The Role of Heterogeneity in Early Warning Systems for Sovereign Debt Crises*

Ana-Maria Fuertes, Elena Kalotychou†

Faculty of Finance, Cass Business School, City University
106 Bunhill Row, London EC1Y 8TZ

October 2005

Abstract

This paper compares rival sovereign default models that differ in how unobservable country, regional and time heterogeneity are treated. The analysis is based on panel logit specifications for a sample of 96 developing economies 1983-2002. Inference-based criteria and the plausibility of estimates overwhelmingly favour more complex models which allow the link between the default probability and the fundamentals to be time- and country-specific. An out-of-sample forecast evaluation exercise is conducted that draws on several loss functions, equal-predictive-ability tests and various naïve benchmarks. Simplicity beats complexity in forecasting. The parsimonious pooled logit model produces the most accurate forecasts and outperforms the various benchmarks.

Keywords:Credit risk; Default probability; Emerging markets; Loss function; Panel logit; Predictive performance; Unobserved heterogeneity.

JEL Classification: C15; C32; C33

*This is a revised version of the Cass Centre for Econometric Analysis discussion paper 11-04 and SSRN working paper 634061. We appreciate the helpful comments of Roy Batchelor, Chris Brooks, Jerry Coakley, William Greene, Ron Smith, Sotiris Staikouras, Giorgio Valente, Frank Packer at the *Bank for International Settlements,* Enrica Detragiache at the IMF, Marcel Peter at the IMF and participants at the 2004 *Society for Computational Economics* Conference at the University of Amsterdam, the 2004 *ESEM-EEA* Conference at Universidad Carlos III de Madrid, the 2004 *Australasian Finance and Banking* Conference at University of New South Wales in Sydney, and seminar participants at the University of Cambridge, Faculty of Economics and Politics. The usual disclaimer applies.

†Correspondence to: Elena Kalotychou, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, e-mail: e.kalotychou@city.ac.uk, Tel: +44 (0)20 7040 8647.

1. INTRODUCTION

A number of sovereign debt crises in recent decades have led to a new emphasis on credit risk which has become one of the most intensely researched areas in international finance. The available studies can be grouped into two broad types. One type exploits option pricing models while the other directly focuses on modeling default risk using structural models or panel discrete-choice models. This study belongs to the latter group. Financial institutions make use of default probability estimates to price loans and bonds, to determine adequate concentration limits and as inputs for Value-at-Risk analyses. This interest has been reinforced by the new Basel Capital Accord which allows banks to use internal ratings and default rates to decide their minimal regulatory capital.

A large empirical literature analyses the determinants of sovereign default, the results however are mixed. Some studies find that liquidity or global business cycle indicators are crucial and others not. The evidence on the importance of structural economic conditions also varies across studies. Most extant analyses build on pooled logit models for a large number of countries. However, the validity of the implicit full homogeneity assumption has been questioned.[1] McFadden *et al.* (1985) and Hajivassiliou (1987) point out that it may not be reasonable to expect the link between debt repayment performance and macroeconomic attributes to be stable across countries and over time. For instance, if a country has fewer capital controls or is more open than other countries, the coefficients on indicators of external economic activity may be larger. Unobservables such as colonial histories, types of government and religious institutions are other obvious potential sources of country heterogeneity. In other words, not-easily-quantifiable idiosyncracies may explain why country A defaults but country B services its debt while exhibiting similar economic fundamentals and debt structures. As Schleifer (2003; p.5) puts it: “Sovereign debt markets could not be more different”.

[1] The issue of whether homogeneous or heterogeneous (linear) models provide better forecasts has been examined in the context of US gasoline and cigarette demand (Baltagi and Griffin, 1997; Baltagi et al., 2000).

The few studies that control for country heterogeneity use either fixed or random effects models (Detragiache and Spilimbergo, 2001; Li, 1992; Oral *et al.*, 1992). Evidence from currency crisis studies suggests that the relevant heterogeneity occurs at the regional level (Burkart and Coudert, 2002; Staikouras, 2004; Kalotychou and Staikouras, 2005). For instance, different key leading indicators of currency crises have been identified for Asia and Latin America. In the context of sovereign default, regional differences have been captured using dummy variables (Feder *et al.*, 1981). On the other end, time heterogeneity may reflect the impact of changing world conditions such as the business cycle and the development of international capital markets or the changing nature of emerging markets themselves. Some studies include year-dummy variables (Aylward and Thorne, 1998) or global macroeconomic indicators such as OECD growth (Lee, 1991; Detragiache and Spilimbergo, 2001) to control for time effects that are assumed common across countries.

One question that has been ignored is how important such specification issues are from a forecasting viewpoint. This paper aims to fill this gap by providing a horse race among several panel logit specifications. In particular, we address the practical issue of whether controlling for latent country, regional or time heterogeneity in panel models of sovereign default helps to produce more accurate forecasts. The analysis is of interest to regulators, practitioners and the rating agencies who are most interested in the *when* rather than in the *why* question of default. Regulators rely on default forecasts to monitor the financial health of banks, pension funds and other financial institutions that include sovereign debt in their portfolios. Practitioners feed their theoretical or simulation models with default forecasts to price sovereign debt. Moreover, academic researchers use default forecasts to test various hypotheses such as the conjecture that country risk is priced in stock returns and borrowing costs. Finally, Early Warning Systems (EWSs) are recognized as a potentially fruitful complement to the wider analysis and judgement of decision-makers for identifying looming crises and for ranking the vulnerability of countries (Berg *et al.*, 1999). They have recently attracted considerable attention within central banks as a device to predict the likelihood that a country will face a debt crisis within a given time horizon. In this context, relevant economic indicators pertinent to domestic conditions, the international business-cycle and market sentiment are monitored and fed into statistical models. For further discussion of the models and empirical studies conducted so far within the EWS literature see Fuertes and Kalotychou (2004) and Berg *et al.* (2004).

Against this background, forecasting issues have surprisingly received only a broadbrush treatment in the literature. Most studies have compared sovereign default models just on the basis of their in-sample forecasts (Hajivassiliou, 1987; Detragiache and Spilimbergo, 2001). A few studies conduct out-of-sample evaluation but the forecasts are based on parameters estimated once and are limited to a 1- or 2-year holdout period. The typical forecast accuracy metrics used are the baseline Type I, Type II error or the overall error rates (Feder *et al.*, 1981; Manasse *et al.*, 2003; Oka, 2003; Peter, 2002).[2] Futhermore, those few studies that provide out-of-sample predictions do not confront them with simple benchmarks such as random walk type models. This is particularly important in the present context due to the persistence in debt-servicing behaviour.

This paper contributes to the literature in two respects. First, in order to investigate the effect of unobserved heterogeneity in sovereign-default models, it considers a wide range of logit specifications that differ in how they treat regional, country and time effects. We utilize statistical tests, information criteria and insights on the plausibility of the parameter estimates to gauge the models' ability to describe the data generating process. By capturing, say, time heterogeneity in different ways we seek to assess whether the time effects are genuine or merely an artefact of misspecification. Some of the specifications, such as the random coefficients that allow for time-dependent country-specific slopes and the models that allow for region- or time-specific slopes, have not been utilized in this context as yet. Three novel world variables in the present setting — macroeconomic uncertainty, monetary policy uncertainty and risk aversion — are included as

[2] In the debt crises literature, the Type I error rate refers to the missed defaults (or false negatives) over the realized defaults. The Type II error refers to the false alarms or false positives.

another way to control for time effects.

Second, a comprehensive forecasting analysis is conducted. A 12-year window is rolled forward to generate out-of-sample forecasts over a relatively large period of 5 years. A battery of forecasting tests is run to address the potential key role played by unobserved heterogeneity. Several forecast accuracy measures, probability scores as well as metrics that allow for asymmetric misclassification costs, are evaluated over the entire holdout set and over a positive-directional-change subset. The latter seeks to focus on the models' ability to predict a new crisis rather than a continuing default. Various uninformative benchmarks are considered including random walk type models and the naive models implicit in Pesaran-Timmermann's (1992) and Donkers-Melenberg's (2002) tests.

The statistical tests and model selection criteria indicate that the more complex specifications describe the data better. Unobserved heterogeneity across countries, regions and time is important in modeling sovereign default. By contrast, the forecast race suggests that the relatively parsimonious pooled logit model that imposes full homogeneity appears capable of yielding relatively good out-of-sample predictions and beating the benchmark models. Hence, our findings corroborate in a novel context the well-known limited relationship between in-sample fit and out-of-sample performance. Simple variants of the pooled logit model that allow for fixed regional- or common time-effects also beat the uninformative benchmarks but, interestingly, are unable to improve significantly upon the pooled logit model. Random coefficient specifications that allow for country and time variation in the link between the default-probability and the fundamentals forecast relatively worse than the pooled logit and generally underperform the benchmarks.

The paper is structured as follows. Section 2 describes the data and the endogenous default indicator. Section 3 outlines the models and the inference-based metrics. Section 4 discusses the forecast framework and Section 5 analyses the empirical results. A final section concludes.

2. THE DATA

The analysis is based on annual data (1983-2002) from the *World Bank* for 96 emerging markets and less developed countries from Africa, Asia, Eastern Europe, Latin America and the Middle East. Information on external long-term debt, arrears and rescheduling to official and private creditors is obtained from the Global Development Finance database. Time series for 24 economic and financial indicators are obtained from the World Development Indicators database (for details, see Appendix A).

2.1. Early Warning of Default (EWD) Indicator

What is a debt crisis? The question is not as simple as it may seem. A debt crisis means different credit events to different authors although all extant definitions seek to capture debt-servicing difficulties. There is a growing literature on the issue of which is the most appropriate proxy for the unobservable or latent variable behind debt-servicing difficulties. For instance, Pescatori and Sy (2004) suggest using bond spreads whereas Manasse *et al.* (2003) rely on the size of IMF loans. Our crisis definition follows a traditional strand of the literature that considers large arrears and rescheduling of long-term debt (Peter, 2002; Detragiache and Spilimbergo, 2001). This choice draws upon four arguments: a) data unavailability for arrears/rescheduling on short-term debt, b) bond spread data is not available before 1994 (or not at all) for many countries in our sample, c) large IMF loans are often granted to countries with balance-of-payment problems which do not necessarily coincide with sovereign debt problems,[3] and d) the debt crisis episodes thus identified match quite well the widely-accepted sovereign default events reported by leading rating agencies. Manasse *et al.* (2003) distinguish between three types of debt-servicing difficulties: outright default, semi-coercive (under the implicit threat of default) restructuring and roll-over or liquidity crises pertaining to maturing short term debt. Our crisis definition focuses on the former two.

[3] The post-1994 Mexican and Asian crises arguably are not defaults per se but currency or banking incidents. Investigating the causal relations among default, currency and banking crises goes beyond the scope of this study.

We categorize country i at year t as a debt-crisis ($d_{it} = 1$) or default case if: a) a jump in arrears (ΔA_{it}) exceeds a threshold percentage (δ) of total external debt (D_{it}), where δ is the sample mean of $\Delta A_{it}/D_{it}$ at 2.26%, or b) the total amount of debt rescheduled exceeds the decrease, if any, in total arrears. Accordingly, the default frequency in our sample is about 30% (see Appendix B for details). We identify 175 crisis episodes of which 59 occur post-1994, a period characterised by the rapid development of international capital markets. Our choice of δ follows Peter (2002), however, as a sensitivity check we set δ at one standard deviation from the mean with no substantial difference in the number of identified crises. Notably, the number of defaults per year is quite close to those identified by S&P (2001) for rated and non-rated debt.

The goal is to predict the probability that a debt crisis will occur at any time over a specific time window. As in Peter (2002) and Oka (2003) we adopt a 3-year warning window and define

$$y_{it} = \begin{cases} 1 & \text{if } d_{i,t+k} = 1 \text{ at any } k = 0, 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

A unit value for this forward-looking variable, called the Early Warning of Default (EWD) state, signifies that country i has defaulted at least once over $[t, t+2]$.

2.2. Country Fundamentals and Global Variables

Several macro and financial ratios have been arguably suggested in the literature as determinants of sovereign debt-servicing behaviour. For comprehensive surveys, see Manasse *et al.* (2003) and Heffernan (2004; ch.6). We initially consider 24 domestic signals, $\mathbf{x}_{it}^0$ from five World Bank categories: *i*) external credit exposure, *ii*) external economic activity and financial resources, *iii*) domestic conditions and *iv*) international financial links. The ratios are logged, $sign(x)\ell n(1 + |x|)$, and any remaining outlier is tackled by winsorization: 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. The limit of $\pm 4\hat{\sigma}^c$ is rather conservative to account for the large ratio variations in the heterogeneous country sample employed.

The literature has emphasized that, in addition to country fundamentals, world factors have an impact on the fluctuations of capital flows to emerging markets, and thus on country creditworthiness (Cantor and Packer, 1996; Arora and Cerisola, 2001; FitzGerald and Krolzig, 2003). Higher interest rates and lower capital availability (i.e. business cycle fluctuations) as well as changes in market sentiment and risk aversion in industrialized countries shift the demand of emerging market assets. This, in turn, influences capital flows and lowers FX reserves thereby affecting debt-servicing behaviour. We consider a world regressor vector, $\mathbf{z}_t^0$, that includes two typical variables — the 10-year US Treasury Bond yield (liquidity proxy) and OECD GDP growth (business cycle) — and three proxies for global conditions that are novel in the sovereign default context.

One is a measure of *macroeconomic uncertainty* obtained as the conditional variance of US monthly real GDP.[4] For this purpose, a AR(1)-GARCH(1,1) model is fitted to the first differenced log real GDP since the augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests indicate the presence of a unit root in the level. Second, we adopt a proxy for *monetary policy uncertainty* that is analogously derived from the monthly yield spread between the 3-month US T-Bill and the US Federal Funds target rate. The former is usually adopted as a benchmark for pricing other high-yield assets in world capital markets and is most likely to reflect expected changes in economic conditions whereas the latter is a reasonable barometer of US monetary policy. According to the ADF and PP tests the spread is stationary and so an appropriate AR(1)-GARCH(2,1) is fitted to the levels for this purpose.[5] Arora and Cerisola (2001) motivate this volatility measure as a proxy for changes in the expected stance of US monetary policy and demonstrate that it is positively related to sovereign spreads. Third, the level of *global risk aversion* is proxied by the Sharpe ratio — the monthly average high-yield spread divided by its standard deviation over the last 12 months — based on the Merryl Lynch 175 US Corporate High-Yield index and the 10-year US T-Bond

[4] Quarterly GDP was interpolated into monthly GDP on the basis of the US monthly industrial production using the proportional Denton approach that belongs to a family of LS-based benchmarking methods (Baum, 2001).

[5] GARCH orders are selected using the Ljung-Box test on the squared residuals. Details available upon request.

yield. In so doing, we follow FitzGerald and Krolzig (2003) who argue in favour of this ratio as a conceptually appropriate measure of risk aversion on the basis of a theoretical model and show that it is inversely related to bond flows. All three measures are annualized.

Large models typically have poor statistical properties. In order to preserve degrees of freedom, we pool the data and deploy a cross-validation (jacknife) approach which assesses the relative merit of each regressor on the basis of the in-sample missed default rate (see Appendix C).[6] The retained domestic and global signals are denoted by $\mathbf{x}_{it}$ and $\mathbf{z}_t$, respectively. These are discussed below in Section 5.

3. MODELS AND ESTIMATION

Let the observed EWD indicator, y_{it}, be influenced by a set of exogenous factors as follows

$$y_{it}^* = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\gamma} + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0, \sigma^2), \quad i = 1, ..., N, \quad t = 1, ..., T \tag{1}$$

where y_{it}^* is the latent index such that $y_{it} = 1$ for $y_{it}^* > 0$ and $y_{it} = 0$ otherwise. The noise ε_{it} is assumed independently distributed from the k domestic regressors ($\mathbf{x}_{it}$) and the r world regressors ($\mathbf{z}_t$). We have $p_{it} \equiv \Pr(y_{it} = 1|\mathbf{x}_{it}, \mathbf{z}_t) = \Pr(y_{it}^* > 0)$ and assuming a standard logistic distribution for ε_{it}, then $p_{it} = G(\mathbf{x}_{it}, \mathbf{z}_t) = \frac{\exp(\alpha + \mathbf{x}_{it}'\beta + \mathbf{z}_t'\gamma)}{1+\exp(\alpha + \mathbf{x}_{it}'\beta + \mathbf{z}_t'\gamma)}$. So the response probability is the logit function evaluated at a linear function of $(\mathbf{x}_{it}, \mathbf{z}_t)$.[7] This nonlinear relation can be rewritten linearly for the log-odds ratio as $\ln \frac{p_{it}}{1-p_{it}} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_t'\gamma$. Equation (1) is referred to as the baseline *pooled logit*

[6] Instead one could gear the jacknife toward some other criteria (e.g. the overall error rate) or models which may, of course, lead to a different regressor set. For reasons of tractability, we assume that the choice of $(\mathbf{x}_{it}, \mathbf{z}_t)$ from the first-stage jacknife procedure is independent of the preferred model resulting from the subsequent evaluation of fit and forecast performance.

[7] Both a standard normal and a standard logistic variable have a zero mean but the latter has a variance of $\pi^2/3$. Because the two pdfs are very similar (the logit density has more mass in the tails), if one corrects for the difference in scaling, the probit and logit models typically yield similar results in applied work. The main competitor to logit for classification is discriminant analysis. However, the latter assumes that the country's characteristics are multivariate normally distributed with a different mean vector (but identical variance-covariance matrix) associated to the default and non-default states. Most studies have concluded that logit is superior to discriminant analysis mainly because this normality assumption for the regressors is unrealistic (see Kennedy, 2003; ch. 15).

model (PLOGIT) that assumes full country and time homogeneity in the response y_{it}^* to $(\mathbf{x}_{it}, \mathbf{z}_t)$. The $(1+k+r) \times 1$ coefficient vector $(\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}')'$ is estimated by Maximum Likelihood (ML).

3.1. Country-specific Heterogeneity

The PLOGIT can be extended to allow for unobserved country-specific effects α_i that remain constant over time, e.g. some countries are more likely to default than others in every period. The *fixed effects* model (FE) treats α_i as fixed and so there are $(N+k+r)$ unknown coefficients $\phi^{FE} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}')'$ where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$ are country-specific constants. The error components or *random effects* model (RE) treats α_i as independent random draws from the same distribution with mean α and variance σ_α^2. Formally, $\alpha_i = \alpha + \sigma_\alpha v_i$ where $v_i \sim iid(0,1)$ is independent of $(\mathbf{x}_{it}, \mathbf{z}_t)$. Alternatively, it can be formalized as equation (1) with the composite error $e_{it} = \alpha_i + \varepsilon_{it}$. The $(2+k+r) \times 1$ parameter vector to be estimated is $\phi^{RE} = (\alpha, \sigma_\alpha; \boldsymbol{\beta}', \boldsymbol{\gamma}')'$.

Dependence between α_i and $(\mathbf{x}_{it}, \mathbf{z}_t)$ does not render $\hat{\phi}^{FE}$ inconsistent but the FE model is bedevilled by two issues. One is that the incidental parameters problem — inconsistency of $\hat{\alpha}_i$ for $N \to \infty$ and finite T — is transmitted into the slopes. This problem does not appear in the linear model because the α_i are effectively removed by using data in country-mean deviations. To avoid it, Chamberlain's (1980) conditional ML (CML) estimator integrates the α_i out of the joint density by conditioning on $\sum_t y_{it}$. But then $\hat{\alpha}_i$ cannot be computed nor, in turn, the forecasts $\hat{p}_{it}$. The second problem arises from the fact that the FE model is only identified through the 'within' dimension of the data. If country i has the same status (y_{it}) in every period because, say, it has never experienced default, it is discarded in estimation. This may induce sample selection bias.

The *random coefficients* specification (RC) goes one step further by introducing random country-heterogeneity both in intercepts and slopes. We consider two variants. First, a model (denoted RC$^\beta$) that allows for the link between the domestic signals and the probability to be heterogeneous

$$y_{it}^* = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \mathbf{z}_t'\boldsymbol{\gamma} + \varepsilon_{it}, \ \varepsilon_{it} \sim iid(0, \sigma^2), \ i = 1, ..., N, \ t = 1, ..., T \tag{2}$$

where $\boldsymbol{\delta}_i = (\alpha_i, \boldsymbol{\beta}_i')'$ is a random vector with $E(\boldsymbol{\delta}_i) = (\alpha, \boldsymbol{\beta}')'$ and diagonal covariance matrix $E(\widetilde{\boldsymbol{\delta}}_i \widetilde{\boldsymbol{\delta}}_i') = \Omega$ with $\widetilde{\boldsymbol{\delta}}_i = \boldsymbol{\delta}_i - E(\boldsymbol{\delta}_i)$ and $diag(\Omega) = \{\sigma_\alpha, \sigma_{\beta_1}, ..., \sigma_{\beta_k}\}$. Equivalently, $\widetilde{\boldsymbol{\delta}}_i \equiv \Gamma \mathbf{v}_i$ where Γ is a diagonal matrix such that $\Gamma\Gamma' = \Omega$ and $\mathbf{v}_i$ are $(k+1)$ latent random terms which are $iid(0,1)$ and independent of $(\mathbf{x}_{it}, \mathbf{z}_t)$. These distributional assumptions are introduced for tractability purposes. Allowing for a more general coefficients covariance matrix (i.e. non-diagonal) is unfeasible given the typical time dimension (T) of sovereign debt-crises samples and the binary nature of the dependent variable. The $(2 + 2k + r)$ unknown parameters are $(\alpha, \sigma_\alpha; \boldsymbol{\beta}', \sigma_{\beta 1}, ..., \sigma_{\beta k}; \boldsymbol{\gamma}')$.

Second, we consider a RC$^\gamma$ model where the link between the global signals $\mathbf{z}_t$ and the response probability is country heterogeneous — equation (1) with the random vector $\boldsymbol{\delta}_i = (\alpha_i, \boldsymbol{\gamma}_i')'$. The $(2 + k + 2r)$ parameters to estimate are $(\alpha, \sigma_\alpha; \boldsymbol{\beta}'; \boldsymbol{\gamma}', \sigma_{\gamma_1}, \sigma_{\gamma_2}, \sigma_{\gamma_3})$. The reason for specifying random heterogeneity on the effects of global signals (RC$^\gamma$) and of domestic signals (RC$^\beta$) separately is degrees of freedom and estimation tractability rather than theory. The loss of information when moving from continuous to binary variables requires long time-series to allow for reliable random coefficients estimation on $\boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}_i$ simultaneously. It is worth noting that neither the RE nor the RC models (in contrast with FE) rely on large T for consistency. The FE logit is estimated by (C)ML whereas the RE, RC$^\beta$ and RC$^\gamma$ are estimated by maximum simulated likelihood (MSL).[8] The ML estimator for the FE model is reported only for the sake of completeness in the forecast exercise ($\hat{\alpha}_i$ is needed to generate the forecasts), but the results should be treated with caution as the α_i and β estimators are likely to be inconsistent.

3.2. Time-specific Heterogeneity

Equation (1) controls for the common time effects (e.g. market sentiment) by means of the global

[8] There is no closed form for the log-likelihood of the RC model. MSL involves draws from the multivariate density of $\mathbf{v}_i$. Bhat (1999) suggests $R = 1000$ draws and shows that a smaller number of Halton draws, $H = R/10$, is equally effective and cheaper. Our MSL uses a standard normal and $H = 500$. For the RE model this is asymptotically equivalent to the Hermite quadrature approach for approximating the likelihood (see Greene, 2003).

signals $\mathbf{z}_t$. We consider also a *fixed time effects* (FTE) model that uses period dummies instead

$$y_{it}^* = \alpha_t + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0,\sigma^2),\ i = 1,...,N,\ t = 1,...,T \tag{3}$$

where the $(T+k)\times 1$ vector $\phi^{FTE} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ with $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_T)'$ is estimated by ML.

Alternatively, the data can be conceptualized as a *time sequence of cross-section* (TCS) relations

$$y_{it}^* = \alpha_t + \mathbf{x}_{it}'\boldsymbol{\beta}_t + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0,\sigma^2),\ i = 1,...,N \tag{4}$$

for $t = 1,...,T$. The elements of the $T(1+k)\times 1$ vector $(\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ where $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_T)'$ and $\boldsymbol{\beta}= (\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_T)'$ are obtained sequentially by ML. This approach allows for time variation in the intercept and slopes. Let $\hat{\beta}_{jt}$ denote the slope estimate of regressor j at period t. For forecasting purposes, we define the *mean time* (MT) estimator $\bar{\beta}_j^{MT} \equiv (1/T)\sum_{t=1}^{T}\hat{\beta}_{jt}$ with standard error $SE(\bar{\beta}_j^{MT}) = \sqrt{\frac{SD(\hat{\beta}_{jt})^2}{T}}$ where SD denotes the sample standard deviation.[9] This estimator is just the time counterpart of Pesaran and Smith's (1995) mean group estimator.[10] We should note that the MT estimator provides a measure of $\boldsymbol{\beta} \equiv E(\boldsymbol{\beta}_t)$ whereas the above pooled time-series estimators measure $\boldsymbol{\beta} \equiv E(\boldsymbol{\beta}_i)$. A consensus view is that cross-section data models capture long run relations (Pesaran and Smith, 1995; Kennedy, 2003).

3.3. Region-specific Heterogeneity

In order to control for region-specific heterogeneity, each country is allocated into one of four groups: I) Asia ($N_{\mathrm{I}} = 17$), II) Latin America ($N_{\mathrm{II}} = 26$), III) Africa ($N_{\mathrm{III}} = 36$), IV) East Europe/Middle East/North Africa ($N_{\mathrm{IV}} = 17$).[11] We consider two approaches. First, the pooled equation

$$y_{it,j}^* = \alpha_j + \mathbf{x}_{it}'\boldsymbol{\beta}_j + \varepsilon_{it,j},\ \varepsilon_{it,j} \sim iid(0,\sigma^2),\quad i = 1,...,N_j,\quad t = 1,...,T \tag{5}$$

[9] Alternatively, the forecasts could just be based on the latest available estimates $(\hat{\alpha}_T, \hat{\boldsymbol{\beta}}_T)'$. However, one potential advantage of averaging (MT approach) is that it helps to mitigate the noise in the year-by-year estimates.

[10] If the slopes are random (orthogonal to $\mathbf{x}_{it}$) then $\hat{\beta}_{jt} \to \beta_{jt}$ as $N \to \infty$ and then $\bar{\beta}_j^{MT}$ is consistent as $T \to \infty$.

[11] There are not enough degrees of freedom in the logit model estimation for any of these 3 regions so we group them. They have in common: i) a similar structure of exports given their oil exporting nature, ii) having gained access to international bond markets between 1995-98. Note that introduction of time-heterogeneity (fixed time effects) in the regional models is also precluded due to degrees of freedom.

is fitted to regions $j =$ I,...,IV. This *regional logit* (RLOGIT) with a $4(1+k) \times 1$ parameter vector $(\alpha_j, \boldsymbol{\beta}_j)$ can be seen as treating the regional heterogeneity in intercept and slopes as fixed. Second, a *regional regressor-specific* (RSLOGIT) model with $4 + \sum_j k_j$ parameters is considered where a distinct regressor set, $k_j \leq k$, is allowed for $j =$ I,...,IV.

3.4. Time-dependent Country Heterogeneity

Next we relax the assumption that the random country effects (in intercepts and/or slopes) are time invariant. More specifically, the RC$^{\beta}$ and RC$^{\gamma}$ models are generalized by allowing the coefficients to be time-dependent according to an AR(1) mechanism. Thus we have the RC$^{\beta}$-AR model

$$y_{it}^{*} = \alpha_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}_{it} + \mathbf{z}_t'\boldsymbol{\gamma} + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0, \sigma^2) \tag{6}$$

where $\alpha_{it} = \alpha + \sigma_\alpha v_{it}^{\alpha}$ with $v_{it}^{\alpha} = \rho_\alpha v_{i,t-1}^{\alpha} + e_{it}$, $e_{it} \sim iid(0,1)$ so that $E(\alpha_{it}) = \alpha$ and $V(\alpha_{it}) = \frac{\sigma_\alpha^2}{1-\rho_\alpha^2}$; likewise for $\boldsymbol{\beta}_{it}$. The RC$^{\gamma}$-AR model is analogously formulated. These RC-AR formulations allow for the impact of the fundamentals on the probability of default to vary both across countries and over time. The $(3 + 3k + r)$ parameters of the RC$^{\beta}$-AR logit and the $(3 + k + 3r)$ parameters of the RC$^{\gamma}$-AR counterpart are estimated by MSL.

3.5. Inference-based Metrics for Model Selection

Several metrics are employed to assess model adequacy. First, we use the AIC and BIC which have been shown by Monte Carlo simulation to have good finite-sample properties for a range of panel models (Hsiao and Sun, 2000). A ranking is thus obtained based on $AIC = -MLL + s$ and $BIC = -MLL + 0.5s\ln(NT)$ where s is the number of unknown parameters.

Statistical tests are also deployed. A Hausman test serves to compare the FE and RE by means of the statistic $H = \mathbf{q}'\{V(\mathbf{q})\}^{-1}\mathbf{q} \overset{a}{\sim} \chi^2_{(s)}$ where $\hat{\mathbf{q}} = (\hat{\boldsymbol{\theta}}^{FE} - \hat{\boldsymbol{\theta}}^{RE})$, $V(\hat{\mathbf{q}}) = V(\hat{\boldsymbol{\theta}}^{FE}) - V(\hat{\boldsymbol{\theta}}^{RE})$ and s is the dimension of the slope vector $\boldsymbol{\theta}$. The null is $\hat{\mathbf{q}} = 0$ and a rejection suggests that there are fixed effects and so the RE model is inconsistent. This test can confront any two models such that both are consistent under the null but only the less efficient is consistent under the alternative.

The PLOGIT, RE, RC and RC-AR models are nested. For instance, under $H_0 : \sigma_\alpha = 0$ the RE collapses to the PLOGIT and thus latent country heterogeneity can be tested by a LR statistic (a counterpart of Breusch and Pagan's LM statistic) which is $\chi^2_{(1)}$ distributed. Likewise, the restrictions $\sigma_{\beta_1} = ... = \sigma_{\beta_k} = 0$ (homogeneous slopes) reduce RC$^\beta$ to RE. Under $\rho_\alpha = \rho_{\beta_1} = ... = \rho_{\beta_k} = 0$ (no time effects), RC$^\beta$-AR collapses to RC$^\beta$. The presence of time effects can be tested in the FTE model ($H_0 : \alpha_t = \alpha$) with a LR statistic that has a limit $\chi^2_{(T-1)}$ distribution for large N and finite T. We also test for $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}$ (and $\alpha_t = \alpha, \boldsymbol{\beta}_t = \boldsymbol{\beta}$) in the TCS model by noting that $MLL_{TCS} = \sum_t MLL_{CS_t}$. Regional poolability, $H_0 : \alpha_j = \alpha, \boldsymbol{\beta}_j = \boldsymbol{\beta}$ for $j =$ I,...,IV, is assessed via a LR statistic which follows a $\chi^2_{3(k_j+1)}$.

4. FORECAST FRAMEWORK

A contemporaneous relationship was presented in Section 3 to simplify the exposition. In our analysis 1983-2002 the regressors are lagged one year for forecasting purposes and to mitigate endogeneity bias. In effect, the forward-looking EWD indicator y_{it} refers to $t = 1984, ..., 2000$. Our panel is unbalanced but we refer to the time span as $[1, T]$ for simplicity. The models are estimated over a 12-year window, denoted $[1, T^*]$, and y_{i,T^*+1} is forecasted. This window is rolled forward. Out-of-sample predictions are thus constructed over a period $[T^* + 1, T]$ of $m = 5$ years (1996-2000) for $N = 96$ countries. This facilitates a relatively large holdout sample for the forecast evaluation exercise.[12]

The probability forecasts from, say, the PLOGIT model are $\hat{p}_{i,\tau+1}$, such that $\ln \frac{\hat{p}_{i,\tau+1}}{1-\hat{p}_{i,\tau+1}} = \hat{y}^*_{i,\tau+1}$ and $\hat{y}^*_{i,\tau+1} \equiv \hat{\alpha}_\tau + \mathbf{x}'_{i\tau}\hat{\boldsymbol{\beta}}_\tau + \mathbf{z}'_\tau\hat{\boldsymbol{\gamma}}_\tau$, obtained over $[\tau - T^* + 1, \tau]$ recursively for $\tau = T^*, T^* + 1, ..., T - 1$. To forecast on the basis of the TCS model we recursively compute the MT estimates $\bar{\alpha}_\tau = (1/T^*) \sum_{t=\tau-T^*+1}^{\tau} \hat{\alpha}_t$ and $\bar{\beta}_\tau = (1/T^*) \sum_{t=\tau-T^*+1}^{\tau} \hat{\beta}_t$ and then construct $\hat{y}^*_{i,\tau+1} = \bar{\alpha}_\tau + \mathbf{x}'_{i\tau}\bar{\boldsymbol{\beta}}_\tau$.

[12] Due to missing data we have an unbalanced panel with N_t, $t = 1, ..., m$ or, equivalently, m_i for $i = 1, ..., N$. The ith country forecast loss over the holdout window is computed first, $\bar{L}_i = \frac{1}{m_i} \sum_{t=1}^{m_i} L(y_{it}, \hat{y}_{it})$ and then the overall loss, $\bar{L}$, is obtained by averaging the latter over countries.

The probability $\hat{p}_{i,\tau+1}$ is transformed into an event forecast ($\hat{y}_{i,\tau+1} = 0, 1$) using a cut-off λ_τ which is optimally chosen for each model and loss function (see Fuertes and Kalotychou, 2004).[13]

We adopt a range of forecast metrics which are evaluated both over the Nm points in the *holdout* sample and over a subset called *positive-directional-change* (PDC) sample. The latter excludes year t for country i if $d_{i,t-1} = 1$ so as to focus on the models' ability to predict new defaults or entries to default rather than the persistence in debt-servicing behaviour (see Appendix B).[14]

4.1. Loss Functions

Let the following *pay-off matrix* summarise the decision-making problem at hand[15]

		Actual state	
		$y_{it} = 0$	$y_{it} = 1$
Decision	$\hat{y}_{it} = 0$	ϕ_t^0	θ_t^1
	$\hat{y}_{it} = 1$	θ_t^0	ϕ_t^1

where $\theta_t^j > \phi_t^j, j = 0, 1$; θ_t^1 is the economic loss of a missed default and so forth. We build on Granger and Pesaran's (2000) framework but make three simplifying assumptions: *a*) the cost of a correct forecast is zero, $\phi_t^0 = \phi_t^1 = 0$, *b*) the cost of an incorrect forecast is constant over the holdout period, $\theta_t^0 = \theta^0$, $\theta_t^1 = \theta^1$ and *c*) the forecaster is able to ascertain the relative penalty level $\theta = \frac{\theta^1}{\theta^1+\theta^0}$ that appropriately characterizes the decision-maker or forecast user.

Let $I(\cdot)$ denote an indicator function and λ_t an optimally chosen cut-off over each rolling window. The following *weighted misclassification rate* (WMR) metric

$$WMR_\theta = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m}\theta y_{it}\{1 - I(\hat{p}_{it} > \lambda_t)\} + (1-\theta)\{1 - y_{it}\}I(\hat{p}_{it} > \lambda_t) \quad WMR \in [0, 1] \quad (7)$$

[13] Extant studies use $\lambda = 0.5$ or fix it at the default frequency or at the value that minimizes the Type I and II error sum. For the τ^{th} rolling window, we find the λ_τ that minimizes the chosen loss function and so forth. Note that for tractability reasons, in the first stage of the analysis designed to select variates in $G(\cdot)$, we set λ_τ at 0.5.

[14] In this paper, we do not consider lagged dependent variables (e.g. $y_{i,t-1}$) as regressors. The incidental parameters problem becomes more severe in such dynamic models. The need to integrate out the α_i, in turn, prompts the initial conditions problem (see Greene, 2003; ch. 21). Modeling dynamic effects and initial conditions in panel binary choice models is more complex than in the linear model (see Honoré and Kyriazidou, 2000). By focusing on a default-entry (PDC) validation sample, we obviate the need for the latter.

[15] Granger and Pesaran (2000) define the economic cost of a decision based on the forecast $\hat{p}_{it}$ as $C_{it}(\hat{p}_{it}) = \phi_t^1 y_{it} I(\hat{p}_{it} > \lambda_t) + \theta_t^0(1 - y_{it})I(\hat{p}_{it} > \lambda_t) + \theta_t^1 y_{it}(1 - I(\hat{p}_{it} > \lambda_t)) + \phi_t^0(1 - y_{it})(1 - I(\hat{p}_{it} > \lambda_t))$.

provides a family of economic loss functions for $\theta \in [0,1]$. Each of them gives the overall cost associated to the model predictions ($\hat{p}_{it}$) for a particular decision-maker whose risk-aversion level towards missing a default is θ. The forecast ranking from the widely used *misclassification rate* (MR)

$$MR = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} y_{it}\{1 - I(\hat{p}_{it} > \lambda_t)\} + \{1 - y_{it}\}I(\hat{p}_{it} > \lambda_t), \quad MR \in [0,1] \tag{8}$$

amounts to that from (7) for $\theta = 0.5$ (i.e. $MR = 2WMR_{0.5}$). The latter defines the overall loss as the frequency of incorrect predictions and so it identically penalises missed defaults and false alarms. The hit rate, $HR = 1 - MR = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m}\left[y_{it} \times \hat{y}_{it} + (1 - y_{it}) \times (1 - \hat{y}_{it})\right]$, is the positively orientated version of (8). In practice, these errors do not have the same importance. From investors' viewpoint, misjudging a highly-risky loan may imply a substantial fall in assets, realised losses and reserves increase whereas incorrectly dismissing a good lending opportunity entails a foregone profit. From policymakers' viewpoint, some of the notable repercussions of default are increased borrowing costs, lost reputation and trade/credit sanctions. Hence, in the present context it seems reasonable to assume that on average the cost of a missed crisis is typically higher than that of a false alarm (for further discussion, see Berg *et al.*, 1999). In the literature, Sommerville and Taffler (1995) and Taffler and Abassi (1984) assume a cost-ratio of 3:1 whereas Sommerville (1991) proposes 3.75:1. For completeness we consider several cases, $\theta \in \{0.8, 0.5, 0.2\}$, but $\text{WMR}_{0.8}$ that implies a cost-ratio of 4:1 is clearly the most plausible.

Scoring rules are another set of criteria that directly evaluate probability forecasts and thus do not require λ_t. One is the *quadratic probability score* (QPS) or the Brier score which resembles but is not the direct counterpart of the MSE because it does not compare the event y_{it} with $\hat{y}_{it}$

$$QPS = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} 2(\hat{p}_{it} - y_{it})^2, \quad QPS \in [0,2] \tag{9}$$

Second, the *logarithmic probability score* (LPS) defines the overall loss as

$$LPS = -\frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} y_{it}\ln(\hat{p}_{it}) + (1 - y_{it})\ln(1 - \hat{p}_{it}), \quad LPS \in [0,\infty) \tag{10}$$

and so it penalizes large errors more heavily than the QPS. The functions in (7) and (8) are economic measures of forecast accuracy; conversely, those in (9) and (10) are statistical measures. As pointed out in Granger and Pesaran (2000), in general a simple one to one relationship between the two approaches is not available.

The loss function implicit in (7) accounts for the distinct preferences of different forecast users via the risk-aversion parameter θ. Ideally, one should adopt a decision-based approach to forecast evaluation which is the subject of a growing literature in empirical finance although far less widespread in economics. This approach is, however, not straightforward in the present context because it requires a complete specification of the decision environment of forecast users. Moreover, the statistical theory for decision-based methods to forecasting discrete variables is still not fully developed. For an interesting overview and discussion and an application see Pesaran and Skouras (2002) and Abhyankar *et al.* (2005), respectively.

4.2. Forecast Accuracy Tests

In order to compare rival forecasts, we deploy the Diebold-Mariano (1995) [DM] test. This approach has been shown to be robust to non-normality of the forecast errors and to be applicable to a wide class of loss functions for continuous or binary forecasts (see Diebold and López, 1996; López, 2001). Let $e_{it} \equiv L(y_{it}, \hat{y}_{it}^{A}) - L(y_{it}, \hat{y}_{it}^{B}), i = 1, ..., N, t = 1, ..., m$ denote the out-of-sample loss differential for models A and B. The test statistic is

$$DM = \frac{\bar{e}}{\sqrt{\hat{f}/N}} \overset{a}{\sim} N(0,1) \tag{11}$$

where $\bar{e} = \frac{1}{N}\sum_i \bar{e}_i$ and $\hat{f}/N$ is an estimate of the variance of $\bar{e}$ that accounts for time dependence.[16]

[16] We have the country-mean loss $\bar{e}_i = \frac{1}{m}\sum_t e_{it}$ with variance $\hat{f} \equiv V(\bar{e}_i) = \frac{1}{m^2}\sum_t V(e_{it}) + \frac{2}{m(m-1)}\sum_t \sum_{s>t} cov(e_{it}, e_{is})$ where $V(e_{it}) = \frac{1}{N-1}\sum_i (e_{it} - \bar{e}_t)$ for $t = 1, ..., m$ and $cov(e_{it}, e_{is}) = \frac{1}{N-1}\sum_i (e_{it} - \bar{e}_t)(e_{is} - \bar{e}_s)$. We also deployed the test by computing $DM_t = \frac{\bar{e}_t}{\sqrt{\hat{g}_t/N}}$ where $\bar{e}_t = \frac{1}{N}\sum_{i=1}^{N} e_{it}$ and $\hat{g}_t = V(e_{it})$. If dependence between DM_t and DM_s $(t \neq s)$ is assumed, then $DM = \frac{1}{m}\sum_{t=1}^{m} DM_t \sim N(0, \frac{1}{m})$. The results from the latter are qualitatively similar to those reported from (11) but the statistics are slightly higher. Finally, we

Pesaran and Timmermann (1992) [PT] propose a nonparametric approach to test the null hypothesis that forecasts and realizations are independent. The underlying idea is that the total number of correct out-of-sample predictions (Nm times the hit rate) can be treated as a binomial random variable with mean $Nm\tilde{p}$ and variance $Nm\tilde{p}(1-\tilde{p})$ where $\tilde{p} = \Pr(\hat{y}_{it} = 1, y_{it} = 1) + \Pr(\hat{y}_{it} = 0, y_{it} = 0)$. Under the null, $\tilde{p} = \hat{P}P + (1-\hat{P})(1-P)$ where $P \equiv \Pr(y_{it} = 1) = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} y_{it}$ and $\hat{P} \equiv \Pr(\hat{y}_{it} = 1) = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} \hat{y}_{it}$ are the unconditional probability of observed and forecasted EWD states, respectively. Thus we have

$$PT = \frac{(Nm)HR - (Nm)\tilde{p}}{\sqrt{(Nm)\tilde{p}(1-\tilde{p})}} = \frac{HR - HR^{PT}}{\sqrt{(Nm)^{-1}\tilde{p}(1-\tilde{p})}} \overset{a}{\sim} N(0,1) \tag{12}$$

where $HR^{PT} \equiv \tilde{p}$ is the hit rate of the model under H_0. A significant PT statistic suggests that the forecasts are dependent on the quantities to be predicted. Equivalently, predictive dependence amounts to the model's hit rate (HR) exceeding that of an implicit benchmark (HR^{PT}) that predicts 1 randomly with probability $\hat{P}$. However, as argued by Donkers and Melenberg (2002), predictive dependence does not imply that the model at hand is superior to an uninformative naive model that predicts the outcome that is most often observed in the sample.

Donkers-Melenberg's (2002) [DoM] test $(H_0 : HR = HR^{DoM})$ is based on a naive model that predicts always 0 in our setting, namely, $HR^{DoM} = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m}(1-y_{it})$. It follows that

$$e \equiv HR - HR^{DoM} = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m}(2y_{it}-1) \times \hat{y}_{it}$$

and $\sqrt{Nm}e$ has a limit normal distribution with zero mean and $E\left\{(2y_{it}-1)^2 \times \hat{y}_{it}^2\right\}$ variance under H_0. For binary variables, the latter equals $E(\hat{y}_{it}) = \Pr(\hat{y}_{it} = 1)$. The test statistic is

$$DoM = \frac{HR - HR^{DoM}}{\sqrt{\frac{1}{Nm}[\frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m}\hat{y}_{it}]}} \overset{a}{\sim} N(0,1)$$

considered the test variant $DM = \frac{\bar{e}}{\sqrt{\hat{w}/m}}$ where $\bar{e} = \frac{1}{m}\sum_t \bar{e}_t$ and $\hat{w} \equiv V(\bar{e}_t) = \sum_{k=-w}^{w} C_k(\bar{e}_t)$ for truncation lag $w = m^{1/3}$. Unsurprisingly, the long-run variance $\hat{w}$ is very small ($C_k, k \geq 0$ is computed over just $m \leq 5$ points) and the resulting DM statistics are implausibly large.

and it can be shown that, when a model has positive predictive performance (i.e. it outperforms the DoM naive), the predictions and realizations are dependent while the opposite is not necessarily true. In this regard, the DoM test is more challenging than the PT test.

4.3. Simple Benchmark Models

Four simple uninformative models are adopted as benchmarks. First, we consider a naive model that forecasts 1 for highly risk-averse decision-makers ($\theta > 0.5$), 0 for low risk aversion levels ($\theta < 0.5$) and the most-frequent-outcome (0 in the present sample) for $\theta = 0.5$. Second, due to the inherent state dependence (persistence) in debt-servicing behaviour, it seems natural to consider random walk (RW) type predictions. We include two variants. One is a RW *event* model based on the last observed default/non-default outcome $\hat{y}^{RW}_{i,\tau+1} = d_{i,\tau}$. Another is a RW *probability* model, $\hat{p}^{RW}_{i,\tau+1} = p_{i,\tau}$, where $p_{i,\tau}$ is the unconditional probability of default (frequency of 1s) over the τ^{th} rolling window, namely, $\hat{p}^{RW}_{i,\tau+1} = \frac{1}{T^*}\sum_{t=\tau-T^*+1}^{\tau} y_{it}$. Finally, we consider the two naive benchmarks implicit in the Pesaran-Timmermann (1992) and the Donkers-Melenberg (2002) tests. The relative forecast performance of the various benchmarks is ultimately an empirical question.

5. EMPIRICAL ANALYSIS

5.1. Analysis of Selected Variables

The variables retained by the in-sample cross validation (jacknife) are listed in Table I.[17]

[Table I around here]

The regressor set for the ensuing analysis thus contains $k = 13$ domestic signals $\mathbf{x}_{it}$ and $r = 3$ global signals $\mathbf{z}_t$. A number of regressors between 10 and 15 is the norm in the literature (see Peter, 2002). All the *external credit exposure* ratios play a relevant role as debt crisis signals, the exception being short-term debt to reserves. The remaining $\mathbf{x}_{it}$ are *external economic activity* (2

[17] The empirical analysis is conducted using LIMDEP 8 and GAUSS 3.4.

out of 5), *domestic conditions* (5/10) and *global financial link* (1/3) signals. The retained $\mathbf{z}_t$ are the US macroeconomic uncertainty, monetary policy uncertainty and risk aversion proxies.

Columns 2-4 report for each variable, the sample mean per state and a t-statistic to gauge its discriminatory power. The mean differential is significant in 10/16 cases and the sign is as expected theoretically in all 10 of them, e.g. total external debt/GDP during pending crisis episodes is about twice its level during tranquil periods (the expected signs in the light of the economics of the issue are discussed below). Columns 5-8 denote the regressors that are retained by a second jacknife applied to each of the regional panels starting from the above 16×1 signals. Reassuringly, the variables that are thrown out in all regions are those that appeared unable to discriminate (overall sample t-statistic) between the two states, the only exception being GDP growth.

Eight domestic signals emerge as robust since they are retained both in the world panel and in at least two regions: four debt structure signals (total external debt/GDP, official/total debt, short-term debt/total debt, IMF credit/exports), one measure of macroeconomic control (GNP per capita), one macroeconomic stability ratio (volatility of p.c. GNP growth) and one measure of openness (trade/GDP). In contrast, three domestic signals are deemed weak: trade balance/GDP, GDP growth and the real exchange rate (RER). Per capita GNP emerges as a strong signal, in contrast with GDP growth, perhaps because it reflects wealth. Total trade/GDP, which measures the degree of trade openness, is a good indicator of looming crises in contrast with trade balance/GDP which measures the country's competitiveness (closely linked to the RER) and is reflected in the current account of the balance of payments. The latter finding supports the 'willingness-to-pay' (as opposed to 'ability-to-pay') theory of sovereign default, according to which the opportunity cost of not servicing debt is relatively high for integrated economies. Interestingly the three global regressors are discarded in the regional models as opposed to the world panel. This may suggest that they capture temporal links such as contagion or spill-over mechanisms across regions.

5.2. Inference-based Comparison

Below we compare the models' ability to explain the observed crisis episodes 1984-2000.

Model Ranking by Information Criteria

The AIC can be cast as a discrepancy measure between the true model and a candidate while the BIC approximates the posterior odds probabilities in a Bayesian framework. In the context of nested models, the latter can be interpreted as adjusting the size of a LR test with the number of observations. Table II sets out the results.

[Table II around here]

The BIC ranks top the RC$^\beta$(ng) model that allows for random country-specific effects; hereafter ng stands for 'no global variables included'. The FE model, with or without globals, is generally not favoured by the BIC. However, the FE(ng) and FE are ranked first and third by the AIC which penalises less heavily for the large number of parameters. But the FE model is estimated over a smaller sample of $\tilde{N} = 53$ countries — the units for which there is no variation in y_{it} are thrown out — which, coupled with the incidental parameters problem, calls for caution in comparing its MLL with that of the other models.

The second BIC-best model is the RC$^\beta$ that allows not only for random country effects but also time effects via global regressors. The AIC ranks the RC$^\beta$(ng) and RC$^\beta$ as second and fourth, respectively. The RC$^\beta$-AR and RC$^\gamma$-AR models where the country-specific coefficients are allowed also to change over time fare relatively well. At the bottom of the ranking are the PLOGIT (ng) that assumes full homogeneity and models that control either for time effects only (PLOGIT, FTE, TCS) or for regional effects only (RLOGIT, RSLOGIT).

Importance of Country-specific Effects

Table III presents the results for hypothesis tests regarding unobserved country heterogeneity.

[Table III around here]

The LR statistic for the homogeneity null ($H_0 : \sigma_\alpha = 0$) in the RE model is significant. The estimate $\hat{\sigma}_\alpha = 2.35$ (t-ratio=22.49) also suggests large country heterogeneity. The model measure $\frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + \sigma^2}$ where $\sigma^2 \equiv \pi^2/3$ indicates that 63% of the unexplained variation in debt-servicing performance is due to latent time-constant country heterogeneity. The RE versus RC$^\beta$ comparison ($H_0 : \sigma_{\beta_1} = ... = \sigma_{\beta_k} = 0$) suggests heterogeneity in $\boldsymbol{\beta}$ also, i.e. the domestic fundamentals affect the probability of default differently across countries. Likewise, the RE versus RC$^\gamma$ test ($H_0 : \sigma_{\gamma_1} = ... = \sigma_{\gamma 3} = 0$) indicates that the impact of global conditions is country-specific as well.[18] Caution is needed in interpreting these tests because, for instance, under $\sigma_\alpha = 0$ the parameter is on the boundary of the maintained hypothesis, $\sigma \in R^+ \cup \{0\}$. In such settings, the usual limit distribution may not apply. For a single restriction, an easy correction has been suggested — use the χ^2 critical value for percentile $1 - 2\alpha$ instead of $1 - \alpha$ where α is the nominal level (see Kodde-Palm, 1986). The corrected test for PLOGIT versus RE obviously remains significant. For joint restrictions (e.g. RE versus RC$^\beta$) the correction is more involved. Nevertheless, the test statistics are rather large and so the corrected values are likely to be significant.

Next we compare the PLOGIT model (ML) and the FE model (Chamberlain's CML) using a Hausman statistic. Under $H_0 : \alpha_i = \alpha$, both are consistent (the CML estimator is inefficient because a) it does not use this information, b) it is based on a reduced sample) whereas under the alternative the consistent estimator is CML. The statistic at 32.14 strongly rejects.[19] Moreover, the ML estimates of the fixed effects $\hat{\alpha}_i$ are widely dispersed, ranging from 3.0 to 17.9 with a standard deviation $SD(\hat{\alpha}_i) = 3.4$. The FE versus RE test (Hausman) is insignificant and so the latter is preferred. In the RC$^\beta$ versus RE comparison, on the one hand, and between RC$^\gamma$ and RE, on the other, the more complex RC models are selected.

[18] The $\hat{\sigma}_\beta$ and $\hat{\sigma}_\gamma$ estimates in the RC$^\beta$ and RC$^\gamma$ models, respectively, suggest large country heterogeneity also.
[19] The Hausman test based on Huber-White's robust covariance for unspecified heterogeneity is $31.16(0.01, 16)$.

Importance of the Time-specific and Region-specific Effects

Next we focus on the importance of time effects. These are allowed for in distinct ways: first, by including global regressors ($\mathbf{z}_t$) in the PLOGIT; second, by means of time dummies in the FTE; third, by specifying a RC$^{\beta}$-AR (and RC$^{\gamma}$-AR) model that extends the RC$^{\beta}$ (and RC$^{\gamma}$) model to allow for time-variation in the country-specific slopes; fourth, by estimating different cross-section regressions sequentially in a TCS approach. Table IV reports the results.

[Table IV around here]

The global variables $\mathbf{z}_t$ are clearly significant ($H_0 : \boldsymbol{\gamma} = \mathbf{0}$) in the PLOGIT. A regression of the FTE estimates $\hat{\alpha}_t$ against $\mathbf{z}_t$ indicates that about half of the variation in the former ($R^2 = 46\%$) reflects changes in macroeconomic uncertainty, monetary policy uncertainty and risk aversion. The LR statistic for $H_0 : \alpha_t = \alpha$ in the FTE model is significant at the 10% level.

The LR statistic for $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}$ (or $\alpha_t = \alpha, \boldsymbol{\beta}_t = \boldsymbol{\beta}$) in the TCS model is insignificant but this outcome may be an artefact of the number of restrictions being tested (above 200), given the marked time-heterogeneity in $\hat{\boldsymbol{\beta}}_t$.[20] For instance, the slope estimates on GDP growth and the volatility of GNP p.c. growth 1984-2000 have ranges $[-18.16, 17.84]$, $[-29.21, 27.89]$ and standard deviations of 10.12 and 13.55, respectively (see Appendix D). This suggests that the link between the domestic signals and the likelihood of default is unstable which may reflect structural change possibly due to increasing globalization and trade/financial integration. In addition, large time series variation is expected in developing economy models due to poor data or measurement error. The TCS regressions do not allow for country heterogeneity and this is another potential source of instability in the slopes — if the latent factors responsible for the country heterogeneity (e.g. political conditions, financial regulation) change over time, this would induce different biases in $\hat{\boldsymbol{\beta}}_t$ over $t = 1, ..., T$ which may cancel out when averaged.[21] When time effects are tested

[20] In this sequential cross-section regression approach, the asymptotic distribution of the LR test holds for fixed T and $N \to \infty$. As T gets large, the number of restrictions will get large and the test may not be appropriate.

[21] In a logit, if the true DGP contains x_1 and x_2 but $y^* = \beta_1 x_1 + \varepsilon$ is estimated, then $\text{plim}\hat{\beta}_1 = \delta_1\beta_1 + \delta_2\beta_2$

($H_0 : \rho_\alpha = \boldsymbol{\rho}_\beta = 0$) in the RC$^\beta$-AR model which allows for country heterogeneity, both the LR and Hausman statistics strongly reject and the individual $\boldsymbol{\rho}_\beta$ are significant for 11/13 regressors. This points toward genuine time heterogeneity.[22]

We finally assess whether there is significant heterogeneity at the regional level. The strongly significant LR test for $H_0 : \alpha_j = \alpha, \boldsymbol{\beta}_j = \boldsymbol{\beta}$ for regions j =I,...,IV. indicates that there are regional differences in the probability response to the fundamentals. This is also borne out by the large variation in the estimates. For instance, the regional range of the coefficient on debt/GDP, official/total debt and trade/GDP is $[5.20, 23.03]$, $[4.19, 30.23]$ and $[-12.60, -3.29]$, respectively (see Appendix D). Unsurprisingly, the within-region heterogeneity, as measured by the SD of the regional fixed effects, appears to be less important than world heterogeneity (the overall FE model gives $SD = 3.4$) — we have $SD_I = 1.2$, $SD_{II} = 2.7$, $SD_{III} = 2.9$ and $SD_{IV} = 2.4$.

Plausibility of Estimates

Table V sets the estimation results for all models.[23] The expected sign of each slope coefficient is denoted in parentheses.[24] In five cases, both signs (+/-) can be theoretically justified.[25]

where δ_1 and δ_2 are complicated functions of the unknown parameters. If there is country-heterogeneity, each of the TCS regressions, $y_i^* = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + e_i$, implies $e_i = (\alpha_i - \alpha) + \mathbf{x}_i'(\boldsymbol{\beta}_i - \boldsymbol{\beta}) + \varepsilon_i = \varsigma_i + \varepsilon_i$ where ς_i represents the factors responsible for the heterogeneous responses whereas ε_i are the true innovations.

[22] We also tried the Hausman statistics using Newey-West HAC standard errors (Fuerets and Kalotychou 2004b). The test statistics for RC$^\beta$-AR1 versus RC$^\beta$and for RC$^\gamma$-AR1 versus RC$^\gamma$change from 122 to 277 and from 1816 to 29.17, but remain strongly significant. Thus, residual auocorrelation possibly arising from the construction of the dependent variable is not the explanantion for why the RC$^\beta$-AR1 is not rejected when the TCS model is.

[23] The FTE and PLOGIT slopes are very close so the former are not reported. The t-ratios based on the robust Huber-White covariance are qualitatively similar to those in Table V.

[24] In $y^* = \mathbf{x}'\beta + \varepsilon$ the marginal effect of x_j is $\frac{\partial p}{\partial x_j} = G(\boldsymbol{\beta}'\mathbf{x})[1 - G(\boldsymbol{\beta}'\mathbf{x})]\beta_j$ so the sign of $\frac{\partial p}{\partial x_j}$ is that of β_j.

[25] Eaton and Gersovitz (1981) set out a theoretical framework where the probability of default hinges on the 'willingness-to-pay'. The higher the *volatility of export growth* (and of *GNP pc growth*), the more an exclusion from the international capital markets is feared and so the more willing a country is to honour its debt (–). On the other hand, Peter (2002) amongst others advocates the 'ability-to pay' theory whereby volatile economies typically have large current account deficits (+). A weaker currency (positive *RER deviation from trend*) favours trade competitiveness and hence exports (–) but it means also a high debt burden in home currency and so, if debt is serviced mostly using GDP, the likelihood of default is higher (+); an overvalued currency implies a high risk of a currency crisis and hence of sovereign default (–). The higher the *IMF credit/exports* the higher the likelihood of default since countries with balance of payments problems are more likely to seek IMF help (+). On the other hand, countries in difficulty benefit from last minute IMF bail-outs from default (–).

[Table V around here]

For three indicators, the coefficients bear the correct sign and are significant in all model specifications: *external debt/GDP* (+), *official/total debt* (+) and *trade/GDP* (–).[26] The exception is *official/total debt* in the FE model (t-ratio = 0.7). The *credit to private sector/GDP* is significantly negative in all models (the exception is FE) so this suggests that the ratio may proxy banking development which is linked with increased economic growth (Bekaert *et al.*, 2005). The opposite theory for the latter (–) says that the higher the private-sector indebtness relative to national output, the higher the likelihood of mass private bankruptcies in times of financial distress.

The effect of *GNP per capita* is correctly picked up (i.e. significantly negative) by the RC$^{\beta}$ and RC$^{\gamma}$ models. Only the RC$^{\gamma}$-AR captures the negative effect of *GDP growth*. The positive effect of the *short-term debt* ratio is picked up by the PLOGIT(ng), PLOGIT, TCS and RC$^{\beta}$-AR models and also in the R(S)LOGIT for Africa and East Europe/MidEast/North Africa (see Appendix D). Regarding the (unreported) coefficients of the global variables: *US macro uncertainty* has the expected (+) sign and is significant in all models except for FE; *US monetary policy uncertainty* has the correct (+) sign and is significant in PLOGIT and RE; the *risk aversion* proxy has the correct (+) sign in RC$^{\beta}$, RC$^{\beta}$-AR and RC$^{\gamma}$-AR. Finally, although the TCS slopes show large instability as noted earlier, their average is plausible and comparable to the PLOGIT slopes (see Appendix D). This vindicates our average-based (*mean-time*) forecast approach for TCS.

To sum up, the AIC and BIC suggest that models that allow for country-specific and possibly time-varying slopes fare better than: *i*) the PLOGIT(ng) that assumes full homogeneity, *ii*) models that control for country heterogeneity at the regional level only, *iii*) models that control for common time effects only. In particular, the R(S)LOGIT that exclusively controls for regional heterogeneity and models that simply control for time effects common across countries such as the PLOGIT with

[26] *Total external debt/GDP* signals the ability to pay debt. Countries experiencing severe balance of payments problems are the most likely borrowers from official, multilateral institutions such as the IMF and so their *official/total debt* ratio is high. A large *trade/GDP* ratio signals openness and hence, the opportunity cost of default.

globals, FTE and TCS fare relatively bad. Moreover, for several variables the random coefficient models (RC$^{\beta}$, RC$^{\gamma}$, RC$^{\beta}$-AR, RC$^{\gamma}$-AR) tend to be the only specifications that yield the theoretical signs. The LR and Hausman tests tend to suggest that country, regional and time effects should not be overlooked when modelling the probability of default. Nevertheless, it may also be the case that the typically short T in such studies (annual data) is not sufiicient to allow for an informative random coefficient estimation. If this is the case, one should expect poor out-of-sample performance, an issue to be addressed in the proceeding forecast exercise.

5.3. Out-of-sample Forecast Comparison

In the following, the models are compared on the basis of their ability to predict outside of the estimation sample. Table VI presents different metrics over the entire holdout sample and over the PDC subset. The model with the minimum forecast error (in bold) is contrasted with all other models using the DM test. Asterisks denote a significant loss differential.

[Table VI around here]

Panel A sets out the comparison over the entire holdout sample. The RSLOGIT model provides the best forecasts according to the QPS, LPS and WMR$_{0.8}$ metrics but the DM test suggests that the forecast accuracy of PLOGIT(ng), PLOGIT, TCS and FTE is similar. Interestingly, despite not including $y_{i,t-1}$ as a regressor these models beat the RW-type predictions which suggests that there is enough persistence in the fundamentals so as to capture the state-dependence in debt servicing behaviour. According to the MR metric, the forecasts from TCS, FTE, R(S)LOGIT, RE(ng) and RC$^{\gamma}$ outperform those from any other model but not the RW predictions.[27] The PLOGIT(ng), PLOGIT, FTE, RE and RC$^{\gamma}$ are the best models under WMR$_{0.2}$, but the benchmark models generate equally satisfactory forecasts. We should note that only under the least plausible

[27] The Type I (false alarm) and Type II (missed default) error rates that underlie the MR metric for PLOGIT(ng) are 0.16 and 0.25, respectively. These are analogous to those reported in the literature for this widely used model, e.g. 0.15 and 0.24 in McFadden *et al.* (1985) and 0.09 versus 0.22 in Sommerville and Taffler (1995).

risk aversion level $\theta = \{0.2\}$ or, equivalently, the $\text{WMR}_{0.2}$ metric, are the best-forecasting models unable to beat the simple benchmarks.

Panel B reports the results over the PDC sample. The simplest model, PLOGIT(ng), wins the forecast race according to QPS and LPS and remarkably, it beats the benchmarks also. Under the QPS metric, the PLOGIT, FTE and RSLOGIT models forecast equally well whereas all other models forecast significantly worse. Under the LPS metric, PLOGIT(ng) and PLOGIT forecast significantly better than any other model. On the other hand, according to the $\text{WMR}_{0.8}$ metric the best model is RSLOGIT whereas FE, RC^{β}, RC^{β}-AR, RC^{γ}-AR models and the naive benchmarks all predict significantly worse. This suggests that controlling for country heterogeneity at the regional level appears to suffice for accurate forecasting and is consistent with our earlier finding that country heterogeneity within regions is less marked than the overall (world) country heterogeneity. Under the MR metric, the PLOGIT, TCS, RSLOGIT and RE models forecast better than any other model, including the benchmarks. Under $\text{WMR}_{0.2}$, the best forecasts are obtained from PLOGIT(ng), PLOGIT, FTE and the benchmark models whereas all other forecasts are significantly worse.

It is worth noting that the only metric under which the best forecasting model is unable to beat the RW model both over the holdout sample and PDC subset is the $\text{WMR}_{0.2}$. This is not surprising given that the RW prediction is 0 in most instances (there are more 0s than 1s in sample) and given that $\text{WMR}_{0.2}$ unrealistically penalises less heavily the missed defaults than the false alarms, the RW forecasts ought to fare well. Nevertheless, even in this case the forecast accuracy of the PLOGIT(ng), PLOGIT and FTE is comparable (insignificant DM test) to that of the RW model.

We next turn to the PT and DoM tests that are based on the hit rate $HR = 1 - MR$. The PT statistic is significant at better than the 1% level for all models (except for FE) which suggests positive dependence between realizations and predictions. Over the PDC sample, the largest PT statistics are obtained for PLOGIT, RSLOGIT and TCS. In contrast, not all models pass the DoM test for predictive performance as one would expect since it is relatively more demanding. In fact, the hit rate of the naive predictor implicit in the DoM test is significantly larger than that of the PT test. However, the hit rate of the RW model is above that of the DoM naive model (that always predicts 0) over the holdout sample so beating the former is a more challenging task.[28] Among those models that pass the DoM test, FTE and RSLOGIT produce the largest statistics over the holdout sample and the PLOGIT over the PDC sample. Interestingly, the FE, RC$^{\gamma}$, RC$^{\gamma}$-AR, RC$^{\beta}$ and RC$^{\gamma}$-AR models do not pass the DoM test neither over the holdout nor the PDC samples. Over the PDC sample, only the PLOGIT model passes the DoM test.

In sum, the more complex formulations such as FE, RE, RC$^{\gamma}$(-AR), RC$^{\beta}$(-AR) that allow for unobserved, fixed or random, heterogeneity across countries and possibly over time tend not to predict well out-of-sample despite the fact that they describe the data quite well. In contrast, the parsimonious pooled (all countries or region by region) logit regression with/without global factors forecasts relatively well and outperforms the naive benchmarks. The R(S)LOGIT that controls for regional effects, on the one hand, and the TCS and FTE models that exclusively control for time-specific effects, on the other, also work reasonably well as early warning devices.

6. DISCUSSION

While the empirical literature on sovereign debt crises is vast, only a few studies explicitly focus on specification issues related to the prediction of country default. Model-based early warning systems (EWS) for debt crises are one approach among many used for country risk monitoring. Quantitative EWS can usefully complement the sound judgement and wider analysis of decision-makers by yielding objective measures of country vulnerability to debt crises.

Concerns have been flagged about the challenges that country heterogeneity and instability pose with regard to the implementation of EWS. For instance, in the context of EWS for arrears to the IMF, Oka (2003; p.33) points out that "Temporal stability and country homogeneity that

[28] Over the holdout sample, we have $HR^{PT} < HR^{DoM} < HR^{RW}$. Over the PDC set where the DoM naive coincides with $\hat{y}^{RW}$, we have $HR^{PT} \leq HR^{DoM} = HR^{RW}$.

are assumed under probit estimation using panel data might be problematic." In a similar vein, Berg *et al.* (1999) provide theoretical arguments in favour of using a fairly homogeneous group of countries and sample period in the design of EWS for financial crises. Despite those concerns, a study of the importance of controlling for latent differences in behaviour across countries and through time is lacking in the EWS literature. This paper seeks to fill this gap.

We formulate distinct logit specifications ranging from a simple pooled regression to a random coefficients model where the impact of the macroeconomic indicators on the probability of a debt crisis is both country-specific and time-dependent. The analysis is based on a sample of 96 emerging/developing economies 1983-2002. The observable ratios that emerge as robust leading indicators of sovereign default are external debt to GDP, official debt to total debt, IMF credit to exports, credit to private sector over GDP and trade to GDP. The relative quality of the models is evaluated first on the basis of statistical hypothesis tests, model selection criteria and theoretical judgements on the economic plausibility of the estimates. These metrics corroborate the importance of controlling for latent country, regional and time effects in sovereign default models.

The paper presents also a comprehensive forecast comparison of the models. Out-of-sample forecasts are generated over a 5-year holdout period on the basis of a rolling estimation window. The forecast contest includes several benchmark models and predictive ability tests. Our findings suggest that, by simply exploiting pooled data across a large number of countries in a recursive modeling approach, it is possible to develop a relatively effective EWS of sovereign default that outperforms uninformative benchmark models. Interestingly, despite the theoretical arguments regarding the vulnerability of developing markets to changes in market sentiment and the global environment, the inclusion of either year-dummy variables or proxies for market volatility and risk aversion does not bring significant forecast gains over a simple pooled logit regression on country fundamentals. Moreover, accounting for temporal instability in the coefficients attributed to, say, changes in the level of economic integration and in market structures does not improve predictive performance either. Remarkably also, the pooling of countries at a regional, more homogeneous level does not generally yield any forecast gains over a broader pooling approach. Models that allow for country and time variation in the impact of key economic indicators on the probability of default perform poorly in terms of out-of-sample forecasting.

Perhaps despite the existence of country and time differences in economic structure, institutional development and political conditions, the effect of changes in the fundamentals on the likelihood of default is not so different as to vitiate the effectiveness of an EWS based on pooled data. Or more likely, perhaps the efficiency gains or reduction in forecast uncertainty from pooling key domestic indicators over time for a large number of countries outweighs the possible misspecification problems arising from neglected heterogeneity. It may be that simple (pooling) methods appear more robust in this context because the available data for relevant macroeconomic and financial indicators in developing countries are rather noisy and the impact of these indicators on the likelihood of default may be subject to unpredictable structural changes.

Several challenges remain for the successful prediction of sovereign debt crises. An important aspect that deserves further study is explicitly assessing the economic value of out-of-sample forecasts from competing models to distinct decision-makers such as international investors and policymakers. This would require full articulation of the decision environment of the forecast user and clearing several technical hurdles in the discrete-variable context. The paper focuses exclusively on the importance of unobserved heterogeneity for prediction purposes. There are two key other factors that warrant parallel research: dynamics and contagion. For instance building on the work of Pesaran and Pick (2003), it may be possible to investigate the forecast value of adequately controlling for contagion versus spill-overs in panel data models for sovereign default.

References

[1] Abhyankar A, Sarno L, Valente G. 2005. Exchange rates and fundamentals: evidence on the economic value of predictability. *Journal of International Economics*, forthcoming.

[2] Aylward L, Thorne R. 1998. An econometric analysis of countries' repayment performance to the IMF. *IMF Staff Papers* **45:** 595-619.

[3] Arora V, Cerisola M. 2001. How does US monetary policy influence sovereign spreads in emerging markets? *IMF Staff Papers* **48:** 474-498.

[4] Baltagi BH, Griffin JM. 1997. Pooled estimators versus their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* **77**: 303-327.

[5] Baltagi BH, Griffin JM, Xiong W. 2000. To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* **82**: 117-126.

[6] Baum CF. 2001 A stata module to interpolate a quarterly flow series from annual totals via the proportional DENTON method. Boston college mimeo.

[7] Bekaert A, Harvey CR, Lundblad C. 2005. Does financial liberalization spur growth? *Journal of Financial Economics,* forthcoming.

[8] Berg A, Borensztein E, Milesi-Ferretti GM, Pattillo C. 1999. Anticipating balance of payment crises: the role of early warning systems. IMF Ocassional Paper 196.

[9] Berg A, Borensztein E, Pattillo C. 2004. Assessing early warning systems: how have they worked in practice? IMF Working Paper 221.

[10] Bhat C. 1999. Quasi-random maximum simulated likelihood estimation of the fixed multinomial logit model. University of Texas at Austin, mimeo.

[11] Burkart O, Coudert V. 2002. Leading indicators of currency crises for emerging countries. *Emerging Markets Review* **3:** 107-133.

[12] Cantor R, Packer F. 1996. Determinants and impact of sovereign credit ratings. Federal Reserve Bank of New York Economic Policy Review.

[13] Chamberlain G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* **47:** 225-238.

[14] Detragiache E, Spilimbergo A. 2001. Crises and liquidity: Evidence and interpretation. IMF Working Paper 2.

[15] Diebold FX, López JA. 1996. Forecast evaluation and combination. In: *Statistical Methods in Finance Handbook of Statistics Series,* Vol. 14, Maddala GS, Rao CR (eds). North-Holland: Amsterdam.

[16] Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13:** 134-144.

[17] Donkers B, Melenberg B. 2002. Testing predictive performance of binary choice models. Econometric Institute Report 254.

[18] Eaton J, Gersovitz M. 1981. Debt with potential repudiation: Theoretical and empirical analysis. *Review of Economic Studies* **48:** 289-309.

[19] Feder GR, Just RE, Ross K. 1981. Projecting debt-servicing capacity of developing countries. *Journal of Financial and Quantitative Analysis* **16:** 651-669.

[20] FitzGerald V, Krolzig D. 2003. Modelling the demand for emerging market assets. Oxford University, mimeo.

[21] Fuertes AM, Kalotychou E. 2004. Towards the optimal design of an early warning system for debt crises. Emerging Markets Group, Cass, Working Paper *0304.*

[22] Fuertes AM, Kalotychou E. 2004b.Inferences on the Determinants of Default and the Autocorrelation Problem. Cass Business School, Faculty of Finance, mimeo.

[23] Granger CWJ, Pesaran MH. 2000. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* **19:** 537-560.

[24] Greene WH. 2003. *Econometric analysis*. Prentice Hall: New Jersey.

[25] Hajivassiliou V. 1987. The external debt repayment problems of LDCs: An econometric model based on panel data. *Journal of Econometrics* **36:** 205-230.

[26] Heffernan S. 2004. *Modern Banking*. Wiley: London.

[27] Honoré BE, Kyriazidou E. 2000. Panel data discrete choice models with lagged dependent variables. *Econometrica* **68:** 839-874.

[28] Hsiao C, Sun BH. 2000. To Pool or Not to Pool Panel Data. In: *Panel Data Econometrics: Future Directions: Papers in Honor of Professor Pietro Balestra,* Krishnaknmar J, Rochetti E. (eds). North-Holland, Amsterdam.

[29] Kalotychou E., Staikouras SK. 2005. The banking exposure to international lending: Regional differences or common fundamentals. *Financial Markets, Institutions and Instruments* **14:** 187-214.

[30] Kennedy P. 2003. *A Guide to Econometrics*. Blackwell: Oxford.

[31] Kodde DA, Palm FC. 1986. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* **50:** 1243-1248.

[32] Lee SH. 1991. Ability and willingness to service debt as explanation for commercial and official rescheduling cases. *Journal of Banking and Finance* **15:** 5-27.

[33] Li CA. 1992. Debt arrears in Latin America: Do political variables matter? *Journal of Development Economics* **40:** 349-359.

[34] López JA. 2001. Evaluating the predictive accuracy of volatility models. *Journal of Forecasting* **20:** 87:109.

[35] Manasse P, Roubini N, Schimmelpfennig A. 2003. Predicting sovereign debt crises. IMF Working Paper 221.

[36] McFadden D, Eckaus R, Feder G, Hajivassiliou V, O'Connell S. 1985. Is there life after debt? An econometric analysis of the creditworthiness of developing countries. In: *International Debt and the Developing Countries,* Smith G, Cuddington J. (eds). The World Bank: Washington.

[37] Oka C. 2003. Anticipating arrears to the IMF: early earning systems. IMF Working Paper 18.

[38] Oral M, Kettani O, Cosset JC, Daouas M. 1992. An estimation model for country risk rating. *International Journal of Forecasting* **8:** 583-593.

[39] Pesaran MH, Smith RP. 1995. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* **68:** 79-113.

[40] Pesaran MH, Skouras S. 2002. Decision-based methods for forecast evaluation. In: *A Companion to Economic Forecasting*, Clements MP, Hendry DF. (eds). Blackwell: New York.

[41] Pesaran MH, Pick A. 2003. Econometric issues in the analysis of contagion, University of Cambridge, mimeo.

[42] Pesaran MH, Timmermann A. 1992. A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* **10:** 461-465.

[43] Pescatori A, Sy AN. 2004. Debt crises and the development of international capital markets. IMF Working Paper 44.

[44] Peter M. 2002. Estimating default probabilities of emerging market sovereigns: A new look at a not-so-new literature. Graduate Institute of International Studies Working Paper, Switzerland.

[45] Shleifer A. 2003. Will the sovereign debt market survive? Harvard Institute of Economic Research Working Paper 2000.

[46] Sommerville RA. 1991. Country risk: multivariate models and human judgement. Unpublished PhD thesis, City University, London.

[47] Sommerville RA, Taffler RJ. 1995. Banker judgement versus formal forecasting models: The case of country risk assessment. *Journal of Banking and Finance* **19:** 281-297.

[48] Staikouras, SK. 2004. A chronicle of the banking and currency crises. *Applied Economics Letters* **11:** 873-878.

[49] Standard & Poor's. 2001. Sovereign defaults decline through third-quarter 2000. In: Ratings Performance 2000, default, transition, recovery, and spreads, S&P (eds). Standard & Poor's: New York.

[50] Taffler RJ., 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.

APPENDIX A: THE DATASET

A1. Emerging and developing economies.

Region (number of countries)	Composition
Eastern Europe (7)	Bulgaria, Czech Republic (R), Hungary, Poland, Romania, Russia, Turkey.
South/East Asia (17)	Bangladesh, China, Fiji, India, Indonesia, Korea R, Maldives, Nepal, Pakistan, Papua New Guinea, Philippines, Samoa, Solomon Islands, Sri Lanka, Thailand, Vanuatu, Vietnam.
Latin America (26)	Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican R, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, St. Kitts-Nevis, St. Lucia, Trinidad-Tobago, Uruguay, Venezuela.
Middle East/North Africa (10)	Algeria, Egypt, Iran, Jordan, Lebanon, Morocco, Oman, Syria, Tunisia, Yemen.
Africa (36)	Benin, Botswana, Burkina-Faso, Burundi, Cameroon, Cape Verde, Centr Afr R, Chad, Congo DR, Congo R, Cote d′Ivoire, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Kenya, Lesotho, Malawi, Mali, Mauritania, Mauritius, Mozambique, Niger, Nigeria, Rwanda, Sao Tome Principe, Senegal, Seychelles, Sierra Leone, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe.

A2. Macroeconomic and financial indicators.

Country-specific fundamentals				*Global conditions*
External credit exposure	External econ. activity	Domestic conditions	Intern. fin. links	
Debt/GDP	Export growth[a]	Credit to private sector/GDP	Trade[g]/GDP	World interest rates[i]
Official debt/Total debt	Vol. of export growth[b]	GDP growth[a]	Net bond flow[f,h]	OECD GDP growth[j]
Short-term debt/Reserves[d]	Trade balance[c]/GDP	GNP per capita (1995=100)	Net equity flow[f,h]	US macro. uncertainty
Short-term debt/Total debt	Reserves growth[a,d]	Volatility of GNP pc growth[b]		US monet. policy unc.
Debt service/Exports	Reserves/Imports[d]	Gov. expenditure[e]/GDP		US risk aversion index
IMF credit/Exports		Inflation		
		M2/Reserves[d]		
		Real exch. rate (1995=100)[f]		
		Gross capital formation/GDP		
		Gross domestic savings/GDP		

[a]Annual percentage growth. [b]Volatility proxied by the standard deviation over the last four years.[c]Trade balance is total exports - imports. [d]Foreign exchange reserves, excl. gold. [e]Government exp. on consumption, national security and defence. [f]Deviation from long-run trend ($\tilde{x}_{it} = x_{it} - \bar{x}_{i,t-1}$), undervaluation if $\tilde{x}_{it} > 0$. [g]Trade = exports + imports. [h]US$ billion. [j]GDP weighted lending rate for G7. [k]Growth of real GDP per capita for high-income OECD members (GNP p.c. in 1999 $\geqslant$ \$9, 361).

APPENDIX B: SOVEREIGN DEFAULTS 1984-2002

Country	Entries to Default (Δd_{it}=1)	Average Length	Defaults (d_{it}=1)	Default Episodes
Algeria	1	5.0	5	1994-1998
Argentina	4	2.3	9	1984, 1986, 1988-1992,1994-1995
Bangladesh	0	0.0	0	—
Belize	1	1.0	1	1984
Benin	3	2.3	7	1984-1988,1991,1993
Bolivia	4	2.3	9	1984-1985, 1987, 1991-1993, 1995-1997
Botswana	0	0.0	0	—
Brazil	3	1.7	5	1987, 1989-1991, 1993
Bulgaria	2	2.5	5	1990-1993, 1997
Burkina Faso	3	2.7	8	1986-1987, 1992-1994, 2000-2002
Burundi	0	0.0	0	—
Cameroon	3	4.3	13	1986-1988, 1990-1996, 1998-2000
Cape Verde	3	1.7	5	1989-1990, 1993, 1999-2000
Centr Africa R	5	1.6	8	1989-1990, 1992-1993,1995,1998,2000-2001
Chad	2	3.0	6	1985-1987, 1996-1998
Chile	0	0.0	0	—
China	0	0.0	0	—
Colombia	1	1.0	1	1988
Congo DR	2	4.5	9	1988-1995, 1998
Congo R	3	5.3	16	1985, 1987-1993, 1995-2002
Costa Rica	3	2.0	6	1986-1989, 1991, 1993
Cote D'Ivoire	3	3.7	11	1988-1993, 1995, 1998-2001
Czech Rep	0	0.0	0	—
Dominican R	5	2.2	11	1984-1985, 1987-1990, 1992-1993,1995-1996, 1998
Ecuador	3	3.7	11	1987-1994, 1999, 2001-2002
Egypt	3	4.3	13	1984-1986, 1988, 1992-2000
El Salvador	2	2.0	4	1984, 1989-1991
Eq Guinea	4	2.8	11	1984, 1986-1992, 1994-1995, 1998
Fiji	0	0.0	0	—
Gabon	4	2.8	11	1986, 1989-1993, 1995-1998, 2000
Gambia	1	2.0	2	1984-1985
Ghana	1	2.0	2	2001-2002
Grenada	1	8.0	8	1984-1991
Guatemala	3	1.7	5	1986-1987, 1990-1991, 1994
Guinea	4	2.8	11	1985, 1988, 1990-1994, 1996-1999
Guyana	3	3.3	10	1984-1989, 1994-1996, 1999
Haiti	2	2.0	4	1992-1994, 1996
Honduras	5	2.4	12	1984-1986, 1989, 1992-1994, 1996-1997, 1999-2001
Hungary	0	0.0	0	—
India	0	0.0	0	—
Indonesia	1	4.0	4	1998-2001

(cont.)

Iran	1	3.0	3	1984-1986
Jamaica	2	3.5	7	1986, 1989-1993, 1995
Jordan	2	6.0	12	1989-1992, 1994-2001
Kenya	2	2.0	4	1992-1993,2000-2001
Korea	0	0.0	0	—
Lebanon	1	3.0	3	1988-1990
Lesotho	0	0.0	0	—
Malawi	1	1.0	1	1989
Maldives	0	0.0	0	—
Mali	3	3.3	10	1984, 1989-1992,1994-1998
Mauritania	3	4	12	1984, 1989-1995, 1997-2000
Mauritius	0	0.0	0	—
Mexico	1	4.0	4	1989-1992
Morocco	3	2.0	6	1985, 1987, 1989-1992
Mozanbique	3	5.0	15	1984-1986, 1988-1998, 2000
Nepal	0	0.0	0	—
Nicaragua	2	6.5	13	1985-1994, 1997-1999
Niger	4	2.3	9	1989-1990, 1992-1993, 1995, 1997-2000
Nigeria	3	4.0	12	1988, 1990-1999, 2001
Oman	0	0.0	0	—
Pakistan	1	4.0	4	1998-2001
Panama	2	4.0	8	1987-1991, 1993-1995
Papua New Guinea	0	0.0	0	—
Paraguay	2	2.0	4	1986-1987, 1989-1990
Peru	3	3.7	11	1984-1990, 1993-1995, 1998
Philippines	1	5.0	5	1989-1993
Poland	3	2.7	8	1984, 1986-1991, 1997
Romania	0	0.0	0	—
Russia	2	5.5	11	1990, 1992-2001
Rwanda	2	2.5	5	1994-1995, 1999-2001
Samoa	0	0.0	0	—
Sao Tome and Principe	3	4.3	13	1985-1993, 1997-1999, 2001
Senegal	2	3.5	7	1989-1994, 1997
Seychelles	2	1.0	2	1991, 2001
Sierra Leone	4	3.0	12	1985, 1987-1991, 1993, 1996-2000
Solomon Islands	1	9.0	9	1993-2001
Sri Lanka	1	1.0	1	1996
St. Kitts and Nevis	1	1.0	1	1992
St. Lucia	0	0.0	0	—
Swaziland	0	0.0	0	—
Syria	2	6.5	13	1986,1990-2001
Tanzania	4	3.3	13	1984-1985,1987,1989-1996,1998-1999
Thailand	0	0.0	0	—
Togo	4	2.8	11	1987,1989-1994,1996,1998-2000
Trinidad and Tobago	1	5.0	5	1988-1992

(cont.)				
Tunisia	0	0.0	0	—
Turkey	0	0.0	0	—
Uganda	3	2.7	8	1988-1992, 1998, 2000-2001
Uruguay	0	0.0	0	—
Vanuatu	0	0.0	0	—
Venezuela	1	2.0	2	1984-1985
Vietnam	2	5.5	11	1988-1996, 1998-1999
Yemen	3	3.7	11	1987-1992, 1995, 1998-2001
Zambia	5	2.4	12	1985, 1987-1990, 1992-1993, 1996-1998, 2000-2001
Zimbabwe	1	3.0	3	2000-2002
Total	175		539	
1984-1995	127		383	
1996-2002	48		156	
Rate	10%		30%	
1984-1995	11%		33%	
1996-2002	7%		23%	

The models have the form y_{it}=f($x_{i,t-1}$) so the first relevant year for y_{it} in the analysis is 1984. The reported statistics are for the default series $\{d_{it}\}_{t=1984}^{2002}$ on which the EWS indicator $\{y_{it}\}_{t=1984}^{2000}$ is based, e.g. $y_{i,2000}$=1 if $d_{i,t}$=1 at t=2000,2001 or 2002. A country-period (i,t) case is a 'default entry' if $d_{i,t-1}$=0 and d_{it}=1. The reported default entries in 1984 are cases where $d_{i,1983}$=0. The analysis is based on N=96 countries. There are 1152(=96×12) and 672(=96×7) country-period cases over 1984-1995 and 1996-2002, respectively.

APPENDIX C: VARIABLE CROSS-VALIDATION

In order to preserve degrees of freedom, a jacknife procedure is conducted to reduce the original set of explanatory variables to an optimal smaller set with large predictive power. This jacknife approach is conducted in-sample, i.e. over the 1984-1995 period denoted $[1, T^*]$. It is based on the Type I (TI) error that is computed over $[1, T^*]$ and over a reduced subset that excludes consecutive defaults. The former measure (TI) gives the percentage of missed *defaults* ($\hat{y}_{T_0+1} = 0, y_{T_0+1} = 1$) whereas the latter gives the percentage of mispredicted positive directional changes (PDC) or missed *entries to default* ($\hat{y}_{T_0+1} = 0, y_{T_0+1} = 1, y_{T_0} = 0$).

The pooled logit estimates over $[1, T_0]$ with $T_0 < T^*$ and cut-off $\lambda = 0.5$ are used to generate 1-step-ahead forecasts $\hat{y}_{i,T_0+1}$ for $i = 1, .., N$ (minimum feasible $T_0 = 4$). This modeling and forecasting exercise is repeated iteratively, adding one further observation at a time, until T^* is reached. We compute the following cross-validation (CV) metric for different regressor sets (S)

$$CV\_TI_S = \frac{1}{(T^* - T_0)} \sum_{t=T_0+1}^{T^*} TI_t$$

and likewise for CV_PDC_S. In the first iteration, the baseline regressor set S_0 contains all regressors and S_j is a model that differs from S_0 in that it excludes x_j. Each iteration has 2 steps. First, collect in $\tilde{X}$ the $x_j \in S_0$ that satisfy

$$CV\_PDC_{S_j} \leq CV\_PDC_{S_0}$$

so that PDC is not increased by excluding any of them. Second, collect in $\breve{X}$ the $x_k \in \tilde{X}$ such that

$$CV\_TI_{S_k} \leq CV\_TI_{S_0}$$

so that their exclusion does not increase TI. The regressor set S_r that satisfies

$$S_r = \arg\min_{k \in \breve{X}} (CV\_TI_{S_k} - CV\_TI_{S_0})$$

in the first iteration is the reduced regressor set that gives the minimal TI without increasing PDC relative to that for S_0. Therefore, x_r is dropped from S_0 and the new baseline regressor set for the second iteration is S_r and so forth. The last iteration occurs when $\tilde{X}$ is the null set. We thus end up with a regressor set that gives the smallest possible Type I error over $[1, T^*]$ under the condition that no variable can be removed without increasing the Type I error over the PDC sample.

APPENDIX D

Table DI. Cross-section and regional estimates

	TCS			RLOGIT				RSLOGIT			
Variables	Min	Max	Median	I	II	III	IV	I	II	III	IV
External debt/GDP (+)	4.24	29.41	10.85	23.03	6.17	7.18	5.20	13.75	7.96	5.86	—
	(2.17)	(3.10)	(3.29)	(3.61)	(3.84)	(8.72)	(2.24)	(5.15)	(8.34)	(9.23)	
Offic debt/ Tot debt (+)	-12.71	18.61	8.73	10.26	15.67	4.19	30.23	-0.88	16.94	8.45	37.78
	(-0.97)	(1.07)	(0.68)	(1.48)	(3.27)	(1.06)	(2.43)	(-0.25)	(6.33)	(2.81)	(3.65)
ST debt/ Tot debt (+)	-14.94	16.25	5.10	7.99	2.83	14.63	17.14	-2.55	—	15.13	24.24
	(-1.33)	(2.01)	(0.79)	(0.99)	(0.74)	(3.92)	(1.80)	(-0.53)		(4.92)	(3.04)
Debt serv/Exports (+/-)	-24.72	7.32	-5.22	-3.63	-3.09	-5.92	-7.50	—	-3.50	-2.82	—
	(-2.87)	(1.53)	(-0.96)	(-0.57)	(-1.40)	(-3.39)	(-2.72)		(-2.24)	(-2.16)	
IMF credit/Exports (+/-)	-19.26	10.45	-1.65	-6.04	5.36	-2.07	11.63	-4.62	—	-1.52	14.60
	(-2.50)	(2.27)	(-0.42)	(-1.07)	(1.88)	(-1.73)	(1.69)	(-1.58)		(-1.85)	(2.43)
Vol export growth (+/-)	-12.14	12.89	1.34	-13.48	1.78	0.21	4.65	—	—	—	4.37
	(-2.24)	(1.93)	(0.36)	(-2.13)	(0.85)	(0.17)	(1.78)				(1.79)
Trade balance/GDP (-)	-6.59	8.61	0.83	-2.45	1.72	0.03	-0.74	—	—	—	—
	(-1.16)	(1.94)	(0.17)	(-0.36)	(0.73)	(0.03)	(-0.24)				
Credit private/GDP (+/-)	-14.33	-0.86	-4.50	-1.47	-3.71	-0.81	-3.76	—	—	—	-4.73
	(-2.61)	(-0.33)	(-1.30)	(-0.29)	(-2.92)	(-0.45)	(-1.83)				(-3.01)
GDP growth (-)	-18.16	17.84	-7.89	10.64	0.91	-0.48	1.04	—	—	—	—
	(-1.54)	(1.40)	(-1.00)	(1.18)	(0.25)	(-0.21)	(0.23)				
GNP per capita (-)	-0.36	2.74	0.98	1.12	-0.33	0.63	-1.92	-0.51	—	0.53	-2.93
	(-0.58)	(2.97)	(1.90)	(1.04)	(-0.92)	(2.86)	(-2.47)	(-0.82)		(2.97)	(-4.99)
Vol pc growth (+/-)	-29.21	27.89	-0.73	-0.53	4.91	7.62	-2.64	-3.69	4.53	—	—
	(-1.20)	(1.62)	(-0.05)	(-0.02)	(0.60)	(2.00)	(-0.27)	(-0.29)	(0.84)		
Real exchange rate (-)	-3.38	1.53	0.49	0.54	0.55	-0.18	-0.85	—	—	—	—
	(-2.16)	(1.45)	(0.83)	(0.51)	(1.55)	(-0.67)	(-1.21)				
Trade/GDP (-)	-19.37	0.43	-6.54	-12.60	-5.49	-7.19	-3.29	-6.59	-7.46	-6.76	2.74
	(-2.87)	(0.16)	(-2.18)	(-1.79)	(-3.82)	(-6.64)	(-1.21)	(-2.52)	(-7.39)	(-7.57)	(1.79)

t-statistics are reported in parenthesis. (I) Asia, (II) Latin America, (III) Africa, (IV) East Europe/Middle East/North Africa.

Table I. In-sample variable selection for world and regional panels

World panel	mean		*t*-stat	Regional panels			
				(I)	(II)	(III)	(IV)
N=96	$\bar{x}^0_{it}$	$\bar{x}^1_{it}$	$H_0 : \bar{x}^1_{it} - \bar{x}^0_{it}$	N=17	N=26	N=36	N=17
A) Country-specific indicators							
External credit exposure							
Total external debt/ GDP	0.379	0.679	18.82*	✓	✓	✓	×
Official debt / Total debt	0.568	0.592	4.65*	✓	✓	✓	✓
Short term debt / Total debt	0.120	0.116	-0.79*	✓	×	✓	✓
Debt service / Exports	0.158	0.191	5.13*	×	✓	✓	×
IMF credit / Exports	0.097	0.149	5.86*	✓	×	✓	✓
External economic activity							
Volatility of export growth	0.111	0.136	4.13*	×	×	×	✓
Trade balance / GDP	-0.083	-0.082	0.11	×	×	×	×
Domestic conditions							
Credit to private sector/ GDP	0.264	0.187	-9.76*	×	×	×	✓
GDP growth	0.041	0.021	-6.90*	×	×	×	×
GNP per capita	7.033	6.529	-8.53*	✓	×	✓	✓
Volatil. of GNP p.c. growth	0.046	0.053	4.20*	✓	✓	×	×
Real exchange rate	0.130	0.138	0.26	×	×	×	×
International fin. links							
Trade / GDP	0.532	0.459	-6.20*	✓	✓	✓	✓
B) Global indicators							
Macroeconomic uncertainty	0.248	0.239	-1.65	×	×	×	×
Monetary policy uncertainty	0.288	0.280	-1.83	×	×	×	×
Risk aversion	0.976	0.991	0.73	×	×	×	×

The variable selection is conducted over the [1984, 1995] period using the jacknife on the basis of a pooled logit. (I) Asia, (II) Latin America, (III) Africa, (IV) East Europe/Middle East/North Africa. t-stat is the statistic for the significance of the absolute mean differential over 1985-1995.
*denotes significant at the 1% level.

Table II. Model comparison on the basis of information criteria

				AIC		BIC	
Model type	Controlled effects	MLL	s	statistic	ranking	statistic	ranking
PLOGIT(ng)	—	-606.3	14	620.3	**12**	656.5	**10**
PLOGIT	time	-601.4	17	618.4	**11**	662.4	**12**
TCS	time	-491.8	238	729.8	**15**	1007.5	**15**
FTE	time	-594.4	30	624.4	**13**	702.0	**13**
RLOGIT	regional	-500.1	56	556.1	**10**	660.0	**11**
RSLOGIT	regional	-646.7	30	676.7	**14**	735.1	**14**
RE (ng)	country	-469.8	15	484.8	**8**	523.7	**3**
FE (ng)	country	-399.0	66	405.0	**1**	559.4	**7**
RC^{β} (ng)	country	-399.7	28	427.7	**2**	500.1	**1**
RE	country, time	-467.8	18	485.8	**9**	532.3	**4**
FE	country, time	-366.0	69	435.0	**3**	596.4	**9**
RC^{γ}	country, time	-457.8	21	478.8	**6**	533.1	**5**
RC^{γ}-AR	country, time	-455.7	25	480.7	**7**	545.4	**6**
RC^{β}	country, time	-408.6	31	439.6	**4**	519.8	**2**
RC^{β}-AR	country, time	-425.8	45	470.8	**5**	587.3	**8**

The criteria are AIC=-MLL+s and BIC=-MLL+0.5s ln(NT) where s is the number of estimated coefficients and NT(=1307) is the effective sample size. $\text{AIC}=-\sum_{t=1}^{T}\text{MLL}_t+\sum_t s_t$ and $\text{BIC}=-\sum_t \text{MLL}_t+0.5 s_t \sum_t \ln(N_t)$ for the TCS model, where N_t is the no. of available observations per cross-section. $\text{AIC}=-\sum_{j=1}^{R}\text{MLL}_j+\sum_j s_j$ and $\text{BIC}=-\sum_j \text{MLL}_j+0.5 s_j \sum_j \ln(\text{NT}_j)$ for the R(S)LOGIT, s_j are the number of coefficients, NT_j the data points per region (R=4).

Table III. Statistical significance of country-specific effects

Tests	Model type					
	FE	RE	RC^{β}		RC^{γ}	
A) Likelihood ratio						
null hypothesis	$\alpha_i=\alpha$	$\sigma_{\alpha}=0$	$\sigma_{\alpha}=0$ $\sigma_{\beta_1}=...=\sigma_{\beta_k}=0$	$\sigma_{\beta_1}=...=\sigma_{\beta_k}=0$	$\sigma_{\alpha}=0$ $\sigma_{\gamma_1}=...=\sigma_{\gamma_3}=0$	$\sigma_{\gamma_1}=...=\sigma_{\gamma_3}=0$
null model	PLOGIT	PLOGIT	PLOGIT	RE	PLOGIT	RE
test statistic	–	267.3*** (0.00, 1)	385.7*** (0.00, 14)	118.4*** (0.00, 13)	287.3*** (0.00, 4)	20.0*** (0.00, 3)
B) Hausman-type						
null hypothesis	FE_{CML}=PLOGIT	RE=FE_{CML}	RC^{β}=RE		RC^{γ}=RE	
test statistic	32.14*** (0.01, 16)	7.75 (0.96, 16)	292.2*** (0.00, 16)		78.58*** (0.00, 16)	

The p-values and degrees of freedom of the tests are reported in parenthesis. ML estimation based on the Newton method (PLOGIT) or MSL using Halton draws (RE, RC). FE_{CML} denotes the Chamberlain's conditional ML estimator of the FE model.

*, ** and *** denote significant at the 10% , 5% and 1% level.

Table IV. Statistical significance of time- and regional-specific effects

	Time effects						Regional effects
Tests	PLOGIT	FTE	TCS		RC$^{\beta}$-AR	RC$^{\gamma}$-AR	RLOGIT
A) Likelihood ratio							
null hypothesis	$\boldsymbol{\gamma}$=$\mathbf{0}$	α_t=α	$\boldsymbol{\beta}_t$=$\boldsymbol{\beta}$	α_t=α, $\boldsymbol{\beta}_t$=$\boldsymbol{\beta}$	ρ_{α}=0, $\boldsymbol{\rho}_{\beta}$=$\mathbf{0}$	ρ_{α}=0, $\boldsymbol{\rho}_{\gamma}$=$\mathbf{0}$	α_j=α, $\boldsymbol{\beta}_j$=$\boldsymbol{\beta}$,
null model	PLOGIT(ng)	PLOGIT(ng)	FTE (ng)	PLOGIT (ng)	RC$^{\beta}$	RC$^{\gamma}$	PLOGIT (ng)
test statistic	9.89**	23.82***	205.30	229.12	44.22***	4.03	215.58***
	(0.02, 3)	(0.08, 16)	(0.54, 208)	(0.39, 224)	(0.00, 14)	(0.40,4)	(0.00, 42)
B) Hausman-type							
null hypothesis	PLOGIT=PLOGIT(ng)	FTE=PLOGIT(ng)	—		RC$^{\beta}$-AR1=RC$^{\beta}$	RC$^{\gamma}$-AR1=RC$^{\gamma}$	—
test statistic	4.12	9.93	—		122.20***	1816.17	—
	(0.99,13)	(0.70,13)			(0.00, 16)	(0.00,16)	

The p-values and degrees of freedom of the tests are reported in parenthesis. Models without global variables are signified by ng.

*, ** and *** denote significant at the 10% , 5% and 1% level.

Table V. Parameter estimates of default probability models 1984-2000

		latent effects										
		time		country	regional		country, time					
Variables	PLOGIT(ng)	PLOGIT	TCS	RE(ng)	RLOGIT	RSLOGIT	FE	RE	RC$^{\gamma}$	RC$^{\gamma}$-AR	RC$^{\beta}$	RC$^{\beta}$-AR
External debt/GDP (+)	7.86	8.18	11.81	9.43	10.40	9.19	10.05	9.95	9.83	15.99	15.25	16.20
	(13.8)	(13.9)	(6.8)	(21.9)			(6.6)	(22.4)	(20.8)	(17.7)	(14.2)	(15.5)
Offic debt/ Tot debt (+)	8.41	8.82	7.92	6.22	15.09	15.57	2.85	7.04	6.91	8.98	11.52	18.65
	(4.7)	(4.9)	(3.7)	(5.9)			(0.7)	(6.4)	(5.7)	(3.5)	(9.5)	(4.6)
ST debt/ Tot debt (+)	5.44	5.41	4.38	-1.31	10.65	12.27	-6.59	-1.02	-1.75	2.69	1.52	8.31
	(3.3)	(3.3)	(2.0)	(-1.2)			(-1.8)	(-0.9)	(-1.4)	(1.11)	(0.5)	(2.2)
Debt serv/Exports (+)	-3.42	-3.80	-6.14	-2.69	-5.03	-3.16	-3.61	-3.01	-2.47	-2.07	1.29	-5.72
	(-3.6)	(-3.9)	(-2.9)	(-4.1)			(-1.92)	(-4.4)	(-3.5)	(-1.7)	(1.0)	(-3.0)
IMF credit/Exports (+/-)	-1.92	-2.11	-2.44	-2.34	2.22	2.82	-4.00	-2.70	-3.17	-3.02	6.70	11.20
	(-2.6)	(-2.8)	(-1.24)	(-4.2)			(-2.1)	(-4.8)	(-4.9)	(-3.2)	(4.3)	(11.7)
Vol export growth (+/-)	2.23	1.97	1.09	3.31	-1.71	4.37	3.21	2.93	2.50	-1.55	-4.07	-4.88
	(2.6)	(2.3)	(0.6)	(6.7)			(2.4)	(5.6)	(4.3)	(-1.8)	(-3.3)	(-3.5)
Trade balance/GDP (-)	1.23	1.05	0.73	3.80	-0.36	—	3.02	3.68	4.4	9.46	5.26	-0.91
	(1.6)	(1.4)	(0.7)	(6.1)			(1.5)	(5.7)	(6.2)	(7.7)	(3.7)	(-0.5)
Credit private /GDP (+/-)	-3.24	-3.49	-4.67	-3.66	-2.44	-4.73	-1.52	-3.81	-3.19	-15.61	-12.71	-16.53
	(-5.1)	(-5.3)	(-5.0)	(-7.8)			(-1.0)	(-8.2)	(-6.2)	(-12.6)	(-9.5)	(-14.8)
GDP growth (-)	-1.55	-1.24	-2.50	0.12	3.03	—	2.25	0.48	1.09	-8.91	9.17	9.42
	(-1.0)	(-0.8)	(-1.0)	(0.1)			(0.9)	(0.4)	(0.85)	(-5.2)	(3.74)	(3.9)
GNP per capita (-)	0.49	0.54	0.83	0.18	-0.13	-0.97	-1.97	0.26	-0.36	1.31	-2.30	0.95
	(4.8)	(5.2)	(5.1)	(2.1)			(-1.6)	(3.0)	(-3.5)	(8.1)	(-10.3)	(3.8)
Vol GNP pc growth (+/-)	4.55	3.51	-0.57	4.07	2.34	0.42	1.38	3.22	1.61	6.71	-7.16	-15.09
	(1.8)	(1.4)	(-0.2)	(2.2)			(0.3)	(1.7)	(0.8)	(1.6)	(-1.65)	(-2.52)
Real exchange rate^{c} (+/-)	0.03	0.01	-0.16	0.15	0.01	—	0.14	0.13	0.12	0.81	-2.02	3.64
	(0.2)	(0.1)	(-0.4)	(1.4)			(0.6)	(1.1)	(0.9)	(7.2)	(-5.9)	(11.7)
Trade/GDP (-)	-4.70	-4.86	-7.31	-6.03	-7.14	-4.52	-6.36	-5.90	-5.74	-16.12	-7.58	-17.79
	(-8.0)	(-8.1)	(-4.8)	(-13.2)			(-3.1)	(-12.8)	(-11.7)	(-14.6)	(-7.7)	(-13.8)

The expected sign according to economic theory is in parenthesis after the variable name. The t-ratio of each coefficient is in parenthesis except for the RLOGIT and RSLOGIT for which the average across regions is reported (see Appendix D for details).

Table VI
Out-of-sample forecast analysis

Model	$WMR_{0.2}$	MR	$WMR_{0.8}$	QPS	LPS	PT	DoM
A: *holdout sample*							
naive	0.0819	0.4097**	0.1172**	–	–	0	0
RW naive	**0.0643**	**0.2014**	0.1372**	0.4117*	1.1588**	5.00**	3.73**
PLOGIT(ng)	0.0754	0.2700*	0.0860	0.3275	0.5130	4.41**	1.90
PLOGIT	0.0712	0.2562*	0.0927	0.3289	0.5168	4.51**	2.20*
TCS	0.0973*	0.2450	0.0776	0.3435	0.5631	4.67**	2.39*
FTE	0.0703	0.2354	0.0971*	0.3367	0.5353	4.53**	2.87**
RLOGIT	0.1004*	0.2314	0.0835	0.3275	0.5904**	4.79**	2.72**
RSLOGIT	0.0960*	0.2320	**0.0760**	**0.3060**	**0.5043**	4.68**	2.82**
RE(ng)	0.0938*	0.2434	0.0948	0.3520*	0.6267*	4.36**	2.77**
RE	0.0897	0.2514*	0.0948	0.3530*	0.6335*	4.24**	2.59**
FE	0.1725**	0.3898**	0.1344	0.5251**	0.8874**	1.72*	0.29
RC^{γ}	0.0836	0.2492	0.0943	0.3629*	0.6672**	4.51**	1.37
RC^{γ}-AR	0.1388**	0.3068**	0.1236**	0.5239**	2.2647**	3.61**	1.46
RC^{β}	0.1581**	0.3220**	0.1588**	0.6231**	3.1384**	3.06**	1.37
RC^{β}-AR	0.1456**	0.3088**	0.1594**	0.5752**	2.6447**	3.25**	1.61
B: *PDC sample*							
naive	0.0671	0.3354**	0.1329**	–	–	0	0
RW naive	**0.0671**	0.3354**	0.2683**	0.4418**	1.2197*	0	0
PLOGIT(ng)	0.0819	0.3292**	0.0970	**0.3489**	**0.5469**	3.38**	0.07
PLOGIT	0.0776	0.2400	0.1023	0.3491	0.5508	3.79**	1.92*
TCS	0.1055**	**0.2294**	0.0960	0.3713*	0.6153**	3.92**	1.63
FTE	0.0817	0.3228**	0.1047*	0.3621	0.5788**	3.16**	0.16
RLOGIT	0.1193**	0.2716**	0.1037	0.3992**	0.7290**	3.71**	1.00
RSLOGIT	0.1233**	0.2474	**0.0887**	0.3685	0.5935*	4.26**	1.32
RE(ng)	0.1064**	0.2648**	0.0909	0.3776*	0.6718**	2.83**	1.22
RE	0.0996*	0.2492	0.0951	0.3743*	0.6730**	3.76**	1.46
FE	0.2091**	0.3642**	0.1691**	0.6321**	1.0806**	0.77	-1.12
RC^{γ}	0.0900*	0.2788**	0.0947	0.3916**	0.7158**	3.11**	0.84
RC^{γ}-AR	0.1523**	0.3172**	0.1280**	0.5483**	2.1234**	2.89**	0.24
RC^{β}	0.1782**	0.3174**	0.1225**	0.5994**	2.9334**	3.12**	0.25
RC^{β}-AR	0.1470**	0.3096**	0.1658**	0.5774**	2.5995**	2.16*	0.41

Bold denotes the minimum loss. **, * denote significant (1-tail) at the 1% and 5% levels. Under (W)MR, QPS or LPS, asterisks indicate that the forecasts are worse than those of the best model according to a Diebold-Mariano test. PT is the Pesaran-Timmerman test, DoM is the Donkers-Melenberg test. Under PT and DoM, a 0 denotes that the model at hand coincides with the naive model that underlies these tests. 'Naive' denotes the model that predicts 1 for $\theta > 0.5$, 0 for $\theta < 0.5$ and the most-frequent-observed event for $\theta = 0.5$. 'RW naive' is the random walk event or random walk forecast model.