

Beckmann, Daniela; Menkhoff, Lukas; Sawischlewski, Katja

Working Paper

Robust Lessons about Practical Early Warning Systems

Diskussionsbeitrag, No. 322

Provided in Cooperation with:

School of Economics and Management, University of Hannover

Suggested Citation: Beckmann, Daniela; Menkhoff, Lukas; Sawischlewski, Katja (2005) : Robust Lessons about Practical Early Warning Systems, Diskussionsbeitrag, No. 322, Universität Hannover, Wirtschaftswissenschaftliche Fakultät, Hannover

This Version is available at:

<https://hdl.handle.net/10419/22434>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Robust Lessons about Practical Early Warning Systems

Daniela Beckmann, University of Hannover, Germany

Lukas Menkhoff, University of Hannover, Germany

Katja Sawischlewski, University of Hannover, Germany¹

Discussion Paper No. 322

Oktober 2005

ISSN: 0949-9962

Abstract:

Early warning systems (EWSs) are subject to restrictions that apply to exchange rates in general: fundamentals matter but their influence is small and unstable. Despite this limitation four major lessons emerge: First, EWSs have robust forecasting power and thus help policy-makers to prevent crises. Second, policy-makers must decide about some EWSs' elements, such as the sensitivity of the forecasts. Third, EWSs' performance is increased by taking a logit model, shorter samples and a regional approach. Fourth, the finding of contagion may motivate policy to shield its economy against inefficient international financial markets.

JEL-Classification: F33, F37, F31

Keywords: early warning system, currency crises, emerging markets

Corresponding: Lukas Menkhoff, Department of Economics, University of Hannover, Königsworther Platz 1, D-30167 Hannover, Germany, menkhoff@gif.uni-hannover.de

¹ We would like to thank Marcel Fratzscher, Torben Lütje, Stefan Niermann and Hans-Herrmann Steinbeck for their very helpful notes. Moreover, we are grateful to all comments and advices received by participants of the 12th Annual Conference on Pacific Basin Finance, Economics, Accounting, and Business in Bangkok/Thailand, the Research Seminar at the University of Hannover/Germany, the Göttinger Workshop Internationale Wirtschaftsbeziehungen 2005 in Göttingen/Germany, the Annual Conference Research Committee Development Economics 2005, Verein für Socialpolitik at the Institute for World Economics in Kiel/Germany, the 20th Annual Congress of the European Economic Association (EEA) in Amsterdam/Netherlands, the 14th World Congress of the International Economic Association (IEA) in Marrakech/Morocco, as well as the critical study and discussion by students of the Master programme in Economics 2004/2005 at University of Minho in Braga/Portugal.

Robust Lessons about Practical Early Warning Systems

1 Introduction

Financial crises have not declined in number, frequency or severity over the last two decades, rather the contrary (Bordo et al., 2001). Each crisis causes enormous costs in the countries concerned. Even if many crises may help to promote overdue structural change, they are costly and it is a worthwhile objective to realize adjustments without this heavy toll. Thus, international financial institutions invest in researching early warning systems (EWS).² There is now a wide range of studies available, however, without real converging results: studies vary in coverage of countries and time, they apply different methods and they may even define crises quite differently. So are there any robust findings that might help policy makers? This paper does not add another EWS to the existing body of literature but systematically analyzes the robustness of main approaches and aims at contributing to the actual policy discussion. It is shown that, indeed, most findings critically depend on their empirical design. EWSs are a somewhat dubious subject from this perspective. However, there also emerge robust lessons that help us to better understand and implement EWSs.

The modern approaches to understanding determinants of currency crises started in the early 1990s. The largely unexpected shake-up of the European Monetary System in 1992/93 motivated Eichengreen, Rose and Wyplosz (1995, 1996). Thereafter, attention focused more on developing countries (Frankel and Rose, 1996), spurred by the Mexican crisis (e.g. Sachs, Tornell and Velasco, 1996) and, above all, by the Asian crisis (e.g. Kaminsky, 1998, Kaminsky, Lizondo and Reinhart, 1998, Berg and Pattillo, 1999). More currency crises followed, e.g. major ones in Russia and Argentina, stimulating ongoing research in common factors and an optimal EWS design. In the beginning, efforts were directed into analyzing an increasing set of variables, collecting larger samples and introducing multivariate techniques. Recent trends in this field point towards further modifications in method, such as applying a multinomial logit approach (Bussiere and Fratzscher, 2002), a Markov

² See for example Berg et al. (2000) as well as the many studies conducted by IMF staff (e.g. Kaminsky, Lizondo and Reinhart, 1998, Berg and Pattillo, 1999, 1999a, Edison, 2003, Abiad, 2003, and Sy, 2004). Furthermore, see Kamin, Schindler and Samuel (2001), representing research by the Federal Reserve Board. Bussiere and Fratzscher (2002) provide an example of ECB work on this topic.

Switching approach (Martinez-Peria, 2002, Abiad, 2003) and extreme value theory to identify crises (Pozo and Amuedo-Dorantes, 2003). Another trend is towards focusing on regional studies, such as Esquivel and Larrain (1999), Brueggemann and Linne (2002) or Pasternak (2003).

The universe of proposed EWSs thus stretches over five dimensions: variables, sample period, sample countries, crisis definition and method applied. Each EWS is a combination regarding these five dimensions, so that – due to focus and product differentiation – usually results cannot be compared to each other. Moreover, most studies do not present extensive robustness checks, so it is not clear which improvement claimed will last under different conditions. There is thus an obvious need to make results comparable in order to derive useful policy lessons. It is, however, also obvious that due to the complexity involved, analytical restrictions are needed. Fortunately, several earlier studies have carefully analyzed the usefulness of variables and found that variables often point in the same direction, in particular in a multivariate approach. As, moreover, all variables are rather more able to give a warning signal of a possible upcoming crisis than predict the timing and exact occurrence of the latter, a selection of few meaningful fundamental indicators will work (Salvatore, 1999). In this respect we follow the suggestion of Berg and Pattillo (1999, p. 573 et sqq.) to rely on their handful of variables only.

Regarding the issue of robustness, Edison (2003) is the most complete study published. She basically extends the leading indicator work of Kaminsky, Lizondo and Reinhart (1998) – in short: KLR – in four of the five dimensions mentioned above, although with different intensity. Our study also departs from KLR, who presented the broadest approach at that time and has thus been a benchmark study until today. Going beyond Edison (2003), we systematically apply several methods to generate warning signals, we analyze the influence of crisis definition more thoroughly and we consider some recent proposals to improve forecasting power of EWSs.

We find, as did Berg and Pattillo (1999) and Edison (2003), that it becomes more difficult over time to replicate earlier crisis dates. So the dependent variable may change from study to study even if all parameters are kept the same. Fortunately, this haziness in data does not seem to influence results too much.

Berg and Pattillo (1999) show that a logit approach can improve results of EWSs. To some extent, this holds for our extensions, too. We test their logit model

with variables given as number as well as with variables given in a reduced 0-1-form. The two approaches do not differ that much although the reduced form yields somewhat better results. One surprise is, however, the implementation of the multinomial logit proposed by Bussiere and Fratzscher (2002). This technique does not provide the expected useful forecasting results when applied out-of-sample but seems to depend decisively on the underlying data.

Regarding the crisis definition, it is found that the choice between the several definitions offered by the literature is non-trivial. Comparing the KLR approach with the narrower Frankel and Rose (1996) – in short: FR – proposal and the broader Bussiere and Fratzscher (2002) – in short BF – measure indicates: It is the KLR measure which seems most practical.

Further examinations on EWS parameters in order to improve forecasting ability reveal that there are indeed options available: we find that the recent trend in the literature towards focusing on regional EWS has some justification in the sense that it helps to improve forecasts. Finally, due to instability over time, shorter samples also tend to provide better forecasts. One should be aware, however, that regional EWSs as well as shorter samples tend to homogenize the problem: explanations are thus "better" because they are easier and not necessarily because one would understand better.

In summary, our understanding of EWSs – which basically aim at predicting larger exchange rate changes – may be guided by core insights from the empirical exchange rate literature: fundamentals matter but their influence is small and unstable (Frankel and Rose, 1995, Sarno and Taylor, 2002). Keeping these limitations in mind, there emerge four major lessons on the elements of robust EWS and on appropriate policy design: First, EWSs have robust forecasting power. Changes in core fundamentals signal future problems and thus help policy-makers to prevent crises. Second, policy makers must decide about pros and cons of competing crisis definitions as well as about the calibration of EWS (false alarms vs. missed crises). Third, EWSs' performance is increased by taking a logit model to condense information from various fundamental variables and by taking shorter samples and regional EWSs to address instability over time and countries. Fourth, the robust forecasting contribution of a regional contagion dummy indicates that international financial markets do not function always efficiently. Policy has to take this into account and may

want to have instruments at hand to partly disintegrate its economy from detrimental shocks.

The remainder of this paper is organized as follows: Section 2 presents our extension of the KLR approach over time and countries. Section 3 examines the advantageousness of several logit approaches compared to the composite indicator. The pros and cons of various crisis definitions are analyzed in Section 4. Further approaches to improve forecasts are discussed in Section 5. Section 6 concludes by shedding light on the design of robust EWSs on the one hand and resulting policy conclusions on the other.

2 KLR-extensions over time and countries

2.1 Reproducing the KLR approach

In order to tie up to previous research and assure data consistency, we begin our robustness analyses with a replication effort of the benchmark KLR approach. Starting with the crisis definition, KLR use an index of exchange market pressure (in short: EMPI) for each country that is constructed as a weighted average of monthly percentage changes in the nominal exchange rate (NER) and in gross international reserves (RES). These two components are weighted in such a way that they have the same conditional variance. Reflecting the pressure on the currency as an increase of the index, either due to a currency's depreciation or due to a loss of international reserves, a crisis (C) is considered a period in which the EMPI is more than three standard deviations above its country-specific mean.

$$\text{EMPI}_t^i = \frac{\text{NER}_t^i - \text{NER}_{t-1}^i}{\text{NER}_{t-1}^i} \cdot \frac{\sigma_{\Delta\text{NER}}^i}{\sigma_{\Delta\text{RES}}^i} \cdot \frac{\text{RES}_t^i - \text{RES}_{t-1}^i}{\text{RES}_{t-1}^i}; \quad C_t^i = \begin{cases} 1, & \text{if } \text{EMPI}_t^i > \overline{\text{EMPI}}^i + 3\sigma_{\text{EMPI}}^i \\ 0, & \text{otherwise} \end{cases}$$

Periods of hyperinflation, with price level increases of more than 150% in the previous six months, are treated separately, forming sub-samples with extra EMPI and crisis calculations for each country. The chosen sample contains 20 countries including 15 developing and 5 industrialized countries over the sample period from January 1970 to April 1995. The first four columns of [Annex 1](#) give an overview of identified crises by KLR and Edison (2003) as well as the results of our own replications.³

³ All data used derive from the IMF international financial statistics (IFS). When trying to reproduce the crises, 2.5 standard deviations instead of 3 above the EMPI mean appear to be most applicable. Our

After determining the crisis dates, the next step consists of choosing meaningful indicators to give a warning signal when crossing a formerly designed threshold value. KLR identify 13 out of 16 tested variables to be appropriate with the help of the so-called (adjusted) noise-to-signal ratio. This ratio contains the number of false signals as a share of possible false signals ($B/B+D$), divided by the number of good signals as a proportion of cases in which good signals could have been issued ($A/A+C$). Good and bad signals are distinguished according to the following matrix.

	Crisis (Within Next 24 Months)	No Crisis (Within Next 24 Months)
Crisis Signal	A	B
No Crisis Signal	C	D

With: A: Number of months in which the indicator issued a good signal

B: Number of months in which the indicator issued a false alarm or "noise", and thus a bad signal

C: Number of months in which the indicator failed to issue a signal (which would have been a good one)

D: Number of months in which the indicator desisted from issuing a signal (which would have been a bad one)

In order to create an easy-to-handle model we choose the set of 6 variables proposed and statistically verified by Berg and Pattillo (1999), namely deviation of real exchange rate from trend, growth of international reserves, of exports, and of M2 to international reserves, as well as current account to GDP and M2 to international reserves in levels (for details about the construction of the variables, see [Annex 2](#)).

Following Eichengreen, Rose and Wyplosz (1995), [Figure 1](#) sheds light on the average movements of the chosen indicators in a 48-month event window around the identified crisis dates and their average course in tranquil periods respectively. All indicators exhibit clearly perceptible patterns showing that their course is different before and after the crisis.

proceedings align with Edison (2003, p. 11). Nevertheless, the results are quite astonishing: KLR identify 75 crises, Edison finds 70 in the same sample, but only 47 of them match exactly with the ones of KLR. Our replication identifies 79 crisis, 36 of them correspond exactly to the original ones, 45 to the replication by Edison. Berg and Pattillo (1999, p. 565), who replicate the KLR benchmark as well, assign these discrepancies to differences in the raw data resulting from data revisions and individual data "cleaning". Taking into account that the latter fail to match 14 out of the 75 KLR crises only one year later while Edison misses 28 three years afterwards, our results five years later appear reasonable.

Table 1 (columns 1-2) presents information on the individual indicator's performance of our replication. For all six variables we reproduce noise-to-signal ratios smaller than one and thus consider all variables to contribute meaningfully to an EWS. The real exchange rate variable clearly performs best. Our results are in line with KLR and former replications. In order to assess the meaning of an indicator's signal better, we also calculate the so-called conditional probability ($A/A+B$) that indicates the probability of the occurrence of a crisis within a 24-month window after the respective indicator has issued a signal. Again, the real exchange rate variable proves to be the top performer with a conditional crisis probability of 67% when signalling.

2.2 Extending KLR over time

We then extend the KLR sample to a period from January 1970 to December 2002. This has different implications: First, the EMPI is highly sample-dependent. An extension changes not only the country-specific mean but also the calculated standard deviation. As a consequence, we identify slightly different crisis periods than before (see Annex 1, column 5).⁴ Taking a closer look at the countries concerned, a further conclusion emerges: As countries hit by the Asian financial crisis in 1997/1998 seem to be especially affected by shifts in crisis periods, one might suppose that heavy crises influence mean and standard deviation in a way that weaker crisis signals are no longer identified (e.g. in Malaysia). Thus, the identification of a currency crisis with the help of the KLR definition depends critically on the country's historical experiences and the considered time period respectively.

Second, all six indicators keep their forecasting power with a noise-to-signal ratio smaller than one (see Table 1, columns 1-3). Furthermore, indicators mainly keep their performance relative to each other with the real exchange rate variable staying the best performer.

2.3 Expanding KLR over countries

Following Edison (2003), we also consider 8 additional countries which experienced at least one currency crisis. Differently from extensions in a time dimension, additional countries have no influence on the crisis identification as the respective

⁴ Adding up to 93 crisis periods, 13 of the old ones can no longer be detected. Instead, 26 new crises are identified, but only 18 of them can be ascribed to the additional months considered.

EMPI and standard deviations are calculated separately for each country. [Annex 1](#) shows the identified crisis periods for the expanded sample: 104 for the shorter period (1970-1995) and 126 for the extended period (1970-2002).

[Table 1](#) (columns 4-5) sheds light on the indicators' performance after also including additional countries. Again, to a great extent, all indicators keep both a noise-to-signal ratio smaller than one as well as their performance order. The worse results for the four variables, real exchange rate, exports, M2 / reserves and M2 / reserves (level), support the idea that the more (heterogeneous) countries are examined, the more difficult it gets to achieve meaningful signals. Nevertheless, all chosen variables seem to be adequate.

2.4 Forecasting out-of-sample

Up to now, we have only considered the performance of univariate indicators in-sample. However, in order to assess the applicability of our variables to form an EWS, we should assess two further aspects. First, it seems reasonable to analyze the indicators' information jointly, as currency crises are often the result of simultaneously appearing weaknesses. Second, reliability out-of-sample is most important when aiming for a forecast.

Therefore, in a first step, we define crises for the whole period from 1970 to 2002, but rely on the period from 1970 to 1995 for crisis identification as well as for the design of the single indicator thresholds. The remaining 92 months from May 1995 to December 2002 represent the out-of-sample test period.

In a second step, we follow the idea of Kaminsky (1998, p. 14 et sqq.). Having looked at each indicator separately so far, we combine the information provided by all variables to a composite indicator:

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

Accordingly, the different indicator signals are summed up ($\sum S$) and weighted with their inverse noise-to-signal ratio ($1/\omega$) each). Depending on the number of variables signaling, the composite indicator can take different values. As pure values of the composite indicator do not say anything about the crisis probability, the associated (conditional) probabilities have to be calculated.

In a third step, finally, the in- and out-of-sample performance can be assessed. When comparing the composite indicator with the best univariate indicator, the real

exchange rate, two aspects emerge: First, in general, results out-of-sample are (as expected) worse in comparison to the in-sample consideration. Second, the composite indicator clearly outperforms the univariate benchmark, especially out-of-sample. As presumed, it is thus worthwhile to continue to work with the combined information of different indicators. Our results align with those of Goldstein, Kaminsky, and Reinhart (2000, p. 64 et sqq.), and are hence not explicitly displayed here.

3 From composite indicators to logit models

In this section, we go beyond the univariate KLR indicator approach and the composite indicator as a step towards a multivariate framework. We analyze whether multivariate logit models outperform former approaches and which among them might be most appropriate for practical purposes.

3.1 Two logit model alternatives

Describing the crisis probability as a non-linear function of different variables, logit and probit models feature the possibility of analyzing their statistical significance in a multivariate approach. Following Berg and Pattillo (1999; in short BP), we consider the 6 BP variables over the original KLR time and country sample. The variables enter the logit model in both the binary signal 0-1-form and with continuous values, alternatively. So far, we have considered a crisis window of 24 months, i.e. a signal is considered to be a good one if an indicator issues a signal and a crisis follows within the following 24 months. For the logit approach we thus have to transpose the dependent crisis variable (see also Bussiere and Fratzscher 2002, p. 10):

$$Y_t^i = \begin{cases} 1 & \text{if a crisis follows within 24 months} \\ 0 & \text{otherwise} \end{cases}$$

Table 2 shows the two estimated logit models. In both models, the same four variables are significant at the 1% level and their coefficients show the expected and correct signs.⁵ The odds ratio also displayed permits us to draw conclusions with re-

⁵ M2 / reserves as well as CA / GDP are excluded from the logit models due to their lack of significance. In the model with variables entering as binary signals, positive coefficients imply a rising crisis probability whenever the indicators cross their formerly designed thresholds and issue crisis signals. In the model with continuous values, coefficients' signs vary according to the definition of the respective variable. A positive coefficient for the ratio of M2 / reserves (level) e.g. implies a rising crisis probability whenever the ratio heightens. By contrast, expressed by the negative coefficient, decreasing reserves growth heightens the crisis probability.

gard to the conditional probability. If in the binomial logit model with binary signals e.g. the exports variable, emits a signal, the crisis probability is about 62.1 % higher than without the signal. Again, the real exchange rate variable shows the highest conditional probability.

As before, we are not only interested in in-sample, but also in out-of-sample results: Our performance assessment both in- and out-of-sample of the two logit alternatives in comparison to the composite indicator can be found in [Table 3](#). In order to assure comparability, we reduce the latter to exactly the same four variables entering the binomial logit models. Following Berg and Pattillo (1999), we take the two alternative cut-off probabilities of 25% and 50% to issue a signal. As the objective of an EWS consists in properly forecasting upcoming crises, the percentage of crises called correctly can probably be considered the most important criterion. However, without taking into account the number of false alarms as well, a performance assessment is only half-hearted.

At first sight, the trade-off between correctly called crises and false alarms becomes obvious and can be seen as valid over all different approaches. Hence we also return to this issue in our policy-related conclusions. At this point in time, let us take a closer look at the in-sample performance of the binary logit with variables e.g. entering in the form of binary signals: using the 50% cut-off, the last model provides an obviously better performance compared to the other two models as far as correctly signaled crises as well as false alarms are concerned. Bringing the cut-off probability down to 25%, the number of correctly called crisis periods is augmented to 53% – but at the cost of more false alarms. The models are thus difficult to compare. Out-of-sample, performance results once again worsen compared to in-sample estimations. Furthermore, with a 50% cut-off, the composite indicator shows the best performance. However, applying a 25% cut-off, it is the binomial logit model with signals which shows a higher percentage of correctly called crises as well as less false alarms. So taking the four cases considered in [Table 3](#) as the benchmark, each of the three models is best in one case but only the binary logit model is never the worst. Though performance results are not unambiguous, the binary logit model seems to perform most robustly.

3.2 The multinomial logit

So far, the dependent variable could either take a value of 1, representing a crisis period, or 0 for tranquil times. We thus assumed variables to behave differently in crisis and non-crisis times. However, reality is more complex. Variables seem not only to deviate from their normal level before but also noticeably after a crisis (see [Figure 1](#)). Accordingly, instead of comparing only two states (crisis and non-crisis times), an EWS could also address what can be called a "post-crisis bias".

To our knowledge, Bussiere and Fratzscher (2002) represent the pioneering work in this direction and thus form our reference. Instead of a binomial logit model they estimate a multinomial one with the following three different states:

$$CC_t^i = \begin{cases} 1 & \text{if a crisis is upcoming in the next 24 months ([pre-]crisis period)} \\ 2 & \text{crisis and the further 24 months after a crisis (recovery period)} \\ 0 & \text{otherwise (tranquil period)} \end{cases}$$

Taking three states into account would presuppose that a mean value comparison of the possible three states displays significant differences. The results of our analyses for the six indicators shown in [Table 4](#) confirm this. Considering, for example, the real exchange rate variable 24 months before and after a crisis respectively, a significant difference in mean values becomes obvious. Whereas the average value in crisis times is -7.654 and thus implies a negative deviation from the trend, representing an overvalued local currency, the mean value for the crisis itself and the 24 months after the crisis is 10.608. Alternatively, the time window is shortened to two times 12 months, but results stay consistent over all variables (and are thus not presented here).

Aiming to assess whether a multinomial logit can really improve crisis recognition and forecast, [Table 5](#) sheds light on both in- and out-of-sample performance of the different types of logit models. In order to make results comparable, we use the same four variables for the multinomial as for the binomial logit model.⁶

Three aspects emerge when comparing the multinomial logit model with the binomial one: First of all, considering both the 25% and 50% cut-off probabilities in

⁶ For the construction of the multinomial logit model, we rely on Bussiere and Fratzscher (2002, p. 19 et sqq.). Like them, we choose the tranquil regime as the base regime. We estimate the probabilities of being rather in the tranquil period or in a (pre-)crisis or a recovery period according to changes in the independent variables. Each observation can be assigned to the state with the highest probability. However, in order to take different cut-off probabilities into account, we also include different cut-off thresholds for signalling a pre-crisis state. In [Table 5](#), the performance results of our estimated multinomial logit model are thus compared with those of the binomial logit model using a 25% and a 50% cut-off probability respectively.

sample, it is worth doubting whether a multinomial logit model generally improves forecasting performance. Different from Bussiere and Fratzscher (2002), we do not find superior performance results in comparison to our binomial logit model. Second, out-of-sample using a 50% cut-off, the multinomial model seems slightly better than the binomial model regarding crisis periods correctly called and false alarms. However, results are not at all convincing when considering the 25% cut-off probability. Third, a binomial logit model is easier to implement as an EWS and we thus decide to continue our work relying on the binomial logit with variables entering in binary form to be the best performing method. We now explicitly consider different crisis definitions presented in the literature.

4 The impact of different crisis definitions

Obviously, the robustness question also applies to the kind of crisis definition used. Though this part of an EWS can be considered as a more philosophical issue as either someone believes something is a crisis or not, there can be at least three different kinds of crisis definitions found in the literature, which comprise between one and three variables. Moreover, the interpretation of concrete values to be reached before a crisis is called, differ from study to study. We compare and analyze the proposals of Frankel and Rose (FR, 1996) and Bussiere and Fratzscher (BF, 2002) with KLR concerning robustness and stability.

FR concentrate explicitly on currency crashes. In contrast to KLR, who include both international reserves and the exchange rate – and thus account for averted as well as successful attacks on the domestic currency, FR focus on the latter. According to their definition, a crisis is considered a period of a nominal exchange rate (NER) devaluation of at least 25%. Additionally, in order to take hyperinflation into account, at the same time this has to comply an exceeding of the previous year's change in the exchange rate by at least 10%.⁷

⁷ FR admit that their cut-off points are somewhat arbitrary but sensitivity analyses support the settings. The idea of adding a second criterion is supported by Esquivel and Larrain (1999, p. 8) who also try to ensure in their work that only meaningful currency devaluations (in the sense of those effecting the purchasing power parity) are considered as crises. When applying the FR definition, due to our use of monthly instead of annual data, we adjusted crisis definition and the proposed 3-year windowing by FR (1996, p. 358) to monthly reference. The transformation from yearly to monthly data is again arbitrary. Godfjan and Valdés (1998) e.g. take a 2-month window when reproducing FR.

$$\text{Crisis} = \begin{cases} 1 & \text{if } \frac{\text{NER}_t^i - \text{NER}_{t-1}^i}{\text{NER}_{t-1}^i} > 0.25 \quad \cup \quad \frac{\text{NER}_t^i - \text{NER}_{t-1}^i}{\text{NER}_{t-1}^i} - \frac{\text{NER}_{t-1}^i - \text{NER}_{t-2}^i}{\text{NER}_{t-2}^i} > 0.1 \\ 0 & \text{otherwise} \end{cases}$$

Whereas FR reduce the KLR definition to the exchange rate focus, BF enlarge it by additionally considering interest rate changes. They base their decision on the idea that a central bank has mainly three options to respond to speculative attacks: In order to stabilize the exchange rate the bank can either augment the interest rate or sell international reserves. A third option would be to let market forces work and thus the currency devaluate. All three components are included in the EMPI of BF.

$$\text{EMPI}_t^i = \frac{1}{\sigma_{\text{REER}}^2} \left(\frac{\text{REER}_t^i - \text{REER}_{t-1}^i}{\text{REER}_{t-1}^i} \right) + \frac{1}{\sigma_r^2} (r_t^i - r_{t-1}^i) - \frac{1}{\sigma_{\text{RES}}^2} \left(\frac{\text{RES}_t^i - \text{RES}_{t-1}^i}{\text{RES}_{t-1}^i} \right)$$

Aiming to avoid sample splits due to hyperinflation periods, they concentrate on the real effective instead of the nominal exchange rates (REER). Accordingly, they consider real interest rates (r). The single components are weighted with their inverse sample variance over all countries for the whole period: the higher the variance, the smaller their weight. Originally, a crisis is considered a period when the EMPI is at least two standard deviations above its country average.⁸

Annex 1 (columns 6 to 10) shows the results of applying both the FR and the BF definition to the original KLR sample and its variations. The respective results of our binomial logit models are displayed in Table 6.

The in-sample FR performance results are convincing, whereas out-of-sample the FR logit model is beaten by the KLR model. Moreover, two aspects should be considered most skeptically: The performance tests for FR are based on a small sample of quite homogenous countries. Neither the EMS nor the Asian crisis is largely identified and thus a considerable number of countries that did not experience any crisis had to be dropped from the analysis. We doubt that an exclusive concentration on currency crashes is warranted for a practical EWS covering a wide set of heterogeneous countries.

In contrast, the BF definition shows a high stability over time and country variations. Admittedly, as EMPI variances are calculated over the whole sample period, the crisis definition seems less sensitive to country specific characteristics. Perform-

ance results, however, especially the percent of crises correctly called, are not fully convincing. Furthermore, restricted data availability and the lack of consistency in the interest rate data should be taken into account when choosing the BF definition.

In summary, the choice of a crisis definition is in the last instance a policy decision. Under many circumstances, however, the KLR definition seems to be most practical.

5 Further lessons for improving forecasts

So far, we have analyzed different methodological approaches as well as three crisis definitions over different sample periods and countries. Three further aspects emerge: There can be observed a recent trend to downsize analyzed samples by focusing on selected regions when scrutinizing EWSs (see Esquivel and Larrain, 1999, Brueggemann and Linne, 2002). Others apply different approaches to one country (see e.g. Alvarez-Plata and Schrooten, 2003, 2004) or choose single examples of few countries to assess the proposed technique (see Abiad, 2003, for example). It seems interesting to examine what happens if the sample is split up into regional sub-samples. Furthermore, we want to assess whether shortening samples could improve results. The third aspect addresses another growing body of empirical literature: the danger of contagion.

5.1 Regional sample splits

The KLR sample comprises a very heterogeneous set of countries. Applying the FR definition has shown that different types of crisis seem to be attributable to different countries. We thus analyze the robustness of former results with the help of the KLR crisis definition and the expanded sample across three different regions: Latin America, Asia and Europe.

As before, we focus exclusively on the multivariate approach.⁹ The outcome is shown in [Table 7](#). Latin American results, in- and out-of-sample, outperform those of the other two regions. Especially out-of-sample, an explicit application of the pre-

⁸ For our reconstruction we slightly modify the definition by taking real exchange rates. Furthermore, when trying to reproduce the results of BF, we find a 2.5 standard deviation to be most appropriate.

⁹ Regional performance results of the individual indicators are not explicitly displayed here but can be resumed briefly. All three regions have the top indicator in common: It is once more the exchange rate variable. All the same, differences across the regions become also evident with one aspect which is worth mentioning: all Latin American variables display consistently better noise-to-signal ra-

sented model on Asia or Europe does not seem to make sense as crisis periods are no longer recognized and in Europe even false alarms skyrocket. Taking into account that the FR crisis definition, relying exclusively on the exchange rate, recognizes more Latin American crises than the other two definitions, and the real exchange rate variable proves its star performer quality one more time, results are not surprising.

To conclude, deviations between different regions exist. In order to improve an EWS' forecasting performance while still relying on a broad set of countries considered, some of the regional commonality might also be brought to an EWS by adding a regional dummy variable.

5.2 Shortening the sample

We continue our robustness analyses with a split in time dimension. Based on Choueiri and Kaminsky (1999), who find supporting evidence for a change in the nature of crises over time, the hitherto considered sample period of 25 years (1970 to 1995) might appear too long for reliable coefficients over time. Supposing shorter sample periods to reveal somewhat better results due to a higher crisis homogeneity, we divide the extended sample period from 1970 to 2002 into three sub-periods: the first one ranging from 1970 to 1980, the second from 1981 to 1990 and the third from 1991 to 2002.

Table 8 displays the multivariate in-sample performance with the help of well-known binomial logit models. We take the logit model for the period from 1970 to 1995 as a benchmark to compare it with the full sample (1970-2002) and with distinctively shorter samples. The assumption that shorter samples might improve performance can be confirmed, especially for a 25% cut-off probability. Nevertheless, performance results vary clearly with the periods considered. The overall most convincing performance is reached for the 1970s. Independently of the chosen cut-off-probability the logit model for the period between 1970 and 1980 outperforms the other ones. For the two other periods, between 1981 and 1990 as well as the 1990s up to 2002, results once again face the problem of setting the right threshold and balancing correctly called crises and false alarms.

5.3 Grasping contagion

tios than in Europe and Asia. These results are in line with Edison (2003) who assesses differences between Asia and Latin America and also gets clearly better performance results for the latter.

The idea of capturing the risk of one country being affected by another country's crisis – either due to regional proximity (like the experience of the Asian crisis of 1997 has shown) or because of similar characteristics that seem to overcome regional distance (e.g. the affection of Russia by the Tequila crisis in 1994) – is one of the subjects given the most emphasis in current research on financial crisis. Efforts to characterize, to prove and to analyze financial contagion are numerous (for a survey, see Moser, 2003, Karolyi, 2003 or Kaminsky, Reinhart and Vegh, 2003). Mainly three channels of transmission can be identified in the literature: Firstly, effects can spill over due to bilateral or multilateral trade linkages (see Eichengreen, Rose and Wyplosz, 1996, Glick and Rose, 1999). Secondly, those countries either exposed to the same environmental conditions and shocks (see Masson, 1999, Moreno and Trehan, 2000), or having economic and political similarities (see Sachs, Tonell, Velasco, 1996, Hartmann, Straetmans and de Vries, 2004) are more likely to experience a crisis with one another. The third channel consists in tight financial linkages, arguing e.g. that investors might adjust their portfolio allocation in case of a financial crisis in one country. Having the same major creditors in common with the crisis country could thus lead to higher vulnerability in the respective country as well (see Kaminsky and Reinhart, 2000, Van Rijckeghem and Weder, 2001, or Caramazza, Ricci and Salgado, 2004). A thorough assessment of how to measure and include contagion in EWSs is offered by Fratzscher (2003).

In order to assess possible effects on performance results, we add two contagion variables to our analyses, one for capturing global, the other for regional linkage. In contrast to other studies, however, which consider quite complex measures of financial and real contagion, we decide to lighten the gathering of contagion. We therefore rely on Eichengreen, Rose and Wyplosz (1996) and construct the new variables in the following manner:

Global Contagion	Regional Contagion
$D_t^{G,j} = \begin{cases} 1, & \text{if } \sum_{i=1}^n C_t^i > 0, \text{ with } i \neq j \\ 0, & \text{otherwise} \end{cases}$	$D_t^{R,j} = \begin{cases} 1, & \text{if } C_t^i = 1 \text{ for } i \neq j \text{ but } i, j \in \text{same region} \\ 0, & \text{otherwise} \end{cases}$

The contagion variable takes the value of one, if in another country of the sample or another country of the region, respectively, a currency crisis occurs; otherwise it takes the value of 0.

Table 9 shows the results of our logit model estimation including contagion. Only the regional contagion variable is significant at the 1 % level and its inclusion seems to improve the estimated logit model slightly. Table 10 confirms this assumption. Taking a look at the in- and out-of-sample performance of two logit alternatives with and without contagion, the trade-off between a slightly higher percentage of correctly called crises at the cost of slightly more false alarms is often recognizable. In one out of our cases, however, there is an unequivocal improvement.

We conclude that the inclusion of a simple contagion dummy based on geographic proximity can capture some country linkages and induced contagion risks to improve model performance. Therefore, the possibility of grasping contagion should not be ignored when constructing an EWS.

6 Conclusions

Financial crises are among the most unwanted manifestations of the present international financial markets. Thus, much emphasis has been given to the development of EWSs in order to understand sources of crises or even better to avoid their outbreak. This research has been so fruitful that there are nowadays plenty of approaches available. Which approach, however, is most appropriate? Or to be less ambitious: what can we learn of this work, what should policy-makers do in this respect?

6.1 Elements of practical early warning systems

This paper addresses exactly the above raised questions by conducting robustness tests that go beyond earlier tests of such kind. In a nutshell, we find that reliable progress of new approaches is much smaller than usually claimed. A plausible explanation for this fact may be that findings are quite sensitive to their exact empirical specification: depending on the sample, the time period, the crisis definition, the method of variable aggregation and the variables chosen, there is always a different model at the top. In other words: there is no single best performing model for all circumstances. So, what can users of an EWS do?

We assess the appropriateness of competing approaches by taking the viewpoint of a user who has two requirements which seem to be self-evident: Any EWS applied should be robust and the approach has to be practical. Under these conditions we come to five insights: First, EWSs are useful as they help to identify and

even forecast fundamental origins of currency crises. Second, the crisis definition should not be too narrow and not too complicated, which is why we favor the KLR definition. Third, multivariate methods underlying EWSs do not differ too much regarding their performance, but a binary logit model seems slightly superior to others and at the same time easy to handle. Fourth, there is contagion in international financial markets and a practical way to roughly capture this effect is adding a regional contagion variable to the set of fundamental variables. Fifth, instability in fundamental relations can be addressed by downsizing samples, for example, by the application of regional or time period restrictions. This will help to get somewhat better results.

6.2 Policy discussion

Having these elements of practical EWSs in mind, what can economic policy do or should avoid doing? In the following we discuss four policy lessons:

First, the forecasting power of EWSs reveals that core macroeconomic fundamentals weaken before a crisis emerges (see Figure 1). Thus, macroeconomic policy should take information from these variables seriously and act in time. Real exchange rate appreciation, (relatively) declining reserves and (increasing) current account deficits indicate clearly that the economy may come into troubled waters.

Second, as shown, available EWSs seem to be – despite some existing forecasting power – too unreliable in their performance that international financial institutions, such as the IMF, should publish their signals. It has to be expected that such signals serve as a coordination device, similar to downgradings of rating agencies, and thus create a self-fulfilling prophecy. The potential damage due to false signals can hardly be compensated by the benefits. Nevertheless, there are good reasons to publish the information underlying EWSs as early as possible and to implement EWSs at the national or private institutions level as a monitoring instrument.

Third, any EWS involves a calibration issue regarding the identification of false alarms versus missed crises (longer in Bussiere and Fratzscher, 2002). Each institution implementing an EWS must thus make a decision on this issue. A more sensitive calibration, leading to more false alarms, seems appropriate if it is used to possibly identify fundamental weaknesses, whereas a very cautious calibration seems advisable if results would be made public (in order to avoid unjustified crises).

The research on EWSs demonstrates moreover that international financial markets do not show textbook-like behavior. Instead, the robust contagion dummy indi-

cates, that sometimes countries fall into a crisis without respective developments of the fundamentals. This finding of inefficient international financial markets is corroborated by recent experimental (e.g. Hey and Morone, 2004) as well as institutional evidence (e.g. Gehrig and Menkhoff, 2005). We know that even professional actors in financial markets often behave short-term oriented and decide in a herding manner. So signals from markets may be wrong and policy should not "believe" in them. Instead, economic policy may want to realize some room for maneuver, keeping signals from international financial markets at bay, for example by disintegrating the economy to some degree from unwanted shocks of the global market.

Overall, it may be thus useful to put the research on EWSs into a broader perspective: EWSs try to forecast larger exchange rate changes by applying a fundamental model. Empirical exchange rate research has shown, however, that rewards to this task can only be meager, as there are no reliable exchange rate models for horizons up to a year or even longer (Sarno and Taylor, 2002). Seen from this perspective, the performance of EWSs is surprisingly good. An intuitive explanation may understand this as an analogy to the finding that fundamentals matter for exchange rates as well in the case of a huge inflation differential between countries (Frankel and Rose, 1996a). This analogy, however, indicates a reverse side to the coin: the impact of fundamental influences is mostly complex, i.e. exchange rates do not always adjust immediately to even huge inflation differentials.

Consequently, the forecasting power of any EWS must be quite limited, as is the case with its robustness, too. One may be tempted to fine-tune EWSs more and more but our results provide evidence against high expectations in this respect. Instead, it seems advisable to be aware of the structural instability in this field – partly reflecting instability of the international financial markets – and to practice the robust and practical insights that can be gained from earlier work.

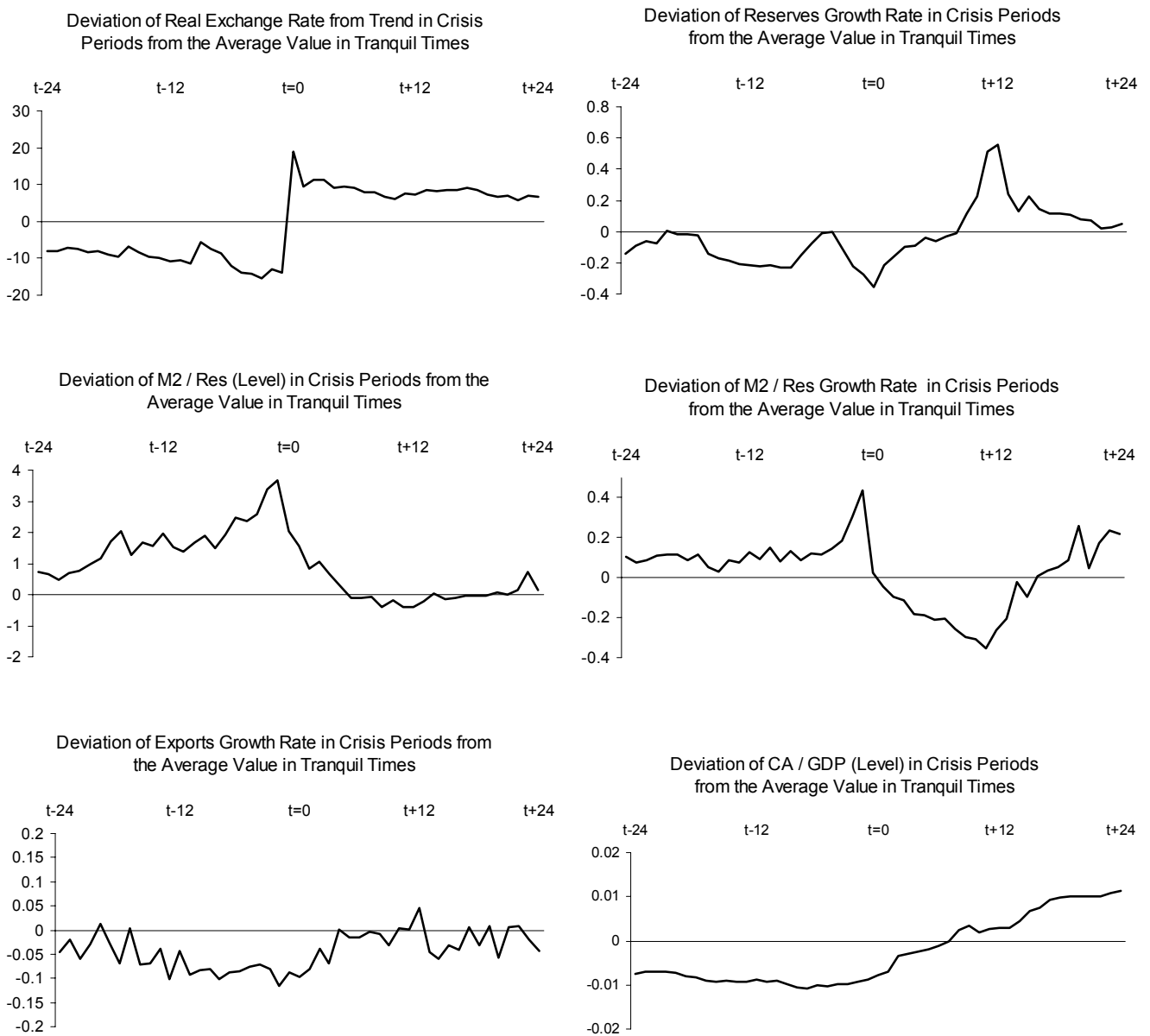
References

- Abiad, Abdul (2003), Early Warning Systems: A Survey and a Regime-Switching Approach, *IMF Working Paper* 03/32.
- Alvarez Plata, Patricia and Mechthild Schrooten (2003): The Argentine Currency Crisis: A Markov Switching Model Estimation, *DIW Discussion Paper* 348.
- Alvarez Plata, Patricia and Mechthild Schrooten (2004): Misleading Indicators? The Argentinean Currency Crisis, *Journal of Policy Modeling*, 26:5, 587-603.

- Berg, Andrew, Eduardo Borensztein, Gian M. Milesi-Ferreti, and Catherine Pattillo (2000), Anticipating Balance of Payments Crises: The Role of Early Warning Systems, *IMF Occasional Paper* 186.
- Berg, Andrew and Catherine Pattillo (1999), Predicting Currency Crises: The Indicators Approach and an Alternative, *Journal of International Money and Finance*, 18:4, 561-586.
- Berg, Andrew and Catherine Pattillo (1999a), Are Currency Crises Predictable? A Test, *IMF Staff Papers*, 46:2, 107-138.
- Bordo, Michael, Barry Eichengreen, Daniela Klingebiel and Maria Soledad Martinez-Peria (2001), Is the Crisis Problem Growing More Severe?, *Economic Policy*, 16:32, 52-82.
- Brueggemann, Axel and Thomas Linne (2002), Are the Central and Eastern European transition countries still vulnerable to a financial crisis? Results from the signals approach, *BOFIT Discussion Papers* 5/2002, Bank of Finland.
- Bussiere, Matthieu and Marcel Fratzscher (2002), Towards a New Early Warning System of Financial Crises, *ECB Working Paper* 145.
- Caramazza, Francesco, Luca Ricci and Ranil Salgado (2004), International Financial Contagion in Currency Crises, *Journal of International Money and Finance*, 23:1, 51-70.
- Choueiri, Nada and Graciela L. Kaminsky (1999), Has the Nature of Crises Changed? A Quarter Century of Currency Crises in Argentina, *IMF Working Paper* 99/152.
- Edison, Hali J. (2003), Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System, *International Journal of Finance and Economics*, 8:1, 11-53.
- Eichengreen, Barry, Andrew K. Rose and Charles Wyplosz (1995), Exchange Rate Mayhem: The Antecedents and Aftermath of Speculative Attacks, *Economic Policy*, 21, 249-312.
- Eichengreen, Barry, Andrew K. Rose and Charles Wyplosz (1996), Contagious Currency Crises: First Tests, *Scandinavian Journal of Economics*, 98:4, 463-484.
- Esquivel, Gerardo and Felipe Larrain (1999), Currency Crises: Is Central America Different?, *CID Working Paper* 26, Center for International Development at Harvard University.

- Frankel, Jeffrey A. and Andrew K. Rose (1995), Empirical Research on Nominal Exchange Rates, *Handbook of International Economics*, Vol. 3, 1689-1729.
- Frankel, Jeffrey A. and Andrew K. Rose (1996), Currency Crashes in Emerging Markets: An Empirical Treatment, *Journal of International Economics*, 41:4, 351-366.
- Frankel, Jeffrey A. and Andrew K. Rose (1996a), A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries, *Journal of International Economics*, 40:1-2, 209-224.
- Fratzscher, Marcel (2003), On Currency Crises and Contagion, *International Journal of Finance and Economics*, 8:2, 109-129.
- Gehrig, Thomas and Lukas Menkhoff (2005), The Rise of Fund Managers in Foreign Exchange: Will Fundamentals Ultimately Dominate? *The World Economy*, 28:4, 519-540.
- Glick, Reuben and Andrew K. Rose (1999), Contagion and Trade: Why Are Currency Crises Regional?, *Journal of International Money and Finance*, 18:4, 603-617.
- Goldfajn, Ilan and Rodrigo O. Valdés (1998), Are Currency Crises Predictable?, *European Economic Review*, 42:3-5, 873-885.
- Goldstein Morris, Graciela L. Kaminsky and Carmen M. Reinhart (2000), *Assessing Financial Vulnerability – An Early Warning System for Emerging Markets*, Institute for International Economics, Washington, D.C.
- Hartmann, Philipp, Stefan Straetmans and Caspar G. de Vries (2004), Fundamentals and Joint Currency Crises, *ECB Working Paper Series 324*, March.
- Hey, John D. and Andrea Morone (2004), Do Markets Drive out Lemmings – or Vice Versa? *Economica*, 71, 637-659.
- Kamin, Steven B., John W. Schindler and Shawna L. Samuel (2001), The Contribution of Domestic and External Factors to Emerging Market Devaluation Crises: An Early Warning Systems Approach, *International Finance Discussion Paper 711*, Board of Governors of the Federal Reserve System.
- Kaminsky, Graciela L. (1998), Currency and Banking Crises: The Early Warnings of Distress, *International Finance Discussion Paper 629*, Board of Governors of the Federal Reserve System.
- Kaminsky, Graciela L., Saul Lizondo and Carmen M. Reinhart (1998), Leading Indicators of Currency Crises, *IMF Staff Papers*, 45:1, 1-48.

- Kaminsky, Graciela L. and Carmen M. Reinhart (2000), On Crises, Contagion, and Confusion, *Journal of International Economics*, 51:1, 145-168.
- Kaminsky, Graciela L., Camen M. Reinhart and Carlos A. Végh (2003). The Unholy Trinity of Financial Contagion, *Journal of Economic Perspectives*, 17:4, 51-74.
- Karolyi, Andrew G. (2003), Does International Financial Contagion Really Exist? *International Finance*, 6:2, 178-199.
- Martinez-Peria, Maria Soledad (2002), A Regime Switching Approach to Studying Speculative Attacks: A Focus on EMS Crises, *Empirical Economics*, 27:2, 299-334.
- Masson, Paul (1999), Contagion: Macroeconomic Models with Multiple Equilibria, *Journal of International Money and Finance*, 18:4, 587-602.
- Moreno, Ramon and Bharat Trehan (2000), Common Shocks and Currency Crises, *Working Paper Federal Reserve Bank of San Francisco* 00-05.
- Moser, Thomas (2003), What is International Financial Contagion?, *International Finance*, 6:2, 157-178.
- Pasternak, Christoph (2003) The Signals Approach as an Early Warning System for Currency Crises. An Application to Transition Economies – With Special Emphasis on Poland, *International Finance Paper* 0304001.
- Pozo, Susan and Catalina Amuedo-Dorantes (2003), Statistical Distributions and the Identification of Currency Crises, *Journal of International Money and Finance*, 22, 591-609.
- Sachs, Jeffrey, Aaron Tornell, and Andrés Velasco (1996), Financial Crises in Emerging Markets: The Lessons from 1995, *Brookings Papers on Economic Activity*, 27:1, 147-199.
- Salvatore, Dominick (1999), Could the Financial Crisis in East Asia Have Been Predicted?, *Journal of Policy Modeling*, 21:3, 341-347.
- Sarno, Lucio and Mark P. Taylor (2002), *The Economics of Exchange Rates*, Cambridge et al., Cambridge University Press.
- Sy, Amadou N. R. (2004), Rating the Rating Agencies: Anticipating Currency Crises or Debt Crises?, *Journal of Banking and Finance*, 28:11, 2845-2867.
- Van Rijckeghem, Caroline and Beatrice Weder (2001), Sources of Contagion: Is it Finance or Trade? *Journal of International Economics*, 54:2, 293-308.

Figure 1: Movements of different variables in crisis times¹

Notes: (1) The graphs show the deviation in course of the respective variable in a 24-months window before and after the crisis from its average value in calm periods. Results are calculated concentrating on the original KLR sample, but they are similar when enlarging the sample to 28 countries as presented later on (see 2.3). Except for M2 / reserves (level) and CA / GDP (level) as well as the real exchange rate, the variables indicate growth rates in reference to the same month of the previous year. The real exchange rate is expressed as deviation of its deterministic trend.

Table 1: In-sample¹ performance of univariate indicators:
KLR definition and sample variations

Sample	Original KLR Sample 1970-4/1995		Time Extension 1970-2002 Original KLR Sample		Country Extension Max. Sample ² Original 1970-4/1995		Time and Country Extension 1970-2002 Max. Sample ²	
Variable ³	Noise / Signal ⁴	P(Crisis Signal) ⁵	Noise / Signal ⁴	P(Crisis Signal) ⁵	Noise / Signal ⁴	P(Crisis Signal) ⁵	Noise / Signal ⁴	P(Crisis Signal) ⁵
Real Exchange Rate	0.18	0.67	0.17	0.65	0.25	0.56	0.23	0.55
Exports	0.59	0.38	0.66	0.31	0.65	0.33	0.73	0.28
Reserves	0.50	0.41	0.46	0.39	0.48	0.40	0.49	0.36
M2 / Reserves	0.56	0.38	0.50	0.37	0.58	0.35	0.55	0.33
M2 / Reserves (Level)	0.55	0.39	0.53	0.35	0.57	0.34	0.58	0.32
CA / GDP (Level)	0.51	0.45	0.49	0.37	0.47	0.43	0.43	0.39

Notes: (1) The sample comprises the original KLR sample, i.e. data from 20 countries for the period from January 1970 to April 1995. (2) The extended sample contains Argentina, Bolivia, Brazil, Chile, Denmark, Finland, Greece, India, Indonesia, Israel, Colombia, Korea, Malaysia, Mexico, Norway, Pakistan, Peru, Philippines, Portugal, Sweden, Singapore, Spain, Sri Lanka, South Africa, Thailand, Turkey, Uruguay and Venezuela. (3) Except for M2 / reserves (level) and CA / GDP (level) as well as the real exchange rate, the variables indicate growth rates in reference to the same month of the previous year. The real exchange rate is expressed as deviation of its deterministic trend. (4) The noise-to-signal ratio shows the number of bad signals as a share of possible bad signals, divided by the number of good signals as a share of possible good signals. (5) The conditional probability indicates the probability of a crisis in a 24 months window after the respective indicator issued a signal.

Table 2: Two logit model alternatives relying on the KLR sample

Variable ¹	Binomial Logit Model with Variables Entering in Form of Binary 0-1 Signals			Binomial Logit Model with Variables Entering in Form of Their Continuous Values		
	Coefficient	Odds Ratio e^{β}	Significance	Coefficient	Odds Ratio e^{β}	Significance
Real Exchange Rate	1.976	7.213	0.000	-0.019	0.981	0.000
Exports	0.483	1.621	0.000	-0.408	0.665	0.001
Reserves	0.685	1.983	0.000	-0.368	0.692	0.000
M2 / Reserves (Level)	0.474	1.606	0.000	0.038	1.039	0.000
No. of Observations	5065			5065		
Pseudo-R ²	0.14			0.09		

Notes: (1) Exports and reserves are expressed as growth rates in reference to the respective month of the previous year, whereas M2 / reserves is a level consideration. The real exchange rate is expressed as deviation of its deterministic trend.

Table 3: In- and out-of-sample¹ performance
of two logit model alternatives and a composite indicator benchmark

	In-Sample			Out-of-Sample		
	Composite Indicator ²	Binomial Logit Model with Variables in Form of Binary 0-1 Signals	Binomial Logit Model with Variables in Form of Continuous Values	Composite Indicator ²	Binomial Logit Model with Variables in Form of Binary 0-1 Signals	Binomial Logit Model with Variables in Form of Continuous Values
Goodness-of-Fit (50% Cut-Off)						
Percent of Crisis Periods Correctly Called ³	23	27	10	24	20	2
False Alarms as Percent of Total Alarms ⁴	40	34	23	76	76	90
Percent of Tranquil Periods Correctly Called ⁵	94	95	99	83	79	94
Percent of Observations Correctly Called	70	77	75	73	64	71
Goodness-of-Fit (25% Cut-Off)						
Percent of Crisis Periods Correctly Called ³	30	53	75	25	39	70
False Alarms as Percent of Total Alarms ⁴	45	56	63	77	72	72
Percent of Tranquil Periods Correctly Called ⁵	92	76	53	82	67	38
Percent of Observations Correctly Called	76	70	59	72	60	46

Notes: (1)) The period from January 1970 to April 1995 is considered as in-sample. The out-of-sample period ranges from May 1995 to December 2002. (2) In order to assure comparability, the composite Indicator contains the same 4 variables that enter the logit models. (3) A crisis signal is considered as correct, if the estimated crisis probability is above the cut-off probability, and a crisis is coming up in the next 24 months. (4) A false alarm is an observation where the estimated crisis probability is above the cut-off probability, but no crisis is coming up in the next 24 months. (5) A signal for a tranquil period is considered as correct, if the estimated crisis probability is underneath the cut-off probability, and no crisis is coming up in the next 24 months.

Table 4: Mean value comparison of key indicators
(24-months windows)¹

Variable ²	Average All Periods	Average Tranquil Periods	Average Crisis Periods	Average Post-Crisis / Recovery Periods	Significance in differences? (ANOVA)
Real Exchange Rate	1.221	2.028	-7.654	10.608	Yes/0.000
Exports	0.153	0.176	0.110	0.150	Yes/0.000
Reserves	0.290	0.313	0.176	0.373	Yes/0.000
M2 / Reserves	0.082	0.064	0.196	-.0173	Yes/0.000
M2 / Reserves (Level)	3.067	2.570	4.363	2.775	Yes/0.000
CA / GDP (Level)	-0.022	-0.020	-0.029	-0.018	Yes/0.000

Notes: (1) The mean value comparison comprises data of the original KLR sample. We consider 24 months before and after a crisis, respectively. Alternative shortenings of the time window to two times 12 months unveils similar results. (2) Analogous to Table 1, Note 3.

Table 5: In- and out-of-sample¹ performance:
A binomial and a multinomial logit model in comparison

	In-Sample		Out-of-Sample	
	Binomial Logit Model ²	Multinomial Logit Model ² (24 Months ³)	Binomial Logit Model ²	Multinomial Logit Model ² (24 Months ³)
Goodness-of-Fit (50% Cut-Off)				
Percent of Crisis Periods Correctly Called ⁴	27	27	20	21
False Alarms as Percent of Total Alarms ⁵	34	33	76	75
Percent of Tranquil Periods Correctly Called ⁶	95	94	79	75
Percent of Observations Correctly Called	77	71	64	59
Goodness-of-Fit (25% Cut-Off)				
Percent of Crisis Periods Correctly Called ⁴	53	30	39	21
False Alarms as Percent of Total Alarms ⁵	56	38	72	75
Percent of Tranquil Periods Correctly Called ⁶	76	92	67	75
Percent of Observations Correctly Called	70	71	83	59

Notes: (1) As in Table 3. (2) The independent variables enter the binomial and the multinomial logit model in binary 0-1 form. Calculations with independent variables entering in form of continuous values reveal weaker results and are thus not explicitly considered here. (3) Shortening the considered time windows to 12 months before and after the crisis, respectively, disclose similar, though even slightly worse results that are not shown here. (4)-(6) Analogous to Table 3, Notes (3)-(5).

Table 6: In- and out-of-sample¹ performance of binomial logit models:
Different crisis definitions and KLR sample

	In-Sample		Out-of-Sample	
	FR Definition	BF Definition	FR Definition	BF Definition
Goodness-of-Fit (50% Cut-Off)				
Percent of Crisis Periods Correctly Called ²	35	4	17	3
False Alarms as Percent of Total Alarms ³	33	30	89	50
Percent of Tranquil Periods Correctly Called ⁴	93	100	70	99
Percent of Observations Correctly Called	75	81	61	84
Goodness-of-Fit (25% Cut-Off)				
Percent of Crisis Periods Correctly Called ²	54	37	29	16
False Alarms as Percent of Total Alarms ³	49	64	86	89
Percent of Tranquil Periods Correctly Called ⁴	78	84	63	76
Percent of Observations Correctly Called	71	75	57	67

Notes: (1) The sample only comprises countries which experienced a crisis. Applying the FR definition, these are Argentina, Bolivia, Brazil, Chile, Indonesia, Israel, Colombia, Mexico, Peru, Philippines, Turkey, Uruguay and Venezuela of the (small) original KLR sample, plus Pakistan, Sri Lanka and South Africa, being part of the expanded one. In-sample analyses refer to the period from January 1970 to April 1995. The preceding 92 months till December 2002 are considered as out-of-sample. (2)-(4) Analogous to Table 3, Notes (3)-(5).

Table 7: In- und out-of-sample¹ performance of a binomial logit model after sample split up²,
KLR definition and KLR sample

	In-Sample			Out-of-Sample		
	Latin America	Asia	Europe	Latin America	Asia	Europe
Goodness-of-Fit (50% Cut-Off)						
Percent of Crisis Periods Correctly Called ³	40	5	20	37	0	0
False Alarms as Percent of Total Alarms ⁴	39	26	33	79	0	100
Percent of Tranquil Periods Correctly Called ⁵	92	100	96	65	100	99
Percent of Observations Correctly Called	80	80	74	59	69	79
Goodness-of-Fit (25% Cut-Off)						
Percent of Crisis Periods Correctly Called ³	60	39	89	41	27	92
False Alarms as Percent of Total Alarms ⁴	53	60	70	79	48	80
Percent of Tranquil Periods Correctly Called ⁵	78	85	18	61	89	8
Percent of Observations Correctly Called	74	75	38	57	69	25

Notes: (1) As in Table 3. (2) The expanded KLR sample is split into three regional sub samples. "Latin America" covers Argentina, Bolivia, Brazil, Chile, Colombia, Mexico, Peru, Uruguay and Venezuela, "Asia" contains India, Indonesia, Korea, Malaysia, Philippines, Thailand, Pakistan, Sri Lanka as well as Singapore, and Denmark, Finland, Greece, Norway, Portugal, Sweden, Spain and Turkey are considered separately as "Europe". Israel and South Africa are dropped from consideration. (3)-(5) As in Table 3.

Table 8: In-sample¹ performance of a binomial logit model:
Shortening samples to ten year periods

	1970-1995	1970-2002 ²	1970-1980	1981-1990	1991-2002 ²
Goodness-of-Fit (50% Cut-Off)					
Percent of Crisis Periods Correctly Called ³	27	25	44	22	3
False Alarms as Percent of Total Alarms ⁴	34	47	34	42	52
Percent of Tranquil Periods Correctly Called ⁵	95	90	94	93	99
Percent of Observations Correctly Called	77	69	83	71	68
Goodness-of-Fit (25% Cut-Off)					
Percent of Crisis Periods Correctly Called ³	53	40	54	52	98
False Alarms as Percent of Total Alarms ⁴	56	57	40	54	67
Percent of Tranquil Periods Correctly Called ⁵	76	84	90	73	4
Percent of Observations Correctly Called	70	73	82	67	35

Notes: (1) The in-sample period comprises each indicated range. (2) Crises are calculated according to the KLR crisis definition for the sample from January 1970 to April 1995. For the period from May 1995 to December 2002, thresholds of the original crisis definition are used. (3)-(5) Analogous to Table 3, Notes (3)-(5).

Table 9: Binomial logit model¹ including contagion

Variable	Coefficient	Odds Ratio e ^β	Significance
Real Exchange Rate	1.962	7.115	0.000
Reserves	0.426	1.531	0.000
Exports	0.661	1.937	0.000
M2 / Reserves (Level)	0.624	1.866	0.000
Regional Contagion	0.397	1.488	0.002
No. of Observations	4761		
Pseudo-R ²	0.15		

Notes: (1) The variables enter the logit model in the binary 0-1 form.

Table 10: In- and out-of-sample¹ performance
of logit model alternatives with and without contagion

	In-Sample		Out-of-Sample	
	Logit Without Contagion	Logit With Regional Contagion Dummy	Logit Without Contagion	Logit With Regional Contagion Dummy
Goodness-of-Fit (50% Cut-Off)				
Percent of Crisis Periods Correctly Called ²	27	28	20	21
False Alarms as Percent of Total Alarms ³	34	35	76	74
Percent of Tranquil Periods Correctly Called ⁴	95	95	79	79
Percent of Observations Correctly Called	77	77	64	64
Goodness-of-Fit (25% Cut-Off)				
Percent of Crisis Periods Correctly Called ²	53	47	39	33
False Alarms as Percent of Total Alarms ³	56	53	72	70
Percent of Tranquil Periods Correctly Called ⁴	76	81	67	73
Percent of Observations Correctly Called	70	73	60	62

Notes: (1)-(4) Analogous to Table 3, Notes (1) and (3)-(5).

Annex 1: Overview of different identified crises¹

Source	Original KLR	Replications by Edison	Own Replications	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations
Crisis Definition	KLR	KLR	KLR	KLR	FR	BF	BF	BF	BF
Period	1970-1995	1970-1995	1970-1995	1970-2002	1970-2002	1970-1995	1970-2002	1970-1995	1970-2002
Country Sample	Small	Expanded	Expanded	Expanded	Expanded	Small	Small	Expanded	Expanded
Argentina	Jun 70	Mar 75	Mar 75	Mar 75	Mar 75	Jun 75	Jun 75	Jun 75	Jun 75
	Jun 75	Jul 82	Jun 75	Jun 75	Jun 75	Dec 83	Dec 83	Jul 82	Jul 82
	Feb 81	Apr 89	Nov 76	Nov 76	Nov 75	Apr 89	Apr 89	Apr 89	Apr 89
	Jul 82	Dec 89	Jul 82	Jul 82	Mar 76	Dec 89	Dec 89	Dec 89	Dec 89
	Sep 86		Dec 83	Dec 83	Nov 76				
	Apr 89		Apr 89	Apr 89	Apr 81				
	Feb 90		Feb 90	Feb 90	Jan 82				
					Jul 82				
					Oct 87				
					Apr 89				
					Jul 89				
					Dec 89				
					Jan 91				
					Jan 02				
Bolivia	Nov 82	Nov 82	Nov 82	Nov 82	Oct 72	Nov 82	Nov 82	Nov 82	Nov 82
	Nov 83	Nov 83	Sep 85	Apr 84	Feb 82	Feb 85	Feb 85	Apr 84	Apr 84
	Sep 85	Sep 85		Aug 84	Nov 82	Sep 85	Sep 85	Feb 85	Feb 85
				Nov 84	Nov 83			Sep 85	Sep 85
				Sep 85	Apr 84				
					Aug 84				
					Nov 84				
					Feb 85				
					May 85				
					Sep 85				

Source	Original KLR	Replications by Edison	Own Replications	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations
Crisis Definition	KLR	KLR	KLR	KLR	FR	BF	BF	BF	BF
Period	1970-1995	1970-1995	1970-1995	1970-2002	1970-2002	1970-1995	1970-2002	1970-1995	1970-2002
Country Sample	Small	Expanded	Expanded	Expanded	Expanded	Small	Small	Expanded	Expanded
Brazil	Feb 83	Sep 82	Sep 82	Sep 82	Dec 79	Sep 82	Sep 82	Sep 82	Sep 82
	Nov 86	Mar 90	Nov 89	Nov 89	Feb 83	Jan 83	Jan 83	Nov 89	Nov 89
	Jul 89	Nov 90	Nov 90	Feb 90	May 87	Jan 87	Jan 87	Nov 90	Nov 90
	Nov 90	Oct 91		Nov 90	Jun 89	Jan 90	Nov 89		Sep 98
	Oct 91			Jan 99	Dec 89	Nov 90	Feb 90		Jan 99
					Oct 91		Nov 90		
Chile	Dec 71	Jul 71	Jul 71	Jul 71	Jul 71	Jul 71	Jul 71	Jul 71	Jul 71
	Aug 72	Sep 72	May 73	May 73	Sep 72	Dec 74	Dec 74	May 73	May 73
	Oct 83	May 73	Oct 73	Oct 73	May 73			Dec 74	Dec 74
	Dec 74	Dec 74	Jul 85	Jul 85	Oct 73				
	Jan 76	Jul 85			Dec 74				
	Aug 82				Mar 75				
Colombia	Mar 83	Jan 85	Jan 85	Jan 85		Jan 84	Sep 74	Jan 84	Jan 84
	Feb 85		Apr 85	Apr 85		Apr 84	Jan 84	Apr 84	Apr 84
				Aug 95		Jan 85	Apr 84	Jan 85	Jan 85
				Sep 98			Jan 85		
Denmark	May 71	Jun 73	Mar 80	Jul 75		Jul 76	Jul 76	Jul 76	Jul 76
	Jun 73	Nov 79	Mar 91	Nov 78		Dec 84	Dec 84	Dec 84	Dec 84
	Nov 79	Aug 93		Mar 80			Jan 93		
	Aug 93			Mar 91					

Source	Original KLR	Replications by Edison	Own Replications	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations
Crisis Definition	KLR	KLR	KLR	KLR	FR	BF	BF	BF	BF
Period	1970-1995	1970-1995	1970-1995	1970-2002	1970-2002	1970-1995	1970-2002	1970-1995	1970-2002
Country Sample	Small	Expanded	Expanded	Expanded	Expanded	Small	Small	Expanded	Expanded
Malaysia	Jul 75	Nov 73	Nov 73	Jul 97		Dec 94	Feb 82	Dec 92	Dec 92
		Nov 78	Jul 75	Dec 97			Dec 92	Dec 94	Dec 94
		Mar 80	Nov 78				Dec 94		Jul 97
		Feb 85	Mar 80				Jul 97		Jan 98
		Dec 92	Feb 85	Dec 92					
Mexico	Sep 76	Sep 76	Sep 76	Sep 76	Sep 76	Sep 76	Sep 76	Sep 76	Sep 76
	Feb 82	Feb 82	Feb 82	Feb 82	Feb 82	Jun 82	Feb 82	Feb 82	Feb 82
	Dec 82	Dec 82	Dec 82	Dec 82	Dec 82	Dec 94	Jun 82	Jul 82	Jul 82
	Dec 94	Dec 94	Dec 87	Dec 87	Dec 87		Mar 90	Dec 82	Dec 82
			Dec 94	Dec 94	Dec 94	Dec 94	Dec 94	Dec 94	Dec 94
Norway	Jun 73	Jun 73	Nov 78	Nov 78		Nov 77	Nov 77	Nov 77	Nov 77
	Feb 78	Feb 78	Jan 81	May 86		Nov 92	Nov 92	Nov 78	Nov 78
	May 86	May 86	May 86	Mar 91			Dec 96	Nov 92	Nov 92
	Dec 92	Nov 92	Mar 91	Nov 92			Dec 97		Dec 96
			Nov 92	Dec 97			Jun 99		Dec 97
				Dec 01			Oct 00		Jun 99
							Jul 01		Oct 00
							Dec 01		Jul 01
								Dec 01	
Pakistan		May 72	May 72	May 72	May 72			May 72	May 72
								Oct 90	Oct 90
Peru	Jun 76	Jun 76	Sep 88	Sep 88	Jun 76	Sep 88	Sep 88	Sep 88	Sep 88
	Oct 83	Oct 83			Oct 77	Aug 90	Aug 90	Aug 90	Aug 90
					Oct 87				
					Sep 88				
					Jan 89				
Philippines	Feb 70	Oct 83	Feb 70	Feb 70	Feb 70	Sep 83	Sep 83	Sep 83	Sep 83
	Oct 83	Jun 84	Oct 83	Oct 83	Oct 83	Jan 90	Feb 85	Jan 90	Jun 84
	Jun 84	Feb 86	Jun 84	Jun 84	Jun 84		Jan 90		Jan 90
			Feb 86	Feb 86	Feb 86				
				Dec 97	Dec 97				

Source	Original KLR	Replications by Edison	Own Replications	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations	Own Calculations
Crisis Definition	KLR	KLR	KLR	KLR	FR	BF	BF	BF	BF
Period	1970-1995	1970-1995	1970-1995	1970-2002	1970-2002	1970-1995	1970-2002	1970-1995	1970-2002
Country Sample	Small	Expanded	Expanded	Expanded	Expanded	Small	Small	Expanded	Expanded
Thailand	Nov 78	Nov 78	Nov 78	Nov 78		Nov 78	Nov 78	Nov 78	Nov 78
	Jul 81	Jul 81	Jul 81	Jul 81		Oct 79	Oct 79	Oct 79	Oct 79
	Nov 84	Nov 84	Nov 84	Jul 97		Oct 80	Oct 80	Oct 80	Oct 80
			Feb 85	Nov 97		Apr 81	Apr 81	Apr 81	Apr 81
						Jul 97			Jul 97
						Nov 97			Nov 97
Turkey	Aug 70	Aug 70	Aug 70	Aug 70	Aug 70	Jan 80	Jan 80	Jan 80	Jan 80
	Jan 80	Jan 80	Jan 80	Jun 79	Mar 78	Dec 81	Dec 81	Dec 81	Mar 94
	Mar 94	Mar 94	Mar 91	Jan 80	Jun 79	Mar 94	Jul 83	Mar 94	Feb 01
			Mar 94	Mar 94	Jan 80		Mar 94		
			Feb 01	Feb 01	Apr 94				
					Feb 01				
Uruguay	Dec 71	Mar 72	Mar 72	Mar 72	Mar 72	Nov 82	Nov 82	Nov 82	Nov 82
	Oct 82	Nov 82	Nov 82	Nov 82	Nov 82	Nov 84	Nov 84	Nov 84	Nov 84
				Jul 02	Jul 02		Jul 02		Jul 02
Venezuela	Feb 84	Mar 84	Feb 84	Feb 84	Feb 84	Feb 84	Feb 84	Feb 84	Feb 84
	Dec 86	Dec 86	Dec 86	Dec 86	Dec 86	Dec 86	Dec 86	Dec 86	Dec 86
	Mar 89	Mar 89	Mar 89	Mar 89	Mar 89	Mar 89	Mar 89	Mar 89	Mar 89
	May 94			May 94	May 94	May 94	May 94	May 94	May 94
	Dec 95			Dec 95	Dec 95		Apr 96		Dec 95
				Feb 02	Apr 96		Feb 02		Apr 96
				Feb 02				Feb 02	
					May 02				
Total Number of Crisis Periods for									
- Small Sample:	75	70	79	93	84	55	83	--	--
- Expanded				126					
Sample:	--	91	104		88	--	--	76	101

Notes: (1) Upon the existence of consecutive signals, only the first one is displayed and counted as a crisis. Moreover, according to the proposal of FR (1996, p. 358), there must be at least 2 months in between two signals before being counted as two different crisis periods.

Annex 2: Definition, background information, data sources and construction of different indicators

Indicator	Definition	Background Information to Explanatory Power	Threshold Percentile	Data Sources	Construction Particularities ¹
Real Exchange Rate	<p>Deviation of real exchange rate (RER) from its deterministic trend over the whole sample period:</p> $Dev_{RER_t^i} = \frac{RER_t^i - Trend_t^i}{Trend_t^i} * 100$	An overvalued exchange rate can often be seen as an indication of a possible devaluation of a currency and thus raises the risk of an upcoming currency crisis.	10	IFS line AE IFS line 64	<p>The real exchange rate is the result of multiplying the nominal exchange rate in price quotation, (indicating local currency per USD, line AE) with the consumer price index (CPI, line 64) of the USA divided by the local CPI (line 64).</p> <p>For calculating the trend, we optionally used either a linear, an exponential or a logarithmic function, taking the best fit into consideration.</p>
Exports	Growth rate in reference to the respective month of the preceding year	A weak export sector implies a higher crisis probability as the government might have an incentive to devaluate in order to empower the exports.	10	IFS line 70_d	--
Reserves	Growth rate in reference to the respective month of the preceding year	A lost of international reserves stands for central bank market intervention in order to stabilize the exchange rate. Thus a small or even negative growth rate heightens the risk of a crisis.	10	IFS line 1LD	--
M2 / Reserves (Level)	Level of the ratio of M2 to reserves	This indicator is considered in order to assess the short-term liquidity and convertibility of a country's currency. An increase of the ratio can occur due to a loss of international reserves or to extraordinary monetary expansion. Both can be considered as risk factors.	90	IFS line 34 IFS line 35 IFS line 1L.d	The sum of money (line 34) and quasi-money (line 35) divided by international reserves (line 1L.d).

Indicator	Definition	Background Information to Explanatory Power	Threshold Percentile	Data Sources	Construction Particularities ¹
M2 / Reserves	Growth rate of the ratio of M2 to reserves in reference to the respective month of the preceding year	Similar to M2 / reserves as a level consideration, this indicator also measures whether and to which degree a country is able to convert domestic into foreign currency. An exceptionally high growth rate shows upcoming imbalances and thus heightens the crisis risk for a country.	90	IFS line 34 IFS line 35 IFS line 1L.d	The sum of money (line 34) and quasi-money (line 35) is considered to form M2. After dividing M2 by international reserves (line 1L.d) the growth rate in reference to the respective month of the preceding year is calculated.
Current Account / GDP (Level)	Ratio of current account to GDP in reference to the respective month of the preceding year	Emerging Markets often exhibit high current account deficits. In order to finance these deficits they depend on foreign capital inflows. High deficits make the respective country vulnerable to expectation shifts and losses of confidence which might imply a sudden turn in capital inflows and thus sharpens the risk of a crisis.	10	IFS line 78AL IFS line 99B	Both variable series are firstly linearly interpolated from annual and quarterly data, respectively, to monthly data. After dividing current account (line 78AL) by GDP (line 99b), a moving average of the previous for quarters is calculated. The variable is measured in levels only.
Real Interest Rate ²	Level of real interest rate	---	---	IFS line 60 IFS line 60B IFS line 60L IFS line 60P IFS line 64	Use of money market rate, discount rate, deposit rate or lending rate according to data availability with prioritization on the money market rate. Real interest rates are calculated with the help of the CPI (line 64).

Notes: (1) All time series were transferred into Millions of USD. Replication efforts imply the reference to details given by Edison (2003, p. 47) and Berg and Pattillo (1999, p. 567). (2) The variable is not used as an indicator but enters the crisis definition of Bussiere and Fratzscher (2002).