Towards the optimal design of an Early Warning System for sovereign debt crises*

Ana-Maria Fuertes and Elena Kalotychou†
Faculty of Finance, Cass Business School, City University London

10 December 2004

Abstract

This paper considers three different approaches — K-means clustering and logit regressions based on macrovariables or credit ratings — to develop an Early Warning System for sovereign default. The data pertain to 75 emerging and developing countries. The optimal choice of key elements, such as the logit cut-off probability and the number of clusters, is shown to depend on the decision-maker's preferences. The latter are encapsulated in a loss function and a degree of risk aversion parameter. The out-of-sample forecast evaluation suggests that the classifiers have different strengths in terms of missed defaults and false alarms and that their forecast ranking is unstable. We show that forecast combining pays and that the weighting scheme for this purpose should also be chosen optimally to account for the decision-maker's preferences.

Keywords: Debt crises; K-means clustering; logistic regression; bank internal ratings; loss function; forecast combination.
JEL Classification: C15; C22; C52

*We appreciate the comments of Roy Batchelor, Christopher Baum, Chris Brooks, Jerry Coakley, Ron Smith and participants at the 10th Meeting of the *Society for Computational Economics* on Computing in Economics and Finance, University of Amsterdam and seminar participants at Cass Business School. We are responsible for any errors.

†Correspondence: Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, e-mail: a.fuertes@city.ac.uk, Tel: +44 (0)20 7040 0186.

1 Introduction

The financial turmoil in emerging and developing markets during the last decade has stressed the need for accurate country risk assessment. A number of studies have focused on the so-called twin crises, namely, banking and currency crises (Frankel an Rose, 1996; Berg and Pattillo, 1999; Kaminsky and Reinhart, 1999; Kumar et al., 2003). As more countries are moving toward flexible exchange rates the latter are becoming less frequent events. But sovereign debt crises remain a matter of concern for international financial markets and economic policymakers.

The sovereign default literature is very prolific. Most studies have focused on identifying the main determinants of default among fundamentals of the domestic economy and indicators of the international business-cycle and market sentiment. For this purpose, different classification techniques have been used. However, scant attention has been paid to forecasting issues and to the optimal design of an Early Warning System (EWS) tailored to the decision-maker's preferences.

Several studies have applied *linear discriminant analysis* which assumes multivariate normal regressors with equal covariance matrices in the default and non-default states (Frank and Cline, 1977; Taffler and Abassi, 1984). These assumptions have been shown to be rather strong. The most recent research is based on nonlinear panel logit/probit models (Peter, 2002). Non-parametric classification techniques such as *clustering* and *recursive tree* analysis, albeit popular in other areas, have received little attention in this literature. One exception is Manasse et al. (2003) who apply both a logit and a recursive tree analysis. Moreover, most of the models thus developed are based on arbitrary choices for the *cut-off* probability and *warning horizon* or crisis window. These ad hoc choices may not necessarily be optimal for the problem at hand. For instance, a low cut-off rate and a long warning horizon may be better choices for a highly risk averse (towards default) decision maker since they lead to more default signals and vice versa.

The credit ratings provided by leading agencies and bankers have also been found to contain

predictive power regarding sovereign debt crises and to Granger-cause the spreads of sovereign bonds (Reinhart, 2002; Rojas-Suárez, 2001; Larrain et al., 1997). Moreover, the New Basel Accord allows banks to use internal ratings for calculating capital requirements. The Institutional Investor ratings can be regarded as consensus internal ratings from major international banks. The upshot is that it is unclear which method and information set one should adopt to develop an EWS of sovereign default. This indirectly stresses the potential importance of forecast combining, an issue that has received scant attention in this literature.

The contribution of this paper is twofold. First, it provides a framework for the optimal design of an optimal EWS for sovereign default. For this purpose, it implements three forecasting tools: *i*) logit based on macrovariables, *ii*) K-means clustering of macrovariables and *iii*) logit based on Institutional Investors' credit ratings (LOGIT-R). The latter two classifiers have not been utilized in the present context as yet. The sample pertains to 75 emerging and developing economies over 1983-2000. We show how the loss function and degree of risk aversion of the decision-maker can be accounted for in order to optimally choose key elements of an EWS such as the logit cut-off rate and the number of clusters. The calibration of the classifiers is conducted in-sample recursively over a 12-year rolling window. The assessment of their forecast ability over a 5-year holdout period focuses on the anticipation of default *entry* events.

Second, the paper delves into several forecast combining issues. It assesses the relative strengths of the classifiers for different decision-makers and the stability of the ranking over the holdout years. In order to adequately gauge the gains from forecast combining, it is shown that the choice of weighting scheme should also be tailored to the loss function and degree of risk aversion.

Section 2 outlines the background literature. Section 3 describes the methodology and Section 4 introduces the data. Section 5 then illustrates empirical issues regarding the optimal calibration of classifiers while our forecast combining analysis is presented in Section 6 before concluding.

2 Elements in the design of an optimal EWS

The goal of an EWS is to issue signals of pending debt repayment difficulties. Hence, the variable of interest takes a value of one at year t if a default occurs any time within a $[t, t+h]$ window

$$y_{it} = \begin{cases} 1 & \text{if } d_{i,t+k} = 1 \text{ at any } k = 0, 1, ..., h-1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and the classification problem at hand is formalized as $y_{it} = f(\mathbf{x}_{i,t-1})$ where $\mathbf{x}_{i,t-1}$ represents the available predictors at $t-1$. The forward-looking variable y_{it} is called the EWS indicator.

2.1 Warning horizon for a crisis signal

It is reasonable to expect that signs of economic deterioration will emerge years before a debt crisis occurs. The warning horizon is the time interval within which the EWS should anticipate the occurrence of a crisis. If the warning horizon chosen is, say, $h = 3$ years then the forecast $\hat{y}_{i,t+1} = 1$ indicates that a debt crisis will occur sometime during $[t+1, t+3]$. Choosing h requires a trade-off. The longer h is the less missed defaults but the more false alarms and vice versa.

Currency crises studies typically set $h = 24$ months (Berg and Pattillo 1999; Kamin, 1999; Kumar et al., 2003). Bussière and Fratzscher (2002) find empirically the optimal banking-crisis warning horizon according to a specific loss function. Burkart and Coudert (2002) carry out a sensitivity analysis of the Type I and II error rates associated to currency-crisis warning horizons from 1 to 4 quarters. However, the choice of warning horizon has received scant attention in the sovereign default literature. Most studies arbitrarily set $h = 1$ year. Sommerville and Taffler (1995) argue that this may be too short for bankers' purposes. Peter (2002) and Oka (2003) arbitrarily set $h = 3$ years in modeling default on external creditors and arrears to the IMF, respectively.

2.2 Cut-off probability rate

In order to assess the adequacy of an EWS, the probability forecasts are usually transformed into event forecasts and compared with the EWS indicator y_{it}. For this purpose, the decision-

maker has to adopt a cut-off or threshold probability λ. Mascarenhas and Sand (1989) show that the overall error rate of a sovereign default predictor based on discriminant analysis varies with λ.

In the financial crises literature, λ is most often arbitrarily set at 0.5, 0.25 or 0.10 (Kumar et al., 2003; Peter, 2002; Berg and Pattillo, 1999; Frankel and Rose, 1996), fixed at the in-sample frequency of defaults (Demirguc-Kunt and Detragiache, 1999; Detragiache and Spilimbergo, 2001; Manasse et al., 2003) or at the level that balances the Type I and II errors (Burkart and Coudert, 2002). Only three sovereign default studies account for the loss function in choosing this parameter (Taffler and Abassi, 1984; Sommerville and Taffler, 1995; Oka, 2003).

2.3 Decision-maker's loss function

The loss function facilitates the expected cost of mispredicting. On this basis, the forecaster can optimally calibrate several parameters of an EWS such as the cut-off rate and the warning horizon. The literature typically treats the Type II error (false alarms) as less worrisome than the Type I error (missed crises), mainly for two reasons. First, the costs of missing investment opportunities or those of taking pre-emptive policy measures in the case of a false warning are often less severe than the losses, reflected in the lender's balance sheet, or the welfare cost of an unanticipated default. Second, false alarms are not always 'errors' as such in that they may not stem from predictive failure of the model but simply reflect that, although economic vulnerabilities might have been severe, appropriate policy actions were taken and a debt crisis was avoided.

Suppose that an EWS is developed using the warning horizon h and the cut-off λ. On the basis of its forecasts, different error measures can be computed. Let $E_0(\lambda, h)$ and $E_1(\lambda, h)$ denote the number of false warnings ($\hat{y}_{it} = 1, y_{it} = 0$) and missed defaults ($\hat{y}_{it} = 0, y_{it} = 1$), respectively. Let $C_0(h)$ and $C_1(h)$ denote the total number of tranquil ($y_{it} = 0$) and debt crisis ($y_{it} = 1$) cases, respectively. The available sample has $C = C_0 + C_1 = NT$ cases where T denotes time periods and N denotes countries. The Type I error probability (P_I hereafter) is estimated by the percentage of

missed defaults, $E_1(\lambda,h)/C_1(h)$. The Type II error probability (P_{II}) gives the likelihood of a false alarm and it can be estimated as $E_0(\lambda,h)/C_0(h)$. Finally, let θ denote the degree of risk aversion (toward missing a crisis) of the decision-maker. Below we outline three loss functions that have been widely used in the broad financial crisis literature.

Kaminsky et al. (1998) introduce the *noise-to-signal* loss (NS) for currency crisis forecasts

$$NS(\lambda,h) = \frac{P_{II}(\lambda,h)}{1 - P_I(\lambda,h)}, \quad NS \in [0,1] \tag{2}$$

defined as the ratio of the probability of a false alarm over the probability of a correct crisis warning. Other currency crises studies adopt this loss function (Berg and Pattillo, 1999; Burkart and Coudert, 2002). But it has not yet been used in the sovereign default literature. It can be estimated by $\widehat{NS}(\lambda,h) = c(h) \times \frac{E_0(\lambda,h)}{C_1(h) - E_1(\lambda,h)}$ where $c(h) \equiv C_1(h)/C_0(h)$. Hence, optimizing λ according to (2) amounts to minimizing the ratio of false alarms to correct alarms.

Another typical loss function, that we call *investor's loss* (IL), is defined as

$$IL(\theta,\lambda,h) = \theta P_I(\lambda,h) + (1-\theta)P_{II}(\lambda,h), \quad IL \in [0,1] \tag{3}$$

and it can be estimated by $\widehat{IL}(\theta,\lambda,h) = \theta \frac{E_1(\lambda,h)}{C_1(h)} + (1-\theta)\frac{E_0(\lambda,h)}{C_0(h)}$. The cost attached to a missed default relative to that of a false alarm is captured by the risk-aversion parameter θ, e.g. $\theta = 0.8$ reflects that the cost ratio for the decision-maker is 4 to 1. Equation (3) represents a family of loss functions parameterized by θ. Oka (2003) and Burkart and Coudert (2002) adopt it for $\theta = 0.5$.

Other studies have employed what we call the *policymaker's loss* (PL) function defined as

$$PL(\theta,\lambda,h) = \theta P_I(\lambda,h) + (1-\theta)P_W, \quad PL \in [0,1] \tag{4}$$

which is a weighted sum of the probability of missing a default and the probability of issuing an early warning. One implicit assumption is that the latter triggers some policy action or structural reform (e.g. a reduction in public-sector pay and employment) that maybe costly for policymakers in terms, for instance, of social unrest and not being reelected. Thus an optimal EWS for policymakers

should not trigger too many warning signals.[1] In contrast, the loss function (3) presumes that correct alarms have negligible costs and so it is thought to be more representative of investors.[2]

Note that $P_W \equiv \Pr(\hat{y}=1) = \Pr(\hat{y}=1 \cap y=0) + \Pr(\hat{y}=1 \cap y=1)$ so that

$$P_W = \Pr(\hat{y}=1|y=0)\Pr(y=0) + \Pr(\hat{y}=1|y=1)\Pr(y=1)$$

which can be estimated by $P_W = \frac{E_0}{C_0}\frac{C_0}{C} + (1-\frac{E_1}{C1})\frac{C_1}{C}$. Thus we have

$$\widehat{PL}(\theta,\lambda,h) = (1-\theta)\hat{p}\left\{[\frac{\theta}{(1-\theta)\hat{p}} - 1]\hat{P}_I(\lambda,h) + \frac{(1-\hat{p})}{\hat{p}}\hat{P}_{II}(\lambda,h) + 1\right\} \tag{5}$$

which reveals that PL is also a linear function of P_I and P_{II} but, in contrast with IL, the weights reflect not only the risk aversion θ but also the prior probability of default or in-sample default frequency $\hat{p} = \frac{C_1}{C}$. The lower $\hat{p}$ the heavier the penalty for the Type II error ceteris paribus. The PL metric is adopted by Bussière and Fratzscher (2002) and Demirguc-Kunt and Detragiaghe (1999). The former show how the optimal (λ, h) depends on θ whereas the latter focus on λ.

2.4 Forecast combining schemes

Bates and Granger's (1969) seminal paper sets out the concept of forecast combination. It urges that, when alternative forecasts are available, it may pay to combine this information rather than to opt for one of the alternatives. Combination permits the blend of forecasts from a range of sources. Combination has been shown to be effective not only when the forecasts are obtained from widely heterogenous methods but more generally also (Montgomery et al., 1998; Winkler and Makridakis, 1983; Clemen et al. 1995). Instability in the ranking of forecasts provides another rationale for combination (Stock and Watson, 2001; Aiolfi and Timmermann, 2003).

The extensive and continued interest in forecast combination is in large part explained by the wealth of evidence from empirical studies on its merits (see Newbold and Harvey, 2004). Surpris-

[1] The cost of a false alarm is simply the cost of the preventive policy action $(1-\theta)$. Strictly speaking, the cost of a correct alarm is the cost of the preventive action minus the benefit from preventing the default $(1-\theta)-\zeta$.

[2] A correct default warning entails transaction costs for investors. Nevertheless, labelling (3) as IL and (4) as PL is mainly for ease of exposition. This terminology does not imply loss of generality in the ensuing analysis.

ingly, this concept has received scant attention in the sovereign default literature. Sommerville and Taffler (1995) compare judgmental forecasts (bankers' ratings) and parametric forecasts (logit and linear discriminant analysis) based on macrodata but do not assess the merits of combined forecasts. Mascarenhas and Sand (1989) investigate the accuracy of discriminant analysis based on three information sets — credit ratings, macrovariables or both. They find that combining credit rating and macroeconomic forecasts using Gupta and Wilton's (1988) Bayesian odds-matrix method outperforms the individual forecasts. But they do not explore alternative weighting schemes in order to find the 'optimal' combination for the problem at hand. Manasse et al. (2003) compare the forecasts from a logit regression and a non-parametric recursive tree based on macrovariables and find that the latter yields less missed defaults but more false alarms. Using two distinct event weightings (dummies in the logit that incorporate information from the tree nodes and the unanimity principle) it is shown that forecast combining improves accuracy.

3 Competing classifiers

Three forecast methods are considered. First, a pooled logit model for macrovariables (LOGIT-M)

$$\log\left[\frac{p_{it}}{1-p_{it}}\right] = \alpha + \boldsymbol{\beta}^{'}\mathbf{x}_{i,t-1},\ i = 1,...,N, t = 1,...,T \tag{6}$$

that implies the nonlinear relation, $p_{it} = \frac{e^{\alpha+\boldsymbol{\beta}^{'}\mathbf{x}_{i,t-1}}}{1+e^{\alpha+\boldsymbol{\beta}^{'}\mathbf{x}_{i,t-1}}}$, where $p_{it} \equiv \Pr(y_{it} = 1)$ and $\mathbf{x}_{i,t-1}$ is an $s \times 1$ vector. The coefficient, $\beta_j, j = 1,...,s$ estimated by maximum likelihood represents the marginal effect of the macrovariable $x_{it,j}$ on the log-odds ratio $\log\left[\frac{p_{it}}{1-p_{it}}\right]$ ceteris paribus.[3]

Second, the analysis relies also on sovereign credit ratings (z_{it} hereafter) that reflect consensus bankers' judgment. Several studies have found these internal ratings to be correlated with default signals such as GDP per capita, inflation, external debt, economic development and the actual default history (Lee, 1993; Cantor and Packer, 1996). Futhermore, these credit ratings incorporate

[3] The forecasts from the pooled logit model are shown to outperform those from more sophisticated specifications such as random coefficients logit under several loss functions (see Fuertes and Kalotychou, 2004a).

important qualitative information on default risk such as the effects of social, political and cultural conditions. A univariate logit transformation $\log\left[\frac{p_{it}}{1-p_{it}}\right] = \alpha + \beta z_{i,t-1}$ is used to generate default forecasts based on the bankers' ratings. We refer to the latter as LOGIT-R forecasts.[4]

As discussed in Section 2, a warning horizon h is embedded in the definition of y_{it}. In a logit framework, a cut-off probability is required to transform the probability estimates into EWS signals, i.e. $\hat{y}_{it} = 1$ if $\hat{p}_{it} > \lambda$ and $\hat{y}_{it} = 0$ if $\hat{p}_{it} \leq \lambda$. Hence, optimal logit calibration implies finding the (λ, h) combination that is 'best' according to the decision-maker's preferences, namely, her loss function and degree of risk-aversion θ. The following two-step optimization approach is proposed:
1) For each (θ, h) pair, compute the loss associated with the λ candidates so that $\lambda^*_{\theta h} \equiv \min_{\lambda} L(\theta, \lambda, h)$ gives the optimal cut-off rate. This facilitates a set of optimal cut-off rates denoted $\{\lambda^*_{\theta h}\}$.
2) For each θ, calculate the loss associated with $h = \{1, 2, ..., h_{\max}\}$ so that

$$h^*_\theta = \min_h [L(\theta, \lambda^*_{\theta h}, h)] \tag{7}$$

is the optimal warning horizon. This facilitates an optimal horizon and cut-off pair $(h^*_\theta, \lambda^*_{\theta h})$ for each θ. The latter is denoted (h^*, λ^*) for simplicity.

The third classification technique we employ is K-means clustering.[5] The inputs or cases are the NT observation vectors, $\mathbf{x}_{it} = (x_{it,1}, x_{it,2}, ..., x_{it,s})$, where s is the number of macrovariables. These are allocated in clusters so as to maximise within-cluster similarity and between-cluster discrepancy. The outputs are K *clusters* labelled as either default ($\hat{y} = 1$) or non-default ($\hat{y} = 0$) according to an *assignment rule*. An unseen or out-of-sample case $\mathbf{x}_{it}$ is classified as default/non-default depending on the cluster whose centroid is closer. (See Appendix A). The choice of K does not follow from the algorithm and is often made subjectively. Hence, optimal clustering calibration requires finding the 'best' assignment rule and K according to the decision-maker's preferences.[6]

[4] The finite-sample properties of different estimation approaches to generate rating migration probabilities from the external ratings provided by Moody's are explored in Fuertes and Kalotychou (2004b) by Monte Carlo simulation.

[5] K-means clustering was chosen over hierarchical clustering techniques, such as nearest neighbour or average linkage, because for large datasets these are computationally and storagewise rather expensive.

[6] The proposed approach can be applied to different warning horizons to optimize h also.

We propose below an approach which is discussed without loss of generality for the IL function.

For a given K, the assignment rule can be optimized as follows.[7] Let $n_c(1)$ be the number of default cases (vectors $\mathbf{x}_{i,t-1}$ such that $y_{it} = 1$) in cluster c. Likewise for $n_c(0)$. Let C_1 (and C_0) denote the total number of default (non-default) cases. The loss implied by labelling cluster c as non-default is $L_{0,c}(\theta) = \theta \times P_I$ where $\hat{P}_I = \frac{n_c(1)}{C_1}$ is the estimated probability that a default case falls in cluster c. Likewise, $\hat{L}_{1,c}(\theta) = (1-\theta) \times \hat{P}_{II} = (1-\theta)\frac{n_c(0)}{C_0}$. The optimal rule for cluster c is

$$y_c^* = \underset{y \in \{0,1\}}{\operatorname{argmin}} L_{y,c}(\theta) \tag{8}$$

with loss $L_c^*(\theta)$. The minimal loss for the overall clustering is $L(\theta, K) = \sum_{c=1}^{K} L_c^*(\theta)$.[8]

Large-K clustering characterizes the sample rather well, but not necessarily the population and so it may produce poor out-of-sample forecasts. The optimal K can be found by a method introduced by Altman et al. (1985) to correct for overfitting bias in recursive partitioning. Consider $K \in \{2, ..., K_{\max}\}$ and define $L(\theta, K, \delta) = L(\theta, K) + \delta \times K$ where $L(\theta, K)$ is the minimal loss for a given K defined as above and $\delta \geq 0$ is an overfitting penalty. For each $\delta \in \{\delta_1, .., \delta_n\}$ we find $\tilde{K}_\delta = \underset{K}{\operatorname{argmin}} L(\theta, K, \delta)$. This yields a set $\left\{\tilde{K}_{\delta_1}, ..., \tilde{K}_{\delta_n}\right\}$ from where K^* is found using a cross-validation. The sample $\{\mathbf{x}_{it}\}$ is randomly partitioned into V equally-sized groups. For say δ_1, we leave out one group and cluster the remaining cases using the above $\tilde{K}_{\delta_1}$ that was selected using the whole sample. The group cases left out are then assigned to the existing clusters. The procedure is iterated by leaving out a different group each time. The cross-validated loss associated to $\tilde{K}_{\delta_1}$ is the average loss over the V iterations $CV[L(\theta, \tilde{K}_{\delta_1})] = \frac{1}{V}\sum_{j=1}^{V} L(\theta, \tilde{K}_{\delta_1})_j$ where j denotes the group left out in iteration j. The optimal K^* minimizes the cross-validated loss

$$K^* = \underset{\tilde{K}_{\delta_j}}{\operatorname{argmin}} CV[L(\theta, \tilde{K}_\delta)],\ j = 1, 2, ..., n.$$

[7] This is akin to finding the optimal cut-off rate λ^* in the logit framework.

[8] For the PL function, $L_0(c) = (1-\theta) \times P_W$ where $P_W \equiv \Pr(\hat{y} = 1)$ is estimated as the number of cases in the cluster over all sample cases, $\frac{n_c(1)+n_c(0)}{C}$. Likewise, we have $L_1(c) = \theta \times P_I$.

In this paper, we set $K_{\max} = 10$, $\delta \in \{0.001, 0.002, ...0.01\}$ and $V = 5$.[9]

The main advantage of clustering over logit is its non-parametric nature, namely, it does not require the forecaster to formalize the relation between the exogenous macrovariables and the default event. But clustering has some pitfalls. First, it does not provide a continuous scoring scale such as the posterior probability of default and so the countries cannot be ranked in terms of default risk, which is important for international investors. Second, the main aspects of the default clusters (e.g. low trade/GDP) are often not clear-cut particularly when many variables are used and so one cannot identify the determinants of default, which is important for policymakers.

3.1 Combining the forecasts from LOGIT-M, LOGIT-R and K-clustering

Let $\{\hat{y}^m_{i,t+1}\}_{m=1}^M$ denote M rival forecasts formed at period t and $\hat{y}^C_{i,t+1} = \mathbf{R}(\hat{y}^1_{i,t+1}, ..., \hat{y}^M_{i,t+1})$ the combined forecast where $\mathbf{R}$ is a mapping or transformation. We consider two mappings: a) *logit* for mixed probability and event forecasts and b) *voting rules* for event forecasts.

Kamstra and Kennedy's (1998) [KK] logit regression method is simple to apply and it permits the combination of probability, event forecasts or a mix. It can be extended to polychotomous and ordered classification problems using multinomial or ordered logits, respectively. It has been shown that KK-logit can beat the equal-weights approach in large samples (Kamstra et al., 2001).

According to the KK-logit approach, we fit by OLS a regression of the EWS indicator (y_{it}) on a constant, the log-odds ratio forecasts ($\log\frac{\hat{p}_{it}}{1-\hat{p}_{it}}$) from LOGIT-M and LOGIT-R and the event forecasts ($\hat{y}_{it}$) from K-clustering. The coefficient estimates are the combining weights. To allow for time-variation, this approach is recursively applied in-sample over a 12-year rolling window. Thus we have weights $\mathbf{w}_\tau \equiv (w^1_\tau, w^2_\tau, w^3_\tau)$ for each set of out-of-sample forecasts, $\tau = 1996, ..., 2000$.

A nice property of KK-logit is that it enables forecast encompassing tests (Fair and Schiller, 1990). We conduct a LR test for $H_0 : w^m_\tau = 0$ for each $m = 1, 2, 3$ and the mth forecast for

[9] Simulations have shown that $V = 5$ works well to calibrate the number of nodes in classification trees (Breiman et al., 1984). The change in IL or PL for successive K is of order $\mathbf{10^{-2}}$. The latter drives our choice of range for δ.

year τ is discarded if the statistic is insignificant. The out-of-sample combined forecasts, $\hat{p}^C_{i\tau}$, are transformed into event forecasts by means of a cut-off rate λ^*_τ. The latter is chosen optimally for each $\tau = 1996,...,2000$ as described in Section 3 according to the decision maker's preferences.

Event type forecasts can be combined using voting rules such as:

$$\hat{y}^C_{i,t+1} = \begin{cases} 1 & if \ \sum_{m=1}^{M} \omega^m_{t+1} \hat{y}^m_{i,t+1} \geq R \\ 0 & otherwise \end{cases} \tag{9}$$

where R is the voting parameter. There is a large literature on the properties of such combining schemes, mostly in the context of recursive trees and neural networks. The results depends on the choice of ω^m_{t+1} — weighting equally or according to cross-validated past performance — and the empirical evidence is inconclusive (Alpaydin, 1998; Ali and Pazzani, 1995). A key role is also played by R. The most frequent choices are $R = 1/2$ and $R = 1$ which, alongside the constraints $\sum_{m=1}^{M} \omega^m_{t+1} = 1$ and $\omega^m_{t+1} = \frac{1}{M}$, yield the Majority Rule (MR) and Unanimous Rule (UR), respectively. The combined forecast is thus the event predicted by the *majority* of the classifiers or by *all* of them, respectively. The empirical evidence on the relative performance of these schemes is conflicting. Battiti and Colla (1994) support the MR whereas Albanis and Batchelor (1999) support the UR. We consider both schemes.

4 The data

The analysis is based on $N = 75$ emerging and developing countries over 1983-2000.[10] Since the predictors (explanatory variables) are lagged 1 year, the effective sample for y_{it} spans the 1984-2000 period: 1984-1995 is the initial 12-year rolling window and 1996-2000 is the holdout period.[11]

The default indicator $\{d_{it}\}_{t=1984}^{2000}$ is based on *World Bank* data for external debt, principal and interest arrears to official/private creditors and amounts of principal rescheduled. A given

[10] See Appendix B. The regions (number of countries parenthesis) are East Europe (7), Asia (12), Latin America (22), Middle East/North Africa (9), Africa (25). For more details, see Fuertes and Kalotychou (2004).

[11] LIMDEP 8 and SPSS 10 are used in the subsequent empirical analysis.

country-year is a 'default' case ($d_{it} = 1$) if: a) There is an increase in total arrears that exceeds a threshold percentage of total external debt ($\Delta A_{it}/D_{it} > \delta$), b) A debt rescheduling agreement was reached and the total amount of debt rescheduled exceeds the decrease (if any) in the arrears stock. The threshold δ is set at the mean of $\Delta A_{it}/D_{it}$ giving 2.26%. The default events that follow from this definition correspond closely to those in Standard and Poor's (2001). See Appendix B.

Data on 25 macroeconomic and financial ratios are obtained from the *World Bank* database. These include short-term financial solvency and liquidity proxies and longer-term structural macroeconomic signals (see Appendix C). To reduce the degree of skewness and kurtosis in the ratios and the number of outliers, these are logged using $sign(x)\ell n(1 + |x|)$. Any remaining outlier in each default/non-default group is reduced by windsorizing the ratios as follows. A data point x_{it} is indexed by $c \in \{0, 1\}$ according to whether it pertains to a tranquil ($y_{it} = 0$) or default ($y_{it} = 1$) window. If x_{it}^c falls outside $\bar{x}^c \pm 4\hat{\sigma}^c$, it is replaced by the appropriate interval limit. This regressor set was reduced using an in-sample jacknife or cross-validation approach.[12] The thirteen variables thus retained (see Appendix C) are denoted $\mathbf{x}_{it} \equiv (x_{it,1}, x_{it,2}, ..., x_{it,13})'$.

Country credit ratings are obtained from the *Institutional Investors* database. These are available for 67 countries in 1984 rising to a maximum of 83 by 2000.[13] They seek to capture the perception of worldwide bankers regarding a country's ability and willingness to service its financial obligations. In particular, our ratings series (z_{it}) is an index based on the weighted scores assigned to the countries by the 100 largest international commercial banks. The latter closely monitor the observance of standards — whether a country has published an IMF Article IV or ROSC and met the SDDS specifications.[14] It varies in a 1-100 scale with 100 representing low

[12] The jacknife is based on a logit regression over 1984-1995 . A variable is dropped if, in doing so, the cross-validated loss (a conservative IL or PL with $\theta = 1$, $\lambda = 0.5$ and $h = 3$) does not increase. For details, see Fuertes and Kalotychou, 2004a. The variable selection could be cast as another element in the optimal design of an EWS, namely, one could select the set that is 'best' according to the decision-maker's preferences.

[13] In contrast, external credit ratings from Moody's and S&P's are unavailable for many countries in our sample.

[14] The Special Data Dissemination Standards (SDDS) was designed for countries with, or seeking access to, international capital markets. It sets macro data definitions, in particular, reserves. It also sets minimum timeliness

default-risk countries. The bankers' ratings are updated semi-annually and our LOGIT-R classifier is based on end-of-year data. To avoid sample selection bias in comparing the classifiers — LOGIT-M, LOGIT-R and K-clustering — the country-period cases used in the analysis are those for which both $\mathbf{x}_{it}$ and z_{it} are available. This leads to 45 countries in $t = 1984$ and 69 in $t = 2000$.

We should stress that sovereign debt crisis typically last longer than one year in contrast with banking crises. About 30% of all country-period cases over 1984-2000 are defaults ($d_{it} = 1$) whereas about 10% are default entries ($\Delta d_{it} = 1$). The average length of a debt crisis is around 3 years. The real challenge for an EWS of sovereign default is to predict a new default *entry* (turning point) rather than a perpetuating default. In order to develop a powerful EWS in the above sense, the loss functions will be evaluated over an *entry* set defined as follows. Default year t is excluded for country i if it was in default at year $t-1$, i.e. $d_{it} = 1$ is excluded if $d_{i,t-1} = 1$.[15]

5 Optimal calibration of forecasting tools

This section discusses the in-sample calibration of the classifiers over the first 12-year window 1984-1995. Unless otherwise noted, y_{it} is based on $h = 1$ and the risk aversion parameter is set at $\theta = 0.5$. Asterisks denote optimal values.

5.1 Balancing the missed defaults and the false alarms

As noted earlier, the cut-off rate (λ) and warning horizon (h) parameters of an EWS are often chosen subjectively. An objective choice requires a trade-off between Type I and Type II errors. Figure 1, Panel A, illustrates the latter for the LOGIT-M classifier.[16] We consider cut-off rates

and frequency standards for data releases. The Reports on the Observance of Standards (ROSC) are voluntary and refer to transparency, financial market regulation and corporate governance issues. (See Glennerster, 2004).

[15] Some studies use 'exclusion windows' whereby consecutive default years within a certain time window are excluded from the empirical analysis; the length of the window is arbitrary and studies have used about 1-3 years (Detragiaghe and Spilimbergo, 2001; Frankel and Rose, 1996). However, in using this reduced sample for the model specification and estimation one may discard important information. In this paper, the *entry* set is used to assess the models' out-of-sample forecast accuracy. The estimation and clustering are based on all the in-sample cases.

[16] The LOGIT-R calibration raises similar issues to that of the LOGIT-M so we focus on the latter. Clustering

$\lambda \in (0,1)$ and warning horizons $h = \{1,2,...,h_{\max}\}$ with $h_{\max} = 3$.

[Figure 1 around here]

A higher λ or a lower h yield fewer false alarms at the expense of more missed defaults. From the perspective of creditors or investors, the latter means realised losses (balance sheet), reserve holdings will need to rise and cash flows and asset values will also be adversely affected. From the perspective of policymakers, the experience of the 1990s has suggested that output contractions, rising unemployment and poverty rates are some of the repercussions of sovereign debt crises.

Lowering λ or raising h will induce the opposite trade-off, namely, less missed crises at the expense of more false alarms. The latter means foregone profit opportunities for investors and unnecessary policy actions which may be costly, for instance, in terms of social unrest. Forecasters should use the (λ, h) combination that are 'best' according to the decision-maker's preferences.

5.2 Well-behaved loss functions

Figure 1, Panel B, shows the interaction between the losses and the cut-off $\lambda \in (0,1)$. The NS loss function monotonically falls as λ increases. The minimum NS is achieved at a relatively high $\lambda^* \geq 0.724$. The total number of default signals, $E_0 + (C_1 - E_1)$ at this λ^* is rather small and the model misses most of the defaults ($\hat{P}_I = E_1/C_1 = 85\%$). The intuition is that, as λ increases, both the rate of false alarms, E_0/C_0, and the rate of correct warnings, $(C_1 - E_1)/C_1$, fall but the former does so faster because $C_0 > C_1$.[17] This suggests that the NS loss function may be unsuitable in this context because it leads to a debt-crisis EWS with a very high probability of missed defaults.

The main pitfall of NS is that it only accounts for the Type I and II error rates in relative terms. For instance, $NS = 1/9$ could stem from $\hat{P}_{II} = \frac{E_0}{C_0} = 10\%$ and $\hat{P}_I = \frac{E_1}{C_1} = 10\%$ or from

methods have been shown to work better when the variables are mapped to the [0,1] interval. Hence, for K-clustering the N points for each year, $\{x_{it}\}_{i=1}^{N}$, are transformed using $\tilde{x}_{it} = (x_{it} - \min\{x_{it}\})/(\max\{x_{it}\} - \min\{x_{it}\})$. The same 13 regressors are used for both the LOGIT-M and K-clustering classifiers to make the comparison more informative. Another rationale is that the determinants of default should not depend on the model employed.

[17] Over the 1984-1995 period, the probability of default *entry* is $\hat{p} = C_1/(C_1 + C_0) = 71/423 = 17\%$.

$\hat{P}_{II} = 1\%$ and $\hat{P}_I = 91\%$. Oka (2003) and Mulder et al. (2002) pose a similar criticism for the NS in the context of arrears to the IMF and currency crises, respectively. Moreover, the NS loss function, in contrast to IL and PL, does not allow the forecaster to control for the decision-maker's degree of risk aversion (θ) in the design of an EWS. Hence, we shall focus on IL and PL hereafter.

4.3 Optimal cut-off and warning horizon combination

Which are the optimal warning horizon and cut-off for an EWS of sovereign default? To answer this question, we deploy the optimization approach suggested in Section 3 for $h \in \{1,2,3\}$ and $\lambda \in [0.17, 1)$.[18] Figure 2 illustrates the in-sample calibration of LOGIT-M for representative risk-affine ($\theta = 0.2$), risk-neutral ($\theta = 0.5$) and risk-averse ($\theta = 0.8$) decision-makers under the IL or PL loss function. Table 1 sets out the results. The entries are the optimal values.

[Figure 2 around here]

[Table 1 around here]

For the IL function, a risk-neutral decision maker ($\theta = 0.5$) would need $\lambda^* = 0.641$ and $h^* = 3$. The best design for a highly risk-averse user ($\theta = 0.8$) corresponds to $\lambda^* = 0.205$ and $h^* = 2$ whereas for the risk-affine user ($\theta = 0.2$) the optimal choice is $\lambda^* = 0.727$ and $h^* = 3$. Regarding the PL function, for the risk-neutral user ($\theta = 0.5$) the best parameters are $\tau^* = 0.270$ and $h^* = 1$ whereas for the risky user ($\theta = 0.2$) we have $\lambda^* = 0.724$ and $h^* = 1$. These findings illustrate that:

Result 1. *The optimal cut-off and warning horizon parameters of a logit EWS depend on the decision-maker's preferences (loss function, risk aversion).*

Figure 3, Panel A, shows the optimal cut-off for different risk aversion (towards default) levels.

[Figure 3 around here]

[18] The LOGIT-M for $h = \{2,3\}$ and $\lambda < 0.17$ predicts 1 nearly all the time. The step size for λ is 10^{-4}.

For a given horizon h, a higher θ needs a lower λ^* to achieve the best balance between the two errors. Hence, more risk-averse decision-makers would need lower cut-off rates. Thus we have:

Result 2. *For a given warning horizon, the optimal cut-off rate decreases with the decision-maker's degree of risk-aversion towards missing defaults.*

Figure 3(A) reveals also that for a given θ, a longer h implies a higher λ^* and vice versa. The intuition is that the longer the warning horizon h ceteris paribus, the more alarms are issued. The latter implies more Type II errors ($E_0 \uparrow$) but less Type I errors ($E_1 \downarrow$) with the former increasing at a faster rate and so a higher cut-off is needed to achieve the desired balance. Thus we have:

Result 3. *For a given risk aversion level, the optimal cut-off increases with the warning horizon.*

The relation between the risk aversion level and the warning horizon is shown in Figure 3, Panel B. For any given $\theta \leq 0.7$, the IL function yields $h^* = 3$. For $\theta > 0.7$, it results in $h^* = \{2, 3\}$. Intuitively, the benefit for institutional investors from using a relatively long horizon stems from missing less defaults ($P_I \downarrow$) which outweighs the opportunity cost of more false alarms ($P_{II} \uparrow$).

Result 4. *The optimal warning horizon is relatively long for international investors.*

According to the PL function, for risk aversion $\theta \leq 0.7$ the logit EWS should be based on $h^* = 1$, for $0.75 \leq \theta \leq 0.85$ the best choice is $h^* = 2$ whereas for $\theta \geq 0.9$ it is $h^* = 3$. Thus we have:

Result 5. *The optimal warning horizon increases with the degree of risk-aversion for policymakers.*

In sum, the optimal horizon depends on the decision-maker's preferences. The contrast between Result 4 and 5 can be rationalized as follows. Since policymakers assign a cost to default alarms (whether correct or false), a longer horizon would only be optimal for those policymakers with high risk-aversion so that the benefit of missing less defaults outweighs the cost of too many alarms.

4.4 Optimal number of clusters

We now turn to the K-means clustering classifier. We consider $K \in \{2, 3, ..., K_{\max}\}$ with $K_{\max} = 10$. The optimal number of clusters under both the IL and PL functions with risk aversion $\theta \in \{0.3, 0.5, 0.8\}$ are set out in Table 1.[19] The optimization results are very similar under both loss functions. Figure 4(A) illustrates the calibration under the IL function with $\theta = 0.5$.

[Figure 4 around here]

For the low risk aversion level $\theta = 0.3$, the optimal number of clusters is $K^* = 8$ whereas for $\theta = 0.5$ and $\theta = 0.8$ we have $K^* = 7$ and $K^* = 6$, respectively. Figure 4(B) also shows the relation between the optimal number of clusters and degree of risk aversion which, interestingly, is roughly V-shaped. A similar relation is found under the PL function. The main finding is

Result 6. *The optimal number of clusters depends on the decision-maker's degree of risk-aversion.*

The calibration of the warning horizon in K-clustering raises similar issues as in the logit. For instance, a higher θ leads to a longer optimal h in both classifiers.[20] The choice of assignment rule for the final clusters into default/non-default is akin to the choice of cut-off in the logit. For a high θ, the preferred assignment rule assigns relatively more clusters to the default state.

6 Optimal forecast combination

The above results suggest that the decision-maker's preferences should be accounted for in the design of an EWS, namely, in choosing parameters such as the warning horizon, cut-off rate (logit) and number of clusters (K-clustering). Another potential way to improve the performance of an EWS is by combining the strengths of different classifiers. Since many weighting schemes are

[19] We adopt $\theta = \mathbf{0.3}$ as *low* risk aversion level here because under $\theta = \mathbf{0.2}$ all clusters are labelled as non-default.

[20] In contrast with the logit estimation, changing h (or the definition of y_{it}) does not change the clustering of the $\mathbf{x}_{i,t-1}$ cases. However, changing h will affect the optimal K and assignment rule for a given loss function.

possible, we suggest to choose among them optimally according to the decision-maker's preferences. To simplify the exposition, we take y_{it} as defined in (1) for $h = 1$ as the event to be forecasted and focus primarily on the IL function.[21] The results for the PL function are outlined below.

We start by comparing the out-of-sample predictive ability of the classifiers. First, the classifiers are calibrated — the cut-off for LOGIT-M and LOGIT-R and the number of clusters and assignment rule for clustering — over the 1984-1995 window. The logit estimates and final clusters thus obtained are used to generate out-of-sample forecasts for 1996. This calibration and estimation/clustering is reconducted over 1985-1996 to generate out-of-sample forecasts for 1997 and so on. A country-mean loss is obtained for each out-of-sample year and then averaged over years.

Table 2 sets out the comparison across individual classifiers over the holdout sample. The entries are the Type I and II error rates and the overall loss (IL) for several risk aversion levels θ.

[Table 2 around here]

The Type I error rate from LOGIT-R is lower than that from LOGIT-M virtually for all θ. The exceptions are $\theta = 0.2$ and marginally $\theta = 0.6$. So LOGIT-R outperforms LOGIT-M regarding missed default entries whereas the opposite holds for false alarms. This indirectly suggests that the bankers' judgments implicit in the ratings are relatively pessimistic about country creditworthiness.

The Type I error rate of K-clustering is essentially higher than that of LOGIT-R or LOGIT-M for low to moderate risk aversion levels $\theta < 0.55$. For higher θ, clustering gives few missed defaults (2-6%) albeit at the expense of many false alarms (69-88%). LOGIT-M classifies the non-defaults relatively well, that is, it dominates the other classifiers in terms of the Type II error rate.[22]

The ranking of the classifiers, in terms of the overall loss (IL), follows from their relative Type I and Type II error strengths for each risk aversion level. For instance, at the low level $\theta = 0.3$

[21] The same warning horizon has to be adopted in all classifiers because, otherwise we would be combining forecasts for different dependent variables (y_{it}). The calibration of h for the combined classifier can be carried out as in Section 5, namely, by choosing $h \in \{1, 2, ..., h_{\max}\}$ so as to minimize the overall loss of the combined forecasts.

[22] The exceptions are $\theta = \{\mathbf{0.2}, \mathbf{0.6}\}$ for which the credit ratings (LOGIT-R) yield the smallest Type II error rate.

the minimal loss is that of the LOGIT-M because of its relatively small Type II error rate despite having a large Type I error rate. However, as the risk aversion (Type I error penalty) increases, the LOGIT-R beats the LOGIT-M.[23] The overall loss of the non-parametric (clustering) classifier is relatively large except for very high risk aversion levels $\theta \geq 0.85$. These considerations prompt the thought that there may be gains from combining the forecasts of the three classifiers.

We now assess the stability of the out-of-sample forecast ranking. Table 3 indicates year-by-year the best classifier and the associated minimal loss for several risk-aversion levels.

[Table 3 around here]

The forecast ranking changes over time which further motivates the forecast combination. For instance, LOGIT-R stands out over 1996, 1998 and 2000 whereas LOGIT-M essentially excels over 1997. The forecast instability pattern can be explained in terms of the relative strengths (Type I and Type II errors) of the classifiers. In particular, LOGIT-R is relatively 'pessimistic' toward country creditworthiness and so it does quite well in 1996 and 1998 where a relatively large number of defaults occurred. Interestingly, relatively few default entries occurred in 1997.

To combine the forecasts we use the (parametric) KK-logit regression and the two non-parametric schemes — the majority rule (MR) and unanimous rule (UR). In contrast to the latter, KK-logit accounts for the historical (in-sample) forecast performance of the rival forecasts. Table 4 reports the KK-logit weights of the out-of-sample forecasts for year 1996.

[Table 4 around here]

These combining weights are obtained via a logit regression of the indicator $\{y_{it}\}_{t=1984}^{1995}$ on the rival forecasts from LOGIT-M, LOGIT-R and K-clustering. The latter change with the optimal number of clusters and assignment rule which, in turn, depend on the decision-maker's risk aversion θ as

[23] The only exceptions occur at $0.65 \leq \theta \leq 0.7$.

the foregoing analysis has shown. Thus a new set of K-clustering forecasts is obtained as θ varies and so the forecast combining weights vary also. The cut-off rates (for the LOGITs) play no role in this combining exercise because $\hat{p}_{it}$ is directly used.

The properties of the combined forecasts are set out in Table 5.

[Table 5 around here]

Regarding false alarms, the best results stem from the UR for nearly all risk aversion levels. This is expected given that, in order to signal a default, the UR requires all individual classifiers to predict a default. In terms of missed defaults, KK-logit excels for low risk-aversion levels $\theta \leq 0.45$ whereas MR leads the race for $\theta > 0.5$.[24] The UR scheme performs rather poorly relative to individual or rival combined forecasts in terms of missed defaults.

Regarding the overall loss (IL), the best combined forecasts stem either from KK-logit for risk-aversion $\theta \leq 0.75$ or from MR for $\theta > 0.75$. We conduct a Diebold-Mariano (1999) [DM] test to compare the best combined forecaster and the best individual forecaster.[25] For several risk aversion levels, it pays to combine the classifiers. For instance, for $\theta \in \{0.2, 0.25, 0.55, 0.65, 0.7, 0.75\}$ the minimal loss from KK-logit combining is significantly smaller than the losses from either of the individual classifiers. For $\theta \in \{0.55, 0.85\}$ the gains from MR combining are significant. Figure 5, Panel A, illustrates the comparison graphically.

[Figure 5 around here]

It shows that either the individual LOGIT-R forecasts or the combined forecasts using KK-logit or MR produce the best out-of-sample forecast performance.[26]

[24] The Type I and Type II error rates for the KK-logit forecasts essentially stabilize for $\theta \geq \mathbf{0.6}$.

[25] We compute $DM_t, t = 1, ..., m$, where $DM_t \overset{a}{\sim} N(\mathbf{0}, 1)$. Under independence between the test statistics, it follows that $DM = \frac{1}{m}\sum_{t=1}^{m} DM_t \overset{a}{\sim} N(\mathbf{0}, \frac{1}{m})$.

[26] For the baseline *overall error* rate reported in many studies — errors over sample cases $(E_0 + E_1)/C$ — the UR forecasts essentially beat all other forecasts despite their relatively high Type I error rate (E_1/C_1). This is because this metric does not account for E_1/C_1. Due to the small number of 1s in the sample, C is much larger than C_1 and so E_1/C and E_0/C appear small relative to E_1/C_1. Hence, like the NS ratio this metric can be misleading.

The forecasts should also be compared with naive predictions. Our uninformative, naive model predicts 1 for highly risk-averse decision-makers, $\theta > 0.5$, 0 for $\theta < 0.5$ and the most frequently observed event in-sample (here 0) for $\theta = 0.5$. Table 6 reports the ratio of the overall loss for each classifier (IL) relative to that of the naive predictor (IL^n). A DM test is conducted to compare the minimal loss among the classifiers, individual or combined, with that of the naive predictor.

[Table 6 around here]

It turns out that for all risk-aversion levels the best forecasting model significantly outperforms the naive predictor. The highest gains relative to the naive come from LOGIT-R for $\theta = 0.5$ with the smallest ratio $\frac{\text{IL}}{\text{IL}^n} = 0.453$, followed by the KK-logit or MR combined forecasts for $\theta = 0.55$ with a ratio of 0.462. Interestingly, the gains of the best-performing model relative to the naive $(1 - \frac{\text{IL}}{\text{IL}^n})$ monotonically increase with θ up to $\theta = 0.5$ and then decrease thereafter.

Next we reconduct the above steps in the optimal EWS design (calibration, estimation/clustering, forecast combination and evaluation) on the basis of the PL function. The comparison of individual and combined forecasts over the holdout sample is set out in Figure 5, Panel B. The results suggest that forecast combining brings gains for a wide range of risk-aversion levels albeit not for all. More specifically, the KK-logit combined forecasts achieve the minimal loss for $\theta < 0.7$ whereas the LOGIT-M and LOGIT-R forecasts are ranked best for $0.7 \leq \theta \leq 0.8$. Clustering significantly outperforms all other individual and combined methods for $\theta > 0.8$.

7 Concluding remarks

This paper highlights the importance of and illustrates how to optimally design an EWS for sovereign default according to the decision-maker's preferences. Debt crisis forecasts are obtained from two different methods based on the same macrovariables — a logit regression (LOGIT-M) and clustering — and a logit regression based on bank-internal ratings (LOGIT-R). The data pertains

to 75 emerging and developing economies over 1983-2000.

First, the study shows how to recursively calibrate in-sample these classifiers according to the decision-maker's preferences. The latter are formalized by means of a loss function and risk aversion parameter. Second, we discuss forecast combining issues. For this purpose, we consider a regression framework that exploits information on the classifiers' past forecast ability and two non-parametric voting rules based on equal weights.

The results suggest that the decision-maker's preferences influence the optimal warning horizon, cut-off probability, assignment rule and number of clusters. These key parameters have mostly been chosen in an ad hoc manner in the literature. The optimally calibrated classifiers show different strengths in terms of missed defaults and false alarms. In particular, the LOGIT-M classifier outperforms the non-parametric (clustering) and judgmental (LOGIT-R) classifiers in terms of false alarms. On the other hand, judgmental and non-parametric classifiers dominate LOGIT-M in terms of missed defaults. Moreover, there is instability in the out-of-sample forecast ranking. Overall these findings vindicate a forecast combining exercise. The latter reveals that the best combining scheme depends on the decision-maker's preferences. In most cases, the combined forecasts significantly outperform the individual forecasts and uninformative naive forecasts.

The findings in this paper should have strong implications in applied work on credit risk prediction. In practice, many EWS for sovereign default are based on ad hoc parameter choices. The optimal recursive in-sample calibration of the forecasting tools that underlie such EWS, including the forecast-combining weigthing scheme, is strongly recommended.

References

[1] Aiolfi, M., & Timmermann, A. (2003). Persistence in forecasting performance. *Mimeo,* University of California at San Diego. Available at: http://econweb.rutgers.edu.

[2] Albanis, G., & Batchelor, R.A. (1999). Combining heterogeneous classifiers for stock selection. *Mimeo*, Cass Business School, London. Available at: http://www.staff.city.ac.uk.

[3] Ali, K.M., & Pazzani, M.J. (1995). Error reduction through learning multiple descriptions, *Machine Learning,* 24, 173-2002.

[4] Alpayadin, E. (1998). Techniques for combining multiple learners, *Proceedings of Engineering of Intelligent Systems Conference,* 2, 6-12.

[5] Altman, E. I., Frydman, H., & Kao, D.L. (1985). Introducting recursive partitioning for financial classification: the case of financial distress. *Journal of Finance*, XL, 269-191.

[6] Bates, J.M., & Granger, C.W.J. (1969) The combination of forecasts, *Operational Research Quarterly,* 20, 451-468.

[7] Battiti, R., & Colla, A. M. (1994). Democracy in neural nets: voting schemes for classification. *Neural Networks*, 7, 691-707.

[8] Berg, A., & Pattillo, C. (1999). Predicting currency crises: The indicators approach and an alternative, *Journal of International Money and Finance*, 18, 561-586.

[9] Breiman, L., Friedman, J.H., Olsen, E.A., & Stone, C.J. (1984). *Classification and regression trees*, Belmont, CA: Wadsworth International Group.

[10] Burkart, O., & Coudert, V. (2002). Leading Indicators of currency crises for emerging countries. *Emerging Markets Review*, 3, 107-133.

[11] Bussière, M., & Fratzscher, M. (2002). Towards a new early warning system of financial crises, *European Central Bank Working Paper* 02/145. Available at: http://www.ecb.int.

[12] Cantor, R., & Packer, F. (1996). Determinants and Impact of Sovereign Credit Ratings, *FRBNY Economic Policy Review*, October, 37-52.

[13] Clemen, R.R, Winkler, R., & Murphy, A. (1995). Screening probability forecasts: contrasts between choosing and combining, *International Journal of Forecasting,* 11, 133-146.

[14] Detragiache, E., & Spilimbergo, A. (2001). Crises and liquidity: Evidence and interpretation. *IMF Working Paper* 01/2.

[15] Demirguc-Kunt, A., & Detragiache, E. (1999). Monitoring banking sector fragility: A multivariate logit approach, *IMF Working Paper* 99/147.

[16] Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics,* 13, 134-144.

[17] Fair, R. C., & Shiller, R.J. (1990). Comparing information in forecasts from econometric models, *American Economic Review*, 80, 375-89.

[18] Frank, C.R., & Cline, W.R. (1971). Measurement of debt-servicing capacity: An application of discriminant analysis, *Journal of International Economics*, 1, 327-344.

[19] Frankel, J.A., & Rose, A.K. (1996). Currency crashes in emerging markets: An empirical treatment, *Journal of International Economics*, 41, 351-366.

[20] Fuertes, A.M., & Kalotychou, E. (2004a). Modeling sovereign default using panel models: a comparison. *Mimeo*, Cass Business School, London. Available at: http://www.ssrn.com.

[21] Fuertes, A.M., & Kalotychou, E. (2004b). A comparative analysis of sovereign credit migration estimators. *Mimeo*, Cass Business School, London. Available at: http://www.ssrn.com.

[22] Glennerster, R. (2004). Transparency and standards: evaluating the effect of institutions, Unpublished PhD thesis, Birkbeck College, University of London.

[23] Gupta, S., & Wilton, P. (1988), Combination of economic forecasts: an odds-matrix approach, *Journal of Business and Economic Statistics*, 6, 373-379.

[24] Kamin, S.B. (1999), The current international financial crisis: how much is new?, *Journal of International Money and Finance*, 18, 501-514.

[25] Kaminsky, G., & Reinhart, C.M. (1999). The twin crises: The cause of banking and balance of payments problems. *American Economic Review*, 3, 473-500.

[26] Kamstra, M., & Kennedy, P. (1998). Combining qualitative forecasts using logit, *International Journal of Forecasting,* 14, 83-93.

[27] Kumar, M.S., Moorthy, U., & Perraudin, W. (2003). Predicting emerging market currency crashes, *Journal of Empirical Finance*, 10, 427-454.

[28] Lee, S.H. (1993). Are the credit ratings assigned by bankers based on the willingness of the LDC borrowers to repay? *Journal of Development Economics*, 40, 349-359.

[29] Manasse, P., Roubini, N., & Schimmelpfennig, A. (2003). Predicting sovereign debt crises, *IMF Working Paper* 03/221.

[30] Mascarenhas, B., & Sand, O. C. (1989). Combination of forecasts in the international context: predicting debt reschedulings, *Journal of International Business Studies*, 20, 539-552.

[31] Montgomery, A.L., Zarnowitz, V., Tsay, R.S., & Tiao, G.C. (1998). Forecasting the US unemployment rate, *Journal of the American Statistical Association*, 93, 478-93.

[32] Mulder, C., Perilli, R., & Rocha, M. (2002). The role of corporate, legal and macroeconomic balance sheet indicators in crisis detection and prevention, *IMF Working Paper* 02/59.

[33] Newbold, P. & Harvey, D.I. (2004). Forecast combination and encompassing. In: M.P. Clements & Hendry, D.F. (Eds.), *A Companion to Economic Forecasting* (pp. 268-83). Blackwell: UK.

[34] Oka, C. (2003). Anticipating arrears to the IMF: Early warning systems, *IMF Working Paper* 03/18.

[35] Peter, M. (2002). Estimating default probabilities of emerging market sovereigns: A new look at a not-so-new literature, *HEI Working Paper* 02/6. Available at: http://www.hei.unige.ch.

[36] Reinhart, C.M. (2001). Default, currency crises and sovereign credit ratings, *World Bank Economic Revie*w, 16, 151-70.

[37] Rojas-Suárez, L. (2001). Rating banks in emerging markets, *Institute for International Economics Working Paper* 01/6. Available at: http://www.ssrn.com.

[38] Sommerville, R.A., & Taffler, R.J. (1995) Banker judgement versus formal forecasting models: The case of country risk assessment. *Journal of Banking and Finance*, 19, 281-297.

[39] Standard & Poor's (2001). Sovereign defaults decline through third-quarter 2000. In: S&P's (Eds.), *Ratings Performance 2000.* Available at: http://www.standardandpoors.com.

[40] Stock, J.H., & Watson, M. (2001). A comparison of linear and non-linear univariate models for forecasting macroeconomic time series. In: Engle, R.F. & White, H. (eds.). *Festschrift in honour of Clive Granger.*

[41] Taffler, R.J, & Abassi, B. (1984). Country risk: A model for predicting debt-servicing problems in developing countries. *Journal of the Royal Statistical Society, Series A*, 147, 541-568.

[42] Winkler, R.L., & Makridakis, S. (1983). The combination of forecasts, *Journal of the Royal Statistical Society, Series A*, 146, 150-57.

Biographies: Ana-Maria FUERTES (BSc Eng, MSc Control Eng, PhD Economics) is currently Reader in Econometrics at Cass Business School, City University, London. She was previously Senior Lecturer at London Metropolitan University where she worked as a Research Assistant. Her research interests are in time series analysis, panel data methods and forecasting. Her work has been published in the *Journal of Economic Dynamics and Control*, *Economics Letters*, *Journal of International Money and Finance* and *International Journal of Finance and Economics.*

Elena KALOTYCHOU recently obtained a PhD in Finance from Cass Business School, City University, London. At present, she holds a Research Associate position at Cass. She took a degree in Maths (BSC, Cantab) and another in Operations Research (MSc, London School of Economics). Her principal research interests are in credit risk modeling, quantitative methods and applied econometrics.

Appendix A: Clustering by K-means algorithm

K-means clustering belongs to the non-hierarchical clustering class of methods. In our context, a case is an observation vector $\mathbf{x}_{it} = [x_{it,1}, x_{it,2}, ..., x_{it,s}]'$ where s is the dimension of the macrovariable set; $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ denote country and period, respectively. The number of sample cases is $M = NT$.

K-means clustering consists of comparing the distances of each observation vector from the mean vectors of each of K proposed clusters in the sample of M observations. The observation $\mathbf{x}_{it}$ is assigned to the cluster with nearest mean vector. The distances are recomputed and reassignments are made as necessary. This process continues until all observations are in clusters with minimum distances to their mean vectors. The algorithm in steps is as follows:

1. Take the first K cases in the sample as the initial cluster centroids ($\mathbf{c}_1^0 \equiv \mathbf{x}_{11}, \mathbf{c}_2^0 \equiv \mathbf{x}_{21}, ..., \mathbf{c}_K^0 \equiv \mathbf{x}_{K1}$).

2. Assign case $\mathbf{x}_{it}$ to the cluster whose centroid is closer

$$\mathbf{c}_j^0 = \underset{q=1,2,...,K}{\text{argmin}} D(\mathbf{x}_{it}, \mathbf{c}_q^0)$$

$i = 1, 2, ..., N, t = 1, 2, ..., T$ where $D(\mathbf{x}_{it}, \mathbf{c}_q^0)$ denotes the Euclidean distance between the it^{th} case and q^{th} cluster centroid, given by

$$D(\mathbf{x}_{it}, \mathbf{c}_q^0) = \sqrt{\sum_{l=1}^{s} (x_{it,l} - c_{q,l}^0)^2}$$

Thus the outcome of Step 2 is a set of K clusters of observation vectors. Let $m_1, m_2, ..., m_K$ denote the number of observation vectors in each cluster such that $\sum_{q=1}^{K} m_q = M$.

3. The centroid of the new qth cluster is given by its mean observation vector

$$\mathbf{c}_q^1 = \left[\begin{array}{c} \frac{1}{m_q} \sum_{it} x_{it,1}, ..., \frac{1}{m_q} \sum_{it} x_{it,s} \end{array} \right]'$$

which facilitates a measure of the change in the cluster centroids, $\Delta S_q = D(\mathbf{c}_q^1, \mathbf{c}_q^0)$, $q = 1, 2, ..., K$.

4. If $\Delta S_q < \varepsilon$ for all $q = 1, 2, ..., K$ the algorithm terminates. Otherwise it goes to Step 2. We set $\varepsilon = 0.01$. The output of the procedure is the set of K clusters obtained at iteration j such that $\Delta S_q = D(\mathbf{c}_q^{j+1}, \mathbf{c}_q^j) < \varepsilon$ for all $q = 1, 2, ..., K$.

Appendix B: Historical sovereign debt crises per year, 1984-2000

Year	Default entries ($\Delta d_{it}=1$)	Countries
1984	15 (7)	Argentina, Benin, Bolivia, Dominican R, Egypt, El Salvador, Grenada Honduras, Iran, Mali, Mozanbique, Peru, Poland, Tanzania, Venezuela
1985	6 (3)	Congo R, Guinea, Morocco, Nicaragua, Sierra Leone, Zambia
1986	9 (7)	Argentina, Burkina Faso, Cameroon, Costa Rica, Gabon, Guatemala, Jamaica, Paraguay, Syria
1987	11 (9)	Bolivia, Brazil, Congo, Dominican R, Ecuador, Morocco, Panama, Sierra Leone, Tanzania, Togo, Zambia
1988	12 (7)	Argentina, Colombia, Congo DR, Cote D'Ivoire, Egypt, Guinea, Lebanon, Mozanbique, Nigeria, Trinidad-Tobago, Uganda, Vietnam
1989	15 (10)	Brazil, El Salvador, Gabon, Honduras, Jamaica, Jordan, Malawi, Mali, Mexico, Morocco, Paraguay, Philippines, Senegal, Tanzania, Togo
1990	7 (4)	Bulgaria, Cameroon, Guatemala, Guinea, Nigeria, Russia, Syria
1991	4 (3)	Benin, Bolivia, Costa Rica, Seychelles
1992	8 (5)	Burkina Faso, Dominican R, Egypt, Haiti, Honduras, Kenya, Russia, Zambia
1993	6 (5)	Benin, Brazil, Costa Rica, Panama, Peru, Sierra Leone
1994	5 (5)	Algeria, Argentina, Guatemala, Jordan, Mali
1995	6 (6)	Bolivia, Congo R, Cote D'Ivoire, Dominican R, Gabon, Jamaica,
1996	7 (7)	Guinea, Haiti, Honduras, Sierra Leone, Sri Lanka, Togo, Zambia
1997	4 (4)	Bulgaria, Nicaragua, Poland, Senegal
1998	11 (11)	Cameroon, Congo DR, Cote D'Ivoire, Dominican R, Indonesia, Pakistan, Peru, Tanzania, Togo, Uganda, Vietnam
1999	2 (2)	Ecuador, Honduras
2000	7 (7)	Burkina Faso, Gabon, Kenya, Mozanbique, Uganda, Zambia, Zimbabwe
	Total	Rate
1984-2000	135 (102)	10% (10%)
1984-1995	104 (71)	11% (11%)
1996-2000	31 (31)	8% (8%)

Data on debt levels, arrears and reschedulings for N=75 countries over T=17 years (1984-2000) is used to define NT=75×17=1275 cases for the default indicator d_{it}. Numbers in parentheses pertain to the effective country-period cases ($\widetilde{NT}$=1017) used in the analysis; $\widetilde{NT}$ is dictated by the missing cases for X (=macrovariables or credit ratings). The following countries did not experience default: Bangladesh, Bostwana, Chile, China, Czech R, Ghana, Hungary, India, Korea, Mauritus, Nepal, Oman, Papua Guinea, Romania, Swaziland, Thailand, Tunisia, Turkey, Uruguay.

Appendix C: Variable reduction by cross-validation (jacknife)

	mean		t-statistic	Jacknife
	x^0_{it}	x^1_{it}	$H_0 : \bar{x}^0_{it} - \bar{x}^1_{it} = 0$	result
External credit exposure				
Total external debt/ GDP	0.3879	0.7032	-23.76^*	✓
Official debt/Total debt	0.5633	0.5933	-6.91^*	✓
Short term debt/Reserves[a]	0.5221	1.4231	-17.96^*	×
Short term debt / Total debt	0.1244	0.1109	2.99^*	✓
Debt service / Exports	0.1492	0.1843	-6.70^*	✓
IMF credit / Exports	0.0828	0.1529	-9.68^*	✓
External economic activity				×
Export growth[b]	0.0691	0.0505	2.52^*	×
Volatility of export growth[c]	0.1072	0.1391	-6.14^*	✓
Trade balance/GDP[d]	−0.0797	−0.0822	0.40	✓
Reserves growth[a,b]	0.2012	0.5585	3.20^*	×
Reserves/Imports[a]	0.0375	0.0248	10.45	×
Domestic conditions				×
Credit to private sector/GDP	0.2746	0.1785	13.41^*	✓
GDP growth[b]	0.0398	0.0247	6.26^*	✓
GNP per capita	7.1236	6.4402	13.87^*	✓
Volatil. of GNP p.c. growth[c]	0.0435	0.0529	-6.35^*	✓
Governm expenditure/GDP[e]	0.1392	0.1369	0.89	×
Inflation	0.1266	0.3106	8.92^*	×
M2/Reserves[a]	0.0479	0.1158	-10.35^*	×
Real exchange rate[f]	0.1273	0.1110	0.74	✓
Gross capital formation/GDP	0.2175	0.1836	9.80^*	×
Gross domestic savings/GDP	0.1407	0.1051	5.42^*	×
Global links				×
Trade/GDP[g]	0.5427	0.4737	6.93^*	✓
Net bond flow[f,h]	1.5185	0.0764	5.64^*	×
Net equity flow[f,h]	1.2649	0.3330	4.69^*	×
FDI/GDP[i]				×

*denotes significant at the 1% level. [a] FX reserves, excl. gold. [b] Annual percentage growth. [c] Volatility proxied by the stdev over the last 4 years. [d] Trade balance=exports-imports. [e] Governm. expenditure on consumption, national security and defense. [f] Deviation from long-run trend. [g] Trade=exports+imports. [h] US$ billion. [i] FDI=Foreign Direct Investment.

Table 1
Optimal calibration of classifiers over estimation window

Risk-aversion	LOGIT-M				K-clustering			
	IL		PL		IL		PL	
θ	(λ^*, h^*)	L^*	(λ^*, h^*)	L^*	K^*	L^*	K^*	L^*
0.2	(0.727, 3)	0.129	(0.724, 1)	0.195	8	0.193	8	0.288
0.5	(0.641, 3)	0.227	(0.270, 1)	0.301	7	0.343	7	0.369
0.8	(0.205, 2)	0.139	(0.205, 2)	0.161	6	0.189	6	0.196

Investors' loss (IL) or policymakers' loss (PL) with risk-aversion parameter θ. The optimal cut-off and warning horizon (λ^*, h^*) or number of clusters (K^*) give the minimal loss L^* over 1984-1995. K-clustering results based on the optimal assignment rule of final K clusters into 1 or 0 for $y_{it}(h{=}1)$.

Table 2
Out-of-sample forecast ability of competing classifiers

	Risk aversion parameter (θ)															
Classifier	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
A: *Type I error* (missed defaults)																
LOGIT-M	**0.844**	0.824	0.730	0.667	0.667	0.607	0.443	0.243	0.137	0.137	0.137	0.137	0.109	0.089	0.089	0.060
LOGIT-R	1.000	**0.779**	**0.721**	**0.564**	**0.357**	**0.207**	**0.187**	**0.187**	0.139	0.099	0.070	0.070	0.070	0.070	0.020	0.020
Clustering	0.971	0.901	0.901	0.687	0.786	0.603	0.493	0.423	**0.060**	**0.060**	**0.060**	**0.060**	**0.060**	**0.020**	**0.020**	**0.020**
B: *Type II error* (false alarms)																
LOGIT-M	0.031	**0.035**	**0.042**	**0.070**	**0.070**	**0.084**	**0.205**	**0.214**	0.350	**0.367**	**0.367**	**0.449**	**0.520**	**0.572**	**0.572**	**0.773**
LOGIT-R	**0.009**	0.048	0.061	0.100	0.131	0.239	0.266	0.266	**0.337**	0.494	0.564	0.564	0.582	0.669	0.788	0.833
Clustering	0.026	0.086	0.086	0.128	0.069	0.190	0.337	0.408	0.690	0.683	0.794	0.794	0.794	0.834	0.834	0.877
C: *Overall loss* (IL)																
LOGIT-M	**0.193**	0.232	**0.248**	0.279	0.309	0.319	0.324	0.230	0.222	**0.218**	**0.206**	0.215	0.191	0.161	0.137	0.096
LOGIT-R	0.207	**0.231**	0.259	**0.262**	**0.221**	**0.227**	**0.227**	**0.223**	**0.218**	0.237	0.218	**0.194**	**0.172**	0.160	**0.097**	**0.061**
Clustering	0.215	0.290	0.330	0.324	0.356	0.376	0.415	0.416	0.312	0.278	0.208	0.243	0.207	**0.142**	0.101	0.063

The reported Type I error is $\hat{P}_I = \frac{E_1}{C_1}$, the Type II error is $\hat{P}_{II} = \frac{E_0}{C_0}$, the overall loss is $\widehat{IL}(\theta) = \theta\hat{P}_I + (1-\theta)\hat{P}_{II}$. E_i and C_i (event i=0,1) are the number of prediction errors and sample cases, respectively. All metrics are evaluated over the holdout entry sample 1996-2000. The out-of-sample forecasts are generated recursively from classifiers calibrated over a 12-year rolling window. Bold denotes the best outcome.

Table 3
Stability of forecast ranking over holdout period

	Risk aversion parameter (θ)															
Year	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
1996	Cluster	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-M	LOG-R	LOG-R	LOG-R	LOG-R	LOG-M
	(0.189)	(0.138)	(0.148)	(0.158)	(0.124)	(0.272)	(0.260)	(0.249)	(0.240)	(0.210)	(0.187)	(0.172)	(0.138)	(0.103)	(0.069)	(0.041)
1997	LOG-M	LOG-R	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-R
	(0.167)	(0.203)	(0.120)	(0.129)	(0.138)	(0.159)	(0.170)	(0.153)	(0.136)	(0.119)	(0.102)	(0.154)	(0.123)	(0.093)	(0.062)	(0.034)
1998	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R
	(0.198)	(0.217)	(0.256)	(0.295)	(0.268)	(0.263)	(0.284)	(0.302)	(0.280)	(0.240)	(0.247)	(0.223)	(0.198)	(0.191)	(0.160)	(0.130)
1999	Cluster	Cluster	Cluster	LOG-R	LOG-R	LOG-R	LOG-R	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-R
	(0.200)	(0.250)	(0.300)	(0.244)	(0.277)	(0.140)	(0.128)	(0.048)	(0.102)	(0.112)	(0.096)	(0.080)	(0.064)	(0.073)	(0.049)	(0.043)
2000	LOG-M	LOG-M	LOG-M	Cluster	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	Cluster	LOG-R
	(0.160)	(0.211)	(0.245)	(0.178)	(0.135)	(0.160)	(0.191)	(0.186)	(0.181)	(0.192)	(0.137)	(0.114)	(0.109)	(0.098)	(0.076)	(0.040)

The Investor's Loss (IL) metric is evaluated over the holdout *default entry* sample 1996-2000. For each out-of-sample year we indicate the best model and the minimal $\widehat{IL}(\theta) = \theta\hat{P}_I + (1-\theta)\hat{P}_{II}$ is reported in parenthesis. The out-of-sample forecasts are generated recursively from classifiers calibrated over a 12-year rolling window. LOG-M and LOG-R denote the logit classifier based on macrovariables and credit ratings, respectively. Cluster denote the K-means clustering classifier.

Table 4
Combination weights based on KK-logit regression

Combining weights (ω_τ^m)	Risk-aversion parameter (θ)									
	0.2	0.25-0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7-0.8	0.85-0.95
LOGIT-M	0.985	1.016	0.989	0.959	0.977	0.959	1.002	1.007	1.005	0.991
	(5.66)	(5.61)	(5.69)	(5.65)	(5.72)	(5.65)	(5.85)	(5.86)	(5.90)	(5.77)
LOGIT-R	0.591	0.605	0.577	0.579	0.577	0.579	0.573	0.590	0.565	0.614
	(3.39)	(3.47)	(3.24)	(3.36)	(3.33)	(3.36)	(3.17)	(3.28)	(3.08)	(3.54)
Clustering	0.390	-0.023	0.204	0.672	0.400	0.672	0.303	0.168	0.357	27.304
	(0.74)	(-0.05)	(0.60)	(2.03)	(1.26)	(2.03)	(0.566)	(0.29)	(0.61)	(0.00)
Intercept	-0.395	-0.330	-0.430	-0.817	-0.614	-0.817	-0.628	-0.500	-0.685	-27.64
	(-1.90)	(-1.53)	(-1.72)	(-2.64)	(-2.08)	(-2.64)	(-1.14)	(-0.82)	(-1.12)	(0.00)

The results are for the $\tau = 1996$ forecasts. In parenthesis, the t-ratio for the coefficient significance.

Table 5
Forecast ability of combined classifiers

Weighting	Risk aversion parameter (θ)															
scheme	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
A: *Type I error* (missed defaults)																
UR	1.000	1.000	1.000	0.921	0.893	0.803	0.729	0.579	0.227	0.387	0.207	0.207	0.179	0.139	0.089	0.060
MR	1.000	0.893	0.843	0.621	0.689	0.486	**0.257**	**0.157**	**0.069**	**0.049**	**0.040**	**0.040**	**0.040**	**0.020**	**0.020**	**0.020**
KK-logit	**0.710**	**0.681**	**0.681**	**0.563**	**0.563**	**0.443**	0.343	**0.157**	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.109
B: *Type II error* (false alarms)																
UR	**0.000**	**0.004**	**0.004**	**0.022**	**0.004**	**0.017**	**0.007**	**0.095**	**0.215**	**0.262**	**0.293**	**0.345**	**0.398**	0.446	0.498	0.671
MR	0.009	0.017	0.026	0.047	0.044	0.141	0.261	0.270	0.411	0.521	0.590	0.616	0.646	0.734	0.786	0.881
KK-logit	0.044	0.048	0.061	0.075	0.070	0.101	0.145	0.270	0.367	0.380	0.393	0.402	0.419	**0.419**	**0.419**	**0.419**
C: *Overall loss* (IL)																
UR	0.200	0.253	0.303	0.337	0.360	0.371	0.399	0.361	0.222	0.343	0.233	0.241	0.222	0.185	0.130	0.091
MR	0.207	0.236	0.271	0.248	0.302	0.296	0.259	**0.208***	**0.206**	0.214	0.205	0.184	**0.161**	**0.127***	**0.097**	**0.063**
KK-logit	**0.177***	**0.207***	**0.247**	**0.245**	**0.267**	**0.255**	**0.244**	**0.208***	0.212	**0.204***	**0.194***	**0.182***	0.171	0.155	0.140	0.124

See footnote to Table 2. The forecast ability (based on the IL function) of the best combined classifier is compared with that of the best individual classifier (reported in Table 2) using a Diebold-Mariano test. ** and * denote significant at the 1% and 5% level, respectively. UR and MR are the unanimous and majority 'voting' rule, respectively. KK-logit is the regression based scheme.

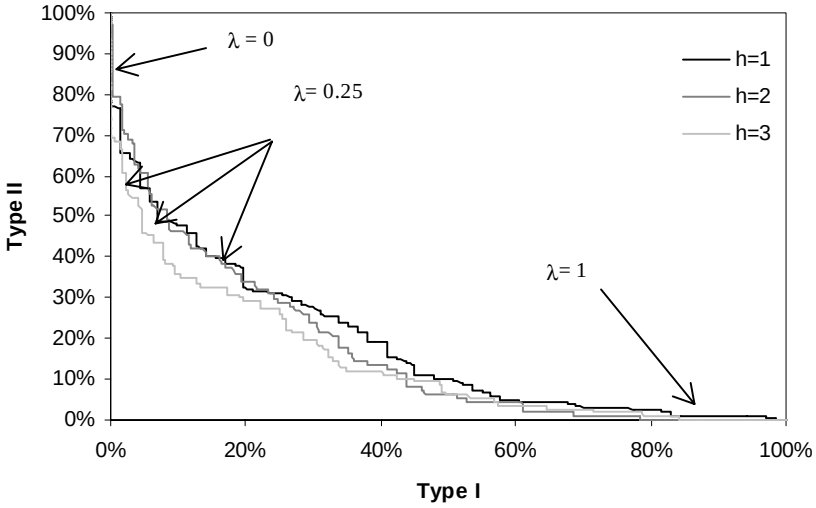
Table 6
Ratio of the classifier's loss to the loss of the naive predictor

	Risk aversion parameter (θ)															
Classifiers	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
A: *Individual*																
LOGIT-M	0.967	0.930	**0.826**	0.798	0.773	0.709	0.648	0.511	0.555	**0.622**	**0.687**	0.860	0.954	1.074	1.369	1.913
LOGIT-R	**1.036**	**0.924**	0.865	**0.750**	**0.553***	**0.500***	**0.453***	**0.495**	**0.545**	0.677	0.728	**0.774**	**0.862**	**1.065**	**0.968**	**1.213**
Clustering	1.074	1.158	1.101	0.925	0.889	0.835	0.830	0.925	0.780	0.794	0.934	0.974	1.034	0.947	1.014	1.257
B: *Combined*																
UR	1.000	1.013	1.010	0.962	0.899	0.824	0.798	0.802	0.555	0.981	0.776	0.966	1.112	1.232	1.295	1.811
MR	1.035	**0.827***	0.904	0.709	0.754	0.658	0.518	**0.462***	**0.514***	0.611	0.683	0.736	**0.806***	**0.848***	**0.966***	**1.261**
KK-logit	**0.886***	**0.827***	**0.745***	**0.701***	**0.668**	**0.566**	**0.487**	**0.462***	0.530	**0.582***	**0.646***	**0.727***	0.853	1.034	1.396	2.482

* indicates that the overall accuracy (IL) of the best forecasts, individual or combined, is significantly better than that of the naive model at the 1% level on the basis of a Diebold Mariano test. The naive forecast is 0 for θ<0.5, 1 for θ>0.5 and the in-sample most frequent event for θ=0.5.

Fig. 1 The Type I and Type II error and the cut-off rate

Panel A: The Trade-off Between the Probability of Type I and Type II Errors (LOGIT-M)



Panel B: The Overall Loss and the Cut-off Rate (LOGIT)

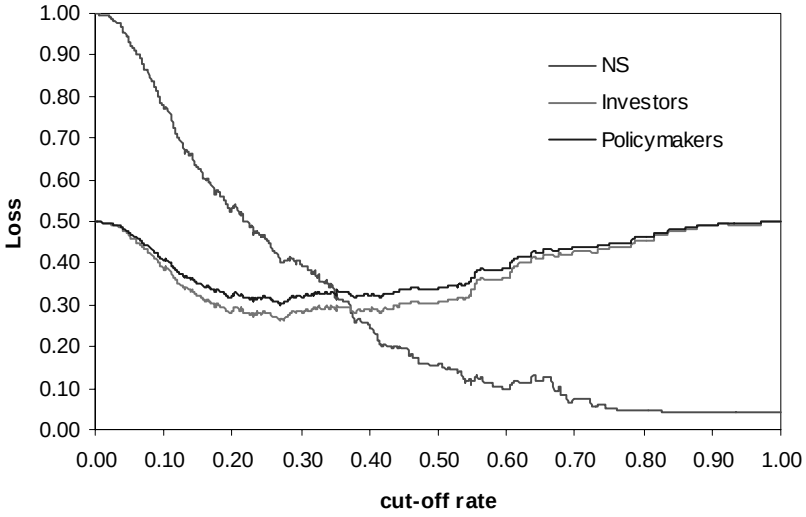
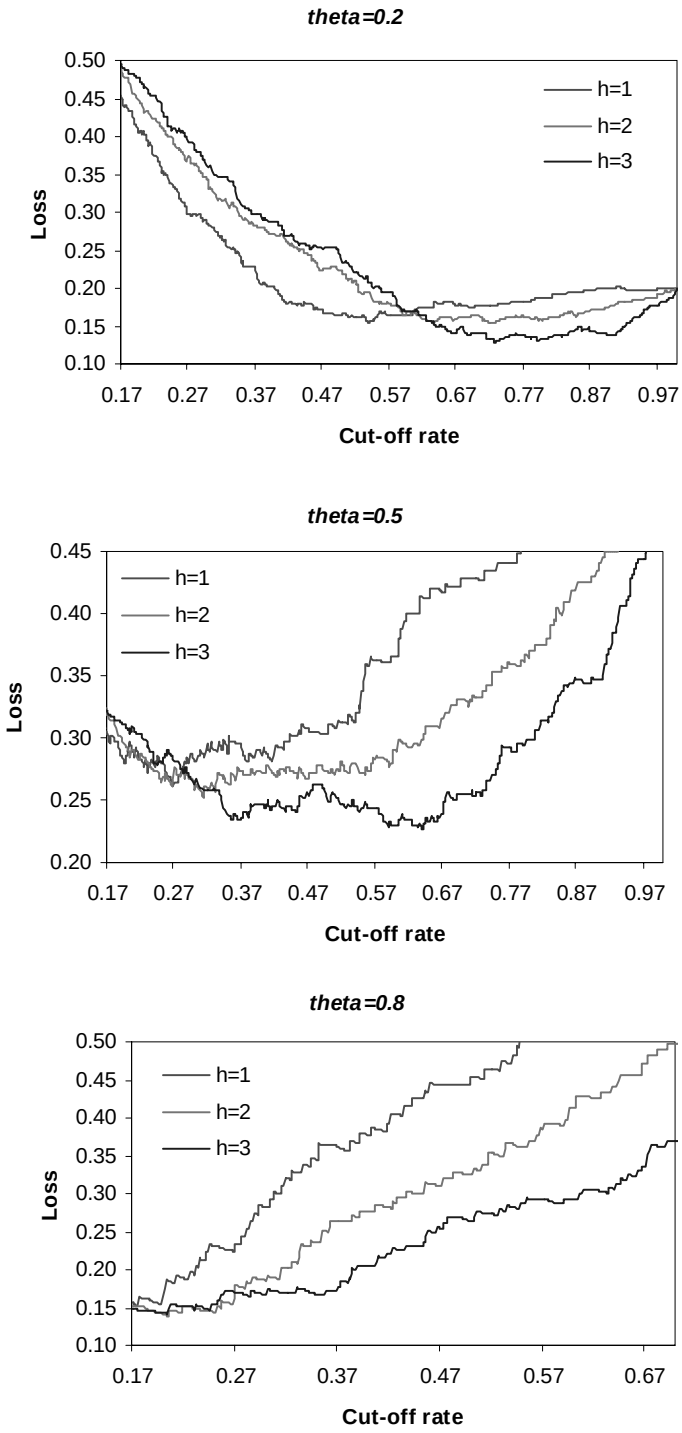


Fig. 2 The Overall Loss for Different Cut-off Rate and Warning Horizons (LOGIT-M)

Panel A: Investors loss function



Panel B: Policymakers Loss function

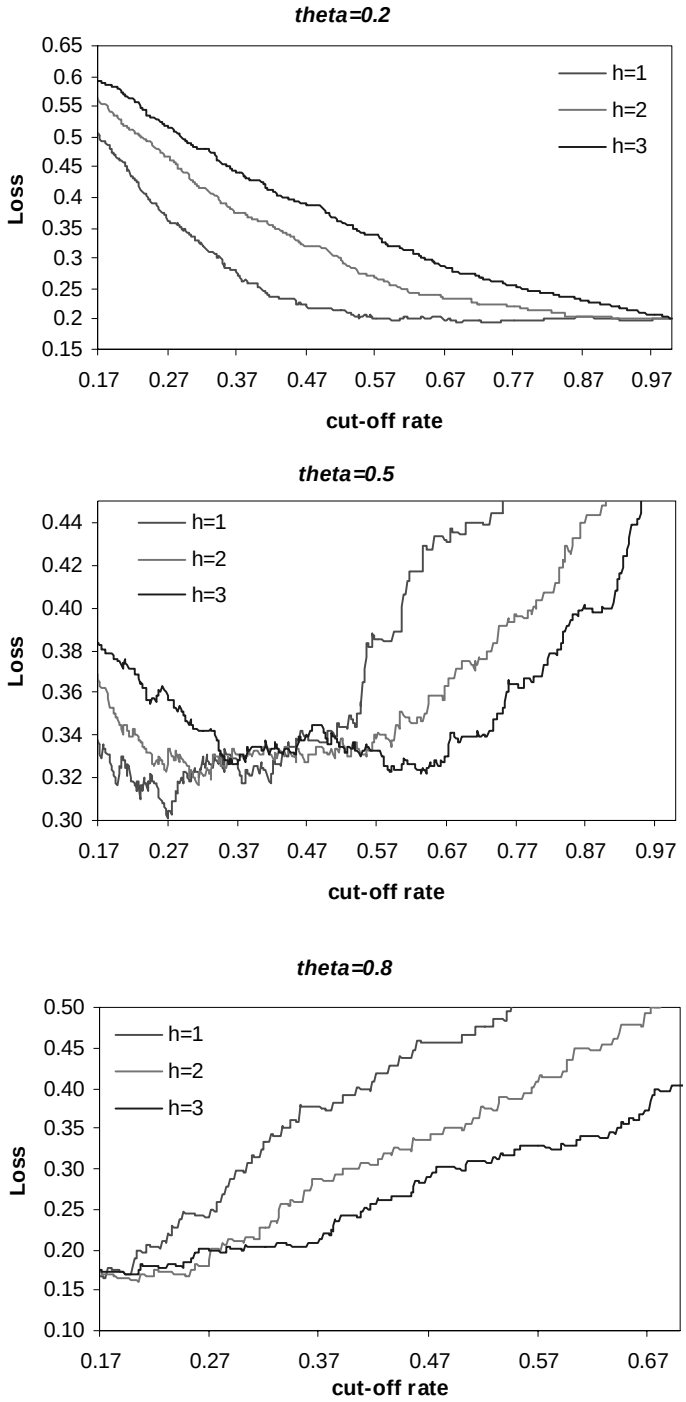
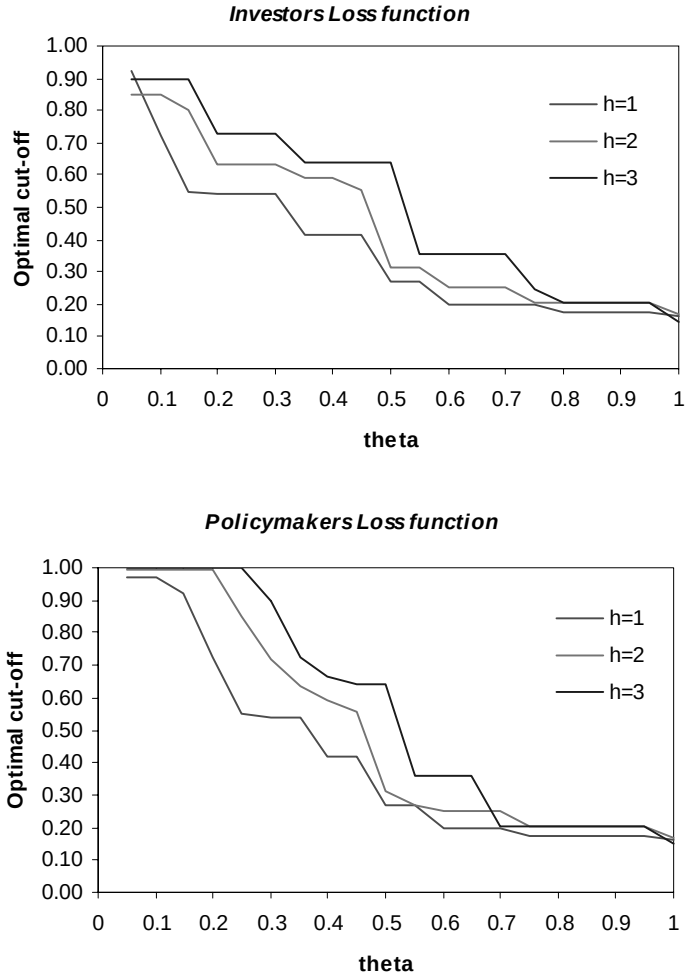
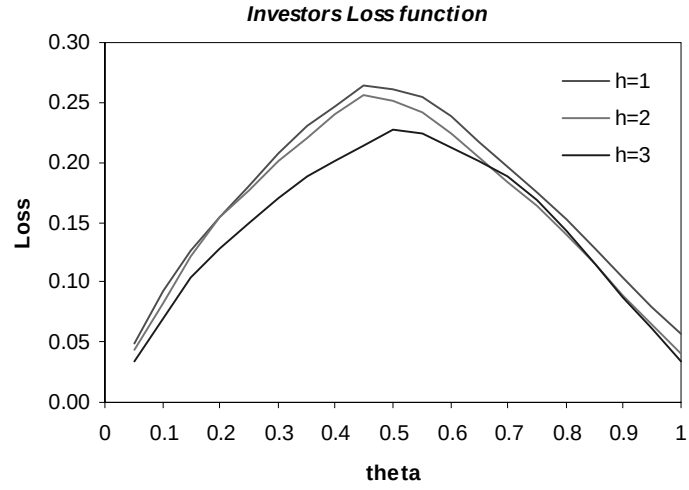


Fig. 3 The Cut-off Rate and the Warning Horizon

Panel A: Optimal Cut-off Rate for Different Warning Horizons (LOGIT-M)



Panel B: The Overall Losses for Different Warning Horizons (LOGIT-M)



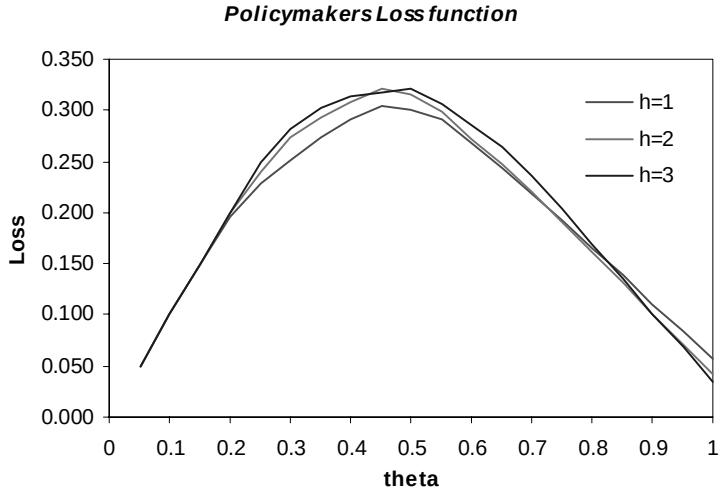
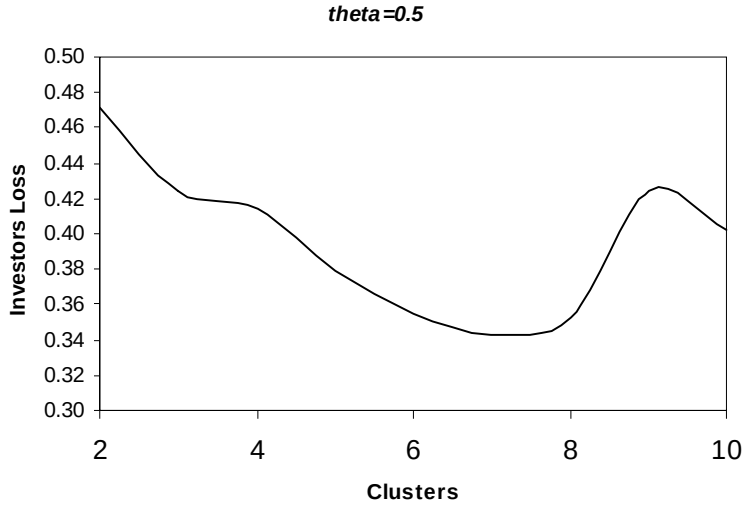


Fig. 4 Optimal Number of Clusters

Panel A: The investor's loss and the number of clusters



Panel B: The optimal number of clusters and the risk aversion level

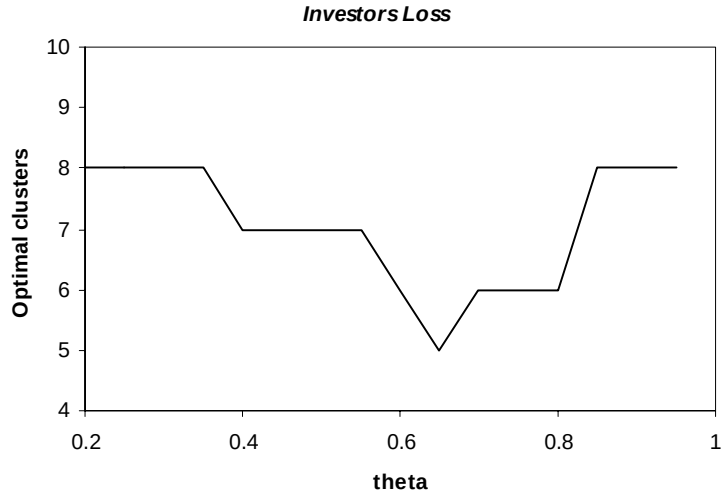
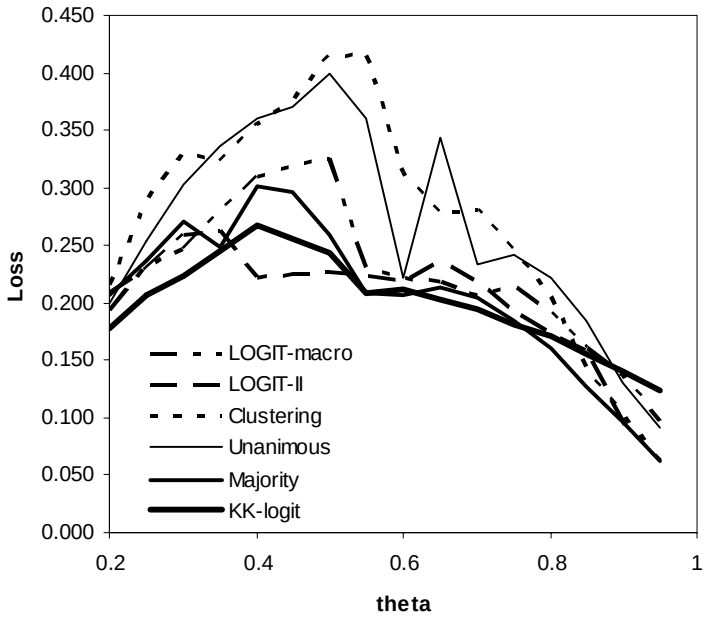


Fig. 5 Overall Loss for Individual and Combined Forecasts

Panel A: IL function



Panel B: PL function