7TH JACQUES POLAK ANNUAL RESEARCH CONFERENCE
NOVEMBER 9-10, 2006

Can We Predict the Next Capital Account Crisis?

Marcos Chamon
International Monetary Fund

Paolo Manasse
University of Bologna

Alessandro Prati
International Monetary Fund

Paper presented at the 7th Jacques Polak Annual Research Conference
Hosted by the International Monetary Fund
Washington, DC─November 9-10, 2006

Can We Predict the Next Capital Account Crisis? [1]

Marcos Chamon
International Monetary Fund

Paolo Manasse
University of Bologna

Alessandro Prati
International Monetary Fund

November 2, 2006

ABSTRACT

This paper uses Binary Classification Trees (BCTs) to predict capital account crises. BCTs successively compare candidate variables and thresholds to split the data into two sub-samples, allowing for a large number of indicators to be considered and complex interactions to emerge in a way that standard regressions cannot easily replicate. We identify a robust leading indicator role for three variables (international reserves, current account balance, and short-term external debt) as well as a reserve cover measure that combines them. External indebtedness and domestic GDP growth forecasts are also important predictors of vulnerability. Out of sample, we are able to capture some of the main emerging market crises with relatively few false-alarms but the overall out-of-sample performance of our forecasts is mixed. Global cyclical variables help explain vulnerability to crises but they are difficult to predict and, therefore, of limited use for forecasting purposes.

I. Introduction

Predicting capital account crises is extremely difficult. Academics and policymakers have identified several factors that contribute to their inception but the imbalances at the origin of each wave of crises, as well as their propagation mechanism, keep changing over time and, with them, the set of potential crisis indicators. A typical (non-comprehensive) list includes

[1] This paper would not have been possible without the contribution of the IMF's Working Group on Vulnerability Indicators (WGVI), IMF desk economists, and several IMF's teams in charge of financial, corporate, and commodity price data. These groups also contributed to the selection of crisis episodes and to the construction of the vulnerability indicators used in this paper. Nonetheless, the methodology proposed in this paper does not correspond to that used by the IMF to assess crisis vulnerabilities. We thank Jonathan Ostry and Antonio Spilimbergo for comments on an earlier draft of this paper. Marcos Souto and Murad Omoev provided excellent research assistance. **The views expressed in this paper are those of the authors and should not be attributed to the International Monetary Fund, its Executive Board, or its management.**

measures of real exchange rate misalignment, terms of trade shocks, international reserves, external and public debt, monetary and fiscal policy, balance sheet mismatches (currency and maturity) in the corporate, banking, and government sectors, political uncertainty, global cyclical and financial conditions, and general market sentiment.

This unwieldy set of potential indicators makes it difficult to compare their information content using standard regression techniques. To remedy this problem and isolate a parsimonious set of robust *leading* indicators of crises within a group of over 100 vulnerability indicators, we use a nonparametric statistical methodology, called Binary Classification Trees (BCTs). BCTs can handle a large set of variables, interact them, and select critical thresholds. For example, we find that three simple conditions—based on international reserve coverage and the level and change of external debt-to-GDP ratios—select a subset of observations in which the frequency of capital account crises is 21.3 percent as opposed to the sample average of 6.1 percent.

We apply BCTs to a new dataset of 34 capital account crises that took place in 49 countries during the period 1994-2005. Capital account crises, or "sudden stops," are defined as large and sudden reversals in net private capital flows. We date all crises on the basis of their *inception* while all indicators are *lagged one year* so that only pre-crisis information is used. We also include one-year-ahead forecasts of contemporaneous variables (e.g., WEO forecasts of GDP growth or current account balances) and market-based forward-looking indicators such as EMBI spreads. We have used previous lists of crises and numerical rules to select and date potential capital account crises, which IMF's country desks have then revised and validated. This last step is important because numerical rules occasionally identify capital flow reversals that have non-crisis explanations that country desks may provide (e.g., the end of a privatization program).

The *in-sample* fit of the BCT estimated on the entire sample is reasonably good with four indicators (and their respective thresholds) breaking down the sample into: (i) a subsample with a frequency of crises 3.5 times as high as in the overall sample; (ii) another subsample with a frequency of crises twice as high as in the overall sample; and (iii) three "safe" subsamples with a minimal frequency of crises (around 1 percent).

BCTs yield mixed results when used *out-of-sample*. A BCT based on information up to 2000 would have predicted correctly three of the five crises in 2001, including Argentina, Turkey, and Lebanon (i.e., it would have classified them as having characteristics typical of countries with a frequency of crises 2.5 times as high as in the estimation sample). A BCT based on information up to 2001 would have missed all crises of 2002 (Brazil, Colombia, Israel, and Uruguay), while a BCT based on information up to 2002 would have perfectly predicted the two crises of 2003 (Dominican Republic and Jamaica)

Would BCTs have predicted the Asian crisis? Given that this crisis took place in 1997 and our sample begins in 1994, we cannot meaningfully estimate a BCT on the previous three years of data. We can, however, estimate a BCT on a sample that excludes all observations corresponding to East-Asian countries and check how it would split the latter into crisis and non-crisis years. On the basis of the level of international reserve coverage (the *lagged* ratio of international reserves to the sum of current account deficits and short-term external debt), this BCT would have initially classified Thailand, Indonesia, Korea, and the Philippines, in a group of countries with a frequency of crises twice as high as in the rest of the sample sample, and considered Malaysia less vulnerable than the average country in the sample. However, all East-Asian countries would have been misclassified as safe once we had taken into account a second set of conditions based on exchange rate overvaluation or fiscal positions, which are good predictors of crises in the subsample without East-Asian countries.[2] This out-of-sample exercise reinforces the view that the East-Asian crises were somewhat special and would have been difficult to predict.

BCTs also allow us to address the issue of the relative role of global economic conditions and country-specific imbalances in capital account crises. Is it true that during global booms or periods of abundant liquidity in capital markets even countries with serious domestic imbalances can remain unscathed? To answer this question, we estimate a variant of the full-sample BCT including *contemporaneous* global indicators (which are exogenous to crisis events in individual countries). We find that two gauges of the global conditions each country faces—commodity export prices and import demand by trading partners—contribute to explain the occurrence of crises. For example, when real commodity prices are at least 13.5 percent below their historical country-specific average, the frequency of crises in countries with low reserve coverage rises from 14 to 22.6 percent; by contrast, when real commodity export prices are higher than this threshold, the frequency of crises drops from 14 to 2.8 percent. In other words, low commodity export prices are a key trigger of crises in countries with low reserve coverage. This finding is, however, of little use for crisis prediction because forecasts, or lagged values, of global indicators are not as good as their contemporaneous values at separating crisis from non-crisis episodes.

The empirical literature on early warning systems (EWS) shares with this paper the focus on crisis prediction. Frankel and Rose (1996), Kaminsky, Lizondo, and Reinhart (1998), and Berg and Pattillo (1999) wrote seminal EWS papers. In the same spirit of this paper, Berg, Borensztein, and Pattillo (2004) analyzed the in-sample and out-of-sample fit of EWS models, showing that the latter varies substantially by model and forecast horizon. The EWS papers differ from our paper for the empirical methodology (probit/logit regressions), the prevalent focus on currency crises, and the monthly frequency of observations. BCTs can assess the predictive power of a much richer set of indicators and experiment with more interactions than EWS. Furthermore, the BCT algorithm selects indicators and thresholds taking into account the preferred trade-off between the cost of missing crises and that of predicting crises erroneously, whereas the EWS' probit/logit models can only be estimated independently from that trade-off. These advantages translate in better in-sample crisis

[2] That is, pre-crisis data of East-Asian countries point to sound fiscal positions and no exchange rate misalignment (although some indication of the latter would emerge if post-crisis information were used).

prediction performance of BCTs with 88 percent of crises correctly called (30 out of 34) as opposed to 60-70 percent in typical EWS models. Comparing out-of-sample performance of BCTs and EWS is much more difficult because of the different periods and crises considered. For example, BCTs predict correctly out of sample the 2001 crises in Argentina and Turkey, which the EWS models considered in Berg et al. (2004) do not try to predict, but they are less successful than EWS models in predicting the Asian crisis.

Few studies have used the BCT methodology. Ghosh and Ghosh (2002) and Frankel and Wei (2004) apply it to currency crises and assess its in-sample forecasting performance. Manasse and Roubini (2006) use BCTs to study the determinants of sovereign crises and to predict them. Van Rijckeghem and Weder (2004) develop a nonparametric technique similar to BCTs to study the political determinants of debt crises. Our paper is the first application of BCTs to capital account crises.

The paper is organized as follows. Section II outlines the BCT methodology. Section III spells out the criteria used to select the crisis episodes (discussed further in Appendix I) and explains the candidate indicators used in the analysis. Section IV presents our baseline BCT, examining its properties and the importance of different classes of indicators. Section V analyzes how BCTs can predict crises out of sample. Section VI compares the relative role of global cyclical indicators and country-specific vulnerabilities. Section VII concludes.

II. METHODOLOGY

BCTs are a nonparametric statistical technique, which is suitable for identifying complex interactions among variables with the objective of predicting binary outcomes (in our case, “crisis inception” or “no crisis inception”). BCTs identify the indicators and their thresholds that can better separate the sample into crisis and non-crisis observations. The order in which the indicators are used in each split allows complex interactions to emerge, in a way that would be difficult to replicate in a standard regression approach.

BCTs’ classification rules are a collection of inequalities, such as: *if* (i) international reserves cover less than 80 percent of the sum of short-term external debt and the current account deficit and (ii) external debt is higher than 24 percent of GDP and (iii) external debt is not falling by at least 3 percent of GDP per year, *then* the frequency of crises next year is 21 percent and the observation is classified as “crisis-prone.”

We compute BCTs using the nonparametric statistical algorithm CART (Classification and Regression Trees, Breiman et al, 1984). In a nutshell, the BCT algorithm computes a score of how well each variable does at separating crisis from non-crisis observations, and splits observations in two groups based on the variable with the highest score. The process continues for each branch of the data and eventually stops according to the criteria used to measure further improvements. Like other nonparametric methods, BCTs are apt tools for detecting nonlinearities, which is critical when an indicator has information content only for values beyond certain thresholds. In theory, standard regression techniques (e.g., probit or logit models) could be used for similar purposes. However, even for a very small set of indicators, it would be impossible to experiment with all possible interactions and thresholds in a single regression.[3] Another important drawback of parametric regression approaches is the need to make assumptions about a functional form. Given the lack of well-established theoretical relationships between the variables used for predicting crises and outcomes, it may become more attractive to rely on a complex set of threshold interactions than on a parametric functional form.

The key elements of the BCT analysis are a set of rules for: (i) splitting each node into two child nodes; (ii) assigning each node to a class outcome (e.g., crisis vs. non-crisis); and (iii) deciding when to stop growing the tree.

The BCT algorithm starts by comparing candidate variables and thresholds to split the sample into two child nodes. All splitting rules are based on whether or not a variable is above or below a threshold. Each split is assigned a score based on how it improves the "purity" of the classification. A variable and a threshold that separate perfectly all crisis observations from all non-crisis observations would yield the "purest" possible classification. In practice, however, each possible split classifies observations in two groups that have both crises and non-crises. The BCT algorithm computes a cost that rises with the extent by which the actual classification departs from the perfect classification and selects the split that minimizes such cost. The tree is grown by repeating this process on the child nodes.

In our baseline classification, we consider misclassifying crises as non-crises twice more costly than the other way around. This means that the BCT algorithm considers an "impure" non-crisis node (i.e., a non-crisis node with a relatively high share of crises) more costly than an equally "impure" crisis node. The rationale behind this parameter choice is a subjective preference for reducing the chance of missing crises. As a consequence of the higher cost of misclassifying crises, the crisis nodes of our BCTs are characterized by low crisis frequency and relatively higher Type II errors. All BCTs presented in this paper are, however, quite robust to perturbations of the relative cost parameter, with the top split remaining unaffected and lower splits changing only for very high levels of the relative cost parameter.

The option of choosing misclassification costs at the outset (i.e., before running the BCT algorithm) to influence the model choice (i.e., the set of indicators and thresholds) is a key difference between BCTs and EWS. In fact, Berg, Borensztein, and Pattillo (2004) use a misclassification cost function (weighting Type I and Type II errors) to identify the probability cutoff point that would best predict crisis and non-crisis events out-of-sample only after having estimated the probit model. As a result, their cost-function cannot influence the choice of the indicators and coefficients in the probit model.

[3] For example, the number of possible interactions of indicators and threshold values in our data exceeds by several orders of magnitude the number of observations.

In BCTs, the prior probability of a crisis observation is another key choice parameter that affects the minimum frequency of crises required to classify a node as crisis-prone. For all BCTs presented in this paper, we choose a prior crisis probability such that the BCT algorithm classifies as crisis nodes those where the crisis frequency is at least twice as high as in the sample. This is achieved by setting the prior to 20 percent.[4] Perturbations of this parameter do not affect the top split of our BCTs but, occasionally, lower-level splits.

We use judgment to decide when data are sufficiently partitioned. This is a critical decision because in BCTs there is not necessarily an "optimal" number of splits. In fact, it is always possible to use a very large set of rules to attain a perfect classification. Increasing the number of splits, however, may lead to poor out-of-sample forecasts, similarly to what happens in regression analysis when the number of regressors increases.

While we use judgment in selecting the size of the trees we present, we also take into consideration the results of a technique called "V-fold cross-validation." This technique amounts to using out-of-sample performance as a guide to select the best number of splits. The sample is divided into 10 parts and, then, each 10 percent of the observations is used, in turn, to test the predictive power of 10 ancillary trees estimated on the remaining 90 percent of observations (in a way that each observation is used once and only once in an ancillary test sample). Based on the out-of-sample performance of these ancillary trees, the algorithm proposes an optimal level of complexity (measured by the number of terminal nodes) for the tree estimated on the full sample. Section V discusses the several reasons why we often overrule the V-fold's proposed pruning. The main reason is that, in many instances, the V-fold cross-validation technique suggests trees with no split[5] or including many splits some of which make no economic sense.

---

[4] The rationale for choosing a threshold that is twice the sample frequency of crises is the following. Crises are relatively rare events in our sample. If we had set, for example, the prior probability of crisis at the sample frequency of 6.1 percent, a very high share of crisis observations in a node (at least 35 percent) would have been necessary to classify it as a crisis node, despite the asymmetric misclassification cost imposed. Since we prefer to err on the side of being conservative, we require a much smaller frequency of crisis observations to classify a node as crisis prone. The threshold around 12.2 percent used for the entire sample allows us to be conservative while still acknowledging that crises are relatively rare events. This is the same logic we applied to the choice of the misclassification cost parameter. As in that case, the option of influencing the model selection by choosing the frequency of crises required to classify a node as crisis-prone is a feature of BCTs that distinguishes them from probit-based EWS.

[5] The V-fold cross-validation methodology assumes that, in the no-split tree, all observations are crises. The no-split tree has, then, a zero Type I error but the highest possible Type II error.

Many indicators have missing values (years, countries, or both).[6] The BCT algorithm does not drop observations for which some indicators are missing unlike, say, a regression where missing observations would be dropped. When a variable with missing observations is used to split the data, the missing observations are assigned to the partitioning of the sample that minimizes the cost function. To prevent this default rule from influencing the selection of the best indicators, we penalize indicators with missing observations. In practice, this choice forces indicators with missing observations to be used for splitting smaller partitions of the data (for which their coverage is reasonable) or not to appear at all.[7]

The BCT algorithm can be applied to a very large number of candidate predictors: unlike in a standard regression, the inclusion of irrelevant indicators, which are not used for splitting the data, does not affect the results. However, when an indicator slightly outperforms another as a "splitter", the latter may never appear in the final tree even though its information content is almost as good as that of the top splitter. To avoid drawing the incorrect inference that all omitted indicators are not "important," we check the competing indicators for the top split.

As a robustness test for our selection of indicators and as a benchmark for out-of-sample prediction, we use a new procedure called "RandomForests" that Breiman (2001)—one of CART's developers—proposed as a way to address the problem of few additional variables or observations changing substantially the BCTs. Adding variables will not change the BCTs if the new variables do not improve any of the splits obtained with the pre-existing variables. However, if one of the new variables is informative enough to replace a pre-existing variable even in a single split, there is a good chance that the branch developing from that split onwards will feature a completely new set of variables and thresholds. Similarly, the introduction of additional years or countries to the sample may lead to substantial changes in the optimal tree if and where changes occur. Breiman proposes an algorithm based on a collection of hundreds or thousands of trees that classifies and predicts each observation according to the response of the majority of trees. A bootstrap procedure over two dimensions selects the sample and the list of variables used to estimate each tree (hence the algorithm's name, RandomForests).

In our application of the RandomForests algorithm, we grow trees on 1000 bootstrapped samples allowing 3 randomly chosen indicators to be used to split the data at each node. By randomizing over the variables, each tree is likely to involve a very large number of different splitters. By randomizing simultaneously over the sample, each tree in the forest analyzes only small portions of data at a time. This process, called "slow learning," highlights

[6] There is substantial variation in data coverage across countries and time. Some indicators are available for only a subset of countries (e.g., corporate vulnerability indicators). Others are not available at the beginning of the sample (e.g., detailed financial vulnerability indicators or data for transition countries).

[7] If the fraction of missing values of an indicator is 50 percent, its improvement score will be multiplied by 25 percent.

different aspects of the data set and reduces the risk of possibly drawing wrong conclusions "too fast" (see Friedman, 2001). If a pattern genuinely exist in the portion of data analyzed, the RandomForests algorithm will detect it repeatedly in different trees; conversely, it will wash out any accidental pattern in the process of averaging the results. While this algorithm can improve predictive accuracy, it has the important drawback of not allowing the researcher to recover thresholds and variable interactions, since it relies upon aggregating many different trees of different shapes. For this reason, we use it only as a robustness check.

III. The Dataset

The sample covers 49 countries during the 1994-2005 period.[8] The countries in the sample are listed in Appendix I. The coverage focuses on emerging market countries that had significant access to private international financial markets and did not have a substantial net foreign asset position. Very small economies (with GDP below 7.5 billion dollars at the end of the sample) were not considered no matter what level of income per capita thay had.

A. Crisis Definition

We define capital account crises as sudden stops in capital flows that are likely to be associated with currency, sovereign, banking, or corporate crises. Table 1 lists crisis episodes for the 49 countries in our sample. Only the first year of a capital reversal (the crisis *inception)* is considered. This selection of the list of crises is the result of a concerted effort by the IMF Working Group on Vulnerability Indicators (WGVI) whose aim was to develop new criteria for rating countries' vulnerability. The following two-stage procedure was followed. A first set of potential crisis episodes was identified on the basis of various definitions of crises, including two measures of sudden stop in net private capital flows,[9] years of high exchange rate pressure as indicated by Early Warning Systems (EWS), sovereign defaults, Fund programs (only years with positive net disbursements), a banking and a corporate crisis indicator.[10] Second, the final set of crisis years was chosen taking into

[8] The period 1994-2005 was chosen because the capital account regime was relatively stable in most countries and to have only post-transition years for Central and Eastern European countries.

[9] There is no standard definition of sudden stop. In some cases, a sharp and sudden reversal in capital flows is easy to classify as a sudden stop (for example, Thailand 1997). In other instances, a steady decline takes place over a prolonged period of time resulting in a crisis (for example, Venezuela from 1998 to 2000). In this latter case, it is not straightforward to determine the inception year. Footnotes 3 and 4 in Appendix Table 1 describe the numerical rules used to address this issue in a systematic manner. Somewhat related rules are used by Catão (2006).

[10] This initial selection of potential capital account crisis years was based mainly on the sudden stop indicators. The other indicators helped select potential crisis episodes that did not translate into a substantial deterioration in net private capital flows or to fine-tune the year of inception of the crisis. Sovereign crises are from Manasse and Roubini (IMF WP 05/42) updated with the sovereign debt default indicator of Debrun (WEO, 2004). The banking crisis indicator is based on Demirguc-Kunt

(continued)

consideration comments from IMF's desk economists. Their suggestions helped solve ambiguities on the timing of the crisis inception, discard years that were identified by some crisis indicator but should not be considered a capital account crisis, and add one episode that no crisis indicator had picked up. A table in Appendix I lists the different crisis indicators, while Appendix II provides country-by-country details on the selected crises.

There are 554 country-year observations, of which 34 observations (6.1 percent) correspond to the year of inception of a crisis, and the rest to non-crisis years. We drop, in fact, from the sample the observations corresponding to years immediately following a crisis because their characteristics are clearly different from non-crisis years. At the same time, these post-crisis years should not be confused with the crisis-inception years because they may be easier to predict using previous-year indicators that already reflect the impact of a crisis. Of course, dropping only the first year after a crisis is a relatively arbitrary way of dealing with this problem. Nonetheless, dropping additional post-crisis years does not change the results.

B. Indicators

The IMF's WGVI also suggested the core set of indicators used in this paper.[11] They cover four sectors:

- ***External Sector***: (i) *gross international reserve coverage* (relative to maturing external debt and the current account deficit); (ii) *current account* balance (in percent of GDP); (iii) real exchange rate *overvaluation*, (iv) rigidity of the nominal *exchange rate regime*; and (v) *external debt* (in percent of GDP).
- ***Fiscal Sector***: (i) *overall balance*; (ii) *primary balance*, including the gap between primary balance and debt-stabilizing primary balance; (iii) *public debt* (in percent of GDP); (iv) maturity of *public debt*; and (v) foreign-currency debt in percent of total debt.
- ***Financial Sector***: (i) *capital adequacy*; (ii) *return on assets*; (iii) *non-performing loans* as a share of total loans; (iv) *growth in private sector credit* (as a ratio to GDP); and (v) the share of *foreign currency loans*.
- ***Corporate Sector***: (i) *default probability* (extracted from a Black-Scholes-Merton formula); (ii) *interest coverage ratio*; (iii) *debt-to-assets* ratio; (iv) *real return on assets*; and (v) a valuation measure based on the *price-to-earnings-ratio*.

---

and Detragiache (IMF Staff papers, 1998) updated by MFD. The corporate crisis indicator is based on Corporate Vulnerability Utility (CVU) developed by the IMF's Research Department.

[11] IMF's country-desk economists provided the historical data going back to 1994 necessary to construct vulnerability indicators of the external and fiscal sectors. IMF's Monetary and Financial Department (MFD) provided most financial sector data (with measures of capital adequacy and non-performing loans beginning in 2000), while Boyd, De Nicolo', and Al Jalal provided data (extracted from BankScope) on return on assets, equity-asset ratio, and loan-to-asset ratio. The Corporate Vulnerability Unit team provided corporate sector indicators.

Whenever data coverage was incomplete, we used close substitutes of these indicators. For example, we used only short-term debt as opposed to short-term debt plus maturing medium- and long-term debt in computing reserve cover. We also constructed a number of alternative measures of financial sector soundness from Boyd, De Nicolo', and Al Jalal (2006). Note that, as previously discussed, the nature of BCTs is such that including additional variables with limited explanatory power does not change the results (unlike in a regression where degrees of freedom would be affected).

Country-specific measures complement these sectoral indicators:

- ***Macroeconomic Conditions***: One-year-ahead WEO forecasts of (i) *real GDP growth* and (ii) *CPI inflation*.
- ***Global Demand Conditions***: (i) One-year-ahead WEO forecasts of growth in *import demand by trading partners* and (ii) levels and changes of *commodity price indices* faced by each particular country.[12] Both measures are country-specific.
- ***EMBI Spreads***.

We did ***not*** include ***country-invariant global macroeconomic and capital markets conditions*** (e.g., global growth or U.S. interest rates) because they could end-up playing the role of yearly dummies.[13] If included, however, they did not show up in any tree.[14] Given that predicting capital account crises is the main goal of our exercise, we used ***lagged values*** for all variables (for example, indebtedness at time *t-1* to predict a crisis at *t*). Moreover, lagged values are more likely to convey useful information since contemporaneous ones would be affected by the inception of a crisis (for example, low levels of reserves could be a consequence of a crisis rather than one of the underlying vulnerabilities that allowed a crisis to happen). The only classification tree that considers contemporaneous variables is the one discussed in Section VI, where contemporaneous global conditions are used to compare country-specific vulnerabilities under favorable and unfavorable global scenarios.

We also considered political-economy related indicators. These indicators tended to have very limited explanatory power possibly because they adjust sluggishly with abrupt movements in political variables often occurring after a crisis.

[12] The IMF's Research Department Commodities Unit constructed these data.

[13] For example, there is a large concentration of crises in the late 1990s, when oil prices were relatively low. In some preliminary versions with an oil price indicator, low oil prices seemed to be harmful for the average country in the sample, most likely because of the association between cheap oil and crises in the late 1990s.

[14] It is possible that the lack of cross-country variation adversely affected their explanatory power in BCTs. Calvo, Leiderman, and Reinhart (1992) found that "push" factors such as international interest rates and the U.S. business cycle explained part of capital flows to Latin America in the early 1990s.

The forecasting nature of the exercise also requires that any ex-post measure of ***real exchange rate (RER) overvaluation*** be excluded as long as it uses unavailable future information to compute real exchange trends. To overcome this problem, we experimented with a number of possible approaches to estimating overvaluation at time *t* using only information available up to that period ("rolling" RER trends). However, those estimates turned out to be very noisy and not informative. For example, a sound economic expansion with rapid productivity growth resulted in an appreciating trend of the real exchange rate just like that of a country teetering on the brink of crisis. Often times, the extrapolation of a trend following a large depreciation would suggest that the currency was overvalued, even when the RER was broadly in line with its equilibrium value (or had overshot it). Using rolling HP filters instead of rolling linear trends did not improve matters much. In the end, our preferred method for determining the level of overvaluation without using ex-post information was to compare the RER with its long-term average since 1960 (where data availability permits).

The ***exchange rate regime*** (e.g., Reinhart and Rogoff's, 2004, *de facto* classification) did not have much predictive power either. This result may be partly due to the fact that, even in countries that pegged the exchange rate and experienced a crisis, no crisis took place during most of the peg years, thus diluting the positive relationship between crisis observations and exchange rate rigidity. Also the ***EMBI spreads*** never appeared in any BCT, which is somewhat surprising because in principle they should be a forward-looking market-based indicator; their limited sample coverage prior to the late 1990s may contribute to this result.

IV. A Baseline Tree that Explains In-Sample Crises

A. Baseline Tree

Figure 1 shows the baseline tree. The BCT algorithm uses only variables dated one year prior to that of the outcome we are trying to predict (crisis and non-crisis). The variable that best splits the sample into crisis-prone and non-crisis-prone observations is the ***lagged reserve cover***, measured as the ratio of gross international reserves to the sum of the short-term external debt (from BIS) and the current account deficit (set to zero if it is a surplus). For example, a reserve cover of 100 percent would allow a country to finance its entire current account deficit plus all short-term external debt maturing within a year by bringing its stock of international reserves to zero. The BCT algorithm selects a threshold of 81 percent, which partitions the sample into 164 crisis-prone observations with a lower lagged reserve cover (of which 23, or 14 percent, are crises, top-left branch) and 390 observations with a higher lagged reserve cover (of which 11, or 2.8 percent, are crises, top-right branch).

The dominant role of reserve cover in predicting capital account crises is very robust. All the BCTs presented in this paper have either lagged reserve cover or its components (the lagged current account balance and the ratio of short-term external debt to reserves) at the top. It is not surprising that countries can forestall capital account crises by accumulating large stocks of international reserves, containing current account deficits, and limiting short-term debt. What is new, however, is how well simple thresholds on the values taken by these variables, or for the reserve coverage measure that combines them, can forecast crises.

The share of countries with reserve cover above the estimated safe threshold has risen from about 40 percent in 1994 to 80 percent in 2005. This is consistent with the lack of crises in recent years. Table 2 shows that, since 2002, between one third and one half of the countries in the sample have had a reserve cover ratio above 200 percent, which is well in excess of the 81 percent threshold selected by BCT. This suggests that motives for reserve accumulation other than crisis prevention might be at play.

Countries with a low reserve cover are not necessarily doomed. In our sample, only one in seven countries with a reserve cover below 81 percent experiences a crisis. The ***level of external debt in relation to GDP*** (left branch of the tree in Figure 1) helps to sharpen crisis prediction in instances of low reserve cover. When lagged external debt is below 24 percent of GDP, no crisis takes place even though reserve cover is below the threshold. The few low-reserve-cover countries in this situation (30 out of 164) have a relatively high fraction of short-term debt or sizable current account deficits but incur no crisis. Conversely, when external debt is above 24 percent of GDP, the frequency of crises among low-reserve-cover countries rises to 17.2 percent (one in six). It is worth noting that, while this 24 percent threshold may seem low, it is based only on external debt as opposed to the entire stock of public debt (which would include also domestic public debt).[15] At the same time, our external debt measure includes the external debt of the private sector. As a result, this split captures an external sector vulnerability rather than fiscal vulnerability.

In our sample, countries with low reserve cover and high external debt can still escape a crisis if external debt has fallen by at least 3.3 percent of GDP in the previous year. Smaller reductions or increases in the external debt to GDP ratio isolate, instead, a crisis-prone group of 108 observations with 23 crises (21.3 percent or one in five).

Returning to the top of the tree and moving down the right branch, we notice that a high reserve cover needs to be combined with a strong growth outlook to shield countries from capital account crises (i.e., to reduce the crisis frequency to about 1 percent). By contrast, when the ***previous year WEO real growth forecast*** is below 3 percent, crises take place with a frequency of 13 percent (7 crises out of 54 observations) even at high levels of reserve cover. Although 3 percent may look like a relatively high threshold for GDP growth, as many as 336 observations end up in the “safe” node with a higher forecasted GDP growth. This reflects the relatively high growth rates of emerging market countries: in our sample, the first quartile of the distribution of real growth forecasts is as high as 3.5 percent (Table 3).

The in-sample fit of the baseline tree is very good. It predicts correctly 30 out of 34 crises (88.2 percent) and wrongly predicts 132 crises out of 520 non-crises observations (a misclassification rate of 25.4 percent). The optimal tree based on the V-fold cross-validation

[15] Reinhart, Rogoff, and Savastano (2003) also found a “safe” threshold for the external debt-to-GNP ratio as low as 15 percent for some developing countries.

technique would have had 30 terminal nodes. This alternative tree would have predicted all the crises and only misclassified 3.7 percent of non-crisis observations.

By comparison, two EWS models estimated on the period 1985-1997 (the DCSD and KLR models considered in Table 4 of Berg et al., 2004) have a poorer in-sample prediction rate (63 and 60 percent, respectively). However, their false alarms as percent of total alarms are lower than those of the BCT of Figure 1 (64 and 71 percent as opposed to 81 percent). The different frequency of the data (monthly in the case of EWS) and sample periods suggest that these comparisons should be interpreted with caution.

B. The Importance of Different Classes of Indicators

The robustness of reserve cover as crisis indicator is confirmed by the fact that the main competitors for the top split are its components (the ratio of short-term external debt to international reserves and the current account-to-GDP ratio) or close substitutes of its components (the WEO forecast of the current account-to-GDP ratio and the ratio of short-term external debt to GDP). Other competitors are indicators that appear further down the baseline tree, such as the change in the external debt-to-GDP ratio and the WEO growth forecast. There are also no surprises among the competitors of the second-level indicators, with the change in the government debt-to-GDP ratio emerging as a possible competitor of the external debt indicators and different WEO vintages of GDP growth forecasts as competitors for the right-hand split.

The lack of an exchange rate overvaluation measure in the baseline tree may look surprising in view of the prominent role this variable played in EWS. This result may reflect, however, the similar information content of current account balances and exchange rate overvaluation measures. The inherent difficulty in constructing an ex-ante indicator of overvaluation using only pre-crisis information can also explain why our simple overvaluation measure—which compares the real exchange rate with its past long-term average—turns out to have less information content than current account balances. Nonetheless, our overvaluation measure is in a second group of competitors for the top split (ranking between fifth and tenth) and emerges as a second-level splitter in the BCT estimated on the subsample that excludes East-Asian countries (Section V.D). Finally, a mixed alternative overvaluation measure—where the simple deviation of the real exchange rate from its long-term average is replaced with its deviation from the equilibrium real exchange rate computed according to the IMF's CGER methodology for all countries in the sample for which the latter is available—would feature as a second-level splitter in the baseline tree in place of external debt.[16]

[16] We did not use the mixed exchange rate overvaluation measure as our main overvaluation measure because the equilibrium real exchange rate is computed as a function of variables such as net foreign assets, relative productivity growth in the traded and non-traded goods' sector, and terms of trade, using parameters that are estimated over the 1973-2004 period and, therefore, on ex-post information for most of our sample. At the same time, relying on such parameters does not create as many problems as proxying equilibrium real exchange rates with country-specific trends because the (continued)

These results shed some light on whether traditional domestic macroeconomic variables—such as growth, real exchange rates, current account deficits, international reserves, and fiscal variables—contain enough information to predict future crises, or micro indicators of imbalances in financial and corporate sectors are needed to improve predictability. In the baseline tree, traditional macroeconomic variables trump financial and corporate sector indicators despite the rich set of candidates for the latter that the BCT algorithm took into consideration (see Section III.B). We also verified that this result was not due to the larger number of missing observations that characterize some financial and corporate sector indicators by re-running BCTs on subsets of observations for which measures of capital adequacy, return on assets, corporate debt-to-asset ratios, and the EMBI index were not missing. In all these instances, only macroeconomic indicators still showed up in the BCTs. The little information content of financial and corporate indicators may, then, reflect the lag with which balance-sheet data record financial and corporate vulnerabilities. Furthermore, we suspect such vulnerabilities play a major role in determining how disruptive capital account crises can be but may play a more limited role in determining whether a capital account crisis takes place to begin with.

V. Out-of-Sample Forecasts

The good in-sample fit of the baseline tree is encouraging but does not actually answer the question in the title of this paper. To have a clue about BCTs' ability to predict crises, we need to consider out-of-sample forecasts. In this section, we estimate BCTs using data up to 2000, 2001, and 2002 to predict crises, respectively, in 2001, 2002, and 2003. We focus on these years because they are the last with crises in our sample. We also present out-of-sample forecasts for East Asian countries based on a sample that excludes them.

Larger trees improve the in-sample fit but may actually worsen out-of-sample performance (as it is the case with standard regressions). The V-fold cross-validation technique described in Section II suggests an optimal "pruning" of trees, which we often override using our judgment. The first reason is that the V-fold cross-validation technique often recommends an uninformative tree with no splits or with too many splits, including some based on statistical correlations that make no economic sense. Secondly, the tree size preferred by the V-fold cross-validation technique is based on the out-of-sample performance of a set of trees that—having been estimated on random subsamples of data—might have little or no resemblance to the tree whose out-of-sample performance we want to assess. Thirdly, the V-fold cross-validation technique may lead to "over-fitting."[17] Despite these reservations, to be fully

---

parameter estimates are not country-specific but panel estimates which are identical for all CGER countries. Moreover, we computed the only country-specific parameter (the fixed effect in the equilibrium real exchange rate equation) in a rolling fashion using only ex-ante information.

[17] Consider a V-fold simulation in which the observation for Indonesia 1997 is randomly selected for out-of-sample testing while the observation for Thailand 1997 is used in-sample. Since those two

(continued)

transparent about our methodology, we always report the size that the V-fold cross-validation would recommend and the associated misclassification rates, together with those of our preferred tree.

In discussing each out-of-sample forecast, we also report out-of-sample results based on the RandomForests algorithm. Despite the drawback of not yielding specific rules and thresholds, the forecasting ability of the RandomForests algorithm is a benchmark against which we can compare that of the out-of-sample trees pruned using our judgment. Overall, the RandomForests algorithm did not yield consistently superior forecasts to those of our trees.

A. Out-of-sample prediction of 2001 crises

Figure 2 shows the tree estimated on data up to 2000 and the nodes in which each country is predicted to end up in 2001. The V-fold cross validation suggested a tree with no splits. We chose to grow the left branch of the tree and have three terminal nodes because the additional split made economic sense and raised the crisis frequency from 12.5 to 16.1 percent.

The variable that best splits the 1994-2000 sample into crisis-prone and non-crisis-prone observations is the ***current account balance over GDP***. The BCT algorithm selects a threshold of minus 2.9 percent of GDP, which partitions the sample into 168 crisis-prone observations with a lower current account balance (of which 21, or 12.5 percent, are crises, top-left branch) and 152 observations with a higher current account balance (of which 2, or 1.3 percent, are crises, top-right branch). The ***ratio of short-term external debt to reserves*** (left branch of the tree in Figure 2) further splits the crisis-prone node. When this ratio is below 41 percent, crises are relatively rare (1 crisis out of 44 observations, or 2.3 percent). High ratios of short-term debt to reserves—combined with large current account deficits—raise, instead, the frequency of crises to 16.1 percent (20 crises out of 124 observations), characterizing the crisis-prone node of this tree. What is interesting is that the two most informative indicators selected using the 1994-2000 sample are the two components of the reserve cover ratio used in the first split of the baseline tree.

The tree in Figure 2 would have successfully predicted the crises in Argentina, Lebanon, and Turkey, but it would have failed to predict those in South Africa and Venezuela. While an error of 40 percent can be seen as high, we find it reassuring that it would have predicted the major crises. The false-positives correspond to 33 percent of non-crisis observations, which is reasonable given the nature of the exercise and the fact that we want to err on the side of being conservative. It is worth noting that one of the false-positives had a crisis in 2002 (Brazil) and two had a crisis in 2003 (Dominican Republic and Jamaica).

---

crises shared similar features, the estimated tree would choose a set of indicators that can predict well Thailand 1997 (and, therefore, probably also Indonesia 1997) but it would be as though we had known ahead of time that a crisis was going to take place in Thailand in 1997.

These out-of-sample results cannot be easily compared with those of the EWS models considered by Berg et al. (2004), who test the DCSD and KLR models on the out-of-sample period from January 1999 to December 2000, which includes only three crises (Brazil, Colombia, and Zimbabwe) and excludes all 2001-2003 crises including Argentina and Turkey. In these EWS modes, the percent of crises correctly called is measured as the number of observations for which the estimated probability of crisis is above the cutoff probability and a crisis ensues within 24 months as a share of all observations for which a crisis ensues within 24 months. Using this measure, Berg et al. find that DCSD predicts correctly 31 percent of the pre-crisis months while KLR predicts correctly 58 percent of pre-crisis months. This compares with the 60 percent out-of-sample prediction rate of the BCT in Figure 2 (3 out of 5 crises).

The RandomForests algorithm also predicts the crises in Argentina, Lebanon, and Turkey and misses those in South Africa and Venezuela, but issues more false-alarms (misclassifying 50 percent of non-crisis). Thus, for this out-of-sample exercise, the forecasting performance of the RandomForests algorithm is worse than that of the tree in Figure 2.

B. Out-of-sample prediction of 2002 crises

Figure 3 repeats the same exercise, this time estimating a tree on the 1994-2001 sample to predict crises in 2002. The V-fold cross validation suggested again a tree with no splits. We chose instead a tree with the same top split of the 1994-2000 tree based on the current account balance. This split fails to predict the crises in Colombia, Israel, and Uruguay, which had a lagged current account balance above the threshold of minus 2.9 percent. The pre-crisis large current account deficit of Brazil places it, instead, in the top-left node with a crisis probability of 12.6 percent. The splits based on the ratio of short-term external debt to reserves and the previous year WEO real growth forecast, which do a good job in isolating crisis observations in-sample, would have, however, misplaced Brazil in a relatively safe node with a crisis frequency of only 4.4 percent.

To put this result in perspective, it is worth mentioning that the tree estimated on the 1994-2001 sample raises the threshold on the ratio of short-term external debt to reserves from the 41 percent level of the 1994-2000 tree to 125 percent. This change improves considerably the in-sample prediction (with the crisis frequency in the crisis-prone node rising to 33 percent) but it makes us “miss” the crisis in Brazil, which had a ratio of short-term external debt to reserves of 119 percent. Moreover, the crises in Colombia and Uruguay had a strong contagion component for which we lack a proper indicator and the crisis in Israel was to a large extent related to security considerations. The false-alarms correspond to only 2.5 percent of the non-crisis observations.

The RandomForests algorithm predicts the crisis in Brazil and misses those in Colombia, Israel, and Uruguay, with a crisis misclassification rate of 75 percent and a non-crisis misclassification rate of 25 percent. Given our preference to err on the side of caution, we would have preferred this performance to that of the tree in Figure 3.

C. Out-of-sample prediction of 2003 crises

Figure 4 shows the tree estimated on data up to 2002 and the nodes in which each country is predicted to end up in 2003. The V-fold cross-validation again suggested a tree with no splits. The tree in Figure 4 has the same left branch of the baseline tree estimated on the full sample, differing from it only for the lack of the split based on the previous year WEO real growth forecast on the right branch. This tree perfectly predicts the two crises of 2003 (Dominican Republic and Jamaica). False-alarms correspond to 16 percent of the non-crisis observations.

The RandomForests algorithm predicts the crisis in Jamaica and misses the one in the Dominican Republic (so 50 percent misclassification of crises), and misclassifies 28 percent of the non-crisis observations. In this case, the performance of the RandomForests algorithm is worse than that of the tree in Figure 4.

D. Out-of-sample prediction of the 1997 East-Asian crisis

Figure 5 estimates a tree based on a sample that excludes all observations corresponding to East-Asian countries (China, Indonesia, Korea, Malaysia, Philippines, and Thailand) over the 1994-2005 period and checks how well it would have predicted the 1997 crisis. This is not a perfect out-of-sample test because it includes post-1997 information from non-East-Asian countries that was unavailable at the time. Nonetheless, it is our only option considering that estimating a BCT on the short 1994-1996 sample would be meaningless. The V-fold cross-validation suggested a tree with no splits. Instead, we chose one with four.

The top split in this tree is the same of the baseline tree (***reserve cover*** at 81 percent). Given their high short-term external debt in relation to reserves prior to 1997, all East-Asian countries except Malaysia would have ended up in the top-left node with a crisis frequency of 12.5 percent. However, the second-level split of the tree would have erroneously classified Indonesia, Korea, Philippines, and Thailand as non-crisis-prone in 1997. For these countries, in fact, there was no sign—based on our measure—of ***real exchange rate overvaluation*** in the year prior to the crisis (i.e., the real exchange rate was less than 12 percent above its country-specific average from 1960 to the previous year). The first-ranked competitor of exchange rate misalignment (the ***government overall balance-to-GDP ratio***) would have also mispredicted the crisis in these four East-Asian countries on the heels of their strong pre-crisis fiscal position. These results highlight the fact that, in the sub-sample without East Asia on which the tree is estimated, crises are unlikely to occur in countries where the reserve cover is low but the real exchange rate is not misaligned or the fiscal position is strong. To predict correctly the crises in Indonesia, Korea, Philippines, and Thailand, we would need to use the second-ranked competitor of the exchange rate overvaluation measure (***the lagged level of external debt***). The estimated tree also fails to predict Malaysia because its reserve cover was relatively high and the ***WEO real growth forecast*** for 1997 was above 3 percent.

In sum, if we had stopped at the first split, a tree estimated on a sample without East Asia would have predicted correctly four crises (Indonesia, Korea, Philippines, and Thailand) and missed only Malaysia (only a 20 percent crisis misclassification rate) and have false-alarms for 15 percent of the non-crisis observations. Using, instead, the entire tree, we miss all crises and misclassify 7.5 percent of the non-crisis observations. The V-fold cross-validation suggested no splits in the tree, classifying all observations as non-crises.

Berg et al. (2004, Table 4) verify how EWS models would have predicted the East-Asian crisis out of sample by estimating the DCSD and KLR models on monthly data over the period December 1985-April 1995 and using them to check in how many months over the period May 1995 – December 1996 the estimated probability of crisis would have been above the cutoff probability for the East-Asian countries that experienced a crisis in 1997. As in previous instances, the comparison with the BCTs results is difficult because the crisis definitions do not match (e.g., EWS models do not consider Philippines a crisis country). Moreover, Berg et al. run a proper out-of-sample exercise based only on pre-crisis information, whereas we use post-crisis experience in other countries to estimate the BCT in Figure 5. In this case, the out-of-sample performance of the EWS is quite good with the percentage of crises correctly called in 24 months at 84 percent for DCSD and 75 percent with KLR. A BCT based only on the first split would have yielded similar results with a prediction rate of 80 percent (four crises out five), whereas the entire tree of Figure 5 would have been much inferior to EWS models.

The RandomForests algorithm predicts the crisis in Korea and misses all other four East-Asian crises, with a crisis misclassification rate of 80 percent and a non-crisis misclassification rate of 29 percent. This performance is marginally preferable to that of the entire tree in Figure 5.

VI. Global Conditions Versus Country-Specific Indicators

This section addresses the question of the role of global economic conditions in capital account crises. So far, we have focused on predicting crises and considered only lagged values of variables to study their leading indicator properties and to avoid endogeneity problems caused by contemporaneous domestic indicators (e.g., an association between low *contemporaneous* reserve cover and crises would be no evidence of the indicator role of reserve cover because reserves typically drop during crises.) We now include contemporaneous values of measures of global conditions that each country faces, such as an export-weighted index of real commodity prices and an index of import demand by trading partners, which are exogenous to contemporaneous developments in individual countries.We find that these indicators play an important role in improving the in-sample classification by isolating subsets of observations with a higher frequency of crises than in BCTs based only on lagged indicators.

Figure 6 shows a variant of the baseline tree of Figure 1 that allows for contemporaneous global indicators. ***Lagged reserve cover*** remains the most important variable with an unchanged threshold of 81 percent. However, on the left (risky) branch of the tree, the ***real level of commodity export prices*** replaces external debt over GDP in splitting observations with low reserve cover. When the real level of commodity export prices is more than 14 percent below its past country-specific average, 22.6 percent of observations are crises (21 out of 93), while only 2.8 percent of observations are crises (2 out of 72) when commodity export prices are above this threshold. A level of ***external debt to GDP*** above 24 percent further raises the frequency of crises to 27.3 percent.

Returning to the top of the tree and moving down the right (safe) branch characterized by high reserve cover, we find the same split of the baseline tree based on the ***previous year WEO real growth forecast***. But the ***contemporaneous growth in (non-oil) import demand by trading partners*** now allows us to split further the node with a weak growth outlook. A strong contemporaneous growth in import demand by trading partners (above 6.4 percent) can offset the impact of the low growth outlook forecasted in the previous year, reducing the frequency of crises to 2.4 percent (1 crisis out of 41 observations). Instead, if import demand by trading partners is weak, the low growth outlook translates into a very high frequency of crises equal to 46.2 percent (6 out of 13 observations).

The in-sample fit of this tree is good. We only fail to predict 7 crises out of 34 (an error of 20.6 percent) and we wrongly predict 62 crises out of 520 non-crisis observations (an error of 11.9 percent). Therefore, false alarms in percent of total alarms drop to 69.7 percent from the 81.5 percent of the BCT in Figure 1.

VII. Conclusions

This paper uses Binary Classification Trees (BCTs) to study the in-sample and out-of-sample forecasting properties of a large set of indicators of capital account crises. While BCTs have been previously used in few studies of currency and sovereign debt crises, this is the first application of BCTs to capital account crises ("sudden stops"). Our results shed light on the relative importance of leading indicators of crises and on their interaction, suggesting that BCTs could be a useful complement to existing crisis prediction methods.

The interaction of international reserves, current account balances, and short-term external debt constitutes the backbone of the BCTs presented in this paper. The evidence supporting the leading indicator role of these three variables is robust. A measure of reserve cover that combines them (the *lagged* ratio of international reserves to the sum of current account deficits and short-term external debt) is the best splitter of the full sample into crisis and non-crisis-prone observations, no matter whether we allow for *contemporaneous* global cyclical conditions or not. In some instances, splits based only on reserve cover or its components lead to better out-of-sample prediction performance than fully-fledged trees. Lagged reserve cover is the preferred leading indicator of crises also in a subsample that excludes the East-Asian countries and in all subsamples that include years from 2002 on. In earlier subsamples (up to 2000 and 2001), the current account balance-to-GDP ratio becomes the top splitter but a combination of high current account deficits and a high ratio of short-term external debt to reserves characterizes crisis-prone observations. This latter evidence suggests that the concurrent build-up of reserves and tranquil international financial markets of recent years is not the only reason why BCTs prefer these three indicators to others.

If our estimates were taken at face-value, they would suggest a much stronger role for macroeconomic variables than for financial sector variables. But caution should be used in reading this suggestive evidence. First, this paper uses BCTs mostly as a forecasting tool. The predominant role of macroeconomic variables is, therefore, only a sign that they are better leading indicators at a one-year horizon, which is not that surprising in view of the balance-sheet, and therefore backward-looking nature, of most financial sector indicators, although the few market-based forward-looking indicators that we consider such as EMBI spreads also do not seem to have much information content. Second, the evidence of this paper leaves it open that, although financial sector variables are not good leading indicators of crisis inception, they might still play a pivotal role in determining how disruptive capital account crises might be once they occur. In other words, financial sector weakness is a vulnerability that raises crisis risks only when it assumes a macroeconomic dimension (e.g., short-term foreign indebtedness of banks and corporations needs to be high not only in relation to other countries but also in relation to international reserves). Further research on the crisis role of financial variables is clearly needed.

BCTs have a clear advantage in permitting analysis of a large number of potential indicators and their interactions but they also have important limitations. One of their unappealing features is potential instability (consider, for example, how our trees change between Figure 3 and Figure 4 only as a result of an additional year of data). In this paper, we use Breiman's RandomForests algorithm—which compares forecasts from a multitude of trees estimated using randomized samples and indicator sets—to verify whether the classification of each observation remains stable as trees change and to check the robustness of our results. We find that the out-of-sample performance of our preferred trees is comparable to that of the RandomForests algorithm.

Another limitation of the BCT algorithm is that each split is considered sequentially without taking into account how it will affect further splits down the tree. That is, in deciding which split to use, the BCT algorithm searches for the indicator and threshold that yield the largest improvement in partitioning a given sub-sample without considering how difficult it would be to partition further the resulting two sub-samples. To remedy this problem, the BCT algorithm should choose splits in a forward-looking manner but the associated computational costs would quickly become prohibitive.

Finally, several crisis episodes have an important contagion component. It is difficult to quantify contagion, let alone predict it. But factoring contagion considerations in crisis prediction models is of critical importance.[18] We plan to experiment with possible contagion indicators in future work.

[18] See, for example, Kaminsky and Reinhart (2000).

Table 1. Capital account crisis episodes by year of inception

Year	Countries					
1994	Algeria	Bulgaria	Mexico	Turkey	Ukraine	Venezuela	
1995	Argentina						
1996	Hungary						
1997	Czech Republic	Indonesia	Israel	Korea	Malaysia	Philippines	Thailand
1998	Brazil	Pakistan	Russia	Ukraine			
1999	Colombia	Ecuador	Lithuania	Romania			
2000							
2001	Argentina	Lebanon	South Africa	Turkey	Venezuela		
2002	Brazil	Colombia	Israel	Uruguay			
2003	Dominican Republic	Jamaica					
2004							
2005							

Table 2. Number of Countries by Reserve Cover Range Over Time

	Reserve Cover				
Year	0 to 0.5	0.5 to 0.812	0.812 to 1	1 to 2	>2
2000	7	11	3	15	13
2001	2	9	8	16	14
2002	6	6	5	15	17
2003	3	7	2	17	20
2004	4	5	6	10	24
2005	3	6	2	21	17

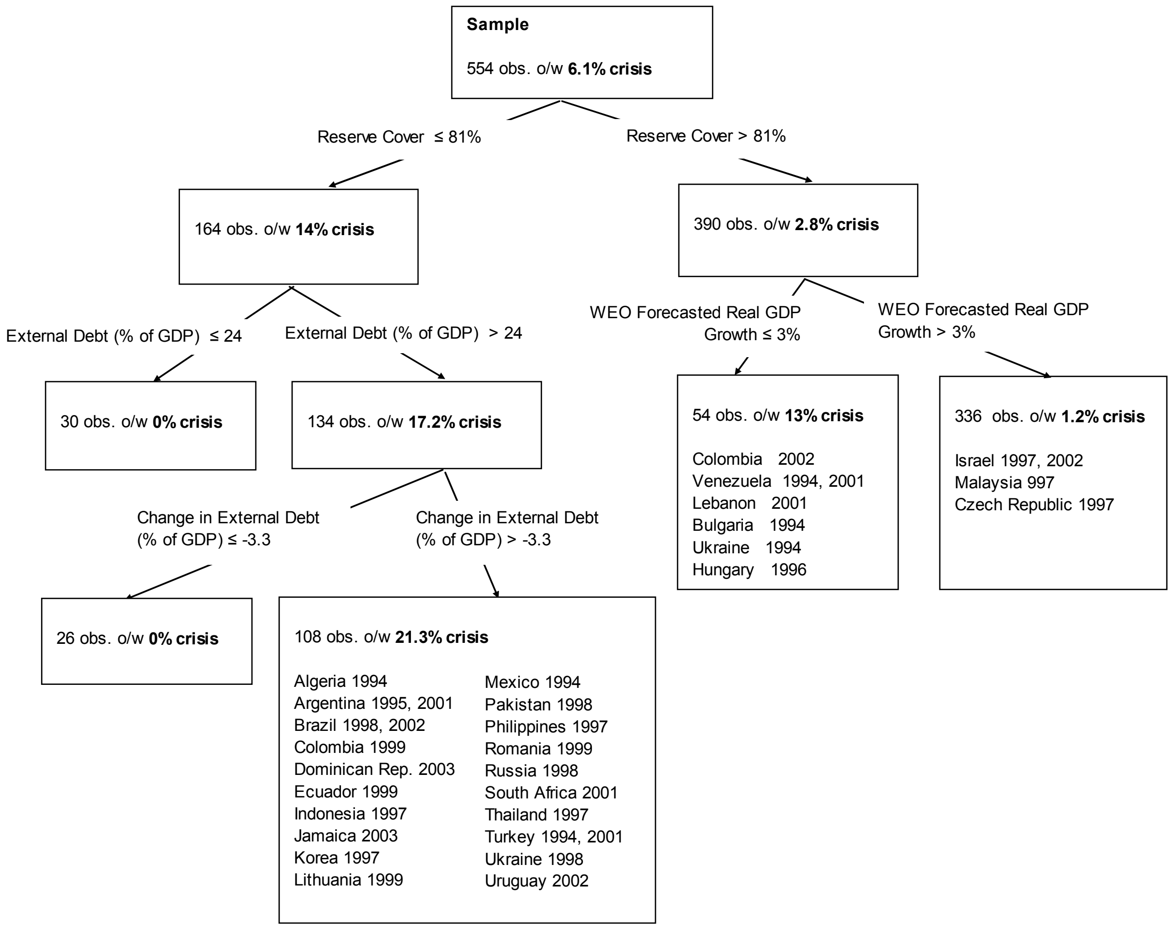
Notes: Reserve Cover defined as as the ratio of gross international reserves to the sum of the short-term external debt (from BIS) and the current account deficit (zero if it is a surplus).

Table 3. Descriptive statistics of key indicators

Variable	Number of Observations	Mean	Std. Dev.	25th Percentile	Median	75th Percentile
Reserve Cover 1/	521	1.836	2.344	0.691	1.133	2.188
External Debt/GDP	496	48.194	22.999	32.549	45.948	59.919
Change in External Debt/GDP	448	-0.179	9.715	-3.578	-0.268	2.735
WEO Real GDP Growth Forecast	526	4.366	1.910	3.500	4.500	5.463
Current Account/GDP	544	-0.024	0.061	-0.052	-0.026	0.002
ST Debt/Reserves	522	1.027	2.053	0.321	0.556	0.980
Fiscal Balance/GDP	495	-3.375	3.835	-5.195	-3.034	-1.177
Exchange Rate Misalignment 2/	546	0.522	23.646	-15.284	0.782	17.829
Deviation of Commodity Prices from past average 3/	542	-9.084	22.914	-21.981	-12.982	-4.039
Change in Real Import Demand By Trading Partners 3/ 4/	542	8.817	4.841	5.673	9.269	12.017

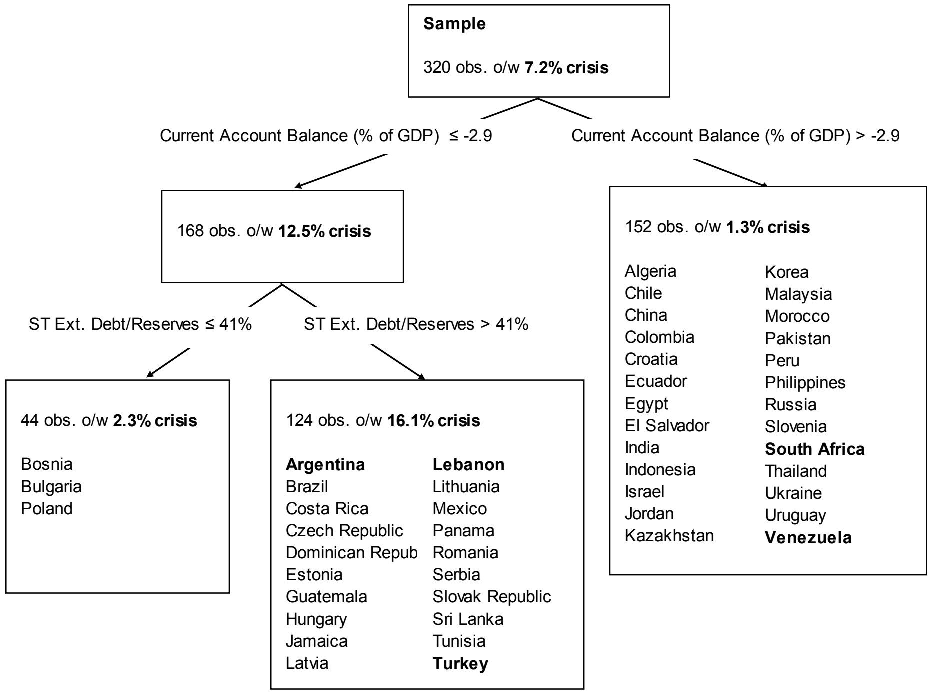
1/ Reserve Cover defined as as the ratio of gross international reserves to the sum of the short-term external debt (from BIS) and the current account deficit (zero if it indicates a surplus).
2/ Misalignment relative to average REER from 1960 up to previous year.
3/ Contemporaneous values.
4/ Excludes Oil

Figure 1. Binary Classification Tree Based on 1994-2005 Sample and Crisis Episodes



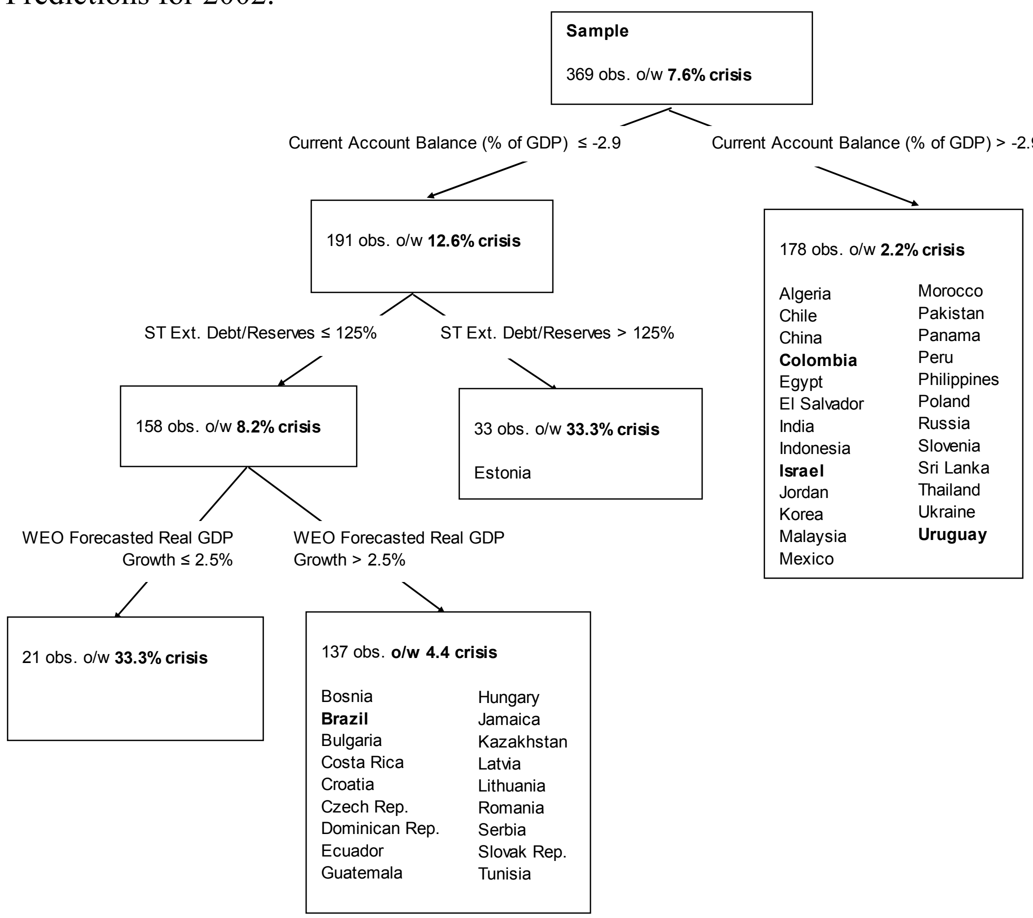
Notes: All variables used are lagged, corresponding to the value in the previous year. Reserve Cover defined as as the ratio of gross international reserves to the sum of the short-term external debt (from BIS) and the current account deficit (zero if it indicates a surplus).

Figure 2. Binary Classification Tree Based on 1994-2000 and Out-of-Sample Predictions for 2001.



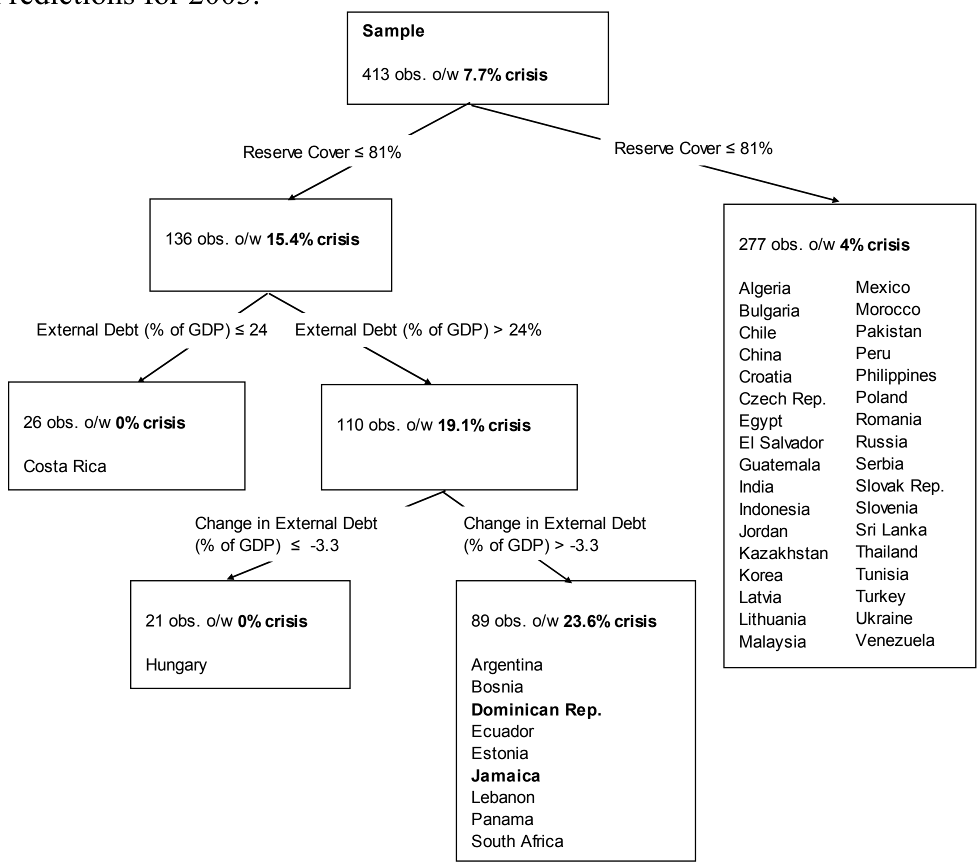
Notes: All variables used are lagged, corresponding to the value in the previous year. Sample frequencies reported correspond to in-sample values for 1994-2000. Countries listed based on their out-of-sample classification for 2001, with crisis episodes in bold.

Figure 3. Binary Classification Tree Based on 1994-2001 and Out-of-Sample Predictions for 2002.



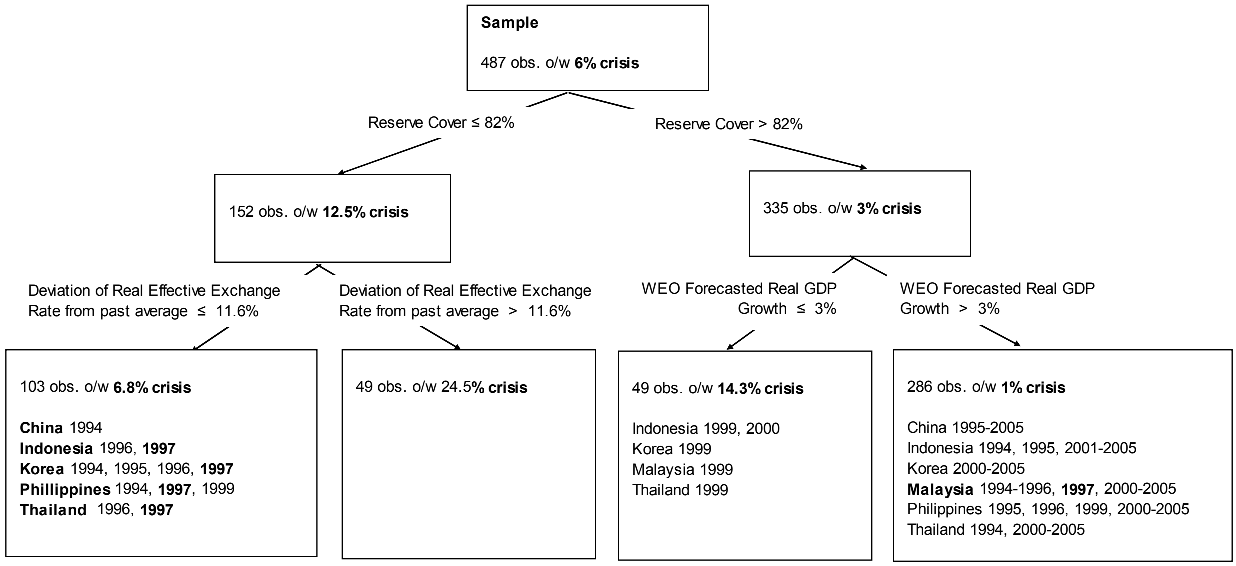
Notes: All variables used are lagged, corresponding to the value in the previous year. Sample frequencies reported correspond to in-sample values for 1994-2001. Countries listed based on their out-of-sample classification for 2002, with crisis episodes in bold. Countries that experienced a crisis in 2001 are excluded.

Figure 4. Binary Classification Tree Based on 1994-2002 and Out-of-Sample Predictions for 2003.



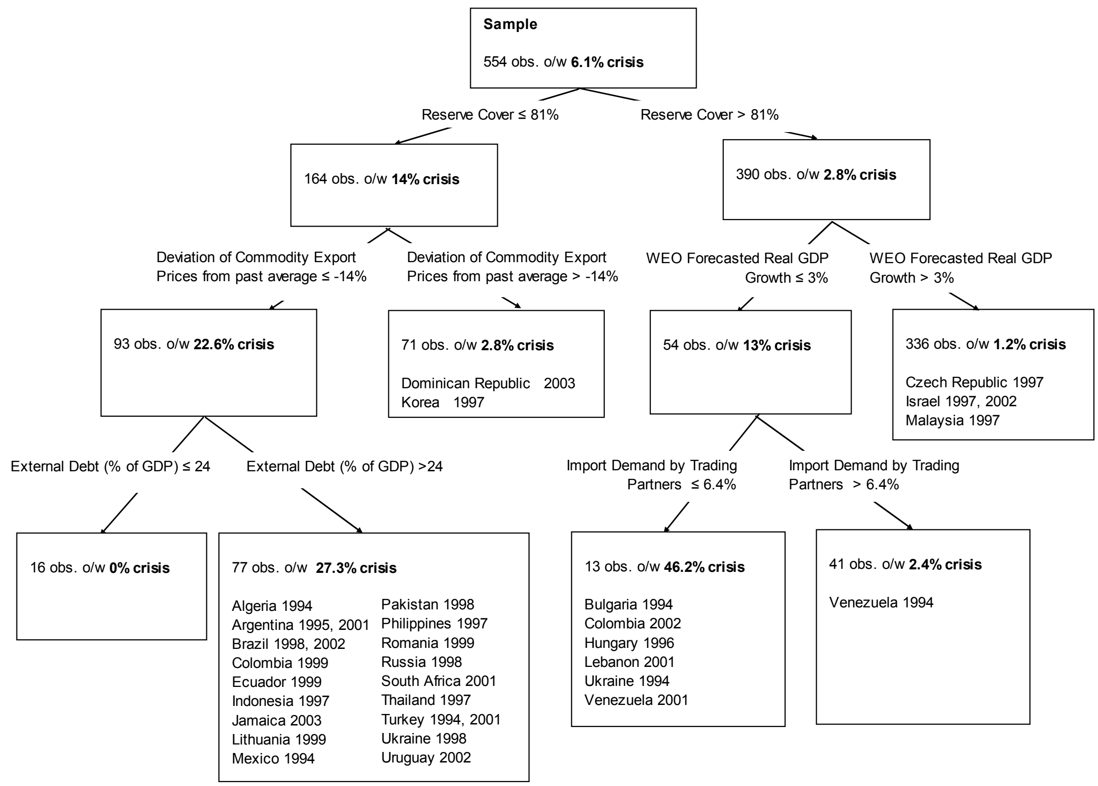
Notes: All variables used are lagged, corresponding to the value in the previous year. Sample frequencies reported correspond to in-sample values for 1994-2002. Countries listed based on their out-of-sample classification for 2003, with crisis episodes in bold. Countries that experienced a crisis in 2002 are excluded.

Figure 5. Binary Classification Tree Based on 1994-2005 Excluding East Asia and Out-of-Sample Predictions for East Asia.



Notes: All variables used are lagged, corresponding to the value in the previous year. Sample frequencies reported correspond to in-sample values for 1994-2005 excluding East Asia. East Asian observations listed based on their out-of-sample classification, with crisis episodes in bold. Countries that experienced a crisis in previous year are excluded.

Figure 6. Binary Classification Tree Based on 1994-2005 Including Contemporaneous Global Demand Variables.



Notes: All variables used are lagged, except for the ones for real commodity prices and import demand which are contemporaneous. Import demand by trading partners indicates the change in the import volume of goods excluding oil.

Appendix I. Capital Account Crises: 1994-2004

Table A1. Capital Account Crises 1994-2004 1/

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Country	Capital account crises[2]	Sudden stop 2[3]	Sudden stop 1[4]	EWS[5]	Sovereign default[6]	Fund program[7]	Banking crisis[8]	Corporate crisis[9]
Algeria	1994	1995	1995, 2004	1993-94	1995-96	1994-97		...
Argentina	1995, 2001	2001	2001	1995, 2001	2001-04	1992-96, 2000-01	1995, 2001-02	2002
Bosnia & Herzeg.		...	...		...	1995, 1998-2002		...
Brazil	1998, 2002	2002	2002	1998, 2000	1983-94	1998-99, 2001-03	1994-99	1995, 1998, 2001-02
Bulgaria	1994	1996	1996	1994	1994	1991-94, 1997-00, 2003		...
Chile		1998, 1999	1998, 1999	1995-96, 1998-99				
China		1998	1998					
Colombia	1999, 2002	1998, 1999	1998, 1999	1993, 1995			1999-00	2000
Costa Rica							1994-97	...
Croatia				1999		1994-95, 1997		...
Czech Republic	1997	1997	1997	1993				
Dominican Rep.	2003	1999, 2003	1999, 2003	1994, 1998-2000	1981-02	1998, 2003-05		...
Ecuador	1999	2000, 2004	2000, 2004	1998-2000	1982-95, 1999-00	1994, 2000-03	1995-02	...
Egypt		1999	1999	1993				
El Salvador		2001	2001	1995, 2002	1981-96			...
Estonia						1995		...
Guatemala		2002	2002	1995				...
Hungary	1996	1996	1996	2003				1995
India				1996-97			1991-94	1993
Indonesia	1997	1998, 1999	1998, 1999	1997-98	1998-2000, 2002	1997-00, 2003	1992-95,1997-02	1997-98, 2000
Israel	1997, 2002	2002	2002	1993, 1998				...
Jamaica	2003	2002	2002	1994, 2003			1996-00	...
Jordan		2002	2002	1993, 1995		1994-99, 2002		...
Kazakhstan				1994		1993-96, 1998		...
Korea	1997	1997	1997	1995-97		1997-98	1997-02	1996-98
Latvia				1994-95, 2003-04		1992-94		...
Lebanon	2001	2000	2000					...
Lithuania	1999	2000	2000			1992-97		...
Malaysia	1997	1998	1998	1994, 1997-98			1997-01	1998
Mexico	1994	1995	1995	1994		1995	1994-97	1998
Morocco								
Pakistan	1998			1993, 1995-96, 1999-2000	1998-99	1994, 1999, 2001-02		
Panama		2000	2000		1983-96	1994, 1996, 1998		...
Peru		1997	1997	1998	1983-97	1997		
Philippines	1997	1998	1998	1993, 1997		1997-00		1997, 2001
Poland				1993-94, 2001		1994		2001
Romania	1999	1999	1999	1999		1994, 1997, 2003		...
Russia	1998	1999, 2000	1999, 2000	1997-98	1995-00	1994-98		1998
Serbia & Mont.		...	...	...	...	2000-03, 2005		...
Slovak Republic						1993-94		
Slovenia		2003	2003	2000	...			
South Africa	2001	2000	2000	1996, 1998				
Sri Lanka		1995	1995, 2003	1993, 1995, 1997-98		1992-94, 2001-03, 2005		...
Thailand	1997	1997	1997	1995, 1997		1997-99	1997-02	1996-97
Tunisia		1995, 2000	1995, 2000	2000			1991-95	...
Turkey	1994, 2001	2001	2001	1993-94, 2000-01		1994-95, 1999-02	1994, 2000-02	1994, 1998, 2000-02
Ukraine	1994, 1998	1995	1995	1994	1998-00	1994-98		...
Uruguay	2002	2002	2002	2002	2002	1998, 2002-04	2002	...
Venezuela	1994, 2001	1999, 2000	1999, 2000	1994, 2002	1995-97		1993-97	1994-95, 1998, 2002

Notes:

1/ Whenever a country is not in the sample used to identify a particular type of crisis, a "..." is placed in the cell. Blank cells mean that there is no crisis for that particular country.

2/ Capital account crises were obtained primarily from sudden stops, but adjusted using information from other crisis definitions, when applicable. The year of inception of the crisis is reported. The final classification was made after accounting for the country desks' comments.

3/ Sudden stop 2: a Sudden stop 1 in which also:
- net private capital flows/GDP have declined by at least 3% from the previous year and 2% from two years before.

4/ Sudden stop 1: a year in which one of the following holds (where means and standard deviations are computed based on the 1993-2004 values deflated by the US CPI):
- net private capital flows are at least 1.5 standard deviations below their mean and have declined by at least 0.75 standard deviation from the previous year; or
- net private capital flows have declined by at least 1.5 standard deviations from the previous year and at least 0.75 standard deviation from two years before; or
- net private capital flows have decline by at least 0.75 standard deviations from the previous year and at least 1.5 standard deviations from two years before.

The source of net private capital flows data is the WEO database.

5/ EWS= 'Early Warning System'. The classification is based on 2-standard-deviation threshold for exchange rate pressure indicator. Source: ICM.

6/ Source: Manasse and Roubini (IMF WP 05/42), updated with sovereign debt default indicator Debrun (WEO, 2004).

7/ Record only the years, for each country, when total disbursements were bigger than total repayments (principal, charges, and interest).

8/ Source: Demirguc-Kunt and Detragiache (IMF Staff papers, 1998), updated by MFD.

9/ Source: RES (based on CVU).

Appendix II. Country-by-country details on the selection of crisis episodes.

Algeria

- **1994** chosen as the inception year because of the EWS and Fund program indicators, while sudden stop and sovereign default indicators pointed to 1995.

Argentina

- **1995** chosen because of EWS and banking crisis indicators and a moderate decline in net private capital flows.
- **2001** chosen because of sovereign default, banking crisis, sudden stop, and EWS indicators.

Brazil

- **1998** chosen because of EWS and Fund program indicators and a moderate cumulative decline in net private capital flows.
- **2002** chosen because of sudden stop and Fund program indicators.

Bulgaria

- **1994** chosen because of sovereign default and EWS indicators, with sudden stop indicators suggesting 1996 instead.

Chile

No capital account crises were identified in this period, even though both sudden stop and EWS indicators suggest a crisis in 1998. Capital outflows during that year were exacerbated by a domestic portfolio reshuffling away from dollar liabilities and toward dollar assets abroad, resulting from liberalization of the capital account and the elimination of the exchange rate band, rather than a loss of access to international capital markets. Thus, that episode was not considered a capital account crisis.

China

No capital account crises were identified in this period, even though both sudden stop indicators suggest a crisis in 1998. It is possible that the observed decline in net private capital flows was due to data problems (much of the flows take the form of trade credit and errors and omissions given the existence of exchange and capital controls), and there was not enough disruption in the economy to justify classifying that year as a capital account crisis.

Colombia

- **1999** chosen because of sudden stop indicators.
- **2002** chosen because of contagion from Brazil (the sovereign spread reached 1100 basis points during that year) and a moderate decline in net private capital flows.

Czech Republic

- **1997** chosen because of sudden stop indicators.

Dominican Republic

- **2003** chosen because of sudden stop indicators and real depreciation. Note that the sudden stop indicators also suggest a crisis in 1999, which was ruled-out because growth remained high, inflation low, and there was no significant change in the exchange rate nor in NIR during that year.

Ecuador

- **1999** chosen mainly because of sovereign default (EWS suggests 1998-2000 and sudden stop indicators suggest 2000 instead). Note that sudden stop indicators also suggest a crisis in 2004. That decline in net private capital flows was due to the completion of a pipeline causing foreign investment to decline to levels close to its historical average.

Egypt

No capital account crises were identified in this period, even though sudden stop indicators suggest a crisis in 1999**.** The decline in net private capital flows following 1998 was the result of a change in the privatization policy, which reduced the supply of assets available to foreigner investors.

El Salvador

No capital account crises were identified in this period, even though sudden stop indicators suggest a crisis in 2001**.** The change in net flows during that year was due to currency substitution and reclassification of assets associated with dollarization.

Guatemala

No capital account crises were identified in this period, even though EWS points to exchange rate pressures in 1995 and the sudden stop indicators suggests a crisis in 2002 respectively**.** The sharp decline in net private capital flows in 2002 was partly due to the disbursements of an Eurobond.

Hungary

- **1996** chosen because of sudden stop indicators.

India

No capital account crises were identified in this period. Despite a moderate decline in net private capital flows in 1995 and EWS pointing to currency pressures in 1996-97, none of these episodes could be described as a capital account crisis.

Indonesia

- **1997** chosen even though net private capital flows deteriorated substantially only in 1998-99. Choice of 1997 is supported by EWS, Fund involvement, banking crisis, and corporate crisis indicators.

Israel

- **1997** chosen because of substantial decline in net private capital flows, with EWS pointing to exchange rate pressures in 1998.
- **2002** chosen because of sudden stop indicators, possibly reflecting an escalation of conflict in the occupied territories.

Jamaica

- **2003** chosen as the crisis year, even though sudden stop indicators suggest 2002. The crisis and real exchange rate depreciation occurred in the first quarter of 2003, which is picked-up by EWS.

Jordan

No capital account crises were identified in this period. The decline in net private capital flows in 2002 is perceived to be driven, at least in part, by the build-up to the Iraq War.

Korea

- **1997** chosen because of sudden stop indicators, currency crisis, and Fund program.

Lebanon

- **2001** chosen as the inception of the crisis even though sudden stop indicators suggest 2000. It was only in 2001 that enough disruption was created to justify a capital account crisis classification (it became very difficult for the government to roll-over its debt and it eventually required exceptional financing under Paris II in 2002).

Lithuania

- **1999** chosen because of contagion from Russia even though sudden stop indicators suggest 2000 as the crisis year.

Malaysia

1997 chosen even though net private capital flows deteriorate substantially only in 1998. Choice of 1997 is supported by EWS and banking crisis indicator.

Mexico

- **1994** chosen mainly because the currency crisis takes place late in that year, with net private capital flows deteriorating substantially only beginning in early 1995.

Pakistan

- **1998** chosen mainly because of sovereign default with EWS pointing to currency pressures in the following two years.

Panama

No capital account crises were identified in this period, despite sudden stop indicators suggesting 2000 as a crisis year. There was a reduction in short-term capital inflows through the banking system during 2000, reflecting temporary concerns about Panama being on the

FATF's list of noncooperative countries (later removed from the list in 2001), which was followed by a rebound.

Peru.
No capital account crises were identified in this period, despite sudden stop indicators suggesting 1997. The observed decline in net private capital flows during that year may be partly due to a debt-restructuring operation.

Philippines
- **1997** chosen even though net private capital flows deteriorate substantially only in 1998. Choice of 1997 supported by EWS, Fund program, and corporate crisis indicators.

Poland
No capital account crises were identified in this period, despite sudden stop indicators suggesting 1994. The decline in net private capital flows in that year may be a result of data problems as there were no capital account pressures in the mid-1990s.

Romania
- **1999** chosen because of sudden stop and EWS indicators.

Russia
- **1998** chosen mainly because of currency crisis even though net private capital flows deteriorate substantially only in 1999-2000. Choice of 1998 is also supported by EWS and corporate crisis indicator.

Slovenia
No capital account crises were identified in this period, despite sudden stop indicators pointing to crisis in 2003. The decline on net private capital flows in 2003 was a correction following 2 years (2001-02) of exceptional inflows.

South Africa
- **2001** chosen despite sudden stop indicators suggesting 2000 instead. Although the decline in net flows started in 2000, the large real exchange rate depreciation only occurs in 2001, which is widely perceived as the crisis year. The exchange rate pressures indicated by EWS in 1996 and 1998 did not generate enough disruption to be considered capital account crisis episodes.

Sri Lanka
- **No capital account crises in this period,** despite sudden stop indicators suggesting 1995 and 2003 as crisis years. None of these episodes created enough disruption to warrant being classified as a capital account crisis.

Thailand

- **1997** chosen because of sudden stop, EWS, Fund program, banking crisis, and corporate crisis indicators.

Tunisia

- **No capital account crises were identified in this period.** The drop in net private capital flows picked-up by the sudden stop indicator in 2000 is a result of the large privatization inflows during 1999.

Turkey

- **1994** chosen because of EWS, Fund program, banking crisis, and corporate crisis indicators.
- **2001** chosen because of the same set of indicators of 1994 plus sudden stop indicators.

Ukraine

- **1994** chosen because of EWS indicator with sudden stop indicators pointing to 1995.
- **1998** chosen because of sovereign default.

Uruguay

- **2002** chosen because of sudden stop, EWS, sovereign default, Fund program, and banking crisis indicators.

Venezuela

- **1994** chosen because of EWS with sovereign default indicator pointing to 1995.
- **2001** chosen as the year the crisis gained momentum, despite sudden stop indicator suggesting 1999 instead.

REFERENCES

Berg, Andrew, and Catherine Pattillo, 1999, "Are Currency Crises Predictable? A Test," *Staff Papers*, Vol. 46 (June), No. 2, pp. 107–38.

———, and Eduardo Borensztein, 2005, "Assessing Early Warning Systems: How Have They Worked in Practice?" *Staff Papers*, International Monetary Fund, Vol. 52, No. 3, pp. 462–502.

Boyd, John, Gianni De Nicolo and Abu Al Jalal, 2006. "Bank Risk Taking and Competition Revisited: New Theory and New Evidence." IMF Working Paper, forthcoming. (Washington: International Monetary Fund).

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

Breiman, Leo, 2001, "Random Forests," *Machine Learning*, Vol. 45, pp. 5–32.

Calvo, Guillermo, Leonardo Leiderman, and Carmen Reinhart, 1992, "Capital Inflows and Real Exchange Rate Appreciation in Latin America: The Role of External Factors," IMF Working Paper 92/62 (Washington: International Monetary Fund).

Catão, Luis, 2006, "Sudden Stops and Currency Drops: A Historical Look," IMF Working Paper 06/133 (Washington: International Monetary Fund).

Demirguc-Kunt, Asli and Enrica Detragiache, 1998. "The Determinants of Banking Crises in Developing and Developed Countries." *IMF Staff Papers*, Vol. 45(1), 81-109.

Frankel, Jeffrey A., and Andrew K. Rose, 1996, "Currency Crashes in Emerging Markets: An Empirical Treatment," *Journal of International Economics*, Vol. 41 (November), No. 3–4, pp. 351–66.

Frankel, Jeffrey A., and Shang-Jin Wei, 2004, "Managing Economic Crises," NBER Working Paper No. 10907 (Cambridge, Massachusetts: National Bureau of Economic Research).

Ghosh, Swati R., and Atish R. Ghosh, 2003, "Structural Vulnerabilities and Currency Crises," *Staff Papers*, International Monetary Fund, Vol. 50, No. 3.

Jeanne, Olivier, and Romain Ranciere, 2006, "The Optimal Level of International Reserves for Emerging Market Countries: Formulas and Applications," IMF Working Paper 06/229 (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), No. 1, pp. 1–48.

Kaminsky, Graciela, and Carmen M. Reinhart, 2000, "On Crises, Contagion, and Confusion," *Journal of International Economics*, Vol. 51 (June), No. 1, pp. 145–168.

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

Manasse, Paolo, and Nouriel Roubini, 2005, "'Rules of Thumb' for Sovereign Debt Crises," IMF Working Paper 05/42 (Washington: International Monetary Fund).

Reinhart, Carmen and Kenneth Rogoff, 2004. "The Modern History of Exchange Rate Arrangements: A Reinterpretation," *Quarterly Journal of Economics*, Vol.119(1), 1-48.

Reinhart, Carmen, Kenneth Rogoff and Miguel Savastano, 2003. "Debt Intolerance," *Brookings Papers on Economic Activity*, v2003(1), 1-74.

Van Rijckeghem, Caroline and Beatrice Weder, 2004. "The Politics of Debt Crises," CEPR. Discussion Paper No. 4683 (London: Center for Economic Policy Research).