis. It shows that *unconditional* thresholds, for example for debt-output ratios, are of little value per se for assessing the probability of default. One country may be heavily indebted but have a negligible probability of default, while a second may have moderate values of debt ratios while running a considerable default risk. Why? Because the joint effects of short maturity, political uncertainty, and relatively fixed exchange rates make a liquidity crisis in the latter much more likely than a solvency crisis in the former, particularly if the large external debt

---

[3] The analysis employs the data mining software CART developed by Salford Systems.

burden goes together with monetary stability, a large current account surplus, and sound public finances.

The plan of the paper is the following. Section II contains a review of the literature. Section III describes the dataset. The Binary Recursive Tree methodology is reviewed in Section IV, and applied to the data in Section V. Section VI discusses a number of refinements, and the main conclusions and policy implications are discussed in Section VII.

II. REVIEW OF THE LITERATURE

The literature on sovereign debt crises falls into four broad categories: (i) theoretical models of sovereign debt and default; (ii) empirical studies of the determinants of debt crisis; (iii) empirical studies of the predictive power of credit ratings; and (iv) empirical studies of the determination of sovereign spreads. Most studies focus on a particular aspect of debt crises or particular determinants of default. This literature suggests a number of macroeconomic and other factors that influence the likelihood of sovereign debt servicing difficulties and default.

The theoretical literature highlights a variety of factors that can trigger sovereign default and debt crises. Thus, we briefly overview in this section what the literature suggests about which factors affect the likelihood of a debt crisis. On the one side, countries may be unwilling to repay their debt, based on a consideration of the relative costs and benefits of default. On the other side, countries may be unable to repay their debt because they are either insolvent or illiquid. In empirical applications, a host of macroeconomic and institutional variables have thus been used to assess willingness to pay, ability to pay and debt servicing difficulties caused by illiquidity.

Starting with the ability to pay, whether a sovereign is insolvent or not depends on its stock of debt relative to its ability to pay, measured, for example, by GDP, exports, or government revenues.[4] A sovereign is solvent, if the discounted value of future primary balances is greater or equal to the current net public debt stock. Likewise, a country is solvent, if the discounted value of future trade balances exceeds the current stock of net external debt. Flow imbalances, such as primary or overall fiscal deficits, or trade and current account imbalances matter as persistent flow imbalances lead to an accumulation of debt and are inconsistent with the intertemporal budget constraint; at some point primary surplus and trade surpluses are necessary to avoid insolvency. So, flow imbalances also affect ability to pay, for any given level of existing debt. GDP growth and terms of trade shocks also affect the ability to pay. The exchange rate regime and exchange rate misalignment impact these debt sustainability considerations because an overvaluation can cause an external imbalance that leads to debt accumulation. Moreover, a currency crisis triggered by overvaluation can lead

---

[4] See Roubini (2001) for a recent overview of debt sustainability and solvency; and Eaton and Fernandez (1995) for a systematic survey of the literature on sovereign debt. Hemming and Petrie (2002) present an extensive and broad discussion of the concept of fiscal vulnerability; the concept includes the failure to avoid excessive deficits and debt. The concept of fiscal sustainability is, for example, discussed in Hemming and Chalk (2000).

to severe balance sheet effects if a large part of the debt is in foreign currency; the stock of debt can sharply increase in real terms after a large currency crisis.

Willingness to pay depends on the relative costs of defaulting or continuing to service the debt.[5] The main costs of defaulting are loss of access to international capital markets and the potential output and trade costs of default. Low output growth does not only affect the ability to pay but also the willingness to pay. When growth is low, being cut off from capital markets is less costly. Openness can affect the costs of default and thus a country's willingness to default or not; more open economies will lose more from the economic disruptions of international trade triggered by default. Measures of macroeconomic policy stability, such as low inflation or low money growth, reflect policy credibility and predictability and thus influence investors' risk attitudes towards a country and their perceptions of the country's willingness to pay. Institutional and political factors affect policy credibility, as well as a government's willingness to pursue policies consistent with a sustainable debt path. Political regime change may lead to the emergence of a political party less committed to service the debt; thus, the nearing of election may trigger investors' flight and increase the likelihood of a crisis. Rule of law and respect of property rights signals that a country's government is more willing to service its debt.

A debt crisis can also occur if a country is illiquid rather than insolvent.[6] Hence, liquidity measures, such as short-term debt over reserves or M2 over reserves, are included in many recent models of currency and financial crisis that stress the risk of a liquidity run.[7] Other measures of debt servicing needs, such as the external financing gap or the interest burden of servicing the debt, may also proxy for liquidity needs and the ability to refinance one's debts.

Regarding the definition of a debt crisis, empirical studies use different crisis definitions depending on the specific research question and the information available in the data source used. A priori, there is no single empirical definition of what should constitute a sovereign default or a debt crisis. Some studies compile a list of debt crisis or default from case studies and anecdotal evidence (e.g., Beers and Bhatia, 1999; or Beim and Calomiris, 2001). Other studies rely on a more quantitative approach. For example, Detragiache and Spilimbergo (2001) define a country to be in a debt crisis if the country has arrears on external obligations towards commercial creditors in excess of 5 percent of commercial debt outstanding or has a rescheduling or restructuring agreement with commercial creditors. This definition does not differentiate between sovereign or private sector arrears and/or rescheduling due to data limitations. Another problem of this quantitative definition is that it might exclude some

---

[5] See Obstfeld and Rogoff (1990) for a more detailed discussion of willingness to pay and the costs of default. Note also that, in general, some variables–such as macroperformance measures and measures of the level and volatility of macropolicies-proxy at the same time for both the ability and willingness to pay.

[6] The two concepts, however, are not necessarily independent: for example Jeanne (2000) suggests that the inability to borrow at short-term maturities may reflect the government's perceived solvency risk.

[7] See Roubini and Setser (2004), chapters 2 and 3, for an overview of such models and a study of the role of illiquidity factors in recent episodes of debt crisis.

incipient debt crises that were only avoided by large-scale financial support from official creditors (IFIs and/or bilateral). A data source that provides uniformly compiled information on sovereign default is Standard & Poor's (2002) that defines a country to be in default as long as the sovereign is not current on any of its debt obligation.

Empirical studies of the determinants of debt crisis are closest in nature to an early warning signal model. Factors influencing the probability of a debt crisis occurring are identified by means of probit/logit regressions or a signal model. Most studies have focused on the debt crisis of the 1980s, but there are also some recent efforts that look at crises occurring in the 1990s.[8] Taken together, measures of solvency, such as the debt-to-GDP ratio, and measures of liquidity, such as short-term debt over reserves or exports and debt service over reserves or exports, are significant explanatory variables in addition to macroeconomic controls, such as real growth, inflation, exchange rate overvaluation, and the fiscal balance. Reinhart (2002) finds that in 84 percent of the cases in her sample, a debt crisis is preceded by a currency crisis. Hence, variables that are well-suited for predicting currency crisis should also have some explanatory power in models for sovereign default. Detragiache and Spilimbergo (2001) carry out a number of interesting tests. They find that short-term debt, debt service, and reserves enter their model separately and the null of equal coefficients is rejected. Using ratios such as short-term debt over reserves, therefore, imposes a restriction that is not supported by the data. They also find that short-term debt is endogenous to the model, as countries find it more and more difficult to borrow long term in the run-up to a debt crisis. While most studies use macroeconomic variables only in levels, Catão and Sutton (2002) also include measures of volatility in their model. Their model in-sample predictive power increases markedly when measures of terms of trade volatility, fiscal policy volatility, monetary policy volatility, and exchange rate policy volatility are added to a model containing real GDP growth, debt service over exports, net international reserves over debt, the fiscal balance, the U.S. interest rate, and the real effective exchange rate. Manasse, Roubini, and Schimmelpfenning (2003) estimate a logit model of sovereign debt crisis that includes a large set of emerging market economies for the 1970–2002 period; thus, they include the sovereign crises of the last decade in their sample. They identify macroeconomic variables reflecting both insolvency, illiquidity and other domestic and external macroeconomic factors that predict a debt crisis episode one year in advance. Their model predicts about three quarters of all crises entries while sending few false alarms.

Taken together, the existing literature suggests several factors that are at the core of an empirical model attempting to predict sovereign default:

Measures of solvency, such as public and external debt relative to the capacity to pay. Liquidity measures such as short-term external debt and external debt service, possibly as a ratio of foreign reserves or exports. Political, institutional and other variables capturing a country's willingness to pay. Macroeconomic variables such as real growth, inflation, exchange rate, etc., capturing both ability to pay and willingness to pay. Measures of external

---

[8] See for example Detragiache and Spilimbergo (2001) for a study including recent episodes.

volatility and volatility in economic policies. We, thus, use these various variables and measures in the empirical study in this paper.

While the tree methodology used in this paper has been used in a limited number of economic studies and even applied to the case of currency crises (see Ghosh and Ghosh, 2002, and Frankel and Wei, 2004) no previous study has used this methodology to assess the determinants of sovereign crises and to predict them.

A. The Data

The full dataset includes annual information on 47 economies with market access from 1970 to 2002 (Table 1).[9] The debt crisis indicator is derived from data provided by Standard & Poor's and data on IMF lending. Data on external debt and public debt is taken from the World Bank's Global Development Finance database (GDF), as well as from IMF sources. Data on public finance and other macroeconomic variables are taken from the IMF's World Economic Outlook database, as well as the Government Finance Statistics database (GFS).

A country is defined to be in a debt crisis if it is classified as being in default by Standard & Poor's or if it receives a large nonconcessional IMF loan defined as access in excess of 100 percent of quota. Standard & Poor's rates sovereign issuers in default, if a government fails to meet principal or interest payment on external obligation on due date (including exchange offers, debt equity swaps, and buy back for cash). A potential problem with this information is that it may not capture near-defaults or coercive debt restructurings that were only prevented through an adjustment program and a large financial package from the IMF.[10] We therefore augment the information obtained from Standard & Poor's with data on IMF nonconcessional lending from the IMF's Finance Department.[11] We use information on the loans approved, approval dates and the actual disbursement of the loans. Based on the information on IMF lending, a country is classified as being in debt crisis if a large nonconcessional loan is approved and a disbursement under this loan is actually made in the first year. The definition of debt crisis thus encompasses actual defaults on debt recorded by Standard & Poor's and "incipient" defaults that were avoided only through a large scale financial support from the IMF. Based on this definition, a country can be in debt crisis for an extended period of time. We define a large IMF loan as being in excess of 100 percent of quota; this threshold selects the top 10 percent of loans when ranked by the loan to quota ratio.

---

[9] For transition economies, the sample period is 1995–2002. Not every variable is available for all countries or for the full-time period.

[10] Recent examples of near-default avoided via a large IMF package are Mexico in 1995, Brazil in 1998 and 2001, and Turkey in 2000.

[11] Mainly lending via Stand By Arrangements (SBA) and Extend Fund Facilities (EFF).

Table 1. Countries and Default Episodes in the Full Sample

	Number of Crisis	Average Length	Years in Crisis	Crisis Episodes (entry–exit)
Algeria	1	6.0	6	1991–97
Argentina	3	5.0	15	1982–94, 1995–96, 2001–
Bolivia	2	6.5	13	1980–85, 1986–94
Brazil	3	5.3	16	1983–95, 1998–00, 2001–
Chile	1	8.0	8	1983–91
China	0	…	0	
Colombia	0	…	0	
Costa Rica	1	10.	10	1981–91
Cyprus	0	…	0	
Czech Republic 1/	0	…	0	
Dominican Republic	1	22.	22	1981–
Ecuador	2	8.0	16	1982–96, 1999–01
Egypt	1	1.0	1	1984–85
El Salvador	1	16.	16	1981–97
Estonia 1/	0	…	0	
Guatemala	1	1.0	1	1986–87
Hungary 1/	0	…	0	
India	0	…	0	
Indonesia	2	2.5	5	1997–01, 2002–
Israel	0	…	0	
Jamaica	3	4.7	14	1978–80, 1981–86, 1987–94
Jordan	1	5.0	5	1989–94
Kazakhstan 1/	0	…	0	
Korea, Rep. of	2	2.0	4	1980–82, 1997–99
Latvia 1/	0	…	0	
Lithuania 1/	0	…	0	
Malaysia	0	…	0	
Mexico	2	5.0	10	1982–91, 1995–96
Morocco	2	3.0	6	1983–84, 1986–91
Oman	0	…	0	
Pakistan	1	2.0	2	1998–00
Panama	1	14.	14	1983–97
Paraguay	1	7.0	7	1986–93
Peru	3	6.3	19	1976–77, 1978–81, 1983–98
Philippines	1	10.	10	1983–93
Poland 1/	0	…	0	
Romania 1/	0	…	0	
Russia 1/	1	3.0	3	1998–01
Slovak Republic 1/	0	…	0	
South Africa	4	1.8	7	1976–78, 1985–88, 1989–90, 1993–94
Thailand	2	1.0	2	1981–82, 1997–98
Trinidad and Tobago	1	2.0	2	1988–90
Tunisia	1	1.0	1	1991–92
Turkey	2	3.5	7	1978–83, 2000–02
Ukraine 1/	1	3.0	3	1998–01
Uruguay	3	2.0	6	1983–86, 1987–88, 1990–92
Venezuela	3	3.3	10	1983–89, 1990–91, 1995–98
			26	
Total	54	5.5	1	

Sources: IMF, Standard & Poor's, World Bank, and authors' calculations.
1/ Transition economy countries are included only from 1995 onwards.

B. Descriptive Statistics

The explanatory variables can be grouped into three sets: (i) macroeconomic fundamentals; (ii) variability indicators; and (iii) political economy variables. As to the former, we use various measures of external debt and public debt, measures of solvency and liquidity, regressors included in the IMF's early warning signals model of currency crises as there is a possible link between currency crisis and sovereign debt crisis, other macroeconomic variables, as well as fiscal flow variables. Table 2 gives the respective mean of the macroeconomic variables used in the analysis, distinguishing between full sample, no crisis episodes, years before a country enters a debt crisis, in-crisis years, and years before a country exits a crisis. In general, the path of means from no crisis to entry into crisis and finally exit from crisis is as expected.

The various measures of external debt (including debt servicing) are relatively low in no crisis years followed by another no crisis year. They increase in the year before crisis entry, and most measures increase even further within crisis. The measures drop again in the year before a country exits from crisis, though they are still higher than before the crisis. The measures of public external debt follow the same pattern, suggesting that public external debt is a possible driving force behind external debt developments (as in many countries a large fraction of external debt is public external debt).

The macroeconomic variables—including those from the IMF's currency crisis EWS— indicate a worsening of the macroeconomic situation in the run-up to a crisis and within a crisis, and an improvement in the situation when exiting from crisis. For example, the current account deficit increases in the year immediately preceding a crisis entry, stabilizes within the crisis, and improves further in the year before exiting a crisis. Real growth falters in the year before crisis entry while inflation spikes. The overall fiscal balance, as well as primary balance, deteriorate in the run-up to crisis. It is interesting to note that both the LIBOR as well as the U.S. treasury bill rate increase in years preceding a crisis, suggesting that tight monetary conditions in the G-7 area may reduce capital flows to emerging market economies and thus contribute to debt servicing difficulties (as it happened in 1982 for example).

The second set of variables are measures of volatility. We show in Table 2 the coefficient of variation calculated over a moving window of four years, for the surplus/GDP ratio, inflation, nominal and real exchange rate and the terms of trade. Interestingly, the volatilities of the real exchange rate and of inflation rise in the wake of a crisis, and again in the midst of a crisis, while falling on the verge of the exit.

Finally, political economy variables are shown in the bottom part of the table. The indexes of political rights, civil liberties and freedom status, compiled by Freedom House (2002) take value on a scale from one (most "free") to seven (least "free"). There seems to be no significant difference between in/out crises episodes. The same applies to the political constraint indexes (Henisz, 2000). These measure the number of player in the political arena with veto power, who can block reforms. They range from zero (no veto players) to one (impossible to reform the status quo). Again, there seems to be little difference between in

and out crisis episodes. The same applies to the typology of electoral systems. The most frequent electoral system across all cases turns out to be the number one, proportional representation. More action seems to stem from the number of years to next presidential election: entry and staying in crisis are on average associated with upcoming elections, possibly indicating that political uncertainty before elections plays a role in contributing to crises.

III. METHODOLOGY[12]

This section describes the Binary Recursive Tree methodology (BRT) for classification and prediction. This methodology has been applied to several fields, including engineering, medical diagnosis, genetics, meteorology, marketing, insurance, consumer credit. Topics have included market segmentation, credit risk assessment, quality control, spread of cancer, blood cell classification, infant mortality, wildlife management, air pollution alerts, speech recognition, and classification of radar images for the military. Developed by statisticians Breiman, Friedman, Olshen, and Stone (1984), it searches for patterns and relationships in the data, and is particularly suited for uncovering hidden nonlinear structures and variable *interactions* in complex datasets. "Complexity" includes considerations such as: high dimensionality, a mixture of data types, nonstandard data structure and, perhaps more challenging, nonhomogeneity, i.e., *different relationships between variables hold in different parts of the measurement space* (Breiman and others, 1984).

The process is binary because parent nodes (partitions) are always split into exactly two child nodes, and recursive because the process can be repeated by treating each child node as a parent. The key elements of a BRT analysis are a set of rules for: (i) splitting each node into two child nodes; (ii) deciding when to stop growing the tree; and (iii) assigning each terminal node to a class outcome (e.g., crisis vs. noncrisis).

To split a node into two child nodes, BRT always asks questions that have a "yes" or "no" answer. For example, the question for assessing the likelihood of a default on a consumer loan might be: his age ≤ 35? BRT's method is to look at all possible splits for all variables included in the analysis. For example, in a dataset containing 2,000 individuals and 50 observed characteristics, such as a credit score record, age, sex, education, income etc., BRT considers up to 2000 times 50 splits for a total of 100,000 possible splits.

The next activity is to rank order each splitting rule on the basis of a quality-of-split criterion. The default criterion, the Gini rule, essentially measures of how well the splitting rule separates the classes contained in the parent node and produces a more homogeneous subnodes. Once a "best" split is found for a node, BRT repeats the search process for each child node, continuing recursively until further splitting is impossible or stopped. Splitting is impossible if only one case remains in a particular node or if all the cases in that node are exact copies of each other (on predictor variables).

---

[12] The next two sections draw from http://www.salford-systems.com/ and from Breiman and others (1984).

Table 2. Mean of Variables Used in the Analysis

Current year Next year	All All	No crisis No crisis	No crisis Crisis	Crisis Crisis	Crisis No crisis	No. of Obs.
Total external debt over GDP	45.5	37.0	54.7	71.4	63.7	1,051
Total external debt over exports	290.7	239.3	359.3	455.9	350.2	1,053
Total debt service over GDP						
Total debt service over reserves						
Short-term external debt (RM, over GDP)	10.9	9.4	15.0	15.1	15.7	993
Short-term external debt (RM, over reserves)	1.7	1.2	2.9	2.9	2.2	948
Interest on short-term external debt over GDP	0.5	0.5	0.8	0.6	0.7	754
Interest on short-term external debt over reserves	0.1	0.1	0.2	0.1	0.1	754
Debt service on short-term external debt over GDP	5.3	4.8	6.9	6.4	7.1	1,050
Debt service on short-term external debt over reserves	0.8	0.7	1.5	1.2	0.9	1,050
Public external debt over GDP	32.2	25.5	36.4	53.0	46.5	1,051
Public external debt over revenue	1.7	1.3	1.9	3.0	2.3	827
Consolidated central government debt	47.5	46.4	38.2	57.3	54.0	462
Overvaluation	0.0	0.0	0.0	-0.1	0.0	799
Current account balance	-2.7	-2.7	-4.3	-2.6	-1.3	1,231
Reserves growth	19.1	20.8	-5.4	17.8	22.9	1,155
Export growth	12.0	13.8	4.9	6.4	7.8	1,270
M2 over reserves	5.6	5.3	7.9	6.2	6.2	1,185
Financing requirement	1.6	1.3	3.0	2.4	1.4	986
LIBOR	9.7	9.5	10.5	10.5	9.4	1,229
U.S. treasury bill rate	6.4	6.3	7.8	6.9	6.3	1,229
Inflation	54.6	17.5	241.1	169.6	84.9	1,273
Nominal GDP growth	55.9	22.8	249.4	148.6	96.0	1,276
Real GDP growth	4.1	4.8	1.8	2.1	2.2	1,276
REER growth	121.5	124.3	139.7	111.0	109.4	937
Import growth	10.0	12.3	5.3	4.8	6.9	902
FDI over GDP	1.7	1.9	1.1	1.0	1.5	1,024
Openness	71.2	71.3	64.1	72.1	72.5	903
Overall balance	-4.3	-4.4	-6.3	-3.8	-4.1	1,013
Primary balance	0.6	0.3	-0.9	2.0	1.5	616
Primary gap	6.6	5.7	23.1	59.0	-15.0	122
G.Gov revenue over GDP	24.3	25.4	22.7	20.1	24.2	1,013
Overall Balance/GDP variability	1.92	-2.53	-1.77	19.42	-1.70	824
Inflation variability	120.43	10.80	125.37	549.81	83.03	1,082
Exchange rate variability 2/	57.8	56.38	45.62	39.87	159.9	1,059
Real Exchange Rate Variability	2.87	1.49	3.76	6.34	3.39	753
Terms of trade variability	2.24	2.28	2.32	2.23	1.33	1,088
Index of political rights	3.52	3.63	3.44	3.18	3.20	1,130
Index of civil liberties	3.77	3.85	3.75	3.44	3.65	1,130
Index of freedom status	1.21	1.16	1.21	1.39	1.39	1,130
Index of political constraints III	0.24	0.24	0.23	0.26	0.31	1,229
Index of political constraints V	0.35	0.33	0.35	0.37	0.43	1,229
Year of parliamentary elections	0.19	0.18	0.24	0.23	0.20	1,276
Year of presidential elections	0.09	0.06	0.11	0.18	0.15	1,276
Year of parliamentary or presidential elections	0.21	0.20	0.24	0.27	0.24	1,276
Years to next parliamentary elections	2.73	2.94	2.67	2.06	2.16	1,071
Years to next presidential elections	5.02	6.47	3.15	2.18	2.39	660
Years since parliamentary wlections	2.02	1.81	2.47	2.60	2.00	901
Years since presidential elections	2.19	2.15	2.83	2.24	1.60	418
Electoral system 1/	1	1	1	1	1	1,210

Sources: IMF, Standard & Poor's, World Bank, and authors calculations.
1/ Mode of electoral system.
2/ Excludes Turkey.

When does the growing process stops? At each node, the algorithm calculates the gain from further splitting in terms of the reduction in the rate of misclassification. This is simply the percentage of type $i \neq j$ observations that are erroneously classified as type j. This number is compared to the cost of further splitting (proportional to the number of nodes in the tree). If costs exceed benefits, the process stops.[13]

Once a terminal node is found we must decide how to classify all cases falling within it. One simple criterion is the plurality rule: the group with the greatest representation determines the class assignment. The rules of class assignment can be modified from simple plurality to account for the costs of making a mistake in classification, to adjust for priors, and to adjust for over- or under-sampling from certain classes. These concepts are made more precise in the Appendix. There we also discuss an instructive early medical application of the technique for classifying patients suffering from heart attacks.

Properties
The BRT can be thought as a way to let the data select interaction dummy-variables with endogenous threshold values (e.g., age≤55 and sex=female). It displays a number of interesting properties. First, it is well suited for discovering context dependence, interactions and heterogeneity. Datasets with a "complex" structure are easily handled. Unlike parametric models, which are intended to uncover a single dominant structure in data, BRT is designed to work with data that might have multiple structures, in the sense that different relationships hold in different parts of the dataset. The methodology is therefore robust to the effects of outliers. Outliers among the independent variables generally do not affect BRT because splits usually occur at nonoutlier values. Outliers are often separated into nodes where they no longer affect the rest of the tree. This feature is particularly useful when dealing with emerging markets, where, particularly in crisis times, variables such as inflation and exchange rate depreciation take extraordinary values. Second, no model-specification search is necessary. BRT consider any number of candidate variable, and selects the relevant ones and their split values. This feature is also useful, when theory does not precisely identify the variable to be used (should short-term debt over GDP, export, reserves, total debt, etc., be used as an indicator of liquidity constraints?). Third, the procedure is nonparametric, as it does not require specification of a functional form for the exogenous variables.[14] In particular, results are invariant with respect to monotone transformations of the independent

---

[13] The CART (Classification and Regression Tree) algorithm is actually more sophisticated than that, as it does not stop in the middle of the tree-growing process: there might still be important information to be discovered by drilling down several more levels. Hence, first a maximal tree is grown and a set of subtrees are derived from it, by "pruning" branches backward. The best tree is determined by looking for the tree whose misclassification (net) error rates is lowest.

[14] Since the procedure is nonparametric and free from assumptions on probability distributions, no confidence intervals can be meaningfully attached to the threshold values. What can be done, however, is to estimate a logit model with dummy variables corresponding to node inequalities, on the right side, and to test for their significance. Results, generally confirming statistical significance of node dummies, are available upon request from the authors.

variables. Fourth, the methodology can deal with cases where the class structure depends on combinations of variables, since it allows for searching linear combination of splits. Fifth, missing values for predictors are handled very effectively. For each split in the tree, BRT develops alternative splits (surrogates), that is, variables that produce a similar allocation of observations in child nodes. These variables are used when the primary splitting variable is missing. Thus, a missing value does not imply that all the observations on a particular case need to be thrown away, as in regression analysis. The missing value is replaced by the best surrogate. BRT can be, therefore, effectively used with data that have a large fraction of missing values, as often is the case, for example, with fiscal data for developing countries . Finally, the algorithm is designed to avoid "over fitting" the model to the data, so that predictions remain accurate when applied to fresh data. When the data are insufficient for having a separate test sample, BRT proceeds by dividing the sample into 10 roughly-equal parts, each containing a similar distribution for the dependent variable. BRT takes the first nine parts of the data, constructs the largest possible tree, and uses the remaining 1/10 of the data to obtain initial estimates of the misclassification error rate of selected subtrees. The same process is then repeated (growing the largest possible tree) on another 9/10 of the data while using a different 1/10 part as the test sample. The process continues until each part of the data has been held in reserve one time as a test sample. The results of the ten mini-test samples are then combined. This procedure, called cross-validation, implies the procedure has a built-in capability of performing well on completely fresh data, even in the absence of an independent test sample.

The procedure has also a few shortcomings. First, unlike regression or probit/logit analysis, the individual *marginal* contribution of each variable to the probability of belonging to a class cannot be ascertained. This is because, unlike regression or probit/logit analysis, BRT assign a single probability to *all* cases belonging to the same node. Second, the procedure is not well suited to uncover "general" relationships that hold across the whole sample. As pattern discovery become progressively more local, sample-wide information is not used. The existence of a linear relationship between the target y and the predictor x would show up as multiple splits for variable x being selected consecutively.[15] Third, if one variable slightly outperforms another as a split, the latter may never appear in the final tree, despite being possibly more closely associated with class membership than other variables appearing down in the tree. Thus, one may incorrectly deduct that the omitted variable is not "important." This problem ("masking") is somewhat akin to having two significant predictor that happen to be strongly collinear. One drops out from the regression. The consequence for classification is that the variables appearing in a tree may be sensitive to changes in the sample or in the a priori distribution. A way out is to rank each variable by looking at the *potential* effect of the variable on classification, i.e., explicitly accounting for its ability to classify observations even when "masked" by the first-choice split. This produces a measure

---

[15] Related to this is the criticism that the procedure is only "one-step" optimal and not "overall" optimal. For example, suppose that the procedure produces a tree with ten terminal nodes. If one could search all possible partitions of the dataset in ten terminal nodes for the one partition that minimizes the sum of node "impurities" (see Appendix), the two result may be quite different (Breiman and others, 1984).

("importance") that is "robust" to changes in the sample or in the a priori distribution of crises (Section VIII).

IV. THE EMPIRICAL TREE

The BRT methodology selects the following 10 variables out of the 50 candidates listed in Table 2: total external debt in percent of GDP; short-term debt on a remaining maturity basis to foreign reserves; public external debt to government revenue; real GDP growth; inflation; the U.S. treasury bill rate; exchange rate overvaluation; exchange rate volatility; the ratio of external financing requirements to foreign reserves; and the number of years before a presidential election.[16]

The first rule splits the sample into two branches (Figure 1): (i) episodes with high external debt (more than 49.7 percent of GDP) go to the right—here the conditional crisis probability rises from 20.5 percent in the entire sample to 45.4 percent; and (ii) episodes with low external debt to the left—with default probability of 9.7 percent. Episodes of high debt (more than 49.7 percent of GDP) are further split into high/low inflation (larger/smaller than 10.5 percent). The former incur the largest default risk, 66.8 percent: see terminal node 14. More than half of all the crisis episodes in the sample satisfy these two simple conditions. For example, the high external debt plus high inflation criterion was met one year ahead of the crises in Jamaica, Egypt, Bolivia, Peru, Ecuador, Uruguay, Indonesia, Bolivia, Morocco, Turkey, South Africa, Uruguay, Brazil, and Venezuela. Terminal node 7 is second in terms of number of crisis episodes. Despite intermediate external debt levels (between 49.7 percent and 19 percent of GDP), the joint effect of short-term debt (exceeding 130 percent of reserves), relatively rigid exchange rates (low volatility) and political uncertainty (less than five years to the next presidential elections), conjure to raise the crisis probability to 41 percent.

By contrast, going down to the left towards terminal node 3, one finds that the circumstances that are more favorable for reducing risks are low external debt, low short-term debt to reserves on a remaining maturity basis (below 130 percent) and low public external debt to revenue (below 210 percent), coupled with the economy not being in recession. Under these circumstances, the likelihood of being in a crisis episode is just 2.3 percent. About 58.4 percent of all noncrisis episodes satisfy these conditions.

Based on the set of rules of this tree, the observations can be classified as crisis-prone or not crisis-prone. Observations in a particular node are classified as crisis-prone (not crisis-prone), if the within node share of crisis observations is higher (lower) than in the total sample. In our case, since we attach a large cost to missing crisis (see previous footnote), the critical

---

[16] This classification tree was obtained with the following specifications: (i) the chosen criterion to be minimized is the Gini index, discussed in the Appendix; (ii) the a priori distribution of crises is taken to be equal to the sample distribution; and (iii) the cost of missing a crisis is set seven times as large as the cost of missing a noncrisis.

threshold for classifying a node as "crisis" prone turns out to be 11 percent, that is below the proportion of crises in the sample 20.5 percent (see Appendix II in Manasse, Roubini, and Schimmelpfennig, 2003). As can be seen from Table 3, Column 4, any threshold between 2.3 percent and 40 percent would not affect our classification.

The tree has one particularly important implication for sustainability analysis. It shows that *unconditional* thresholds for debt output ratios are of little value per se for assessing the probability of default. Take nodes 7 and 11. The former has only moderate values of debt ratio, between 49.7 percent and 19 percent, but the probability of a crisis is high, 41 percent. The latter has a debt ratio of *at least* 49.7 percent, but the crisis probability is just 2 percent. Why? Clearly, other factors are at play. What makes node 7 risky is the compound effect of short maturity of the debt, political uncertainty and relatively fixed exchange rates. What makes node 11 safe, despite the large debt burden, is monetary stability, a large current account surplus, and relatively large fiscal revenues that guaranteed solvency on public debt.

An application to Colombia
One advantage of our approach is that it can be immediately applied for evaluating default in risks. Take the case of Colombia, 2004 (Arias, 2004). One may start asking:

- "Does total external debt exceed 49.7 percent of GDP?" Since the answer is "no" (its value is 48.6 percent), one moves down to the left and asks:

- "Is short-term debt over reserves above 130 percent?" Again, the answer is "no" (its value is 98 percent), and one proceeds to the left.

- Then "Is public external debt above 215 percent of revenue?," the answer is "no" (it is 100 percent) and one moves to the left.

- "Is the economy growth rate above -5.45?" "Yes" (it is 3.13 percent) and move to the right until terminal node 3 is reached. The result is that Colombia in 2004 is not crisis prone (has a crisis probability of 2.3 percent).

Figure 1. The Empirical Tree
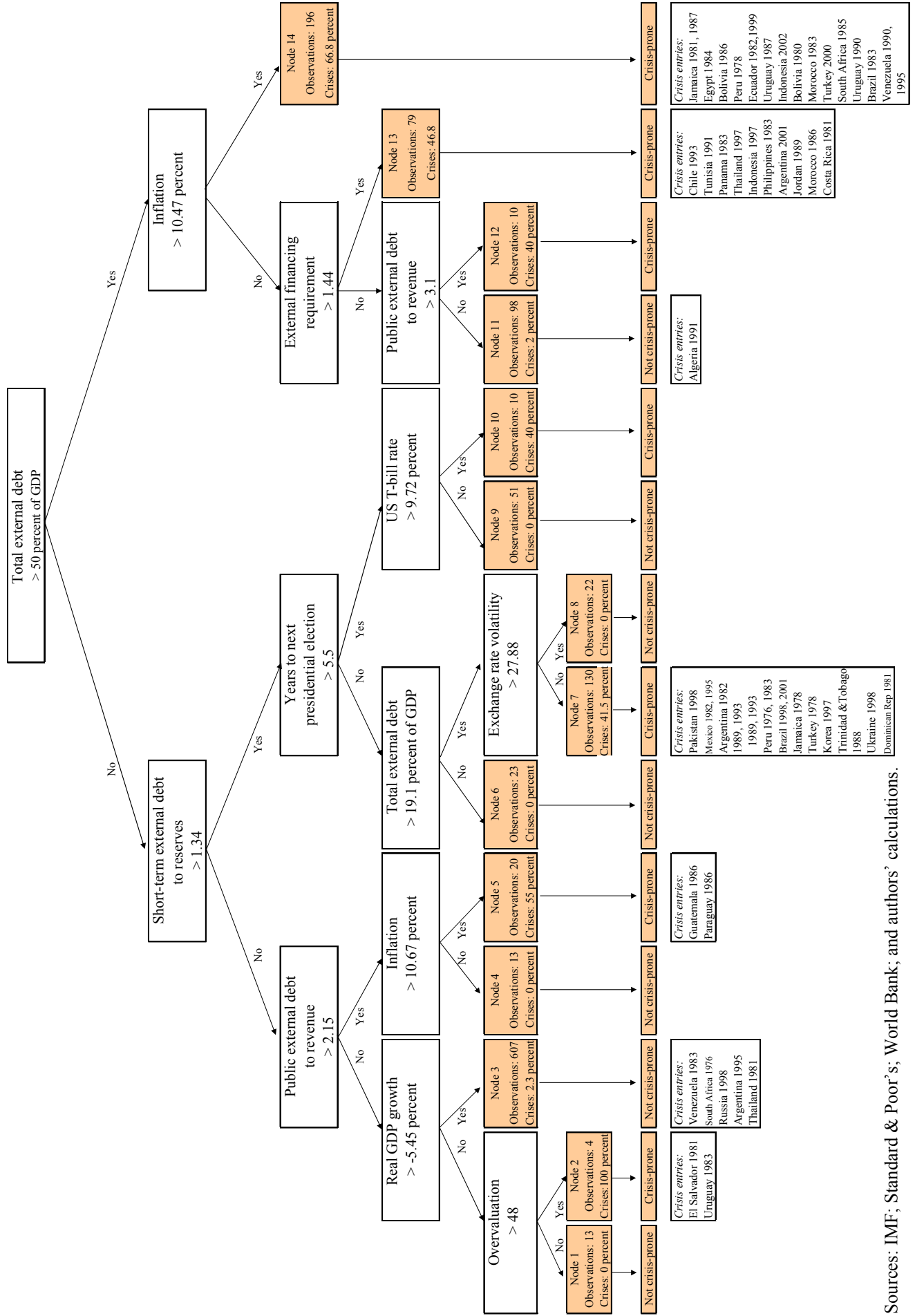
Sources: IMF; Standard & Poor's; World Bank; and authors' calculations.

V. CLASSIFICATION TABLE

The previous observation suggests an effective way to organize the tree's information (Table 3). The first column shows the terminal node number. The second column reports the set of inequalities satisfied by each node's observations: for example, row 14 of the table identifies node 14 containing all the cases (country-year) where total external debt/GDP exceeds the threshold of 49.7 percent and inflation exceeds 10.5 percent. Column 3 counts the number of observations that satisfy the node criteria: for example, looking at the row 14 and column 3, 196 (out of the total 1,276) observations satisfy the two inequalities on debt and inflation. The fourth column reports the probability of a crisis conditional on the node's inequalities being satisfied (i.e., the within node probability): for example the probability of a crisis next year, conditional on external debt and inflation exceeding 49.7 percent and 10.5 percent, respectively, is 66.9 percent. This probability is calculated as the ratio of crisis episodes in the node over the total number of observations in the node. The fifth column reports the probability that a crisis episode satisfies the inequalities in the node: for example, according to the last row and column 5, 50.2 percent of all debt crises satisfy the debt and inflation criteria. This probability is calculated as the ratio between the number of crises in the node over the total number of crises in the sample. The sixth column reports the probability that a no crisis observation falls in the node: for example, from the row 14, column 6, we see that 6.4 percent of no crisis episodes were characterized by large external debt and inflation. This entry is calculated as the ratio between the number of no crisis episodes in the node and the total number of no crisis episodes in the sample. For nodes identified as "crisis nodes" (see last column), this probability can be interpreted as a "type II error." Note, however, that this is an imperfect measure of the "safety" of the node, since a node (e.g., node 3) may contain plenty of noncrises just because it contains a lot of observations. In order to obtain an index of node "safety," therefore, we must normalize each entry by the ratio of observations in the node over total observations in the sample (divide by (607/1276) for node 3). The result is shown in the seventh column of the table. Here we show an index that takes value greater than one when the node is "safer" than the overall sample, and smaller when it is riskier. The index is perfectly (negatively) correlated with the crisis probability of column 4. The eighth column contains the predicted state (crisis=1, no crisis=0) based on the classification exercise.

Interpretation

For interpreting our results, it is useful to regroup the nodes into four blocks, that identify four different typologies. Block A can be interpreted as describing the characteristics of the "(relatively) safe fundamentals" type: observations in that block are classified as noncrisis-prone. Blocks B, C, and D identify potential defaulters, which are prone to risks of different sort. Observations therein are classified as crisis prone.

"Relatively Safe" Fundamentals, type (A). Nodes 1, 4, 6, 8, 9, 11, and 3, in descending order of "safety," show the prerequisites for a relatively risk-free environment. Low total external debt (below 49.7 percent) is the common denominator of these nodes (with the exception of partition 11). Low short-term debt over reserves (below 130 percent), low public external debt over revenue (below 215 percent), low inflation (below 10.5 percent) and not too strong

recession (growth above -5.5 percent) also characterize many (but not all) of these partitions. By far the more representative partition is node 3, that contains 607 episodes and 58 percent of tranquil nodes. This subset is characterized by low total and short-term debt, low public external debt over revenue (below 215 percent) and by not too negative growth (above minus 5.45 percent). Node 11 comes next in terms of observations. Node 11 shows that low external debt is by no means necessary for avoiding a crisis. Despite a large foreign debt, a country may still enjoy a relatively safe environment (2 percent crisis probability), provided inflation is under control (below 10.5 percent), external financial requirements over reserves are not too high (below 140 percent) and public external debt over revenue is not too large (below 310 percent). Node 3 shows that low external debt is by no means sufficient for avoiding a crisis. Despite relatively sound fundamentals as defined in node 3, a number of crises followed (see last column, row 3).

The nature of the inequality constraints in the following "crisis prone" nodes suggests an intuitive classification into solvency (B), liquidity (C), macroexchange rate (D) risks.

*Liquidity Crisis-Prone, type (*B) is described by nodes 7 and 10. Despite relatively low or intermediate external debt ratios, these episodes share a ratio of short-term debt to reserves in excess of 130 percent. This, coupled with political uncertainty (presidential elections in less than five years) and fixed exchange rate (low volatility), in node 7, or with tight monetary conditions in international capital markets (treasury bill rate above 9.7 percent), in node 10, raise the probability of crisis to about 40 percent. Solvency risks of the first type (node 7) make up more than one-fifth (20.7 percent) of all debt crisis episodes.

*Unsustainable Debt Path (Solvency) Crisis-Prone, type (*C) is identified by nodes 5, 12, 13, and 14. In all these cases either external debt exceeds 49.7 percent of GDP, or its public component exceeds 215 percent of revenues. In partition 14, high total external debt, together with inflation above 10.5 percent, raises the likelihood of a crisis from 20.5 percent (entire sample) to almost 67 (55) percent. About half of all crises episodes satisfy these characteristics (fifth column). Note that the high inflation rate could be here a proxy for fiscal problems: inability to reduce deficit may need to monetization of them and seignorage financing of them. Node 5 is similar, but a solvency crises (4.2 percent of total) are signaled by public, rather than total, external debt and high inflation. In node 13 (12), high external debt and high financing external requirement (or fiscal imbalances measured by high public external debt over revenue), raise the default probability to 47 percent (40 percent). These two nodes jointly account for another 15.7 percent of crises. The large external financing need can be partly interpreted as a illiquidity problem: this financing need is high when the current account deficit is high and/or when there is a large amount of debt coming due; in either case, the country may be illiquid if creditors do not provide new financing and/or do not rollover their debt. Thus, crisis episodes in node 13 can be interpreted as cases where a country is not necessarily "insolvent" but rather has an unsustainable—and nonfinanceable— debt path given large stocks of debt and illiquidity measured by large financing needs. Finally, the observations in node 1 should be interpreted as outliers: in a few cases (13) where the exchange rate was not over appreciated, values of external debt were small and there were no liquidity problems, there were no crises *despite* a strong recession

Exchange Rate and Macrocrisis-prone, type (D) A small number of sovereign debt crises (1.5 percent) in node 2 appear to be driven by large exchange rate overvaluation (above 48 percent) and large recession (negative growth below -5 percent). These are the crises of El Salvador, 1981–82, Uruguay,1983 (the collapse of the Tablita, the forward-looking crawling peg regime of some Southern Cone economies in the early 1980s), and Venezuela, 1984.

This classification of crises captures quite well most of the crises of the 1990s. Korea (1997), Mexico (1995), Brazil (1998) and (2001), were in the illiquidity node 7; and indeed these were typical episodes of a liquidity crisis, not insolvency. Pakistan (1998) was also in this node; its debt levels were not excessive but the country was illiquid given large lumpy debt servicing payments coming due and its loss of market access. Pakistan was then forced to restructure its external sovereign debt as, unlike large more systemic countries, it did not receive a large IMF package. Ukraine in 1998 (also in node 7), was a similar case of a solvent but illiquid country that was forced coercively to restructure its external debt by extending its maturities at below crisis level market rates. Clear insolvency cases in node 14 were Ecuador (1999), which defaulted on its external debt, Indonesia (2002), with a very large debt burden (but it is not illiquid given its restructuring of many external debt claims, both public and private), and Turkey (2000) with a very large debt burden. Turkey did not default thanks to an exceptional IMF program, but based on this classification was borderline insolvent.[17] Russia (1998) and Argentina (1995) appear in the noncrisis prone node 3; so, these crises are not well forecast by the tree. In the case of Argentina, the Tequila contagion from Mexico in 1994 played an important role as other fundamentals (debt levels and deficits) were fine at that time. Russia was not insolvent based on debt levels and external flow imbalances (current account deficits) but its failure to make flow fiscal adjustments and large capital flight led to default in 1998.[18] Node 13 of unsustainable debt path with possible illiquidity included Argentina (2001), which did default, Indonesia (1997), which defaulted on many private external debts, and Thailand (1997) that did not default on most external debt (only on some private claims), but had severe debt servicing problems in its private financial and corporate sectors. So, these are cases of outright insolvency or high illiquidity that made the debt path not sustainable/financeable.

---

[17] Given the lack of good data on primary balances, our classification does not capture well cases, such as Turkey, where the country would be deemed as insolvent based on a debt level criteria, but may be solvent as it is running a very large primary surplus (above 5 percent of GDP currently), that is stabilizing and reducing such high debt level.

[18] In principle, we may have included a variable measuring "contagion" effects from crises elsewhere. In practice, however, implementation in our context is problematic, since contagion typically occurs within a year, and we employ (one period lagged) data at annual frequency.

Table 3. Classification Table

Column 1 / Terminal Node	2 Rules and Critical Thresholds	3 N. Episodes In Node	4 Crisis Prob In Node	5 % of Crises In Node	6 % of NonCrisis In Node	7 Relative Safety of Node	8 Classified Crisis (1) NonCrisis(0)	9 Crises (entry)
A. Relatively Safe								
1	TotExtDebt/GDP ≤ 49.7, ShortTDebt/Res ≤ 1.3, PublicExtDebt/Revenue ≤ 2.1, GDPGrowth ≤ -5.5, EchROverV ≤ 48	13	0.0	0.0	1.3	1.26	0	
4	TotExtDebt/GDP ≤ 49.7, ShortTDebt/Res ≤ 1.3, PublicExtDebt/Revenue > 2.1, Inflation ≤ 10.7	13	0.0	0.0	1.3	1.26	0	
6	TotExtDebt/GDP ≤ 19.1, ShortTDebt/Res > 1.3, YearNextPresElect ≤ 5.5	23	0.0	0.0	2.3	1.26	0	
8	TotExtDebt/GDP ≤ 49.7, TotExtDebt/GDP > 19.1, ShortTDebt/Res > 1.3, YearNextPresElect ≤ 5.5, EchRVolatil > 27.9	22	0.0	0.0	2.2	1.26	0	
9	TotExtDebt/GDP ≤ 49.7, ShortTDebt/Res > 1.3, YearNextPresElect > 5.5, USTresBill ≤ 9.7	51	0.0	0.0	5.0	1.26	0	
11	TotExtDebt/GDP > 49.7, Inflation ≤ 10.5, ExtFinReqs/Reserve ≤ 1.4, PublicExtDebt/Revenue ≤ 3.1	98	2.0	0.8	9.5	1.23	0	Algeria (1991)
3	TotExtDebt/GDP ≤ 49.7, ShortTDebt/Res ≤ 1.3, PublicExtDebt/Revenue ≤ 2.1, GDPGrowth > -5.5	607	2.3	5.4	58.4	1.23	0	Venezuela (1983), South Africa (1976), Russia (1998), Argentina (1995), Thailand (1981)
B. Liquidity Risk								
10	TotExtDebt/GDP ≤ 49.7, ShortTDebt/Res > 1.3, YearNextPresElect > 5.5, USTresBill > 9.7	10	40.0	1.5	0.6	0.75	1	
7	TotExtDebt/GDP ≤ 49.7, TotExtDebt/GDP > 19.1, ShortTDebt/Res > 1.3, YearNextPresElect ≤ 5.5, EchRVolatil ≤ 27.8	130	41.5	20.7	7.5	0.73	1	Pakistan (1998), Mexico (1982, 1995), Argentina (1982, 1989, 1993), Peru (1976, 1983), Brazil (1998, 2001), Jamaica (1978), Turkey (1978), Korea (1997), Trinidad & Tobago (1988), Ukraine (1988), Dominican Republic (1981)
C. Solvency Risk								
12	TotExtDebt/GDP > 49.7, Inflation ≤ 10.5, ExtFinReqs/Reserve ≤ 1.4, PublicExtDebt/Revenue > 3.1	10	40.0	1.5	0.6	0.75	1	
13	TotExtDebt/GDP > 49.7, Inflation ≤ 10.5, ExtFinReqs/Reserve > 1.4	79	46.8	14.2	4.1	0.67	1	Chile (1993), Tunisia (1991), Panama (1983), Thailand (1997), Indonesia (1997), Philippines (1983), Argentina (2001), Jordan (1989), Morocco (1986), Costa Rica (1981)
5	TotExtDebt/GDP ≤ 49.7, ShortTDebt/Res > 1.3, PublicExtDebt/Revenue > 2.1, Inflation > 10.6	20	55.0	4.2	0.9	0.57	1	Guatemala (1986), Paraguay (1986)
D. Fiscal, Exchange rate and Macrorisks								
14	TotExtDebt/GDP > 49.7, Inflation > 10.5	196	66.9	50.2	6.4	0.42	1	Jamaica (1981, 1987), Egypt (1984), Bolivia (1986), Peru (1978), Ecuador (1982, 1999), Uruguay (1987), Indonesia (2002), Bolivia (1980), Morocco (1983), Turkey (2000), South Africa (1985), Uruguay (1990), Brazil (1983), Venezuela (1990, 1995)
2	TotExtDebt/GDP ≤ 49.7, ShortTDebt/Res ≤ 1.3, PublicExtDebt/Revenue ≤ 2.1, GDPGrowth ≤ -5.5, EchROverV > 48	4	100.0	1.5	0.0	0.00	1	El Salvador (1981), Uruguay (1983)
Total		1276		100.0	100.0			

Source: Author's Calculations.

VI. PREDICTION

In this section we discuss the prediction accuracy of our classification kit. Table 4 distinguishes between crisis state and crisis entry, the latter being defined as a crisis state preceded by a noncrisis state, and similarly for exits. Given the long persistence of default states, we expected more difficulties in predicting entries/exits rather than states. The tree performs quite well with respect to standard early warning system models: it correctly predicts 82.9 percent of states, 93.9 percent of crisis, and 79.9 percent of noncrisis states; moreover, it does even better in predicting crisis entries (88.9 percent) than states, while sending a limited number of "false alarms".

Table 4. Prediction

	Full Sample	1990's Onwards
Observations	1276	556
Number of crisis episodes	261	114
Number of crisis entry episodes	54	20
	(In percent)	
Correctly called episodes	82.8	78.8
Correctly called entries into default	88.9	85.0
Incorrectly called entries into default	18.5	15.0
Correctly called exits from default	32.0	35.3
Incorrectly called exists from default	4.8	64.7
Correctly called default episodes	93.9	89.5
Correctly called nondefault episodes	79.9	76.0

Sources: IMF, Standard & Poor's, World Bank; and authors' calculation.

VII. WERE THE 1990S DIFFERENT?

One common opinion is that, because of growing financial and goods market integration, the more recent crises in the last decade are "fundamentally" different from those of the 1980s: "capital account" crises rather than traditional "current account" crises. If this were true, our characterization of the critical variables, as well as their thresholds, may not be appropriate, as it was derived from a sample that goes back to the early 1970s. We ask, therefore, how well does our model, based on the entire sample, predict the crises of the 1990s. The second column from Table 5 shows that the model is as good as predicting the most recent episodes as well as those before 1990.

As we discussed in Section III, BRT is in principle suited for dealing with heterogeneous observations: if episodes in the 1990s were systematically different from those in the rest of the sample, they would be separated out in some node, and would not affect the rest of the classification, unlike traditional regression analysis. Thus, we perform another test. First, we check whether the observations from the 1990s fall disproportionately in some node. Second,

we build a classification tree for the more recent period and compare the result with the ones already obtained.

Table 5. Distribution of Observations in Nodes, 1990s vs. Full Sample

Node	Type	N(t)1990/N(t) / [N1990 s/N]
1	Safe	0.936654
2	Solvency	0
3	Safe	0.785689
4	Safe	0.468327
5	Solvency	0.710296
6	Safe	0.882355
7	Liquidity	1.202039
8	Safe	1.383694
9	Safe	0.119377
10	Liquidity	0
11	Safe	1.573834
12	Solvency	1.047987
13	Solvency	0.976175
14	Solvency	0.828334

Source: Authors' calculations.

We start by calculating the percentage of observations from the 1990s in each node *t*, *N(t)1990/N(t)*, and express it as a ratio of the sample observations from the 1990s, *N1990/ N*. This measure equals one if the node contains observations from the 1990s in the same proportion of the whole sample, exceeds one or falls below one if the 1990s are over/under represented in the node. The result is shown in Table 4. Solvency crises (nodes 5, 12, 13, and 14) appear in similar proportions in the 1990s as in the whole sample. Some liquidity crises are over (node 7) and some (10) are under represented, and the same holds for safe zone. Our conclusion is that there seem to be no discernible pattern in terms of over/under representation of the observation from the 1990s.

Next, we want to compare our "full sample" classification trees with one obtained exploiting data only since 1990. This is not a straightforward exercise, however, because of the "masking" problem discussed in Section III. This problem implies that the variable specification in the optimal tree may be sensitive to small changes in the sample, even if the overall classification ability of each variable is approximately constant. Rather than reporting a second tree, we calculate a measure of how well each variable does in separating crises from noncrises in both classifications. The "variable importance" index discussed in Breiman et al (1984), essentially equals the change in the "purity measure" of child nodes with respect to the parent node that is obtained by each variable's split (Appendix).[19] The index, and its

---

[19] Interestingly, both the actual and *potential* effects of the variable on classification is explicitly accounted for in CART, so one variable retains *importance* (and its ability to classify observations is accounted for) even

(continued…)

rank, is shown in Table 6 for the classification tree of the full sample, as well for that calculated with post-1990 data. The most important variable, the one that produces purest child nodes, is normalized to 100. Seven (eight, counting the different short-term debt measures) out the ten most important variables for the full sample, also appear in among first ten variables in the restricted sample. The correlation coefficient between the two measures is 0.47, and the rank correlation is 0.61. We conclude that our general tree deals effectively with potential sample heterogeneity.

VIII.   OTHER EXTENSIONS

A number of possible extensions come to mind, some of which we have tried. First, following the idea the a "history" of past defaults may bear on the credibility of a sovereign and thus affect the default probability (as suggested by Reinhart and others, 2003, in their "debt intolerance" hypothesis), we constructed a new variable taking on incremental values each time a new default episode occurred. Our indicator of (bad) "reputation" did not affect the classification tree.

Finally, one can imagine that the economic fundamentals underlying staying in (out) or getting into (out) a crisis may be different. We have repeated the analysis distinguishing four different states, staying in/out a crisis, and getting into/exiting from a crisis. So far, this approach has not proved particularly effective: in particular, it is not easy to disentangle the staying in (out) from the entry (exit) states. This is consistent with the results in Manasse, Roubini, and Schimmelpfenning (2003), who find, in a logit model, that most variables coefficients for entry (exit) do not differ significantly from those of staying in (out). However, some new interesting, albeit, preliminary conclusion seems to emerge for exits and entries: *exits* seem to be typically associated with (regained) fiscal solvency and floating exchange rates, while *entries* seem to be typically associated with liquidity and monetary problems. These preliminary conclusions clearly deserve more research.

---

when it does not explicitly appears in the classification tree, because *masked* by the first-choice split (see previous discussion and Appendix for details).

Table 6. Variable Importance: Full Sample vs. Post-1990s

	Importance 1970–2002	Rank 1970–2002	Importance 1990–2002	Rank 1990–2002
Tot Ext Debt/GDP	100	1	ExtFinRequir	100	1
Pub Ext Debt/GDP	88.349	2	ShTermExtDebt/Res(o.mat)	72.56	2
Tot Ext Debt/Exp	69.352	3	PubExtDev/Res	71.255	3
Sh TermExtDebt/Res (rem mat)	38.869	4	Inflation	56.891	4
ShTermExtDebt/Res (orig mat)	34.55	5	ShTermExtDebt/GDP(rem.mat)	54.272	5
PubExtDev/Res	33.707	6	ShTermExtDebt/GDP(or.mat)	50.347	6
ShTermExtDebt/GDP(rem.mat)	29.849	7	Pub Ext Debt/GDP	46.717	7
ExtFinRequir	28.233	8	IntShTermExtDebt/Res	42.618	8
Nominal GDPGrowth	25.503	9	Nominal GDPGrowth	38.491	9
Inflation	24.748	10	IntShTermExtDebt/GDP	37.617	10
ServExtDebt/GDP	24.398	11	ReserveGrowth	37.337	11
DebtServExtDebt/Res	20.893	12	Openness	31.866	12
M2/Res	19.728	13	Tot Ext Debt/GDP	23.991	13
Overall Exch Rate	15.562	14	DebtServiceExtDebt/Res	23.372	14
US TreasBill Rate	15.479	15	PoliticConstraintIndex	21.737	15
YearsNextPresElec	14.298	16	FDI/GDP	18.724	16
IntShTermExtDebt/Res	13.333	17	M2/Reserves	15.777	17
ResGrowth	11.946	18	Inflation Vol	15.297	18
DebtServExtDebt/GDP	11.606	19	CurrAcc/GDPY	13.977	19
Exch Vol	11.434	20	TotExtDebt/Exports	13.709	20
Inflation Vol	11.329	21	RealExcfRate Vol	13.667	21
Real GDP growth	8.556	22	ReqAdjPrimBalance	12.249	22
Revenue/GDP	8.346	23	Civil Libert Index	8.847	23
Openness	7.786	24	Rev/GDP	6.78	24
RealEchRateVol	7.573	25	PublicDebt/GDP	6.305	25
Politic Constr Index V	6.996	26	Libor Rate	5.139	26
PolitcConstr Index	6.996	27	USTreas Bill Rate	5.139	27
ExpGrowth	5.975	28	YeardSincePresElect	3.417	28
FDI/GDPY	5.341	29	Political Rights Index	2.893	29
Libore RateLIBOR	4.45	30	Freedom Status Index	1.628	30
Civil Liberty Index	3.466	31	TermsofTrade Vol	1.369	31
M2 Growth	3.202	32	Overval Exch Rate	1.273	32
PoliticalRights Index	2.951	33	YearsNextPresElections	0.947	33
YearsNextParlElect	2.23	34	Exch Rate Vol	0.941	34
Electoral System	2.211	35	ServExtDebt/GDP	0.66	35
IntShTermExtDebt/GDP	2.15	36	PrimaryBalance/GDP	0.002	36
GenGovBal/GDP Volat	1.331	37	Sh ermExtDebt/GDP(rem mat)	0	37
Res Growth	0.494	38	M2 Growth	0	38
YearsSincePresElect	0.306	39	RealExchRate Deprec	0	39
CurrAcc/GDPY	0.097	40	Political Constr Index V	0	40
YearsSinceParlElect	0	41	ParliamElection Year	0	41
ReqAdjPrimBalance	0	42	PresidElectYear	0	42
FreedomStatusIndex	0	43	Parl or Pres ElecYear	0	43
PrimaryBalance/GDP	0	44	Years NextParlElect	0	44
RealExchRateDep	0	45	ResGrowth	0	45
ParlamElectYear	0	46	GenGovBal/GDP Volat	0	46
TermsofTrade Vol	0	47	Exp Growth Rated	0	47
PublicDebt/GDP	0	48	RealGDPGrowth	0	48
PresElectionYear	0	49	YearsSinceParlElect	0	49
Pres orParlElectionYear	0	50	Eectoral System	0	50
Correlation=0.466689			Rank correlation=0.614502	

Source: Author's Calculations.

IX. CONCLUSIONS

In this paper, we applied a new statistical methodology to the question of understanding sovereign debt crises, both in terms of fundamentals that lead to a crisis, and of the factors that allow us to predict such crises. This tree technique allows us to derive endogenously the most important factors that lead to vulnerability in sovereign debt crises and the thresholds that signal greater risk of a crisis.

We find that most debt crises can be classified into three types: i) episodes of insolvency (high debt and high inflation) or debt unsustainability due to high debt and illiquidity; ii) episodes of illiquidity, where near default is driven by large stocks of short-term liabilities relative to foreign reserves; and iii) episodes of macro and exchange rate weaknesses (large overvaluation and negative growth shocks). Conversely, a relatively "risk-free" country type is described by a handful of economic characteristics: low total external debt relative to ability to pay, low short-term debt over foreign reserves, low public external debt over fiscal revenue, and an exchange rate that is not excessively overvalued. Political instability and tight monetary conditions in international financial markets aggravate liquidity problems. The approach suggests that *unconditional* thresholds—for example, looking at debt to output ratios in isolation—are of little value per se for assessing the probability of default; it is the particular combination of different types of vulnerability that may lead to a sovereign debt crisis.

The predictive power of the tree approach is quite good with very few crises missed or mistyped; the model predictive accuracy does not suffer when applied to the post-1990 experience. However, type II errors (false alarms) are somewhat higher than desirable. The tree approach allows us to adjust the relative weights given to type I and type II errors; thus, one future challenge is to calibrate the model to reduce false alarms while maintaining a high predictive ratio for actual crises.

The tree approach is also very useful both to derive rules of thumb or vulnerability thresholds, which may be useful to predict crises early on (an "early warnings signal" model), and in deriving policy adjustment paths, which may reduce the likelihood of a crisis for countries that may be entering in a danger zone. Thus, ideally, this tool can be used for surveillance, crisis prevention, and also crisis resolution.

References

Arias, Andres F., 2004, "Comments on the paper 'Assessing Fiscal Sustainability;' by Enrique G. Mendoza and Pedro Marcelo Oviedo." Available via the Internet http://www.iadb.org/res/centralbanks/publications/cbm35_199.ppt

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone, 1984, *Classification and Regression Trees*, (London: Chapman & Hall).

Beers, David T., and Ashok Bhatia, 1999, "Sovereign Defaults: History," in *Standard & Poor's Credit Week*, December 22.

_____, and John Chambers, 2002, "Sovereign Defaults: Moving Higher Again in 2003?," *Standard & Poor's* (September). Available via the Internet http://www.emta.org/keyper/sov2003.pdf

Beim, David O., and Charles W. Calomiris, 2001, *Emerging Financial Markets*, (New York: McGraw-Hill/Irwin).

Cantor, Richard, and Frank Packer, 1996, "Determinants and Impact of Sovereign Credit Ratings," *Federal Reserve Bank of New York Policy Review* (October), pp. 37–52.

Catão, Luis, and Bennett Sutton, 2002, "Sovereign Defaults: The Role of Volatility," IMF Working Paper 02/149 (Washington: International Monetary Fund).

Corsetti, Giancarlo, Bernardo Guimarães, and Nouriel Roubini, 2003, "The Tradeoff Between an International Lender of Last Resort to Deal with Liquidity Crisis and Moral Hazard Distortions: A Model of the IMF's Catalytic Finance Approach," IMF Seminar Series Research Paper 03/149 (Washington: International Monetary Fund).

Cottarelli, Carlo, and Curzio Giannini, 2002, "Bedfellows, Hostages, or Perfect Strangers? Global Capital Markets and the Catalytic Effect of IMF Crisis Lending," IMF Working Paper 02/193 (Washington: International Monetary Fund).

Dell'Arricia, Giovanni, Isabel Schnabel, and Jeromin Zettelmeyer, 2002, "Moral Hazard and International Crisis Lending: A Test," IMF Working Paper 02/181 (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).

Eaton, Jonathan, and Raquel Fernandez, 1995, "Sovereign Debt," NBER Working Paper 5131, Prepared for *Handbook of International Economics*.

Frankel, Jeffrey, and Shang-Jin Wei, 2004, "Managing Macroeconomic Crises: Policy Lessons," NBER Working Paper 10907, November. Available via the Internet http://ksghome.harvard.edu/~jfrankel/Managing%20Macroeconomic%20Crises%20Policy%20Lessons.pdf

Ghosh, Swati, and Atish Ghosh, 2002, "Structural Vulnerabilities and Currency Crises," IMF Working Paper 02/9 (Washington: International Monetary Fund).

Haque, Nadeem U., Mark Nelson, and Donald J. Mathieson, 1998, "The Relative Importance of Political and Economic Variables in Creditworthiness Ratings," IMF Working Paper 98/46 (Washington: International Monetary Fund).

Hemming, Richard, and Nigel Chalk, 2000, "Assessing Fiscal Sustainability in Theory and Practice," IMF Working Paper 00/81 (Washington: International Monetary Fund).

_____, and Murray Petrie, 2002, "A Framework for Assessing Fiscal Vulnerability," IMF Working Paper 00/52 (Washington: International Monetary Fund).

Henisz, W. J., (2002), "The Institutional Environment for Infrastructure Investment," *Industrial and Corporate Change*, Vol.11, Number 2, pp. 355–89.

Jeanne, Olivier, 2000, "Debt Maturity and the Global Financial Architecture," CEPR Discussion Paper 2520 (London: Centre for Economic Policy Research).

Lane, Timothy, and Steven Phillips, 2001, "Does IMF Financing Result in Moral Hazard?" IMF Working Paper 00/168 (Washington: International Monetary Fund).

Larrain, Guillermo, Helmut Reisen, and Julia von Maltzan, 1997, "Emerging Market Risk and Sovereign Credit Ratings," OECD Development Centre Technical Paper 124 (Paris: Organization for Economic Co-operation and Development).

Lee, Suk Hun, 1993, "Are the credit ratings assigned by bankers based on the willingness of LDC borrowers to repay?," *Journal of Development Economics*, Vol. 40, pp. 349–59.

Manasse, Paolo, Nouriel Roubini, and Axel Schimmelpfennig, 2003,"Predicting Sovereign Debt Crises," IMF Working Paper 03/221 (Washington: International Monetary Fund).

Mody, Ashoka, and Diego Saravia, 2003, "Catalyzing Private Capital Flows: Do IMF-Supported Programs Work as Commitment Devices?" (unpublished; Washington: International Monetary Fund).

Obstfeld, Maurice, and Kenneth Rogoff, 1996, "Foundations of International Macroeconomics," *MIT Press* (Cambridge, Massachusetts).

Reinhart, Carmen M., 2002, "Default, Currency Crises and Sovereign Credit Ratings," NBER Working Paper 8738. Also in *The World Bank Economic Review*, Vol. 16, No. 2, pp. 151–70.

Reinhart, Carmen M., Kenneth S. Rogoff, and Miguel A. Savastano, 2003,"Debt Intolerance," NBER Working Paper 9908 (Cambridge, Massachusetts: National Bureau of Economic Research).

Rojas-Suarez, L., 2001, *Rating Banks in Emerging Markets,* (Washington: Institute for International Economics).

Roubini, Nouriel, 2001, "Debt Sustainability: How to Assess Whether a Country is Insolvent," unpublished (New York: New York University) (December).

Roubini, Nouriel, and Brad Setser, 2004, *Bailouts or Bail-ins? Responding to Financial Crises in Emerging Economies***,** (Washington: Institute for International Economics).

An Application of BRT To Heart Attacks

The following medical application (see Breiman and others, 1984, p.175) illustrates the logic of classification trees. The problem is to identify patients who are at risk of dying within 30 days (High Risks) from among those who have suffered heart attacks and have survived at least 24 hours, past admission to the University of California, San Diego Medical Center. These patients would be placed on an intensive care unit for constant monitoring, while a Low Risk patient could remain on a standard medical unit. The dataset contains medical records on 215 patients, of whom 37 (17 percent) died not more than 30 days following the heart attack, and 178 (83 percent) who survived. The records comprise 19 measurements of each patient, e.g., minimum systolic blood pressure, history of heart attacks, presence of tachycardia, concentration of enzymes, sex , age etc.

The first question that is selected is "Is minimum systolic blood pressure below 91?". [20] If the answer is "Yes" observations move to the left node. Here we find 20 patients. Note that the proportions of survivors and early deaths here have changed dramatically: only six patients are survivors (30 percent), while 14 (70 percent) early deaths. Further splitting does not improve the classification, so the node is classified as High Risk. If the answer to the question is "No" observations move to the right child node, where we find the remaining 195 patients, among which 172 (88 percent) are survivors, and 12 percent are early deaths. For these patients new information is considered. The algorithm picks a new question: " Is age below 62.5 years?" For 104 cases the answer is "Yes," and these observations move to the left, in the relatively "safe" terminal node 2. This contains 102 (98 percent) survivors and only 2 early deaths (2 percent). The node is classified as Low Risk. For the remaining 91 patients aged above 62.5 years, a new question is asked: "Was there sinus tachycardia present?" (Footnote 13). For 28 observations the answer is "Yes," and these move to the left terminal node 3, that contains 28 cases with equal proportions of the two types, and is classified as High Risk. The remaining 63 patients go to the right terminal node, which is classified as Low Risk, since 89 percent of the cases therein are survivors.

When the prior distribution of cases is assumed to be equal to the actual sample distribution (83 percent of Low, and 17 percent of High Risks), the within node frequencies coincide with conditional probabilities. From the tree we can see immediately which types of patient are most at risk of early death: those in terminal node 1, displaying systolic pressure below 91 (the conditional probability of early death is 70 percent) and those in terminal node 3 that, even with "good pressure," are aged more than 62.5 years and present tachycardia (their conditional probability of early death is 50 percent).

---

[20] The systolic blood pressure is the maximum blood pressure that occurs with each hart cycle during contraction of the left-sided pumping chamber. Sinus tachycardia is defined to be present if the sinus node heart rate ever exceeded 100 beats per minute during the first 24 hours following admission to the hospital; the sinus node is the normal electrical pacemaker of the heart and is located in the right atrium (Breiman and others, page 179)

Figure 2. Survivors and Early Deaths
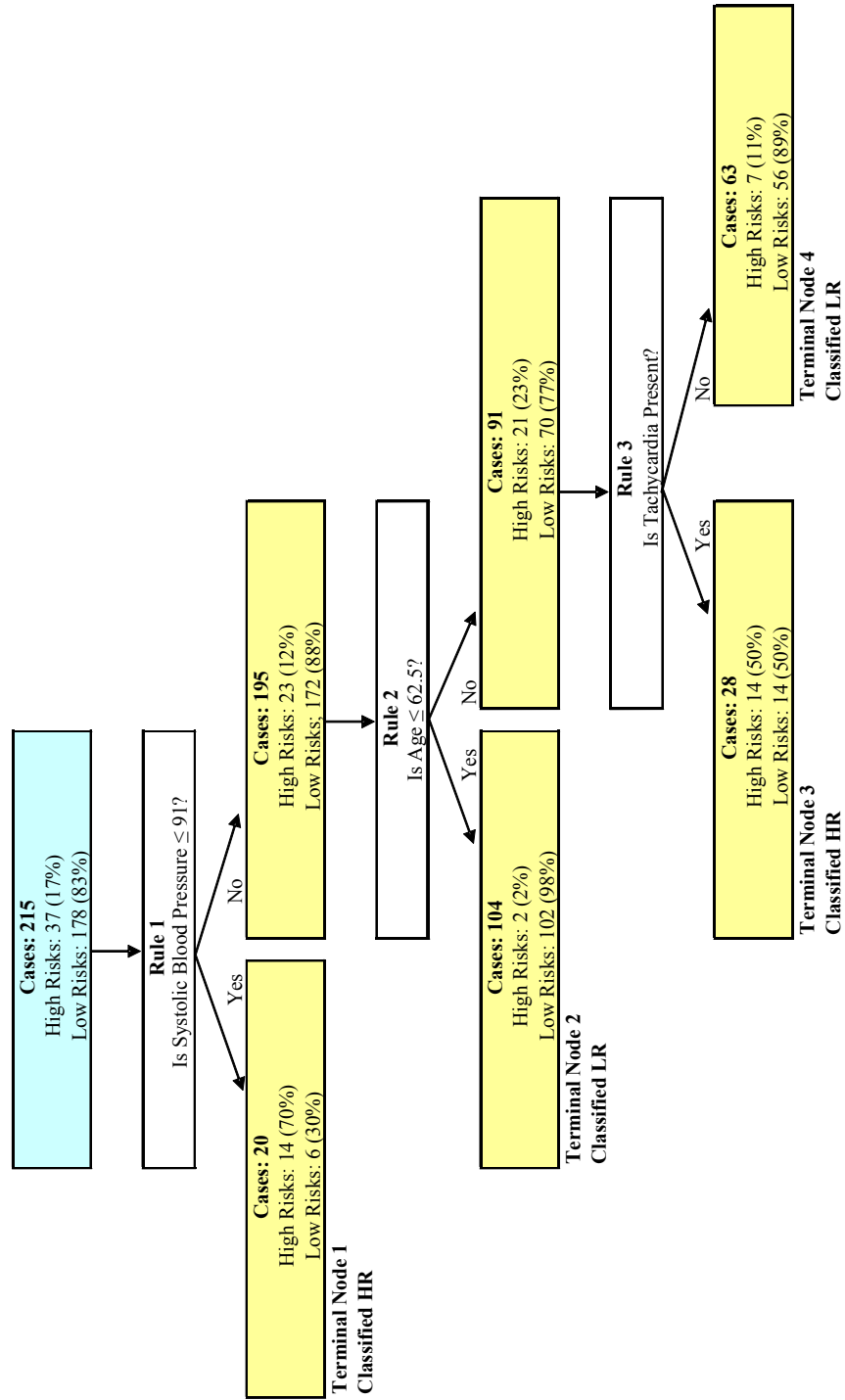
Some Basic Concepts of Binary Classification Trees

Next, we briefly summarize the basic concepts of BRT (see Breiman and others, 1984, for more details). Suppose that observations on variable *Y* must be classified into *j= 1...J* classes. Let *N, N_j, N(t), N_j(t)* represent the number of observations in the sample, the number of class-*j* observations in the sample, the number of observations in node *t,* the number of class-*j* observations in node *t*. Clearly, $N = \sum_j N_j = \sum_t N(t) = \sum_t \sum_j N_j(t)$.

Let *π_j , j=1...J*, denote the a priori distribution for an object to belong to class *j*. The ratio *N_j(t)/N_j* denotes the (empirical) probability that an object of class *j* falls in node *t, p(t|j)*. Thus, *p(j,t)= π_j N_j(t)/N_j* gives the joint probability that an object belongs to class *j and* falls in node *t*. Then the probability than an observation falls in node *t* is equal to *$p(t) = \sum_j p(j,t)$*. Finally, the *conditional* probability of observing a class-*j* individual, given that it has reached node *t* is given by *p(j | t)= p(j,t)/p(t)*. From the above expressions it follows that

$$p(j \mid t)= [\pi_j N_j(t)/N_j]/ [\sum_j \pi_j N_j(t)/N_j], \text{ with } \sum_j p(j \mid t)=1 \qquad (1)$$

When the prior distribution is assumed to coincide with the sample distribution, $\pi_j = N_j/N$, *p(j | t)* simplifies to the frequency of class *j* in the node, *p(j | t)=N_j(t)/N(t)*.

Best Split

At each node different split values are compared on the basis of how pure child nodes are produced. Impurity is a (criterion) function φ(.) defined over the *p(j | t),* such that the node's impurity *measure* i(t) is

$$i(t) = \varphi(p(1 \mid t) \ p(2 \mid t),..., p(J \mid t)) \qquad (2)$$

The function φ must satisfy three properties: (i) achieve a maximum at point (1/J,...,1/J); (ii) achieve its minimum only at points (1,0,...0), (0,1,0,...0), ...(0,0,...1); and (iii) be a symmetric function of the *p(j | t)*.

Suppose that a value *s* (split) for an explanatory variable sends a proportion of data p_R to the right and a proportion p_L to the left.. Then one can measure the reduction in the impurity as

$$\varDelta i(s,t)= i(t) - p_R \ i(t_R)- p_L \ i(t_L) \qquad (3)$$

This is the "goodness of split" criterion. At each node, the best split is the one that maximizes *$\varDelta i(s,t)$*. The Gini rule is a particular function φ satisfying properties (i)-(iii) :

$$G(t) = \sum_{j\neq i} p \ (j \mid t) \ p \ (i \mid t) \qquad (4)$$

One interpretation (see Breiman and others, 1984) is in term of classification error. Imagine you use the rule that assigns an object selected at random from node *t* to class *i* with probability *p (i | t)*. The estimated probability that the item is in effect in class *j* is *p (j | t)*.

Therefore, the estimated probability of misclassification is $G(t)$.[21] Since $G(t)$ can be written as:

$$G(t) = \left(\sum_j p(j \mid t) \right)^2 - \sum_j p^2(j \mid t) = 1 - \sum_j p^2(j \mid t) \qquad (5)$$

it is easy to see that the index has he minimum value of zero if node t contains only one class (for $J=4$, $G(t)=1-(0)^2-(1)^2-(0)^2-(0)^2$), and reaches a maximum when the node has equally frequent classes $G(t)= 1-(1/4)^2-(1/4)^2-(1/4)^2-(1/4)^2=3/4$.

Classification and Stopping Rules

At each node, the observations in node t are assigned to a class j, which is the class with greatest within node probability. Thus, the class assignment rule $j^*(t)$ is given by:

$$\text{if } p(j \mid t) = \max_i p(i \mid t) \text{ then } j^*(t)=j \qquad (6)$$

If the maximum is achieved for two or more different classes, assign j*(t) arbitrarily as any one of the maximizing classes.

For example, in a $j=1,2$ class problem with uniform prior distribution $(1/2,1/2)$, the criterion would be: classify node t as class 1 if $N_1(t)/N_2(t) > N_1/N_2$. If the prior was set equal to the sample frequency, $\pi_j=N_j/N$, the criterion would be the majority rule, i.e., classify node t as class 1 if $N_1(t) > N_2(t)$.

Early methodologies would stop further splitting whenever the reduction on impurity was less than a given value, according to the following stopping rule:

Set a threshold β and declare node t terminal if

$$\max_s \Delta i(s,t) < \beta \qquad (7)$$

The methodology developed by Breiman and others (1984) is more sophisticated: it builds a very large tree and then "prunes" it back. For more details on "pruning" see Breiman and others (1984).

Goodness of Fit

A measure of the probability of misclassification, given that an observation falls into node t is given by

$$r(t) = 1 - \max_i p(i \mid t) \qquad (8)$$

Finally, an estimate for the overall misclassification rate $R(t)$ of the tree classifier is given by

$$R(t) = \sum_{t \varepsilon T} p(t)r(t) \qquad (9)$$

where the summation is over the set of terminal nodes t.

---

[21] The researcher can specify that the cost erroneously classifying object j for an i type, $C(i \mid j)$ is different from misclassifying object i for j, $C(j \mid i)$, in which case the expression for the index is:
$G(t) = \sum_{j \neq i} C(i \mid j) \, p(j \mid t) \, p(i \mid t)$.