Early Warning Systems : A Comparison of Currency Crises Forecasting Methods

Bertrand Candelon*, Elena-Ivona Dumitrescu†, Christophe Hurlin†‡

September 2009

Abstract

This paper proposes a new statistical framework inherited from the traditional credit-scoring literature, to evaluate currency crises Early Warning System (EWS). Applied to evaluate for the predictive power of panel logit and Markov frameworks, it results that the panel logit is outperforming the Markov switching ones. Furthermore, the introduction of forward looking variables clearly improves the forecasting properties of the EWS. It thus confirms the adequacy of the second generation crisis models in explaining the occurrence of crises.

Key words : currency crisis, Early Warning System, credit-scoring.
J.E.L Classification : C33, F37.

*b.candelon@algec.unimaas.nl, Maastricht University, School of Business and Economics, Department of Economics. The Netherlands.
†elena.dumitrescu@univ-orleans.fr, University of Orléans, Laboratoire d'Economie d'Orléans (LEO), France.
‡Corresponding author: christophe.hurlin@univ-orleans.fr, University of Orléans, Laboratoire d'Economie d'Orléans (LEO), France.

1 Introduction

The recent sub-prime crisis has renewed interest for financial crises. In a prophetic paper, Bordo et al. (2001), distinguishing 4 types of financial crises (banking, currency, twin and all crises), already shown that financial crises, in particular currency ones, would become more frequent since the collapse of the Bretton-Woods system in emerging as well as developed countries. This stylized fact should stimulate economists in improving the quality of Early Warning Systems (hereafter EWS), set up to ring before the occurrence of a currency crisis. Such an alarm system constitutes the only tool available for authorities to implement optimal policies to prevent or at least attenuate the impact of a crisis.

The first EWS was proposed by Kaminsky, Lizondo and Reinhart (1998) (KLR hereafter) using a signalling approach. They use a large database of 15 indicator variables covering the external position, the financial sector, the real sector, the institutional structure and the fiscal policy of a particular country. An indicator (i) will signal a crisis, when it will exceed a particular cut-off point (C_i). The estimation of this threshold is at the core of such an analysis. KLR determine it in order to minimize the noise-to-signal ratio (NSR hereafter) *i.e.* such that the probability of occurrence of a crisis becomes maximum after exceeding the cut-off point. The EWS for the country j is then built as the weighting-sum of the individual indicators, the weights being in the inverse of (NSR).

Berg and Patillo (1999) (BP hereafter) compare this signalling method to a panel logit model. It turns out that their in-sample forecasts dominate KLR ones, when considering measures similar to the mean square error, as the quadratic probability score (QPS) and log probability score (LPS), as well as NSR. It thus paved the way to a huge number of empirical studies (Kumar et al., 2003, Fuertes and Kalotychou, 2007, Berg et al., 2008 to name but a few). Nevertheless, in BP and all other studies thresholds beyond which crisis are detected are exogenously fixed (25% and 50%) and no formal statistical framework is proposed to test for the improvement of the NtoRS (a lower one is enough to conclude about the superiority of the model).

Moreover, they do not exploit the fact that turmoils refer to specific states structurally different from the one governed by tranquil periods. Hence, Bussiere and Fratzscher (2006) propose a multinomial logit EWS[1], whereas other studies use Markov-Switching models (see Abiad, 2003 for a survey or Martinez-Peria, 2002 and Fratzcher, 2003). The evaluation of these models is similar to the one proposed in BP, as it relies on the NSR. The determination of the optimal cut-off point is not explicitly tackled, even if Harding and Pagan (2006) showed that it is endogenously determined. Nevertheless, contrary to BP, there is no raison why it should lead to a minimum NRS.

[1] They first only consider two regimes (turmoil and tranquil periods) but also associate a third regime to the post-crisis period, justifying it with recovery specificities.

Hence, even if these approaches seem to be different, they suffer from similar drawbacks related to their evaluation. First, they all use the NSR measure as *in fine* comparison criterion. As noticed by Bussiere and Fratzscher (2006) (p.957) this approach has a clear problem. If the cut-off point decreases, it will lead to a better detection of the coming crises (*i.e.* the type 1 error will decrease), but at the same time it will lead to increase the number of false alarms (*i.e.* type 2 error will increase). To summarize, a more efficient detection of future crises will be at the cost of more frequent false alarms, which may have an efficiency cost in terms of economic policy. Second, no statistical inference is available to test for the forecasting superiority of an EWS compared to another one. It does represent an important issue in light of the previous remark: Does a model exhibiting lower type 1 but higher type 2 error really outperforms the others ?

This paper proposes to tackle both problems by developing a new statistical framework to evaluate EWS. Inherited from the traditional credit-scoring measure (Basel Committee on Banking Supervision, 2005 and Lambert and Lipkovich, 2008 inter alii), it goes beyond the simple analysis of the NSR, and proposes to determine the optimal threshold for each country, relying on the sensitivity-specificity plot and on the accuracy measures. Similarly, we adapt the most important credit-scoring criteria (e.g. AUC, Kuiper Score, Pietra Index, Bayesian Error Rate) using them as evaluating criteria to gauge the forecasting performance of the models. Finally, inference for nested and non-nested hypothesis is developed to identify the optimal specification.

In an empirical part, this framework is implemented to compare the forecasting ability of a fixed effects panel logit and a country-by-country Markov switching model for six Asian and six Latin-American countries. Moreover, the role of forward-looking variables (which have a key role for the theoretical second generation models, see Obstfeld, 1996) is also investigated in both models. Anticipating on our conclusion, it appears that the logit model with market expectation variables works better than the other logit or Markov specifications ; Based on optimally identified thresholds, this model forecasts correctly more than 67.9% of crises and 61% of calm periods in each of the twelve countries, and it is robust to some changes in the construction of the dependent variable like the modification of the pressure index we use or the modification of the crisis definition. In such a framework, the performance assessment criteria proposed and the comparison tests work much better than the usual methods.

This paper is organized as follows: the panel logit as well as the Markov Switching EWS models are presented in Section 2. Section 3 presents our new evaluation framework. The determination of the optimal cut-off as well as the comparison tests are developed. The database as well as the method used to date currency crises are scrutinized in section 4. Empirical results are exposed in section 5 where models are compared. Section 6 concludes.

2 EWS Models : Panel logit and Markov Switching

In this section, the two most well-known EWS models (Panel logit and Markov Switching) are presented. They will be used in an empirical section to apply our new validation framework.

2.1 Panel Logit Model

Following the seminal paper of BP, the fixed-effects panel logit model seems to constitute the adequate model to build an EWS. In a recent paper, Berg et al. (2008) do not advice to include all the countries available in the panel, but instead to consider EWS specific to homogenous clusters (determined in advance). The EWS for countries which cannot be included in any clusters, will rely on simple time series logit regressions.

Let $n = 1, ..., N$ be the number of countries and $t_n = 1, ..., T_n$ the number of time periods considered for the n^{th} country. The dependent variable $y_{n,t_n,j}$ is a binary variable taking the value of 1 if at least a crisis occurs within j periods ahead and 0 if no crisis occurs. j is the forecast horizon and will be removed from the notation for ease of simplicity.

Using a logistic cumulative distribution function, we obtain a conditional logit model for each cluster (Hsiao, 2003):

$$\Pr(y_{n,t_n} = 1) = \frac{\exp(\beta' x_{n,t_n} + f_n)}{1 + \exp(\beta' x_{n,t_n} + f_n)}\ \forall n \in \Omega_h, \tag{1}$$

where Ω_h is the h^{th} cluster, $h \in \{1, ..., H\}$, and $\dim(\Omega_h) = N_h$, so that $\sum_{h=1}^{H} N_h = N$, $\dim(\Omega_h)$ is the number of countries in the h^{th} cluster and where f_n represents the fixed effects (the constant term specific to each country).[2]

This panel so specified exhibits several problems leading to the presence of serial correlation. First, Berg and Coke (2000) shown that when the horizon forecast is larger than 1, it leads to autocorrelation in the crisis variable. Moreover, Harding and Pagan (2006) shown that including a constructed binary variables in a model (as it is the case for y_{n,t_n}) always leads to serial correlation problem. A robust variance-covariance matrix is considered via the sandwich estimator (Williams, 2000).[3]

Nevertheless, the binary choice models have some major drawbacks. First of all, the dependent variable is a binary one, requiring an a priori dating of crisis periods, which is always challenging. Secondly, the threshold used to identify the crisis from non-crisis periods are arbitrarily chosen. Thirdly, a part of the information is lost when the continuous variable

[2]In the case of time-series models there are no individual effects and we can estimate a standard unconditional bivariate logit. The estimation is straightforward.

[3]This method is described in Appendix 1.

is transformed into a qualitative one.

2.2 Markov Switching Model

Markov Switching models do not require prior dating of crises and imposes fewer distributional hypotheses.

Proposed first by Hamilton (1988, 1995) in order to analyse the stance of business cycles, it assumes regime specific relationship between variables. The transition between states is endogenous and depends on fixed transition probabilities[4] which are estimated via a modified Maximum likelihood approach. The model used in the paper allows for change in regime in the regressor but also in the volatility (see Abiad, 2003) and has the following form :

$$KLRm_t = \mu_t(S_t/\Im_t) + \beta(S_t/\Im_t)x_t(S_t/\Im_t) + \epsilon_t(S_t/\Im_t), \tag{2}$$

where $KLRm_t$ is a market pressure index vector, x_t represents the matrix of economic variables, μ is the intercept, ϵ_t is $i.i.d.(0, \sigma^2_{St})$ and $\Im_t$ the information set available at time t. S_t is a latent variable representing the regime, which follows a first order two states Markov chain $\{St\}_{t=1}^{T}$. $S_t = 0$ if there is a crisis and $S_t = 1$ if not. For the ex-post identification of the two regimes, Abiad (2003) considers that the crisis (resp. tranquil) regime is the one having a higher (resp. lower) volatility.

The constant transition probabilities matrix between the regimes from time $t-1$ to t is:

$$P = \begin{pmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{pmatrix} = \begin{pmatrix} P_{00} & 1-P_{11} \\ 1-P_{00} & P_{11} \end{pmatrix}, \tag{3}$$

The model is estimated by Maximum Likelihood method, as described by Hamilton (1995). The initial values of the parameters are obtained from Ordinary Least Squares regressions. For each state the conditional mean and difference to the mean are computed. Next, the normal probability density for each regime η_t can be obtained. Given the initial values of the parameters $(\mu_0,\ \beta_0)$ and of the conditional probability for each regime $\xi_0 = 1/2$, we can iterate from $t = 1$ to T on the following equations:

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}'_{t|t-1} \circ \eta'_t}{1'(\hat{\xi}'_{t|t-1} \circ \eta'_t)} \tag{4}$$

$$\hat{\xi}_{t|t-1} = P\hat{\xi}'_{t-1|t-1}, \tag{5}$$

[4]Time-Varying transition Probabilities have been proposed by Diebold, Weinbach and Lee (1994) but are not considered in the paper.

where $\circ$ denotes element-by-element multiplication, and 1 is the unit vector (column).

The first equation gives the filtered probabilities[5] $\Pr(S_t = i|\Omega_t)$ for each state, while the second one shows the forecasted probabilities of being in a state in the next period $\Pr(S_{t+1} = i|\Omega_t)$.

The conditional normal density η_t and the filtered probabilities allow to compute the conditional log likelihood of the observed data:

$$L(\theta) = \sum_{t=1}^{T} \log(1'(\hat{\xi}_{t|t-1} \circ \eta_t)). \tag{6}$$

Since our objective is to compare the results of the logit and Markov methods, we need to obtain a series of j month ahead forecasts. More precisely, we estimate the probability of observing at least one crisis in the next j periods as follows (Arias and Erlandsson, 2005): [6]

$$\begin{aligned}
\Pr(S_{t+1\dots t+j} = 0|\Im_t) &= 1 - \Pr(S_{t+1\dots t+j} = 1|\Im_t) && (7)\\
&= 1 - \{[P_{01}P_{11}^{(j-1)}\Pr(S_t = 0|\Im_t)] + [P_{11}^{j}\Pr(S_t = 1|\Im_t)]\}. && (8)
\end{aligned}$$

3 A new framework to evaluate EWS models

The evaluation of EWS is particularly important since several models are proposed. However, most of the studies comply with only a few performance assessment procedures based on an arbitrary choice of cut-off and does not offer a framework allowing for statistical inference.

In this section, we offer a new evaluation framework borrowed from the credit-scoring literature. In a first step, assessment criteria are presented and used to determine optimal cut-off points. In a second step, evaluation measure are developed in order to gauge the forecasting quality of the models. Finally a battery of tests is presented to test for the best performing model.

3.1 Performance Assessment and optimal cut-off

This section presents the criteria used to compare the crises probabilities obtained from the EWS model with the actual occurrence of crises within a certain horizon. To be more precise, the EWS models we estimate output crisis probabilities, and in order to shift from these probabilities to crisis forecasts, we have to define an optimal threshold (cut-off) which

[5]Often we are interested in forming an inference about the true regime at date t based on observations obtained through a later date T, denoted $\xi_{t|T}$. These are referred to as "smoothed" probabilities $\Pr(S_t = i|\Omega_T)$ and they are given by Kim's algorithm (1994): $\xi_{t|T} = \xi_{t+1|T}\xi_{t|t}P/\xi_{t+1|t}$.

[6]See Krolzig (2000) for a description of the other estimation methods.

discriminates between predicted crisis periods and predicted calm periods. It is known that if the probability of a crisis oversteps the cut-off, it issues a signal of a forthcoming crisis. Moreover, the lower the threshold, the more signals the model will send (*i.e.* error of type 1 will decrease), but at the same time, the number of wrong signals rises (*i.e.* error of type 2 will increase). Conversely, using a higher threshold level reduces the number of wrong signals, but the number of missing crisis signals increases. Therefore, we can define an indicator variable of predicted occurrence of crisis as following:

$$\hat{I}_t = \begin{cases} 1, \text{if} P_t > C \\ 0, \text{otherwise} \end{cases}, \tag{9}$$

where C represents a fixed cut-off.

We have decided to solve this trade-off by following the methodology used in credit-scoring (Basel Committee on Banking Supervision, 2005). In fact, several evaluation criteria of predictive ability are available in the credit-scoring literature, but we concentrate on the most important ones: Kuiper's score, Quadratic Probability Score, Log Probability Score, The area under the ROC curve (AUC), Pietra Index, and Bayesian Error.

First, we begin with some general concepts and two methods we use to determine the optimum threshold for each country. Finally, we can introduce the performance assessment criteria.

3.1.1 Optimal Cut-off Identification

For a given value of the cut-off C, where $C \in [0, 1]$, Table 1 reports the link between the observed (y_t) and predicted ($\hat{I}_t$) conditions in the following matrix:

TAB. 1 – True versus predicted occurrence of crises

		True value		
		Crisis	No crisis	Total
Predicted result	Crisis	True Positive (TP(C))	False Positive (FP(C))	All predicted crisis
	No crisis	False Negative (FN(C))	True Negative (TN(C))	All predicted non-crisis
	Total	All true crisis (N_D)	All non-crises (N_{ND})	T (sample size)

Note : For a given cut-off we count the number of crises correctly identified and the number of non-crises well predicted by the model (the first diagonal elements of the matrix) and the misidentified crisis and respectively non-crisis (the off diagonal elements). The number of real crises/non-crises periods lies on the 'total' row, while the number of predicted crises/non-crises lies on the 'total' column.

BP consider as the only objective criteria the NSR, defined as :

$$NSR(C) = \frac{FP(C)}{TP(C)}. \tag{10}$$

The optimal threshold (C^*) is obtained as the value that minimized $NSR(C)$ (see BP and KLR).

It is nevertheless possible to go deeper, defining first the hit rate (*sensitivity*) as

$$HR(C) = \frac{TP(C)}{N_D}, \tag{11}$$

where $TP(C)$ is the number of crisis predicted correctly (hits) using a cut-off equal to C , and N_D is the total number of crisis in the sample. At the same time, the false alarm rate ($1 - specificity$) is defined as

$$FAR(C) = \frac{FP(C)}{N_{ND}}, \tag{12}$$

where $FP(C)$ is the number of false alarms (false signals), and N_{ND} is the total number of non-crisis periods. The optimal cut-off rate is determined as to maximize simultaneously and conditionaly *sensitivity* and $specificity$. (see Figure 1)

insert Figure 1

Several additional operating characteristic measures of accuracy are implemented (Lambert and Lipkovich (2008)): The Youden Index (J), Total Accuracy (TA), and Matthews Correlation Coefficient (MCC). Besides, the total and weighted misclassification errors are presented (TME and WME). Similarly to the *sensitivity* and $specificity$, these accuracy and error measures can be used to define the optimal cut-off value. More precisely, the Total Accuracy measure is defined as the proportion of cases correctly predicted:

$$TA = \frac{TP(C) + TN(C)}{T}, \tag{13}$$

the Youden Index by:

$$J = HR(C) - FAR(C), \tag{14}$$

and Matthews Correlation Coefficient as:

$$MCC = \frac{TP(C) * TN(C) - FP(C) * FN(C)}{\sqrt{((TP(C) + FN(C)) * (TP(C) + FP(C)) * (TN(C) + FP(C)) * (TN(C) + FN(C)))}}. \tag{15}$$

At the same time, the misclassification errors are calculated as follows:

$$TME = FN(C) * L_{fn} + FP(C) * L_{fp}, \tag{16}$$

$$WME = \frac{FN(C) + W_{fn}}{T} + \frac{FP(C) + W_{fp}}{T}, \tag{17}$$

where L_{fn} are the losses associated with crisis periods categorized as tranquil ($L_{fn} = 1/N_D$),

L_{fp} are the losses associated with calm periods identified as crises ($L_{fp} = 1/N_{ND}$), and where W_{fn} and W_{fp} are weights based on the relative losses: ($W_{fn} = L_{fn}/(L_{fn} + L_{fp})$, and ($W_{fp} = L_{fp}/(L_{fn} + L_{fp})$).

As a result, it is possible to determine for each country an optimal cut-off for which the accuracy measures are maximized and the error measures are minimized.

3.1.2 Evaluation Criteria

Further, we are giving more details of the performance assessment criteria previously mentioned at the beginning of this subsection. To start with, we can formally introduce Kuiper's score (Granger and Pesaran (2000)), as the difference between the proportion of crises correctly predicted (HR) and the proportion of tranquil periods incorrectly forecasted (FAR):

$$KS = HR - FAR. \tag{18}$$

The model above generates more hits (crises well identified) than false alarms if the value of Kuiper's score is positive.

The Quadratic Probability Score (QPS) is a mean square error measure which compares the predicted probabilities of the two states (crisis/ non-crisis) with the real crisis indicator. It is defined as:

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2(P_t - y_t)^2, \tag{19}$$

where P_t represents the estimated probability of crisis at time t and y_t is the realization of the crisis event at time t. QPS takes values from 0 to 2, with 0 being perfect accuracy.

The Log Probability Score (LPS) loss function penalizes large errors more heavily than QPS:

$$LPS = -\frac{1}{T}\sum_{t=1}^{T}((1 - y_t)\ln(1 - P_t) + y_t\ln(P_t)). \tag{20}$$

It ranges from 0 to ∞, with LPS=0 being perfect accuracy.

The ROC curve is a visual tool whose concavity is equivalent to the conditional probabilities of crisis being a decreasing function of the underlying scores. Its objective is to minimize the misclassification and maximize the overall hit rate of the model. More exactly, the ROC curve is a graphic of the *sensitivity* against $1 - specificity$ for different values of the cut-off and is represented in Figure 2. The curve of the perfect model passes through the point (0,1), recognizing perfectly the crisis and non-crisis periods as they really are.

insert Figure 2

The Area Under the ROC curve, AUC, which ranges from zero to one, provides a measure of the model's overall ability to discriminate between the cases correctly predicted and the false alarms. The larger the AUC, the better the model. This statistic can be calculated as $A = \int_0^1 HR(FAR)d(FAR)$, and its value corresponds to the Wilcoxon-Mann-Whitney ranking statistic. In other words, the AUC estimates the probability that a randomly chosen crisis observation is ranked higher than a non-crisis observation. Thus, a perfect ranking means that all crisis observations are ranked higher than the non-crisis observations, and consequently it implies an AUC equal to 1. By contrast, the expected value of the AUC statistic for a random ranking is 0.5.

The Pietra Index is another ROC measure which quantifies half of the maximal distance between the ROC and the diagonal of the unit square. Like the AUC, the Pietra Index does not depend on the crisis probability of the sample. Geometrically, in the case of a concave ROC curve, we obtain the following representation of the statistic:

$$Pietra\ Index = \frac{\sqrt{2}}{4} \max_C |HR(C) - FAR(C)|. \tag{21}$$

Last but not least, the Bayesien Error rate (or classification error) criteria returns the minimum probability of error for a binary case model. It can be estimated parametrically or non-parametrically, but in the case of a concave ROC curve the error rate can be expressed as:

$$Error\ rate = \min_C (P_D(1 - HR(C)) + (1 - P_D)FAR(C)), \tag{22}$$

where P_D represents the rate of crisis in the sample ($P_D = N_D/T$).

3.2 Comparison Tests

At this stage, we have set up rules to obtain an optimal cut-off, as well as forecast evaluation. While most of the other studies would stop at this stage, and select the model providing the best forecast evaluation, we think it is necessary to develop a suitable statistical framework to test for the forecasting equivalence between 2 models with respect to the evaluation measure.

Accordingly, we will use three such tests, namely the DM statistic for non-nested models (Diebold-Mariano, 1995), Clark and West (2007) statistic 'MSPE-adj' for nested models, and a test of comparison of correlated ROC curves (DeLong et al., 1988).

The first test is the one proposed by Diebold and Mariano (1995) which can be built on any criteria, such as MSFE or MAE, or even loss functions.

Let us consider two forecasts at a horizon k, $\hat{y}_{1,t+k}$, and $\hat{y}_{1,t+k}$, $t \in \{1, ..., T\}$ of the time series y_t , obtained from two concurrent models. Now, let the forecast errors for the two models be $\{\widehat{e_{1,t}}\}_{t=1}^{T}$, and $\{\widehat{e_{2,t}}\}_{t=1}^{T}$. We denote by $g(\widehat{e_t})$ the loss function associated with a forecast e_t. Now, we can formulate the null hypothesis of equal forecast accuracy:

$$H_0 : \mathrm{E}[g(\widehat{e_{1,t}})] = \mathrm{E}[g(\widehat{e_{2,t}})], \tag{23}$$

or

$$H_0 : \mathrm{E}(d_t) = 0, \tag{24}$$

where d_t represents the loss differential: $d_t = g(e_{1,t}) - g(e_{2,t})$.

In this paper Diebold-Mariano's test is based on the loss differential mean $\overline{d} = (1/T)\sum_{t=1}^{T} d_t$. Under the null hypothesis $\overline{d}$ follows a normal distribution with a variance equals to $\sigma^2_{\bar{d},0}/T$, where $\sigma^2_{\bar{d},0}$ is the long term variance of the loss differential:

$$DM = \frac{\bar{d}}{\sqrt{\mathrm{var}(\bar{d})/T}} = \frac{\bar{d}}{\sigma_{\bar{d},0}\sqrt{T}} \xrightarrow[T\to\infty]{d} N(0,1). \tag{25}$$

The long term variance can be obtained by using a kernel estimator which corresponds to a weighted sum of future and past autocovariances of the loss differential d_t. In the case of an uniform kernel, as chosen by Diebold and Mariano(1995), the long term variance estimator has the following form:

$$\sigma^2_{\bar{d},0} = \sum_{t=-(k-1)}^{k-1} \hat{\gamma}_d(j) = \hat{\gamma}_d(0) + 2\sum_{t=1}^{k-1} \hat{\gamma}_d(t), \tag{26}$$

where $\hat{\gamma}_d(t) = (1/T)\sum_{i=t+1}^{T}(d_i - \bar{d})(d_{i-t} - \bar{d})$ is the empirical autocovariance of order t of d_t.

Dielbod and Mariano (1995) highlight that the auto-covariances can be at most of order $k-1$, where k is the forecast horizon. Therefore, in case of a one step ahead forecast the variance estimator is given by the empirical variance of the loss differential $\hat{\gamma}_0(t)$. In addition, since there is no particular loss function that we could use in the case of dichotomous dependent variables, we decide to consider the MSFE measure as the comparison criterion of the models. As a result, we have $g(\hat{e}_{1,t}) = \hat{e}^2_{1,t}$ and $g(\hat{e}_{2,t}) = \hat{e}^2_{2,t}$. The rest of the formulas can be obtained by a simple substitution.

Even if this test is appealing in our analysis, it cannot be implemented when models are nested. In such a case the distribution under the null hypothesis could not be established anymore. An appropriate alternative test has been suggested by Clark et McCracken (2001) and Clark and West (2007).

Let us consider that model 1 is the parsimonious one and model 2 is the larger one, which reduces to model 1 if some of its parameters are set to 0. We remind that the sample size is T. Besides, the k step ahead forecasts of the two models are denoted $\hat{y}_{1,t+k}$ and $\hat{y}_{2,t+k}$.

The null hypothesis is equal MSPE, while the alternative is that the unrestricted model (model 2) has a smaller MSPE than the restricted one (model 1), *i.e.* it performs better than the other one.

Consequently, we can compute Clark and West (2007) MSPE-adj. statistic as:

$$\text{MSPE-adj.} = \frac{\sqrt{T}\bar{f}}{\sqrt{\hat{V}}}, \tag{27}$$

where $\hat{f}_{t+k} = (y_{t+k} - \hat{y}_{1,t+k})^2 - [(y_{t+k} - \hat{y}_{2,t+k})^2 - (\hat{y}_{2,t+k} - \hat{y}_{1,t+k})^2]$, $\bar{f}_{t+k}$ is the sample average of $\hat{f}_{t+k}$ and $\hat{V}$ is the sample variance of $(\hat{f}_{t+k} - \bar{f})$. This one-sided test uses critical values from the standard normal distribution.

The last test to use is DeLong et al. (1988) which presents a nonparametric analysis of areas under correlated ROC curves, by using the theory on generalized U-statistics to generate an estimated covariance matrix. The null hypothesis is the equality of areas under the ROC ($L * \widehat{AUC'} = 0$), *i.e.* none of the models performs better than the others.

Then, we can introduce the test statistic as having the following form:

$$(\widehat{AUC} - AUC)L^{'}[L(\frac{1}{N_D}S_{10} + \frac{1}{N_{ND}}S_{01})L^{'}]^{-1}L(\widehat{AUC} - AUC)^{'}, \tag{28}$$

where the vector of statistics $\widehat{AUC}$ represents the estimated area under the ROC curve for h different models, $\widehat{AUC} = (\widehat{AUC^1}, \widehat{AUC^2}, ...\widehat{AUC^h})$, L is the coefficients vector (row), and S_{10} and S_{01} are two components of the covariance matrix for $\widehat{AUC}$, defined on the basis of Hoeffding (1948)'s theory for generalized U-statistics, so that

$$S = \frac{1}{N_D}S_{10} + \frac{1}{N_{ND}}S_{01}. \tag{29}$$

This statistic has a χ^2 distribution with *df* degrees of freedom, where *df* is given by the rank of $LSL^{'}$. In addition, we remind that in small samples the power of the test is likely to be small.

4 Data

Before going on with the model comparison, it is important to look at the data as well as some specification problems (such as the way to build the binary variable and the data poolability).

4.1 Data

Following Berg et al. (2008), we consider a database of 12 countries[7] for which we have obtained monthly data in US dollars adjusted for seasonality, from January 1985 to January 2005 from the IMF-IFS database or national banks of the countries via Datastream.

We act in accordance with the choices of Lestano et al. (2003) in selecting the economic variables and then we reduce the impact of extreme values as Kumar et al. (2003). Accordingly, we consider explanatory variables from three economic sectors:

1. External sector: one-year growth rate of international reserves, of imports and exports, M2 to foreign reserves, and one-year growth of M2 to foreign reserves.

2. Financial sector: one-year growth of M2 multiplier, one-year growth of domestic credit over GDP, real interest rate, lending rate over deposit rate, one-year growth of real bank deposits, and real interest rate differential.

3. Domestic real and public sector: one-year growth of industrial production.

In addition, we test the capability of the market expectation indicators to explain the occurrence of currency crises by introducing the yield spread and the stock market price index into the model. The term spread is defined as the difference between the money rate and the long term government bonds. In case of missing data, proxies like 'money ten days', 'interbank one year rate' or 'money 364 days' were used. As in Kumar (2003), we dampen every variable using the formula: $f(x_t) = \text{sign}(x_t) * \ln(1 + |x_t|)$. Traditional first generation (Im, Pesaran, Shin, 1997) and MW (Maddala and Wu 1999) and second generation (Bai et Ng, 2001 and Pesaran, 2003) panel unit root tests are performed, leading to the rejection of the null hypothesis of stochastic trend except for the lending rate over deposit rate indicator. Hence, this series is substituted by its first difference.

Finally, we identify the most correlated leading indicators: real interest rate is highly correlated with real interest rate differential, while one-year growth of imports is strongly correlated with one-year growth of industrial production and with one-year growth of exports. Based on the minimization of the AIC and BIC information criteria of the panel data models, we identify the best two variables, namely real interest rate and one-year growth of industrial production. The gaps through the series are replaced with the mean values, but when series revealed missing values at the beginning of the sample, such as "one-year growth of terms of trade" or "yield spread", the corresponding observations are dropped from the analysis, leading to the unbalanced framework mentioned earlier.

[7] Argentina, Brazil, Mexico, Peru, Uruguay, Venezuela, Indonesia, South Korea, Malaysia, Philippines, Taiwan and Thailand

4.2 Dating currency crises

A crisis episode is generally detected if an index of speculative pressure exceeds a certain threshold. Many alternative indexes have been developed and used for identifying crises. But they are all non-parametric termination rules which take into consideration the size of the movements in a combination of a number of series. Lestano and Jacobs (2004) have compared several currency crisis dating methods, aiming to spot the one that recognizes most of the crises categorized by the IMF. Finally, they conclude that KLR modified index, Zhang original index (Zhang, 2001), and extreme values applied to KLR modified index perform best.

Following their results, we identify crisis periods using the KLR modified pressure index (KLRm), which, unlike the KLR index, also includes interest rates:

$$\mathrm{KLRm}_{n,t} = \frac{\Delta e_{n,t}}{e_{n,t}} - \frac{\sigma_e}{\sigma_r}\frac{\Delta r_{n,t}}{r_{n,t}} + \frac{\sigma_e}{\sigma_i}\Delta i_{n,t}, \tag{30}$$

where $e_{n,t}$ denotes the exchange rate (units of country n's currency per US dollar in period t), $r_{n,t}$ represents the foreign reserves of country n in period t, while $i_{n,t}$ is the interest rate in country n at time t. Meanwhile, the standard deviations σ_X are actually the standard deviations of the relative changes in the variables ($\sigma_{(\Delta X_{n,t}/X_{n,t})}$), where X denotes each variable at a time: the exchange rate, the foreign reserves, and the interest rate, and $\Delta X_{n,t} = X_{n,t} - X_{n,t-6}$.[8]

Nonetheless, for both subsamples the threshold equals two standard deviations above the mean[9]:

$$Crisis_{n,t} = \begin{cases} 1, & \text{if } \mathrm{KLRm}_{n,t} > 2\sigma_{\mathrm{KLRm}_{n,t}} + \mu_{\mathrm{KLRm}_{n,t}} \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

From a macroeconomic point of view, it is more important to know if there will be a crisis in a certain horizon than in a certain month, because it allows the state to take steps to prevent the crisis. Consequently, we define for each country $C24_t$ which corresponds to y_t from our general framework, as the crisis dummy variable taking the value of one if there

[8] Additionally, we take into account the existence of a higher volatility in periods of high inflation, and consequently the sample is split into high and low inflation periods. The cut-off corresponds to six month inflation rate higher than 50%.

[9] In the case of KLR the threshold equals three standard deviations, but then Taiwan would never register any currency crises, which historically is not true (e.g. Taiwan was not exempted by the Asian crisis in 1997).

will be a crisis in the following 24 months and taking the value of 0 otherwise[10]:

$$C24_{n,t} = \begin{cases} 1, & \text{if } \sum_{j=1}^{24} Crisis_{n,t+j} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

It is important to note that in the case of Markov switching model, since the two regimes are identified intrinsically by the model, we use the KLRm index as a continuous variable, without transforming it into a binary one. Nevertheless, the identification of crisis dates concerns us from the viewpoint of forecast evaluation criteria and model comparison tests.

4.3 Optimal country clusters

Berg et al. (2008) have pointed out the importance of applying a panel-logit model only on clusters made of statistically poolable countries. In order to identify these countries , we use Kapetanios recursive procedure (Kapetanios, 2003) based on a traditional Hausman test (see Appendix 3).

Unsurprisingly, four optimal clusters are identified: Argentina, Brazil, Mexico, Venezuela ; Peru and Uruguay ; Korea, Malaysia, Taiwan ; Philippines and Thailand and one not poolable country, Indonesia, confirming the fact that in different countries there are different factors explaining currency crises. Actually, we were expecting to see clusters of countries smaller than the regional sample (Latin America and respectively South Asia), since they experienced similar macroeconomic events (high inflation rates, reforms and stabilization plans) and they are linked by strong trading relations.

In search of comparable results, we shall use the five[11] previously identified clusters for all panel and time-series data models developed in this paper.

5 Empirical results

As aforementioned, the aim of this paper is to find the model that best identifies crisis periods as crises and calm periods as calm periods. In order to do this, we use both a panel logit and a Markov switching framework. More exactly, we develop several specifications for each of the two approaches: a logit model without market expectation variables and one including also market expectation variables, two Markov models without market expectation variables (one with intercept and standard error switching coefficients and one

[10]The whole exercise is performed with $C12$ in order to test for the robustness of the approach in appendix 2. It results that outcomes are similar.

[11]We consider the non-poolable country (Indonesia) as an individual cluster. Therefore, five optimal clusters are identified.

including an *international reserves* - hereafter *resg* switching coefficient as well), and three Markov models with market expectation variables (one with intercept and standard error switching coefficients, a second one which also includes an *international reserves* switching coefficient, and a third one including a *yield spread* - hereafter *spread* switching coefficient apart from the two previously mentioned). Then, for each model we opt for the following 4-steps approach:

First, each model is estimated so as to obtain the country probability at time t of having a crisis in the following 24 months.

Second, we apply the backtesting methodology based on credit-scoring criteria so as to identify the optimal cut-off.

Third, the predictive performance of each model is scrutinized.

Finally, comparison tests are implemented to find the outperforming model.

5.1 Estimation, optimal cut-off and performance assessment

In this section we analyze the estimation results of our models, and for a given model we identify the optimal cut-off and compute the values of the performance assessment criteria for each of the twelve countries. Since we study numerous specifications, the results are only partially reproduced in the paper, the rest of them being available upon request.

insert Tables 2 and 3

Tables 2 and 3 show the Maximum Likelihood estimates of the logit model with market expectation variables for each of the five optimal clusters. The goodness of fit indicators reveal that the independent variables have an important explanatory power for all clusters. Moreover, the signs of the coefficients tend to correspond to the a priori expectations. Nonetheless, we encounter some variations from one cluster to another which concerns both the significance and the sign of the parameters. Therefore, we can say that we find evidence of parameter heterogeneity between optimal clusters. To be more precise, only the growth of international reserves and the growth of M2 to reserves coefficients are always negative, indicating that a rise of one unity of their value implies a reduction in the crisis probability. Besides, the signs of the coefficients for variables like growth of domestic credit over GDP, growth of stock market price index, or yield spread are relatively stable across the five clusters (only one of the signs differs).

At the same time, we identify the variables growth of real bank deposits, growth of M2 to reserves and yield spread as being the most significant across clusters. On top of that, the yield spread variable is significant only in the case of the last three clusters (South-Asian countries), which indicates the importance of forward information to predict crises.

These results provide support for second generation models, stressing the importance of non fundamental factors in the occurrence of financial turmoil.

insert Figures 3 and 4

Figures 3 and 4 report the probabilities of crisis obtained from the two logit models. For Brazil, Indonesia, Korea, Taiwan and Thailand the probabilities of crisis obtained using a logit model with market expectation variables are most of the time superior to those obtained from a simple logit model. On the contrary, for Latin American countries (Argentina, Malaysia, Mexico or Venezuela) probabilities of crises are quite close whatever the considered model. Finally, for Peru, Philippines and Uruguay, the comparison is uneasy as each model overcomes its competitor on particular periods. In other words, there are countries for which the model with market expectation variables seems better than the other one, and there are countries for which it is difficult to identify any permanent differences in the forecast of crisis probabilities and consequently it is hard to choose the best model. Nevertheless, the simple logit model is never the most adequate model for the whole period.

Generally speaking, all currency crisis episodes are forecasted quite well by these two models. For example, warning signals of the Asian crisis are already apparent in 1995 in some of the concerned countries, and the 2002 Argentinean depreciation, the Brazilian crisis in 1999, the Mexican peso crisis in 1994-1995, and the Peruvian crisis in 1999 are anticipated with at least one year in advance. On the contrary, neither the 2002 currency crises in Venezuela, nor the 1997-1998 crisis in Thailand, or the 1998-2000 crisis in Philippines were foreseen in due time. We must add, though, that sometimes these models identify depreciations which were not quantified as crises by our KLRm pressure index, since they appear at the beginning of the estimation sample (e.g. January 1995 in Argentina, 1995 in Brazil, 1992-1994 in Malaysia, etc.). Nevertheless, all currency crisis episodes pointed at by these models have already been recorded and analyzed by other studies (see Abiad, 1993, and Dabrowski, 2003).

Last but not least, all variables with the exception of the growth of domestic credit over GDP are significant in at least one model. Moreover, the use of robust standard errors has a certain effect on the significance of the coefficients, since some variables are no more significant whereas others are becoming highly significant.

The results of the simple logit model are quite similar to those of the logit model with market expectation variables in terms of sign and significance of the common coefficients. At the same time, the Markov models are characterized by a large variability in terms of sign and significance of the coefficients of the independent variables [12].

In order to analyze the performance of each model, the optimal cut-off C has to be estimated for each country. As explained in section (3.1), we use two graphical tools: one

[12] These results are available upon request from the authors

based on the sensitivity-specificity indicators, and one based on several accuracy and error measures. Moreover, to compare our results with the existing literature, we indicate in the last column, the cut-off obtained using NSR.

insert Tables 4 and 5

Tables 4 and 5 report the cut-off obtained for each country. We put in bold the ones which maximize both sensitivity and specificity. It is noticeable that most of the time the optimal cut-off is the one obtained by the accuracy measures. Moreover, cut-off obtained by NSR (reported as KLR) is always higher than the one resulting from the two other methods. It reveals that existing studies are much more conservative than the current one. Concerning the models, the panel logit leads to a cut-off varying between 0.08 and 0.38, whereas for Markov models it is higher (0.7-1). These results highlight that the use of a predetermined cut-off may induce some loss of information and disturbances in the measure of the forecasting performance of the models.

Most importantly, the sensitivity and specificity values seem higher for logit models than for Markov ones; in the first case we always deal with more than 60% of cases correctly identified (see table 4), while in the second case, this rate is lower, reaching even 0% in some cases (see table 5). In fact, there are even two models (Markov model with *resg* switching and Markov model with market expectation variables and *resg* switching) delivering a constant cut-off of 0.9995 and with sensitivities (specificities) of 0% (100%) for all countries.

At the same time, the market expectation logit model leads to higher sensitivity and specificity values for the South-Asian countries than the simple logit model, while the simple logit has higher values for the Latin-American countries, emphasizing the idea that the currency crises of the Asian countries can be better predicted by using market expectation variables. It confirms our previous results of the forward looking specificity of the asian crisis. On the other hand, in the case of Markov models, the role of market expectation variables is not so clearly supported by the results.

It seems that logit models, in spite of their autocorrelation problem, appear to over-perform Markov models. Similarly, market expectation variables turn out to have a more clear effect when introduced in a logit models. In the next subsection the proper statistical assessment criteria of the performance of each model is presented.

5.2 Evaluation criteria

The following statistics assess the performance of a model based on the sensitivity-specificity measures (AUC, Kuiper score, Pietra index, Bayesian error rate) or by comparing the forecasts with the realizations of the crisis variable C24 (QPS, LPS). Therefore, the higher the AUC the better the model will be. A positive value of Kuiper's score signifies that the model generates more hits than false alarms and so its predictive performance will increase. Similarly, a higher Pietra index, respectively a lower Bayesian error rate indicate a more stringent model, as well as values of the QPS and LPS closer to zero.

insert Table 6

From Table 6, it turns out that the logit model exhibits correct predictive properties: AUC is higher than 0.731, Kuiper score is always positive, Pietra index values are greater than 0.145, Bayesian error rate is inferior to 0.272, the absolute value of QPS is less than 0.37 and the absolute value of LPS is less than 0.558. The results of the two models (with or without forward looking variables) are quite similar, with small differences from one country to another.

Concerning Markov models, the results are less satisfactory. First of all, the two models with switching *resg* are not very different from a random model in terms of performance[13]: the AUC criteria equals very frequently 0.5, Kuiper's score and Pietra index are very small, while the Bayesian error rate, QPS and LPS are high. Second, the other three models (without *resg* switch) have similar behavior (see table 6), and clearly outperform the first two (with *resg* switch). Nevertheless, we get the overwhelming feeling that Markov specifications are dominated by panel logit in terms of forecasting performances. Still, it has to be confirmed by proper statistical test. This will constitute the last part of our analysis.

5.3 Comparison Tests

In order to determine the best model specification, the three tests presented in section 3.2 (the first one compares the area under the ROC curve, the second one <Clark-West MSPE-adj> compares the forecasts of two competitive nested models, while the third one <Diebold-Mariano DM statistic> compares the forecasts of two non-nested models) are implemented.

The results of the ROC and Clark-West tests comparing the simple logit without and with market expectation variables is presented in Table 7. It corroborates the idea that forward looking variables matter for crisis prediction. In fact, this conclusion is not surprising at all as it matches the results obtained with the predicted probability plots.

insert Table 7

In the case of Markov models, the differences between models are not always clear. For example, for countries like Indonesia, Korea, Malaysia, Mexico, Markov model with market expectation variables is outperforming the one with forward looking variables whereas for Brazil, Indonesia, Peru and Uruguay, an opposite conclusion is reached.

[13]These results are available upon request.

insert Table 8

Finally, the last step in the procedure of predictive ability consists in comparing the logit model with market expectation variables to the Markov model with market expectation variables and *spread* switching. Results reported in Table 8 indicate that the two tests are always significant, leaving no doubt concerning the choice of the panel logit as the best model. Thus, we can clearly conclude that the outperforming model is the logit model with market expectation variables. Moreover, a sensitivity analysis of the Logit model with market expectation variables (see appendix 2) proves the robustness of this result.

6 Conclusion

This paper proposes to tackle both problems by developing a new statistical framework to evaluate EWS. Inherited from the traditional credit-scoring measure (Lambert and Lipkovich, 2008), it goes beyond the simple analysis of the NSR analysis, and proposes a measure of accuracy as well as a sensitivity and specificity analysis. It is then possible to determine the optimal threshold for each country, relying on the sensitivity-specificity plot. Similarly, we adapt the most important credit-scoring criteria (e.g. AUC, Kuiper Score, Pietra Index, Bayesian Error Rate) using them as evaluation criteria of the performance of the model specifications we have developed. Finally, nested and non-nested comparison tests are developed to identify the optimal specification.

It results from this approach an unified framework to compare candidate EWS models. Applied to compare the predictive power of a panel logit and Markov frameworks, it leads to several conclusions. First the Panel logit is outperforming the Markov based EWS. Second, the introduction of forward looking variables (here the term spread) clearly improves the forecasting properties of the EWS. It thus confirms the adequacy of the second generation crisis models in explaining the occurrence of crises. Third, the optimal EWS based on logit model with market expectation variables predicts quite well most currency crises in the specified emerging markets (it forecasts correctly at least 67.9% of crises and 61% of calm periods in each of the twelve countries). The very good forecasting performance of this model and its robustness to some sensitivity analysis provides a dominating position within EWS models for logit model with market expectation variables.

Such a conclusion is of course conditional to our specification. Hence it calls for deeper empirical applications relying on some hypotheses (two-regime model, with autoregressive and volatility regime dependent model,..) and including extra forward looking variable (market feeling,..). Nevertheless, our evaluation procedure provides an unified framework to compare EWS models and clearly indicate a research direction to follow.

Bibliography

(1) Abiad, A., (2003). Early Warning systems: A survey and a regime switching approach, IMF Working Paper.

(2) Arias, G. and Erlandsson, G., (2005). Improving early warning systems with a Markov switching model - an application to south-east Asian crises, Lund University, Working Paper.

(3) Basel Committee on Banking Supervision, (2005). Studies on the Validation of Internal Rating Systems, working paper no.14, Bank for International Settlements.

(4) Berg, A., and Pattillo, C., (1999). Predicting Currency Crises: The Indicators Approach and an Alternative. Journal of International Money and Finance 18, 561-586.

(5) Berg, J.B., Candelon, B, and Urbain, J.P., (2008). A cautious note on the use of panel models to predict financial crises, Economics Letters 101, issue 1, 80-83.

(6) Berg, A., and Cooke, R., (2004). Autocorrelation corrected standard errors in panel probits: an application to currency crisis prediction, IMF Working Paper.

(7) Bussiere, M. and Fratzscher, M., (2006). Towards a new early warning system of financial crises, Journal of International Money and Finance 25, issue 6, 953-973.

(8) Clark, T. E., McCracken, M. W., (2001). Tests of equal forecast accuracy and encompassing for nested models, Journal of Econometrics 105, issue 1, 85-110.

(9) Clark, T. E., West, K. D., (2007). Approximately normal tests for equal predictive accuracy in nested models, Journal of Econometrics 138, issue 1, 291-311.

(10) Diebold, F. X., Lee, J.-H. and Weinbach, G., (1994). *Regime Switching with Time-Varying Transition Probabilities*, in C.Hargreaves (ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press.

(11) Diebold, F. X., Mariano, S., (1995). Comparing Predictive Accuracy, Journal of Business and Economic Statistics, American Statistical Association 13, issue 3, 253-63.

(12) Fratzscher,M., (2003). On currency crises and contagion, International Journal of Finance and Economics, 8(2), 109-129.

(13) Fuertes, A.-M., Kalotychou, E., (2007) Optimal design of early warning systems for sovereign debt crises, International Journal of Forecasting, 23(1), 85-100.

(14) Gould, W., Pitblado, J., Sribney,W., Stata Corporation ,(2005). *Maximum likelihood estimation with stata*, Stata Press.

(15) Granger, C.W.J., and Pesaran, M.H., (2000). Economic and Statistical Measures of Forecast Accuracy, Journal of Forecasting 19, 537-560.

(16) Hamilton, J. D., (1995). *Regime Switching Methods, University of California*, Prepared for: Palgrave Dictionary of Economics.

(17) Hamilton, J. D., (1988). Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates. Journal of Economic Dynamics and Control 12, 385-423.

(18) Harding, D., Pagan, A., (2006). The Econometric Analysis of Constructed Binary Time Series, Working Papers Series 963, The University of Melbourne.

(19) Hoeffding, W. (1948) A class of statistics with asymptotically normal distributions. Annals of Statistics, 19, 293-325.

(20) Hsiao, C., (2003). *Analysis of panel data*. Cambridge University Press.

(21) Im, K.S. & Pesaran, M. H., & Shin, Y., (2003). Testing for unit roots in heterogeneous panels, Journal of Econometrics, 115(1), 53-74.

(22) Jacobs, J., Kuper, G., and Lestano, (2003). Indicators of financial crises do work! An early warning system for six Asian countries, CCSO Centre for Economic Research.

(23) Kaminsky, G., Lizondo, S., Reinhart, C., (1998). Leading indicators of currency crises, IMF Staff Papers 45, issue 1.

(24) Kapetanios, G., (2003). Determining the Poolability of Individual Series in Panel Datasets, Working Paper 499.

(25) Kim, C. J. (1994), Dynamic Linear Models With Markov Switching, Journal of Econometrics 64, 1-22.

(26) Kumar, M., Moorthy, U. and Perraudin, W., (2003). Predicting emerging market currency crashes, Journal of Empirical Finance 10, 427-454.

(27) Lambert, J. and Lipkovich, I., (2008). A Macro for getting more out of your ROC Curve, SAS Global forum, paper 231.

(28) Lestano and Jacobs, J., (2004). A comparison of currency crisis dating methods: East Asia 1970-2002, CCSO Working Papers 200412, CCSO Centre for Economic Research.

(29) Maddala, G.S., and Wu, S., (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test., Oxford Bulletin of Economics and Statistics, special issue, 631-652.

(30) Martinez Peria, M.S.(2002) "A Regime-Switching Approach to the Study of Speculative Attacks : A Focus on EMS Crises", in J.D. Hamilton and B. Raj (eds), Advances in Markov-Switching Models. Applications in Business Cycle Research and Finance, Physica-Verlag, Heidelberg, Germany, 2002.

(31) Williams, R., (2004). A Note on Robust Variance Estimation for Cluster-Correlated Data, Biometrics 56, Issue 2, 645 - 646.

Appendix 1 : A Robust Estimator of the Variance of the Parameters

As previously mentioned, we use a sandwich estimator in order to compute robust estimators of the variance in the case of the logit models. Technically, we know that the matrix of variance of the estimators is asymptotically equal to the inverse of the hessian matrix: $\mathrm{V}(\hat{\beta}) = -H(\hat{\beta})^{-1}$. Still, this works only if we use the real Data Generating Process. Since we want a more permissive method from this point of view, we define the variance vector as follows:

$$\mathrm{V}(\hat{\beta}) = (-H(\hat{\beta})^{-1})\mathrm{V}(g(\hat{\beta}))(-H(\hat{\beta})^{-1}), \tag{33}$$

where $H(\hat{\beta})^{-1}$ is the inverse of the hessian matrix, and $\mathrm{V}(g(\hat{\beta}))$ is the variance of the gradient. Using the empirical variance estimator of the gradient we find that:

$$\mathrm{V}(\hat{\beta}) = -T/(T-1)H(\hat{\beta})^{-1}\{\sum_{t=1}^{T} g_t(\hat{\beta})g_t(\hat{\beta})^{-1}\}(-H(\hat{\beta})^{-1}), \tag{34}$$

which is a robust variance estimator for the time-series model.

The main advantage of this sandwich method is that it can also be applied in the case of grouped data, as in our case. It is important to note that in the current situation each country from a cluster is a group of time-series observations which are correlated. Thus, the observations corresponding to a country are not treated as independent, but the countries themselves, forming the analyzed cluster, are considered independent. Therefore, instead of using $g_t(\hat{\beta})$, we use the sum of $g_t(\hat{\beta})$ for each country and T is replaced by the number of countries in a cluster. These changes ensure the independence of the "superobservations" entering the formula (Gould et al. (2005)).

Appendix 2 : Robustness Check of the Optimal Model

In this section we propose a sensitivity analysis of the performance of our best EWS for currency crises.

Methodology

We decide to consider both a change in the dating currency crises method and in the crisis definition. Therefore, instead of using the KLR modified pressure index (KLRm), we will use Zhang original index (Lestano and Jacobs(2004)) :

$$Crisis_{n,t} = \begin{cases} 1, & \text{if } \begin{cases} \frac{\Delta e_{n,t}}{e_{n,t}} > \beta_1 \sigma'_{e_{n,t}} + \mu_{e_{n,t}} \quad \text{or} \\ \frac{\Delta r_{n,t}}{r_{n,t}} < \beta_2 \sigma'_{r_{n,t}} + \mu_{r_{n,t}} \end{cases} \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

where $\sigma'_{e,t}$ is the standard deviation of $(\Delta e_{n,t}/e_{n,t})$ in the sample of (t-36, t-1), and $\sigma'_{r,t}$ is the standard deviation of $(\Delta r_{n,t}/r_{n,t})$ in the sample of (t-36, t-1). The thresholds are set to $\beta_1 = 3$ and $\beta_2 = -3$. We can notice that unlike the KLRm index, the interest rates are excluded from the Zhang index and the thresholds used are time-varying for each component.

Regarding the crisis definition, this time we will define $C12$ as taking the value of 1 if there is a crisis in the following 12 months, and the value of 0 if there is no crisis in the next 12 months :

$$C12_{n,t} = \begin{cases} 1, & \text{if } \sum_{j=1}^{12} Crisis_{n,t+j} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

Empirical Results

For the first test of robustness, Zhang's currency crisis (ZCC) Index is used instead of KLRm. For the second robustness check, KLRm pressure is maintained, but instead of considering a horizon of 24 months, we fix it to 12 ($C12$). The same evaluation procedure as in the main text is followed. The estimation of the models looks like the one of the original logit model with market expectation variables in terms of global goodness of fit, parameter's sign and significance. Still, there are some differences, especially when considering ZCC index : Significant variables are less numerous and different compared to the original model , i.e. M2 to reserves, Growth of stock market price index, a more intense variability of the parameter's signs from one cluster to another, etc.. On the contrary, the results obtained with $C12$ is identical to the one derived from the initial logit model with market expectation variables.

Moreover, the spread variable is significant for all the Asian clusters, like in the case of the optimal model, whereas in the ZCC experiment, this variable is significant only for the first Asian cluster (Korea, Malaysia and Taiwan).

The optimal cut-off value for each country is reported in Table 9. It appears it is almost always given by the accuracy criteria. Besides, the cut-off values are all inferior to 0.35 (except for Uruguay). The sensitivity and specificity values of the three models are quite similar, and for some countries like Mexico, Peru or respectively Argentina, Korea, Malaysia, Peru, Philippines, Thailand, Uruguay, Venezuela they are even better than those obtained with the original model.

The idea that these two models aiming to check the robustness of our optimal specification are similar and even slightly better (for some countries) than the initial one, is supported by the performance assessment criteria used (see Tables 10 and 11). For countries like Malaysia, Philippines, Taiwan, Thailand, Uruguay and Venezuela, the AUC and the Pietra index for the two "robust" models are constantly higher than for our optimal model, and the Bayesian error rate, the QPS and the LPS are lower. When considering only the C12 method, the following countries can be added to the same category : Argentina, Korea, and Peru.

It turns out that our optimal specification is robust.

Appendix 3 : Kapetanios' (2003) Recursive Procedure

Considering a panel logit data model $\Pr(y_{nt} = 1) = \frac{\exp(\beta' x + f_n)}{1+\exp(\beta' x + f_n)}$ $\forall n \in \Omega_h$, $n \in \{1, ..., N\}$, $t \in \{1, ..., T\}$, we can write the null hypothesis : $\beta_n = \beta$ $\forall n$. Therefore, in terms of estimation we have : $\hat{\beta}_n = \hat{\beta}$, $\forall n$, where $\hat{\beta}_n$ is the $(q, 1)$ vector of the parameters estimates resulted from the logit model for the n^{th} country, while $\hat{\beta}$ is the $(q, 1)$ vector of the parameters estimates obtained from the panel logit model, knowing that q represents the number of explicative variables in the model. Thus, an Hausman type statistic can be written in the following form :

$$S_{T,n} = (\hat{\beta}_n - \hat{\beta})' V(\hat{\beta}_n - \hat{\beta})^{-1}(\hat{\beta}_n - \hat{\beta}). \tag{37}$$

Consequently, we note $S_T^s = sup(S_{T,n})$ the statistic testing our null hypothesis. Since the accuracy of the time series estimators $\hat{\beta}_n$ improves as T rises, for a low T the estimators are not accurate and, on top of that, the variance of the estimators soars. In this case the statistic $S_{T,n}$ tends not to reject the null hypothesis as often as it should, leading to erroneous results. Actually, we face a small sample problem, because the $S_{T,n}$ statistics are sometimes negative, so the asymptotical assumption that $S_{T,n}$ follows a χ_q^2 distribution fails.

The 5% critical values of S_T^s for different n and q is supplied by the author. If the test does not reject the null hypothesis, the dataset is ready to use in a panel framework, but if the null hypothesis is rejected, the country with the highest statistic ($S_{T,n}$ has to be dropped and the procedure must start over again with the remaining countries. For more technical details see Kapetanios (2003).

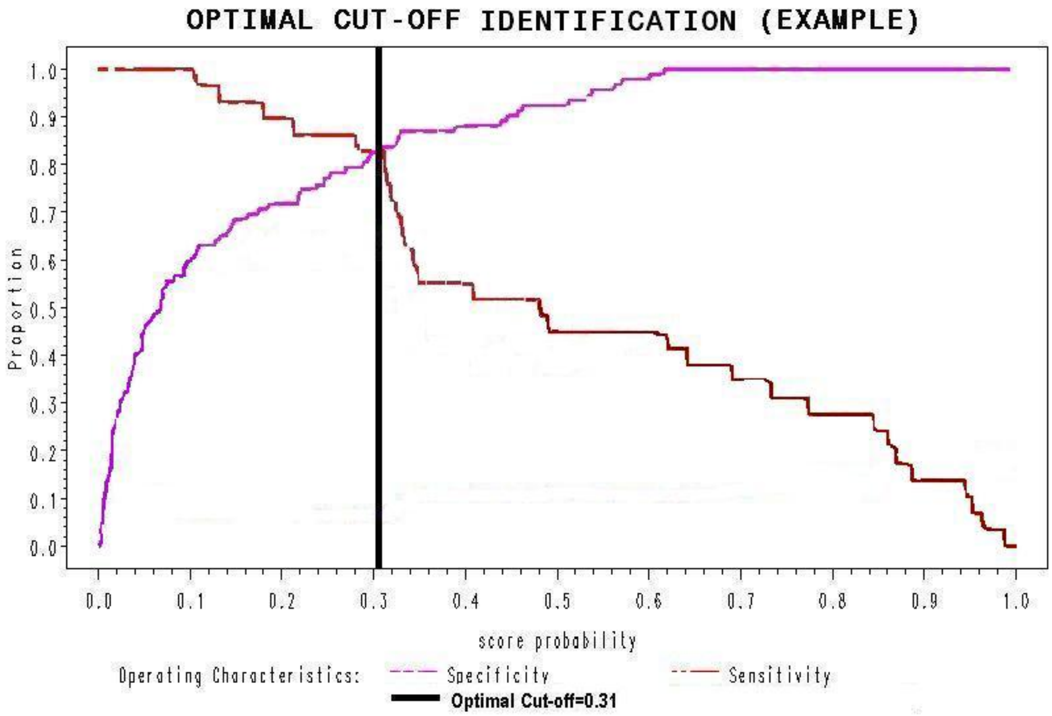


FIG. 1 – Optimal Cut-off determination

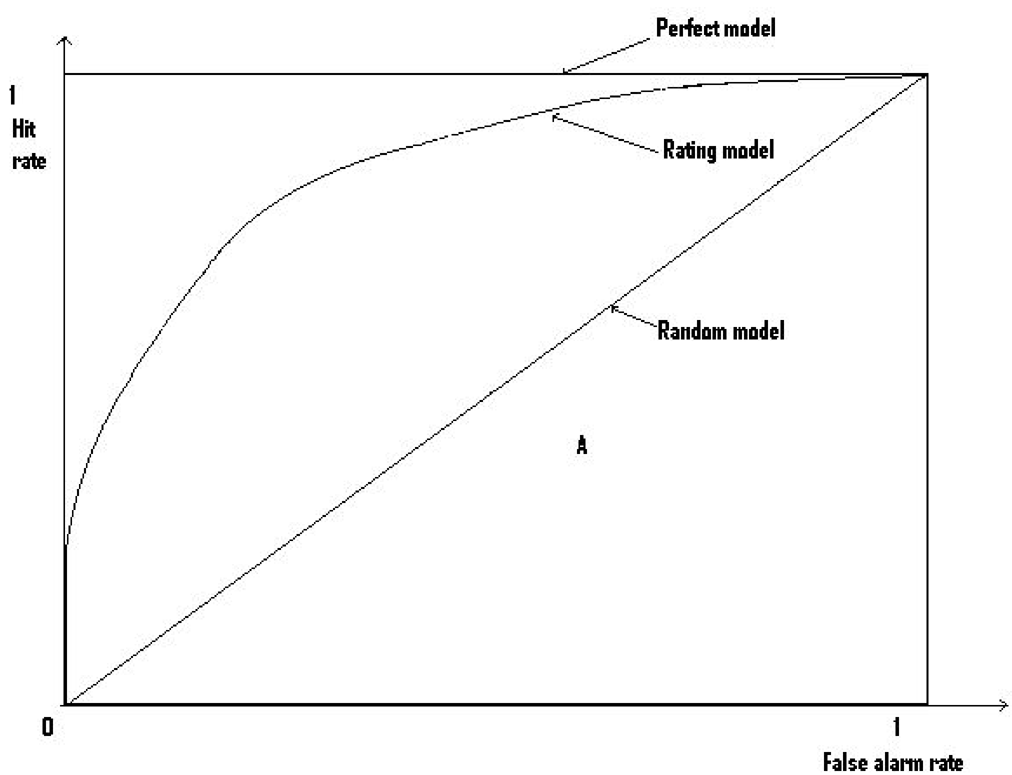
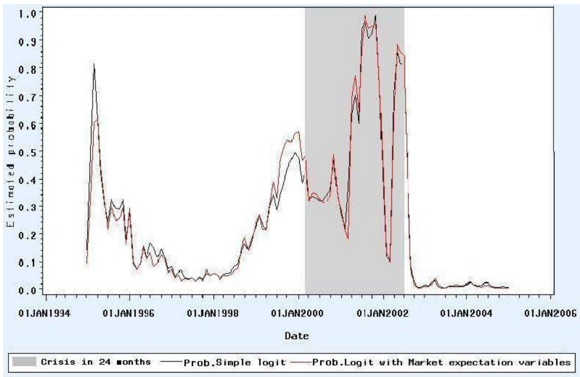
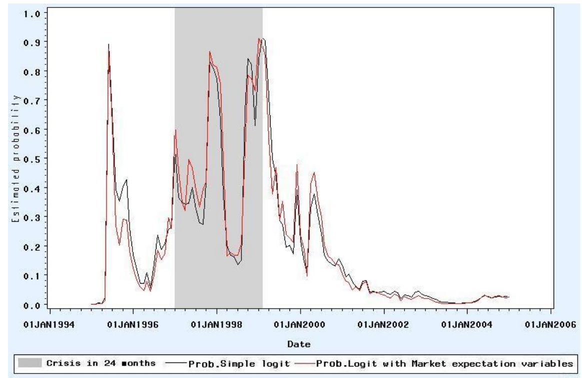


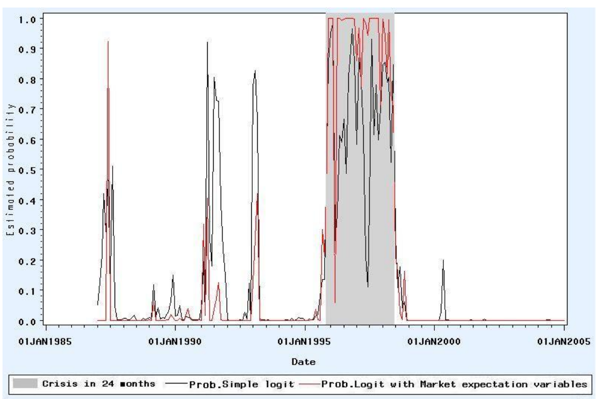
FIG. 2 – The ROC curve



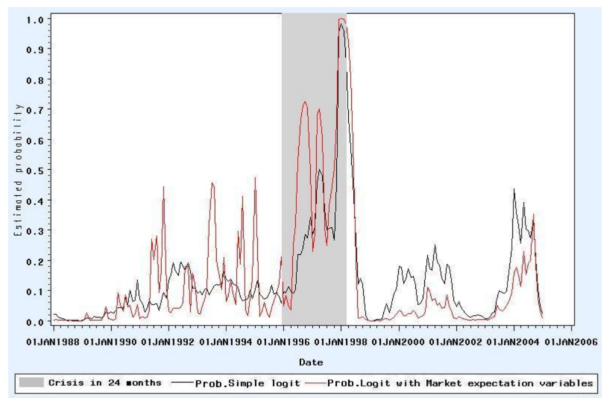
(a) Argentina



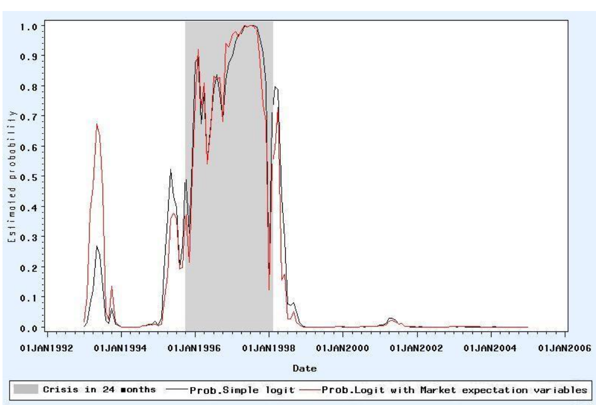
(b) Brazil



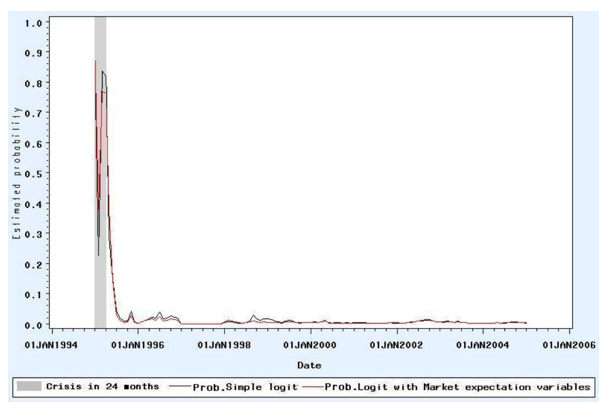
(c) Indonesia



(d) Korea

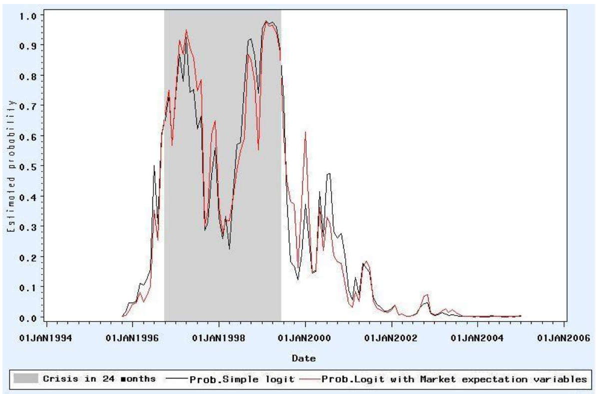


(e) Malaysia

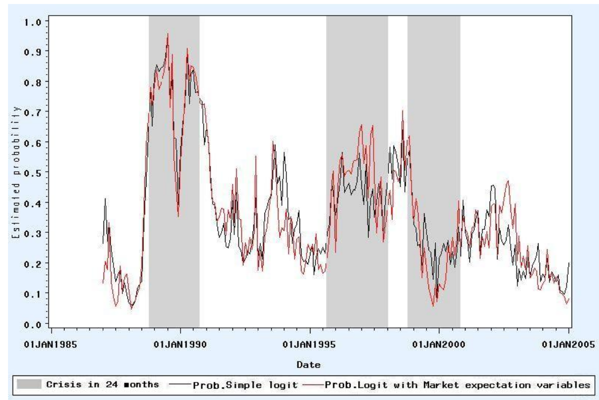


(f) Mexico

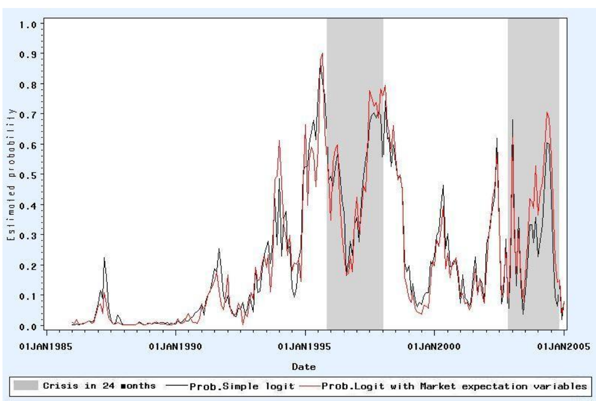
FIG. 3 – Predicted probability of crisis



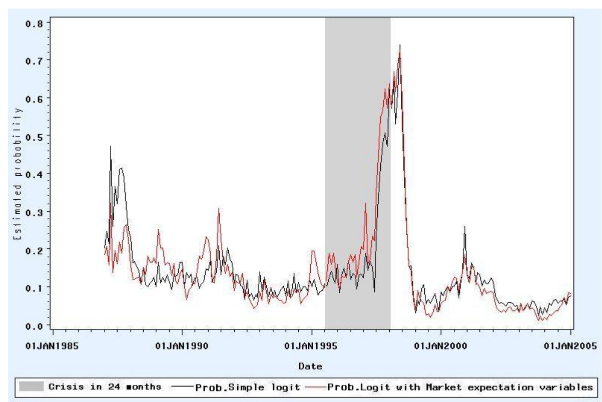
(a) Peru



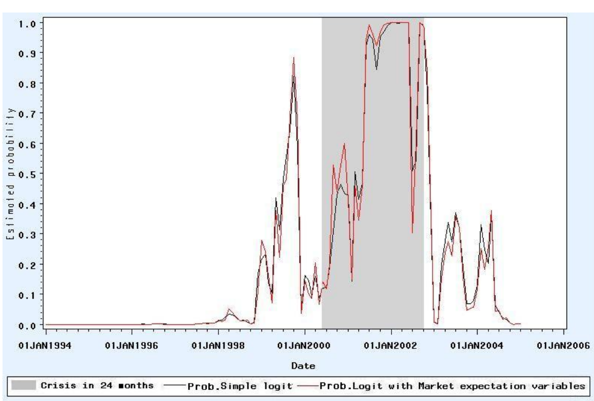
(b) Philippines



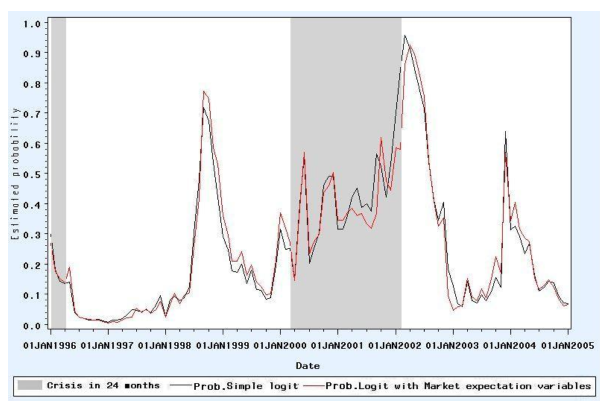
(c) Taiwan



(d) Thailand



(e) Uruguay



(f) Venezuela

FIG. 4 – Predicted probability of crisis (continued)

Tab. 2 – Estimation results of a logit model with market expectation variables

	cluster 1		cluster 2		cluster 3	
Indicator	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Growth of international reserves	-1.426	2.374	-0.817	3.452	-9.190	1.327
Growth of mm2 multiplier	-1.136	4.411	8.192	0.359	-3.678	4.565
Growth of domestic credit over GDP	-2.787	3.770	2.594	5.469	4.756	6.983
Real interest rate	18.72	6.125	41.04	21.37	-4.248	17.56
First difference of lending rate over deposit rate	0.631	0.284	-0.374	0.660	-0.868	0.526
Growth of real bank deposits	-2.809	0.822	-2.699	5.698	4.572	0.260
M2 to reserves	3.454	1.704	-20.35	4.448	25.25	4.002
Growth of M2 to reserves	-.739	0.562	-10.02	1.796	-11.16	1.369
Growth of industrial production	1.769	1.218	-1.813	5.903	5.198	4.655
Growth of stock market price index	0.549	1.068	2.755	0.403	0.376	1.240
Yield Spread	0.299	0.358	-0.234	0.439	-1.545	0.343
Likelihood Ratio	140.644		147.501		234.707	
Log pseudolikelihood	-129.995		-58.3494		-150.614	
Pseudo R2	0.35110		0.55830		0.43790	
AIC	281.991		138.699		323.228	
SC	327.718		177.213		371.202	

Tab. 3 – Estimation results of a logit model with market expectation variables (continued)

	cluster 4		cluster 5	
Indicator	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Growth of international reserves	-0.287	2.308	-12.23	14.06
Growth of mm2 multiplier	1.822	8.984	23.50	11.82
Growth of domestic credit over GDP	0.527	2.389	7.981	5.727
Real interest rate	-1.731	6.259	11.38	20.06
First difference of lending rate over deposit rate	-0.592	0.465	4.410	10.50
Growth of real bank deposits	0.165	1.073	38.47	10.63
M2 to reserves	4882	354.0	$6.14 * 10^6$	$4.6 * 10^6$
Growth of M2 to reserves	-2.107	3.149	-40.04	31.90
Growth of industrial production	-4.242	0.336	-41.16	21.14
Growth of stock market price index	-1.155	2.388	4.716	4.308
Yield Spread	-0.221	0.018	-4.157	1.614
Likelihood Ratio	67.5548		159.626	
Log pseudolikelihood	-191.503		-12.6918	
Pseudo R2	0.14990		0.86280	
AIC	405.006		49.3840	
SC	449.809		89.9420	

Tab. 4 – Optimal cut-off identification in a Logit model with market expectation variables

	Accuracy measures			Sensitivity-Specificity graphic			KLR
Country	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off
Argentina	**0.300**	82.76	82.61	**0.300**	82.76	82.61	0.620
Brazil	**0.160**	100.0	69.47	0.250	80.77	81.05	0.880
Indonesia	**0.200**	96.97	96.20	0.210	96.97	96.20	0.930
Korea	**0.206**	85.71	90.96	0.170	85.71	85.31	0.930
Malaysia	**0.380**	93.10	93.97	0.370	93.10	93.10	0.730
Mexico	**0.379**	100.0	99.15	**0.379**	100.0	99.15	0.390
Peru	**0.260**	100.0	82.72	0.350	87.10	86.42	0.940
Philippines	0.400	64.10	82.01	**0.346**	67.95	68.35	0.730
Taiwan	**0.160**	94.12	65.17	0.228	76.47	76.97	0.670
Thailand	**0.120**	90.32	61.29	0.150	74.19	74.19	0.321
Uruguay	**0.119**	93.33	75.73	0.230	83.33	83.50	0.900
Venezuela	**0.225**	85.71	67.90	0.280	71.43	71.60	0.330

Note : For each country we choose the optimal cut-off from the values obtained by using two different methods (accuracy measures and sensitivity-specificity graphic) as being the one that maximizes both sensitivity and specificity, usually giving more weight to the correct identification of crisis periods (sensitivity). The selected cut-off values are in bold. For comparison reasons, we also present the KLR cut-off which is obtained by minimising NSR (reported as KLR).

Tab. 5 – Optimal cut-off identification in a Markov model with market expectation variables and *spread* switching

	Accuracy measures			Sensitivity-Specificity graphic			KLR
Country	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off
Argentina	**0.990**	51.72	36.96	**0.990**	51.72	36.96	0.987
Brazil	**0.800**	65.38	73.68	0.610	65.38	65.26	0.964
Indonesia	**0.522**	57.58	56.52	**0.522**	57.58	56.52	0.828
Korea	**0.500**	71.43	71.19	**0.500**	71.43	71.19	0.947
Malaysia	**0.967**	100.0	58.62	0.988	68.97	71.55	0.989
Mexico	**0.923**	50.00	51.28	**0.923**	50.00	51.28	0.922
Peru	0.860	41.94	59.26	**0.85**	100.0	<0.001	0.982
Philippines	0.928	60.26	66.19	**0.910**	61.54	61.87	0.926
Taiwan	**0.952**	64.71	42.70	0.980	50.98	50.56	0.953
Thailand	**0.875**	74.19	23.66	0.890	38.71	34.41	0.989
Uruguay	**0.520**	100.0	<0.001	0.527	33.33	26.21	0.524
Venezuela	**0.988**	82.14	29.63	0.989	17.86	58.02	0.988

Note : For each country we choose the optimal cut-off from the values obtained by using two different methods (accuracy measures and sensitivity-specificity graphic) as being the one that maximizes both sensitivity and specificity, usually giving more weight to the correct identification of crisis periods (sensitivity). The selected cut-off values are in bold. For comparison reasons, we also present the Berg-Patillo cut-off which is obtained by minimising the noise-to-signal ratio.

Tab. 6 – Evaluation criteria for a logit model with market expectation variables and a Markov model with market expectation variables and *spread* switching

Country	Model	AUC	Kuiper score	Pietra index	Bayesian error rate	QPS	LPS
Argentina	Logit	0.898	65.37	0.235	0.132	0.215	-0.325
	Markov	0.625	-11.32	<0.001	0.239	1.492	3.558
Brazil	Logit	0.907	69.47	0.249	0.132	0.202	-0.311
	Markov	0.710	39.06	0.142	0.215	0.765	1.057
Indonesia	Logit	0.996	93.17	0.330	0.0138	0.034	-0.058
	Markov	0.685	14.10	0.114	0.129	0.567	0.607
Korea	Logit	0.920	76.67	0.273	0.0780	0.135	-0.228
	Markov	0.748	42.62	0.164	0.136	0.682	0.909
Malaysia	Logit	0.985	87.07	0.311	0.048	0.083	-0.131
	Markov	0.809	58.62	0.207	0.200	1.184	2.165
Mexico	Logit	0.998	99.15	0.350	0.008	0.011	-0.023
	Markov	0.564	1.280	0.081	0.033	1.658	2.526
Peru	Logit	0.947	82.72	0.292	0.107	0.166	-0.266
	Markov	0.529	<0.001	0.027	0.277	1.192	1.939
Philippines	Logit	0.739	36.30	0.163	0.235	0.368	-0.558
	Markov	0.582	23.41	0.093	0.359	1.080	1.756
Taiwan	Logit	0.837	59.29	0.211	0.196	0.270	-0.399
	Markov	0.511	7.410	0.028	0.223	1.467	3.085
Thailand	Logit	0.811	51.61	0.192	0.138	0.218	-0.348
	Markov	0.592	-2.150	0.038	0.143	1.511	2.710
Uruguay	Logit	0.939	69.06	0.257	0.105	0.165	-0.246
	Markov	0.725	<0.001	<0.001	0.225	1.186	2.023
Venezuela	Logit	0.777	53.61	0.189	0.257	0.370	-0.530
	Markov	0.511	11.77	0.042	0.257	1.454	3.361

Note : The AUC criteria takes values between 0.5 and 1, 1 being the perfect model. Kuiper's score should have positive values if the model identifies well the crisis periods. Pietra index takes values from -0.354 to 0.354, the higher its level, the better the model. Bayesian error rate takes values between 0 and 1, 0 corresponding to the perfect model. QPS ranges from 0 to 2, 0 being perfect accuracy, while LPS ranges from 0 to ∞, 0 being perfect accuracy.

Tab. 7 – Comparison tests of Simple logit and Market expectation logit models

	ROC		Clark-West	
Country	test statistic	p-value	test statistic	pvalue
Argentina	0.0301	0.8622	0.1372	0.4454
Brazil	**5.7105**	0.0169	**3.4901**	0.0002
Indonesia	**7.9917**	0.0047	**4.4332**	0.0000
Korea	**4.5357**	0.0332	**3.7746**	0.0001
Malaysia	0.3859	0.5345	0.3288	0.3711
Mexico	<0.001	1.0000	0.6869	0.2460
Peru	0.0028	0.9577	**2.1634**	0.0153
Philippines	0.8738	0.3499	0.8709	0.1919
Taiwan	**10.475**	0.0012	**3.5603**	0.0002
Thailand	**6.9801**	0.0082	**4.5964**	0.0000
Uruguay	0.7443	0.3883	0.6656	0.2528
Venezuela	**6.6647**	0.0098	-2.0740	0.9810

Note : The null hypothesis of the ROC test is the equality of areas under the ROC curve, while the alternative hypothesis is the statistical difference between the two areas. Its statistic follows a normal distribution whose critical values are $\pm$ 1.96 (5%). The null hypothesis of the Clark-West test is the equality of predictive performance of the two models. The alternative indicates that the non-constraint model (the bigger one) is better than the other one. Under the null hypothesis, the MSPE-adj statistic follows a normal distribution with a critical unilateral value of 1.645(5%). Bold entries indicate significance at the 5% level.

TAB. 8 – Comparison tests of market expectation logit and market expectation *spread* switching markov models

	ROC		Diebold-Mariano	
Country	test statistic	p-value	test statistic	pvalue
Argentina	**62.678**	<0.001	**12.965**	<0.001
Brazil	**9.7859**	0.0018	**8.783**	<0.001
Indonesia	**46.529**	<0.001	**29.244**	<0.001
Korea	**9.8754**	0.0017	**12.207**	<0.001
Malaysia	**21.455**	<0.001	**17.066**	<0.001
Mexico	**17.829**	<0.001	**50.850**	<0.001
Peru	**45.942**	<0.001	**12.164**	<0.001
Philippines	**7.4266**	0.0064	**9.7129**	<0.001
Taiwan	**34.195**	<0.001	**16.591**	<0.001
Thailand	**45.902**	<0.001	**18.281**	<0.001
Uruguay	**125.00**	<0.001	**12.877**	<0.001
Venezuela	**17.351**	<0.001	**9.4665**	<0.001

Note : The null hypothesis of the ROC test is the equality of area under the ROC curve, while the alternative hypothesis is the statistical difference between the two areas. Its statistic follows a normal distribution whose critical values are ± 1.96 (5%). The null hypothesis of the Diebold-Mariano test is the equality of predictive performance of the two models. The alternative indicates that the first model is better than the other one. Under the null hypothesis, the teste statistic follows a normal distribution. Bold entries indicate significance at the 5% level.

TAB. 9 – Optimal cut-off identification in a logit model with market expectation variables

	Zhang method						C12 method					
	Accuracy measuress			Sensitivity-Specificity graphic			Accuracy measuress			Sensitivity-Specificity graphic		
Country	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity
Argentina	**0.080**	97.14	58.00	0.550	65.71	70.00	**0.180**	94.12	90.38	0.195	88.24	90.38
Brazil	**0.330**	100.0	68.09	0.438	76.32	78.72	**0.060**	92.86	61.68	0.080	71.43	71.03
Indonesia	**0.200**	100.0	100.0	**0.200**	100.0	1000	**0.3500**	100.0	100.0	**0.3500**	100.0	100.0
Korea	**0.150**	85.19	78.87	0.170	81.48	81.69	**0.110**	93.75	89.95	0.150	93.75	93.65
Malaysia	**0.030**	100.0	96.47	0.038	95.83	96.47	**0.080**	100.0	90.63	0.120	94.12	93.75
Mexico	0.320	79.17	78.69	**0.340**	79.17	80.33	**0.225**	100.0	98.29	**0.225**	100.0	98.29
Peru	**0.020**	75.00	73.61	**0.020**	75.00	73.61	0.110	94.74	91.40	**0.260**	94.74	94.62
Philippines	**0.220**	82.05	85.21	0.200	82.05	82.39	**0.180**	71.43	70.29	**0.180**	71.43	70.29
Taiwan	**0.200**	100.0	95.81	0.220	96.15	95.81	**0.110**	85.19	80.69	0.120	81.48	81.68
Thailand	**0.200**	100.0	64.8	0.350	78.57	78.4	**0.120**	94.74	83.84	0.160	89.47	88.89
Uruguay	0.685	26.79	95.2	**0.700**	91.53	89.47	**0.300**	94.44	97.39	0.120	94.44	93.91
Venezuela	**0.240**	92.86	68.8	0.450	79.31	81.82	**0.060**	100.0	58.06	0.140	75.00	75.27

Note : For each country we choose the optimal cut-off from the values obtained by using two different methods (accuracy measures and sensitivity-specificity graphic) as being the one that maximizes both sensitivity and specificity, usually giving more weight to the correct identification of crisis periods (sensitivity). The selected cut-off values are in bold.

Tab. 10 – Evaluation criteria for a logit model with market expectation variables (ZCC)

Country	AUC	Kuiper score	Pietra index	Bayesian error rate	QPS	LPS
Argentina	0.797	55.14	0.195	0.259	0.389	-0.548
Brazil	0.895	68.09	0.241	0.176	0.274	-0.419
Indonesia	1.000	100.0	0.353	<0.001	0.001	-0.005
Korea	0.887	64.06	0.229	0.094	0.169	-0.282
Malaysia	0.999	96.47	0.341	0.009	0.025	-0.042
Mexico	0.878	59.5	0.213	0.188	0.279	-0.432
Peru	0.892	48.61	0.245	0.052	0.090	-0.154
Philippines	0.914	67.26	0.240	0.105	0.199	-0.339
Taiwan	0.994	95.81	0.339	0.031	0.046	-0.075
Thailand	0.88	64.80	0.229	0.177	0.267	-0.405
Uruguay	0.968	81.00	0.298	0.072	0.129	-0.214
Venezuela	0.897	61.66	0.257	0.151	0.256	-0.395

Note : The AUC criteria takes values between 0.5 and 1, 1 being the perfect model. Kuiper's score should have positive values if the model identifies well the crisis periods. Pietra index takes values from -0.354 to 0.354, the higher its level, the better the model. Bayesian error rate takes values between 0 and 1, 0 corresponding to the perfect model. QPS ranges from 0 to 2, 0 being perfect accuracy, while LPS ranges from 0 to ∞, 0 being perfect accuracy.

Tab. 11 – Evaluation criteria for a logit model with market expectation variables (C12)

Country	AUC	Kuiper score	Pietra index	Bayesian error rate	QPS	LPS
Argentina	0.956	84.50	0.299	0.049	0.095	-0.171
Brazil	0.861	54.54	0.203	0.083	0.153	-0.257
Indonesia	1.000	100.0	0.354	<0.001	0.002	-0.005
Korea	0.973	83.70	0.316	0.044	0.073	-0.113
Malaysia	0.992	90.63	0.327	0.021	0.045	-0.077
Mexico	0.996	98.29	0.347	0.008	0.017	-0.028
Peru	0.987	89.36	0.323	0.036	0.068	-0.115
Philippines	0.755	41.72	0.149	0.152	0.264	-0.425
Taiwan	0.898	65.88	0.233	0.087	0.142	-0.234
Thailand	0.945	78.58	0.279	0.059	0.105	-0.177
Uruguay	0.989	91.83	0.325	0.023	0.045	-0.092
Venezuela	0.823	58.06	0.205	0.138	0.224	-0.338

Note : The AUC criteria takes values between 0.5 and 1, 1 being the perfect model. Kuiper's score should have positive values if the model identifies well the crisis periods. Pietra index takes values from -0.354 to 0.354, the higher its level, the better the model. Bayesian error rate takes values between 0 and 1, 0 corresponding to the perfect model. QPS ranges from 0 to 2, 0 being perfect accuracy, while LPS ranges from 0 to ∞, 0 being perfect accuracy.