CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

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

Modelling Sovereign Debt Crises Using Panels

Ana-Maria Fuertes and Elena Kalotychou

CEA@Cass Working Paper Series

WP–CEA–11-2004

Modeling Sovereign Debt Crises Using Panels*

ANA-MARIA FUERTES and ELENA KALOTYCHOU†
Faculty of Finance, Cass Business School, City University, London

13 December 2004

Abstract

This paper compares rival sovereign default models that differ in how country-, region- and time-specific effects are treated. The quality of the models is gauged using inference-based criteria and the plausibility of estimates. An out-of-sample forecast evaluation framework is deployed based on statistical- and economic-loss functions, naïve benchmarks and equal-predictive-ability tests. The inference metrics overwhelmingly favour more complex models that allow for time-varying country heterogeneity. However, simplicity beats complexity in terms of forecasting. Pooled logit models that simply control either for regional heterogeneity or for time effects produce the most accurate forecasts and outperform the naive models.

Keywords: Panel logit, unobserved heterogeneity, economic loss, predictive performance.
JEL Classification: C15; C32; C33

---

*We appreciate the helpful comments of Roy Batchelor, Jerry Coakley, William Greene, Frank Packer, Ron Smith, Sotiris Staikouras, Giorgio Valente and participants at the 2004 *Society for Computational Economics* Conference at the University of Amsterdam, the 2004 *ESEM-EEA* Conference, Universidad Carlos III de Madrid, the 2004 *Australasian Finance and Banking* Conference, University of New South Wales, Sydney, and seminar participants at the University of Cambridge, Faculty of Economics and Politics. The usual disclaimer applies.

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

1 Introduction

Over the last two decades, a number of sovereign debt crises have led to a new emphasis on credit risk analysis. In the early 1990s the Mexican peso crisis led to several defaults whereas in 1997 the Asian 'meltdown' spread through East Europe and Latin America triggering a chain of unprecedented credit jitters — in 1998, the plummeting of the Russian ruble pushed the government to default on the US dollar denominated 'MinFin' bonds; in 2000, Ecuador declares the first-ever default on Brady Bonds; in 2001, Argentina announces the largest sovereign default in history. Financial institutions and regulators have thereafter become painfully aware of the credit exposure to international lending and international credit risk has been one of the most intensely researched areas in modern finance through two broad approaches: option pricing models, on the one hand, and the direct modeling of the default probability of issuers using panel data on the other. Sovereign default probabilities are of interest to financial institutions in order to price loans and bonds, to determine adequate concentration limits and as inputs for Value at Risk analyses. Futhermore, this interest has been reinforced by the revised Basel Capital Accord, under which banks are allowed to use internal credit ratings and default rates in determining their minimal regulatory capital.

The formulation and estimation of binary-choice models for panel data has been the subject of a rapidly growing literature.[1] Panels can provide insights which are not available in pure cross-section or time-series data (see Baltagi, 2002). The choice of estimator depends on whether the data are conceptualized as repeated cross-sections or a pool of time series. Pure cross-section estimators cannot allow for country heterogeneity. Pooled time series estimators can pick up differences in behavior across individuals not captured by the included regressors. Default on sovereign debt is not a frequent event for a given country and definitely not a short-term situation. As a result, the binary dependent variable representing the latter exhibits small variation. In addition, the available emerging/developing country indicators for such studies are typically annual (moderate T) series over 10 to 20 years, often with missing observations. The above two issues combine to make individual country analyses unfeasible. Researchers have circumvented this problem by jointly exploiting time series on typically 20 to 130 countries (large N panels). Over 80% of these

[1] A comprehensive survey can be found in Arellano and Honoré (2001).

studies employ logit models. Simple pooled logit regressions are common, although the validity of the implicit full homogeneity assumption has been questioned (McFadden et al. 1985, Hajivassiliou, 1987). As Schleifer (2003) puts it: "Sovereign debt markets could not be more different".

The limited number of studies that control for latent country heterogeneity use either fixed or random effects logits. One exception is Oral et al. (1992) who allow for fixed country effects both in the intercept and slope coefficients of the domestic signals. Some evidence in the related scenario of currency crises suggests that the relevant heterogeneity occurs at a broader, regional level (Burkart and Coudert, 2002; Staikouras, 2005). In the context of sovereign default, regional differences have been accommodated via regional fixed effects (Feder et al., 1981). Some studies find significant time effects using year dummies (Aylward and Thorne, 1998) or including global macroeconomic variables such as OECD economic growth (Lee, 1991; Detragiaghe and Spilimbergo, 2001). This stresses the importance of exogenous world shocks on default risk.

The large number of empirical models available have led to at best mixed evidence regarding the determinants of sovereign default. The estimates are based on different samples (countries and time span) and there is no unanimous default definition. Both of these aspects make the model comparison onerous. There is also the non-trivial issue of how to carry out the comparison. One can employ statistical hypothesis tests of parameter restrictions. Alternatively, the 'best' model can be chosen using extant model selection criteria. With a small number of degrees of freedom, this approach can lead to quite close values for the criteria and hence, to model selection instability. A third common approach is to compare the plausibility of the estimates. However, if the ultimate is to design an early warning system, it seems more natural to confront the models on the basis of their forecast ability.[2] Nonetheless, forecasting issues have only received superficial attention.

Some studies generate in-sample forecasts to compare rival models (Hajivassiliou, 1987; Detragiaghe and Spilimbergo, 2001). A few studies conduct out-of-sample evaluation but limited to a 1- or 2-year holdout period and using a fixed estimation window. The typical forecast criteria used in these studies are the Type I, Type II error or the overall error (Feder et al., 1981; Sommerville and Taffler, 1995; Oka, 2003; Peter, 2002).[3] Recent contributions in the forecasting literature

[2] 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).

[3] In this literature, the Type I (Type II) error rate are the missed defaults (non-defaults) over the realized defaults

stress that such metrics may not represent well the decision-maker's problem (Granger and Pesaran, 2000; Pesaran and Skouras, 2002). Futhermore, those studies that provide out-of-sample predictions do not confront them to naive benchmarks. This is important due to the persistence of the external debt repayment behavior. A specific question we address is whether controlling for country, regional or time heterogeneities leads to 'better' sovereign default models.

This paper contributes to the literature in three directions. First, regarding model specification we consider a wide range of panel logits, some of them novel in this context, that treat regional-, country- and time-specific effects in different ways. The analysis is based on data for 96 countries over 1983-2002. The regressor set includes three world variables — macroeconomic uncertainty, monetary policy uncertainty and risk aversion — that, to our knowledge, have not been considered in the literature. Second, the models are compared using various statistical metrics that gauge their ability to describe the data generating process (DGP). These include likelihood ratio (LR) or Hausman type tests and the AIC and SBC model selection criteria. These metrics overwhelmingly suggest that heterogeneity across countries, regions and time should not be overlooked.

Third, a rigorous forecast analysis is conducted. A 12-year estimation window is rolled forward to generate recursive out-of-sample forecasts over a 5-year holdout period. We focus on 1-step-ahead point forecasts. Rather than using a fixed ad hoc cut-off probability, this parameter is optimally calibrated in-sample over each window for the model and loss function at hand. Both statistical and economic loss functions are evaluated over the holdout set and a positive-directional-change subset. The latter allows the emphasis to be on anticipating new (rather than perpetuating) debt crises. The equal-predictive-ability test of Diebold and Mariano (1995) is deployed to compare the models. Simple naive forecasts are also considered. These include random walk predictions and both the random prediction and the most-frequent-event prediction implicit in Pesaran-Timmermann's (1992) and Donkers-Melenberg's (2002) tests, respectively. Models that simply control for fixed regional- or time-effects are capable of yielding relatively good forecasts.

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 results. A final section concludes.

(non-defaults). The overall error is the sum of missed defaults and false alarms over the number of sample cases.

2 The Data

The analysis is based on *World Bank* data for 96 emerging markets and LDCs from Africa, Asia, Eastern Europe, Latin America and Middle East over 1983-2002. Information on external debt, arrears and rescheduling to official and private creditors is obtained from the Global Development Finance database. Macroeconomic and financial time series are obtained from the World Development Indicators database (see Appendix A).

The Early Warning of Default (EWD) indicator

There is no unanimous definition of sovereign default. The rating agencies definition reflects only default on rated sovereign bonds, which is a rare event. Few countries have defaulted on rated bond issues, however, many countries defaulted on their bank debt and trade credit obligations, especially during the 1980s and the early 1990s. By adopting a broader definition that encompasses bank debt and trade obligations we can can base our study on a wider set of countries over a longer period of time.[4] We categorize country i as being in default in year t (denoted $d_{it} = 1$) if any of these conditions are met: a) there is a large jump in total arrears (interest and principal repayments) relative to total external debt $\Delta A_{it}/D_{it} > \delta$ or b) the total amount of debt rescheduled exceeds the decrease, if any, in total arrears.[5]

The goal is to predict the probability of a debt crisis over a specific time window longer than a year, while updating the forecasts annually. We adopt a 3-year warning window and define[6]

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

following Peter (2002) and Oka (2003). 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

[4] Moody's and S&P's sovereign default rates are based on a limited number of defaults on rated sovereign bonds that occurred between 1998-2003. The leading rating agencies also report corporate default rates as proxies for the sovereign default probabilities associated with each sovereign credit-rating. Peter (2002) shows that these proxies are too low on average relative to the estimated default probabilities from a panel logit model.

[5] Some studies rely on reschedulings (Lee, 1991), others focus on arrears (Sommerville and Taffler, 1995; Aylward and Thorne, 1998) and a third group use both (Detragiache and Spilimbergo, 2001; Peter, 2002). Credit rating agencies also typically consider both arrears and reschedulings. Following Peter (2002), we adopt $\Delta A_{it}/D_{it}$ rather than A_{it}/D_{it} so that, say, countries with large arrears but which are reducing their arrears stock relative to external debt are not classified as default. We set the threshold δ for $\Delta A_{it}/D_{it}$ at its 1983-2002 mean $\hat{\delta} = 2.26\%$.

[6] The estimated models thus can be cast as "early warning" devices. Fuertes and Kalotychou (2004b) discuss how to optimally choose the warning horizon.

$[t, t+2]$. The default frequency in our sample is about 30% (see Appendix B). The number of defaults per year is quite close to those identified by S&P (2001) for rated/non-rated debt.

Country-specific and global variables

A number of macroeconomic variables have been identified as determinants of sovereign ratings (Cantor and Packer, 1996), sovereign defaults (Detragiache and Spilimbergo, 2001) and sovereign spreads (Kamin and Kleist, 1999). Building on these findings we consider 25 domestic signals from the World Bank categories: *i*) external credit exposure, *ii*) external economic activity and financial resources, *iii*) conditions of real/public/financial sector and *iv*) global financial links. To reduce the degree of skewness and kurtosis in the ratios and the number of outliers, these are transformed using $sign(x)\ell n(1+|x|)$. Any remaining outlier in each default/non-default group is reduced by windsorizing the ratios as follows. A data point x_{it} is indexed by $c \in \{0,1\}$ according to whether it pertains to a tranquil ($y_{it}=0$) or default ($y_{it}=1$) window. If x_{it}^c falls outside $\bar{x}^c \pm 4\hat{\sigma}^c$, it is replaced by the appropriate interval limit.

The 5×1 world regressor vector, $\mathbf{z}_t^0$, includes two typical variables — the 10-year US Treasury Bond yield and OECD GDP growth — and three global indicators (annualized) that have not been used in the sovereign default context as yet. One is a proxy for *macroeconomic uncertainty* obtained as the conditional variance of the US monthly logarithmic real GDP.[7] For this purpose, an appropriate AR(1)-GARCH(1,1) model is fitted to the detrended (first differenced) real GDP since the augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests indicate the presence of a unit root in the levels. Second, a measure of *monetary policy uncertainty* is analogously derived from the monthly yield spread between the 3-month Treasury Bill and the US Federal Funds Target rate. 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. Arora and Cerisola (2000) find evidence that sovereign bond spreads are significantly related to such uncertainty measure.[8] Third, the level of

[7] The quarterly US real GDP was interpolated into monthly frequency on the basis of the monthly industrial production (indicator series) using the proportional Denton approach that belongs to a family of LS-based benchmarking methods (Baum, 2001).

[8] The plot of the monthly spread suggests that there is an upward pattern during 1994 (Mexican crisis) and in the second half of 1998 (Asian crisis). The order of the GARCH is selected on the basis of a Ljung-Box test on the squared residuals. Details available from the authors upon request.

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 yield. Fitzgerald and Krolzig (2003) find evidence that this ratio captures risk aversion which dampens the demand of emerging market assets. The latter, in turn, influences capital flows and translates into lower FX reserve levels.

Large models typically have poor statistical properties. In order to preserve degrees of freedom, we deploy a cross-validation (jacknife) approach which assesses the relative value of each regressor on the basis of the in-sample Type I error (see Appendix C).[9] The retained domestic and global signals are denoted by $\mathbf{x}_{it}$ and $\mathbf{z}_t$, respectively. These are discussed 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 $p_{it} \equiv \Pr(y_{it} = 1) = \Pr(y^*_{it} > 0)$ and assuming a standard logistic distribution for ε_{it} it follows that $p_{it} = G(\mathbf{x}_{it}, \mathbf{z}_t) = \frac{\exp(\alpha+\mathbf{x}'_{it}\beta+\mathbf{z}'_t\gamma)}{1+\exp(\alpha+\mathbf{x}'_{it}\beta+\mathbf{z}'_t\gamma)}$. The response probability is thus the logit function evaluated at a linear function of the variables.[10] This nonlinear relation between p_{it} and $(\mathbf{x}_{it}, \mathbf{z}_t)$ can be rewritten linearly in terms of 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* model (PLOGIT) which assumes full country and time homogeneity in the response y^*_{it} to $(\mathbf{x}_{it}, z_t)$. The $(1+k+r)\times 1$ coefficient vector $(\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}')'$ can be estimated by Maximum Likelihood (ML).

[9] Instead one could base the jacknife on some other criteria (e.g. the overall error rate) which may, of course, lead to a different regressor set. For our purpose, however, the relevant aspect is to use the same regressor set for all models in order to make the comparison informative with regard to the treatment of unobserved heterogeneity.

[10] 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 pdf's 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. The latter assumes that the country's characteristics are multivariate normally distributed with a different mean vector (but the same variance-covariance matrix) associated to the default and non-default events. 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).

3.1 Country-specific Heterogeneity

The PLOGIT can be extended to allow for unobserved country-specific effects α_i that stay 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 is an unknown $(N+k+r)\times 1$ coefficient vector to be estimated $\phi^{FE}=(\boldsymbol{\alpha}',\boldsymbol{\beta}',\boldsymbol{\gamma}')'$ 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^2_α. 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 Eq. (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. However, the FE logit is bedevilled by two problems. One is 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 this issue, Chamberlain's (1980) conditional maximum likelihood (CML) estimator of the FE logit 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 (linear or nonlinear) 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.[11]

The *random coefficients* parameterization (RC) goes one step further by allowing for random country heterogeneity both in intercepts and slopes. We consider two variants. First, a model (denoted RC$^\beta$) where the coefficients of the domestic regressors are heterogeneous

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

so that $\boldsymbol{\delta}_i=(\alpha_i,\boldsymbol{\beta}'_i)'$ is a random vector with mean $E(\boldsymbol{\delta}_i)=(\alpha,\boldsymbol{\beta}')'$ and diagonal covariance matrix $E(\widetilde{\boldsymbol{\delta}}_i\widetilde{\boldsymbol{\delta}}'_i)=\Omega$ with $diag(\Omega)=\{\sigma_\alpha,\sigma_{\beta 1},...,\sigma_{\beta k}\}$. Thus we can write $\widetilde{\boldsymbol{\delta}}_i\equiv\boldsymbol{\delta}_i-\boldsymbol{\delta}=\Gamma\mathbf{v}_i$ where

[11] In this paper, we do not consider lagged dependent variables $(y_{i,t-1})$ as regressors. The incidental parameters problem becomes more severe in 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 binary choice models is more complex than in the linear model.

$\boldsymbol{\delta}$ is a $(k+1)\times 1$ vector of fixed means, Γ is a diagonal matrix such that $\Gamma\Gamma^{'}=\Omega$, and $\mathbf{v}_i$ contains $(k+1)$ unobservable latent random terms which are $iid(0,1)$ and independent of $(\mathbf{x}_{it},\mathbf{z}_t)$. The $(2+2k+r)\times 1$ parameter vector to be estimated is $(\alpha,\sigma_\alpha;\boldsymbol{\beta}^{'},\sigma_{\beta 1},...,\sigma_{\beta k};\boldsymbol{\gamma}')$.

Second, we consider a RC$^\gamma$ model where the effect of the global signals $\mathbf{z}_t$ on the log-odds ratio is country heterogeneous — equation (1) with the random vector $\boldsymbol{\delta}_i=(\alpha_i,\boldsymbol{\gamma}_i')'$. The $(2+k+2r)\times 1$ parameter vector to be estimated is $(\alpha,\sigma_\alpha;\boldsymbol{\beta}';\boldsymbol{\gamma}',\sigma_{\gamma 1},\sigma_{\gamma 2},\sigma_{\gamma 3})$. 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).[12]

3.2 Region-specific Heterogeneity

In order to control for region-specific heterogeneity, each country is allocated into one of four groups: I) Asia ($N_{\rm I}=17$), II) Latin America ($N_{\rm II}=26$), III) Africa ($N_{\rm III}=36$), IV) East Europe/Middle East/North Africa ($N_{\rm IV}=17$).[13] We consider two approaches. First, the PLOGIT equation

$$y^*_{it,j}=\alpha_j+\mathbf{x}'_{it,j}\boldsymbol{\beta}_j+\varepsilon_{it,j},\ \varepsilon_{it,j}\sim iid(0,\sigma_j^2),\quad i=1,...,N_j,\quad t=1,...,T \tag{3}$$

is fitted to regions $j=$ I,...,IV.[14] This *regional logit* (RLOGIT) with $4(1+k)\times 1$ 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 to preserve degrees of freedom.

3.3 Time-specific Heterogeneity

Equation (1) controls for common time effects (e.g. oil price shocks) by including the global signals $\mathbf{z}_t$. We consider also a *fixed time effects* (FTE) model that uses period dummies instead

$$y^*_{it}=\alpha_t+\mathbf{x}'_{it}\boldsymbol{\beta}+\varepsilon_{it},\ \varepsilon_{it}\sim iid(0,\sigma^2),\ i=1,...,N,\ t=1,...,T \tag{4}$$

[12] 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 (for details, see Greene, 2003).

[13] There are not enough degrees of freedom for the LOGIT estimation of any of these 3 regions so we group them. They share: i) a similar structure of exports given their oil exporting nature, ii) having gained access to international bond markets between 1995-98.

[14] Alternatively, one could estimate $y^*_{it}=\sum_{j=I}^{IV}\alpha_jD_j+\mathbf{x}'_{it}\boldsymbol{\beta}+\varepsilon_{it},\varepsilon_{it}\sim iid(0,\sigma^2)$ by pooling all countries and using regional dummies. This is however a more restrictive version of (3) where $\boldsymbol{\beta}_j=\boldsymbol{\beta}$.

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 sequence of cross-section relations

$$y_{it}^{*}=\alpha_t+\mathbf{x}_{it}'\boldsymbol{\beta}_t+\varepsilon_{it},\ \varepsilon_{it}\sim iid(0,\sigma_t^2),\ i=1,...,N \tag{5}$$

over time $t=1,...,T$. 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. On this basis we define the counterpart of the Pesaran and Smith's (1995) mean group estimator. Let $\hat{\beta}_{jt}$ denote the slope estimate of regressor j at period t. A *mean cross section* (MCS) estimator is defined as $\bar{\beta}_j^{MCS}\equiv(1/T)\sum_{t=1}^{T}\hat{\beta}_{jt}$ with standard error $SE(\bar{\beta}_j^{MCS})=\sqrt{\frac{SD(\hat{\beta}_{jt})^2}{T}}$ where SD denotes the sample standard deviation.[15]

We should note that the mean MCS 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 estimate long run relations (Pesaran and Smith, 1995; Kennedy, 2003).

3.4 Time-varying 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 is analogously formulated.

These RC-AR formulations allow for the effects of the regressors on the log-odds ratio to vary across countries and over time. The $(3+3k+r)\times 1$ parameter vector of the RC$^{\beta}$-AR logit and the $(3+k+3r)\times 1$ vector of the RC$^{\gamma}$-AR counterpart are estimated by MSL.

3.5 Inference-based Metrics for Model Selection

Several metrics are employed to compare the above models. First, we use the BIC and AIC which have been shown by Monte Carlo simulation to have good finite-sample properties for a range of

[15] If the slopes are random (orthogonal to $\mathbf{x}_{it}$), $\hat{\boldsymbol{\beta}}_t\rightarrow\beta_t$ as $N\rightarrow\infty$ and then $\bar{\beta}^{MCS}$ is consistent as $T\rightarrow\infty$.

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 compares the FE and RE using 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 $\boldsymbol{\theta}$. The null is $\hat{\mathbf{q}} = 0$ and a rejection suggests that there are fixed effects and so the RE model is inconsistent.[16] 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 $\sigma_{\beta_1} = ... = \sigma_{\beta_k} = 0$ reduce RC$^\beta$ to the RE. For $\rho_\alpha = \rho_{\beta_1} = ... = \rho_{\beta_k}$, the RC$^\beta$-AR collapses to the RC$^\beta$. The PLOGIT and FE are also nested ($H_0 : \alpha_i = \alpha$) and the LR statistic follows a $\chi^2_{(N-1)}$ for large T and finite N. Caution is required with the latter test for large N.

The significance of the time effects in the FTE model can be tested with a LR statistic that has a limit $\chi^2_{(T-1)}$ distribution under $H_0 : \alpha_t = \alpha$. We also test for $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}$ and for $H_0 : \alpha_t = \alpha, \boldsymbol{\beta}_t = \boldsymbol{\beta}$ in the TCS model (which reduce it to the FTE and the PLOGIT without globals, respectively) using that $MLL_{TCS} = \sum_t MLL_{CS_t}$. The poolability across regions, $H_0 : \alpha_j = \alpha, \boldsymbol{\beta}_j = \boldsymbol{\beta}$ for $j =$ I,...,IV, can be assessed by means of a LR statistic which is $\chi^2_{3(k+1)}$ distributed.

4 Forecast Framework

Although the panel is unbalanced we denote the sample period by $[1, T]$ for expositional simplicity. A static model was used in Section 3 to simplify the theoretical exposition. In our analysis the regressors are lagged one year for forecasting purposes.[17] The last two years are also lost because of the forward-looking nature of y_{it}. Thus, in effect, y_{it} is observed over 1984-2000. The models are initially estimated over the first 12-year window, denoted $[1, T^*]$, and y_{i,T^*+1} is forecasted.

[16] Unlike in the linear case, the RE estimator is inconsistent (in the presence of fixed effects) even if α_i is orthogonal to $(\mathbf{x}_{it}, \mathbf{z}_t)$. This is because in discrete-choice models the ML estimator is generally inconsistent under misspecification such as in the presence of unmeasured heterogeneity, omitted variables (even if they are correlated with the included ones) and any form of heteroskedasticity (Yatchew and Griliches, 1984).

[17] This also helps to mitigate endogeneity bias.

This window is then rolled forward. Out-of-sample predictions are thus constructed over a holdout period $[T^* + 1, T]$ that spans $m = T - T^* = 5$ years (1996-2000) for $N = 96$ countries. This facilitates a relatively large Nm validation set.

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 $\hat{y}^*_{i,\tau+1} \equiv \hat{\alpha}_\tau + \mathbf{x}'_{i\tau}\hat{\boldsymbol{\beta}}_\tau + \mathbf{z}'_\tau\hat{\boldsymbol{\gamma}}_\tau$ is obtained over $[\tau - T^* + 1, \tau]$ recursively for $\tau = T^*, T^*+1, ..., T-1$. To forecast on the basis of the MCS model we recursively compute $\bar{\alpha}_\tau = (1/T^*) \sum_{t=\tau-T^*+1}^{\tau} \hat{\alpha}_t$ and $\bar{\beta}_\tau = (1/T^*) \sum_{t=\tau-T^*+1}^{\tau} \hat{\beta}_t$ and then construct $\hat{y}^*_{i,\tau+1} = \bar{\alpha}_\tau + \mathbf{x}'_{i\tau}\bar{\boldsymbol{\beta}}_\tau$. The probability $\hat{p}_{i,\tau+1}$ is transformed into an event forecast $(\hat{y}_{i,\tau+1} = 0, 1)$ using a cut-off probability λ_τ which is optimally chosen for each model (see Fuertes and Kalotychou, 2004b).[18]

Several forecast metrics are evaluated both over the Nm points in the *holdout* sample and over a subset called *positive-directional-change* (PDC) sample.[19] The latter excludes year t for country i if $d_{i,t-1} = 1$ so as to focus on the models' ability to predict default entry. This is important since debt crises, in contrast with currency/banking crises, are persistent (see Appendix B).

We adopt statistical and behavioral loss functions. The latter can tailor more closely the forecaster's decision-making problem.[20] Pairwise model comparisons are drawn and additionally, each model is confronted to simple benchmarks. One is a RW type *event* model based on the last observed outcome $\hat{y}^{RW}_{i,\tau+1} = d_{i,\tau}$. Another is a RW type *probability* model, $\hat{p}_{i,\tau+1} = p_{i,\tau}$, where $p_{i,\tau}$ is the prior probability of default (the unconditional frequency of 1s) over the τ rolling window, namely, $\hat{p}^{RW}_{i,\tau+1} = \frac{1}{T^*} \sum_{t=\tau-T^*+1}^{\tau} y_{it}$. Finally, we consider the implicit benchmarks in the Pesaran-Timmermann (1992) and the Donkers-Melenberg (2002) tests.

4.1 Statistical Loss Functions

The *misclassification rate* (MR) defines the overall loss as the frequency of incorrect predictions

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

[18] Extant studies use $\lambda = \mathbf{0.5}$ or fix it in-sample at the default frequency or at the value that minimizes the Type I and II error sum. For each rolling window τ, we find the λ_τ that minimizes the chosen loss function.

[19] Due to missing data we have a heterogeneous N_t, $t = 1, ..., m$ (or equivalently, m_i for $i = 1, ..., N$) holdout panel set. 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.

[20] See Diebold and López (1996) for a comprehensive survey.

where $I(\cdot)$ is an indicator function and λ_t is an optimal cut-off probability.

Scoring rules do not require λ_t because the probability forecast $\hat{p}_{it}$ is directly used. One is the *quadratic probability score* (QPS) or the Brier score which, strictly speaking, is not the direct counterpart of the MSE because it does not compare the realized event y_{it} with $\hat{y}_{it}$ but with $\hat{p}_{it}$

$$QPS = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} 2(\hat{p}_{it} - y_{it})^2, \quad QPS \in [0, 2] \tag{8}$$

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

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

and so it differs from QPS in that large errors are more heavily penalized.

4.2 Economic Loss Functions

Let the following *pay-off matrix* summarise the decision-making problem at hand[21]

		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 θ_t^1 is the economic loss of a missed default and so forth $(\theta_t^j > \phi_t^j, j = 1, 2)$. We build on Granger and Pesaran's (2000) framework but make two simplifying assumptions: *a*) the cost of a correct forecast is zero, $\phi_t^0 = \phi_t^1 = 0$ and *b*) the cost of an incorrect forecast is constant over the holdout period, $\theta_t^1 = \theta^1$, $\theta_t^0 = \theta^0$. We define the latter in relative terms, i.e. $\theta = \frac{\theta^1}{\theta^1 + \theta^0}$.

The following *economic missclassification rate* (EMR) measure

$$EMR_\theta = \frac{1}{Nm}\sum_i\sum_t \theta y_{it}\{1 - I(\hat{p}_{it} > \lambda_t)\} + (1 - \theta)\{1 - y_{it}\}I(\hat{p}_{it} > \lambda_t) \tag{10}$$

provides a family of forecast criteria for $\theta \in [0, 1]$, each giving the overall loss associated to the model predictions $(\hat{p}_{it})$ for a particular decision-making scenario (θ). The forecast ranking from $\text{EMR}_{0.5}$ amounts to that from (7) since $\text{EMR}_{0.5} \equiv \frac{1}{2}\text{MR}$. The MR identically penalises missed defaults and false alarms like the hit rate given by HR=1-MR=$\frac{1}{Nm}\sum_{t=1}^{m}\sum_{i=1}^{N}[y_{it} \times \hat{y}_{it} + (1 - y_{it}) \times (1 - \hat{y}_{it})]$.

[21] Granger and Pesaran (2002) define the economic cost of a decision based on the forecast $\hat{p}_{it}$ as $C_{it}(\hat{p}_{it}) = \phi_t^1 y_{it} I(\hat{p}_{it} > \lambda_t) + \theta_t^0(1 - y_{it})I(\hat{p}_{it} > \lambda_t) + \theta_t^1 y_{it}(1 - I(\hat{p}_{it} > \lambda_t)) + \phi_t^0(1 - y_{it})(1 - I(\hat{p}_{it} > \lambda_t))$.

In practice, the Type I and Type II errors need not be symmetric in their relative importance. For instance, from investors' viewpoint misjudging a bad borrower implies a fall in assets (reflected in the balance sheet) whereas incorrectly dismissing a loan applicant as a bad risk just entails a missed profitable lending opportunity. Hence, the average cost of a Type I error is typically higher than that of the Type II error. Nevertheless, we consider $\theta \in \{0.8, 0.2\}$ in the analysis below.

4.3 Forecast Accuracy Tests

Let $e_{it} \equiv L(y_{it}, \hat{y}_{it}^A) - L(y_{it}, \hat{y}_{it}^B), i = 1, 2, ..., N, t = 1, 2, ..., T$ be the loss differential between models A and B. We deploy the Diebold-Mariano (1995) [DM] test statistic

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

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.[22] This test can readily accommodate non-normality of the forecast errors and is applicable to a wide class of loss functions (Diebold and López, 1996). Below we deploy (11) also to confront each model with the appropriate RW type naive model, y^{RW} or p^{RW}, depending on the loss function.

Pesaran and Timmermann (1992) [PT] propose a nonparametric approach to test the null that forecasts and realizations are independent. The test is formulated on the basis of the HR. The total number of correct out-of-sample predictions (Nm times the model's HR), can be treated as a binomial random variable (the number of successes in Nm trials) 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. Testing for independence amounts to comparing the model's HR with that of the

[22] We have $\bar{e}_i = \frac{1}{m} \sum_t e_{it}$. The variance estimator is $\hat{f} \equiv V(\bar{e}_i) = \frac{1}{m^2} \sum_t V(e_{it}) + \frac{2}{m(m-1)} \sum_t \sum_{s>t} cov(e_{it}, e_{is})$ 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 is assumed, then $DM = \frac{1}{m} \sum_{t=1}^{m} DM_t \sim N(\mathbf{0}, \frac{1}{m})$. The results from the latter are qualitatively similar to those reported from (11) but the statistics are slightly higher. Finally, we 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, this long-run variance is very small ($C_k, k \geq \mathbf{0}$ is computed over just $m \leq \mathbf{5}$ points) and the resulting DM statistics are very large.

implicit benchmark that predicts 1 randomly with probability $\hat{P}$. Thus we have

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

where $HR^{PT} \equiv \tilde{p}$ and a significant PT statistic suggests that the forecasts are dependent on the quantities to be predicted. However, as argued by Donkers and Melenberg (2002), predictive dependence does not imply that the model outperforms an uninformative naive model whose out-of-sample forecast is the in-sample outcome that is most often observed.

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

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

and $\sqrt{Nm}D$ follows 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 DoM statistic is

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

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

5 Results and Discussion

The results of the jacknife are reported in Table 1.[23] The first column reports the results based on the world PLOGIT starting from the initial 30×1 regressor vector $(\mathbf{x}_{it}^0, \mathbf{z}_t^0)'$.

[Table 1 around here]

The retained regressor set 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). Nearly all the debt (*external credit exposure*) ratios have good predictive power, the exception being the

[23] The empirical analysis is conducted using LIMDEP 8.

short-term debt/reserves ratio. The remaining $\mathbf{x}_{it}$ indicate *external economic activity* (2 out of 5 variables retained), *domestic conditions* (5/10) and *global links* (1/4). Interestingly, the retained $\mathbf{z}_t$ are the US macroeconomic uncertainty, monetary policy uncertainty and risk aversion proxies.

Columns 2-4 report the means per state and a t-statistic to assess whether it has good discriminatory power. The regressors retained by the jacknife have a significant mean differential, except for the trade balance/GDP, the real exchange rate misalignment (RER) and the three global indicators. In most cases, the sign of the mean differential can be explained theoretically, e.g. the debt/GDP during pending crisis episodes ($y_{it} = 1$) is about twice its level during tranquil periods. The exception is the short term/total debt ratio (liquidity) which is counterintuitive.

Columns 5-8 denote the region-specific regressor sets that are obtained next by applying the jacknife to each of the four RLOGIT equations starting from the above 16×1 variables.[24] Interestingly, the variables that are dropped in all the regional models are those that are unable to discriminate (t-statistic) between the two states, the only exception being GDP growth. Eight indicators are retained both in the world model and in at least two regional models: five debt burden and liquidity indicators (external debt/GDP, official/total debt, short-term debt/total debt, IMF credit/exports), a measure of macroeconomic control (per capita GNP), a macroeconomic stability signal (volatility of p.c. GNP growth) and a measure of openness (total trade/GDP).

Three domestic indicators are deemed weak in terms of predicting default: trade balance/GDP, GDP growth and the RER. The per capita GNP emerges as a strong signal, in contrast with GDP growth, perhaps because it reflects wealth. Total trade/GDP, which measures the country's degree of trade openness, is identified as a good default predictor in contrast with trade balance/GDP which measures the country's competitiveness. The latter is reflected in the current account of the balance of payments and is closely linked to the RER. This finding supports the 'willingness-to-pay' (as opposed to 'ability-to-pay') theory according to which the opportunity cost of not servicing debt is relatively high for integrated economies.

[24] One should ideally apply the jacknife to each RLOGIT equation starting from the initial **30** regressors but this is unfeasible in terms of degrees of freedom. In order to preserve the latter in the regional models, we use a stricter jacknife where a variable is dropped if, in so doing, the Type I error does not increase by more than 1%.

5.1 Inference-based Comparison

Below we compare rival logit models using several statistical metrics evaluated over 1984-2000.

Model ranking by information criteria

The AIC can be cast as a discrepancy measure between the true model and a candidate whereas the SBC 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 sample size. Table 2 reports the MLL, the number of estimated parameters and the AIC and BIC statistics.

[Table 2 around here]

The best model according to the BIC is the RC^{β}(ng) that allows for random country-specific effects but no time effects (*ng* stands for 'no global variables'). The FE(ng) is generally not favoured (with or without globals) by the BIC but is ranked first by the AIC which less heavily penalises for the large number of estimated parameters. Nevertheless, the fact that the FE model is estimated over a distinct, reduced sample $\tilde{N} = 53$ — the countries for which there is no variation in y_{it} are thrown out — calls for caution in comparing its MLL with that of the other models.

The second BIC-best model is the RC^{β} that allows not only for random country effects but also time effects by including global regressors. The AIC ranks the RC^{β}(ng) and RC^{β} as second and fourth, respectively. The RC^{β}-AR and RC^{β}-AR models where the country-heterogeneneous coefficients are allowed to change over time fare better than both PLOGIT variants and that FTE.

At the bottom of the ranking are the models that control for either time effects (PLOGIT, FTE, MCS) or regional effects (RLOGIT, RSLOGIT) but not for country effects. The RLOGIT is preferred over the PLOGIT vindicating the importance of the regional effects. The worst model is the TCS followed by the RSLOGIT.

Importance of country-specific effects

Table 3 presents the test results for hypotheses regarding the country-specific effects.

[Table 3 around here]

The LR statistic for $H_0 : \alpha_i = \alpha$ using the same reduced $\tilde{N}$ sample for the unrestricted (FE) and restricted (PLOGIT) model is clearly significant.[25] The ML estimates of the FE are widely dispersed, $\hat{\alpha}_i$ ranging from 3.0 to 17.9 with a standard deviation $SD(\hat{\alpha}_i) = 3.4$. We calculate $Z(\hat{\alpha}_i) = \frac{\hat{\alpha}_i - \bar{\alpha}}{SD(\hat{\alpha}_i)}$ where $\bar{\alpha}$ is the sample mean of the $\hat{\alpha}_i$. A coefficient estimate is identified as an outlier in the distribution of $\hat{\alpha}_i$ if $|Z(\hat{\alpha}_i)| > 1$ given that $SD(\hat{\alpha}_i)$ is quite large. About 34% of the countries are outliers. Similarly, the FE(ng) model suggests 32% of outliers.

Moreover, the LR statistic for the homogeneity null ($H_0 : \sigma_\alpha = 0$) in the context of the RE model is also significant. The estimate $\hat{\sigma}_\alpha = 2.35$ (t-ratio=22.49) suggests a large dispersion.[26] The measure $\hat{\sigma}_\alpha^2/(\hat{\sigma}_\alpha^2 + \sigma^2)$ where $\sigma^2 \equiv \pi^2/3$ indicates that 63% of the variation in debt-servicing performance that is unexplained by $(\mathbf{x}_{it}, \mathbf{z}_t)$ is due to unobserved (time-constant) country heterogeneity. The RE versus RC$^\beta$ comparison ($H_0 : \sigma_{\beta_1} = ... = \sigma_{\beta_k} = 0$) also suggests country heterogeneity in $\boldsymbol{\beta}_i$. Likewise, the RE versus RC$^\gamma$ test ($\sigma_{\gamma_1} = ... = \sigma_{\gamma 3} = 0$) indicates that the influence of the global factors on the log-odds ratio is country-specific also.[27,28]

Caution is needed in interpreting the latter set of 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 of the test may not apply. For the case of a single restriction, an easy correction has been suggested — to 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). Most obviously, the corrected test for PLOGIT versus RE remains significant. For joint hypotheses (e.g. PLOGIT versus RC$^\beta$) the correction gets more complicated and is not pursued here. Nevertheless, the test statistics are rather large and so the corrected values are likely to remain significant.

Next we compare the PLOGIT model (ML) estimator and the FE model (Chamberlain's CML) estimator using a Hausman statistic. Under $H_0 : \alpha_i = \alpha$ both are consistent (the CML estimator is inefficient because *i*) it does not use this restriction, *ii*) it is based on a reduced sample) whereas under the alternative the consistent estimator is CML. The Hausman statistic at 32.14 strongly

[25] The brute force test for $H_0 : \alpha_i = \alpha$ gives $LR = \mathbf{470.8(0.00)}$. This result must be interpreted with caution because the MLL$_{FE}$ is obtained from a reduced set ($\tilde{N} = \mathbf{53}$) whereas the MLL$_{PLOGIT}$ is based on all N countries.

[26] For the RE(ng) the estimates are similar at $\hat{\sigma}_\alpha = \mathbf{2.37}$ (t-ratio=23.00).

[27] We also tried a RC formulation which allows for heterogeneity in the 17×1 parameter vector $(\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}')$ but the MSL estimates and standard errors were massive. Hence, we discard this model as implausible.

[28] The $\hat{\boldsymbol{\sigma}}_\beta$ and $\hat{\boldsymbol{\sigma}}_\gamma$ estimates in the RC$^\beta$ and RC$^\gamma$ models, respectively, suggest large country heterogeneity also.

rejects. The regressor set without globals gives a qualitatively similar result at $34.46(0.001, 13)$.[29] The Hausman test for FE versus RE is clearly insignificant and so the latter is preferred.[30] In the comparison between RC^{β} and RE, on the one hand, and between RC^{γ} and RE, on the other, the Hausman test selects the RC models.

Importance of the time-specific and region-specific effects

Next we focus on the statistical importance of the time effects which are controlled for in distinct ways. First, a PLOGIT that includes $\mathbf{z}_t$. Second, the FTE that includes year-specific dummies. Third, the MCS that allows for time variation in the intercept and slopes. Fourth, the RC^{β}-AR (and RC^{γ}-AR) model that extends the RC^{β} (and RC^{γ}) formulation to allow for time-variation in the random country-specific slopes. Table 4 reports the results.

[Table 4 around here]

The 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 shocks to global macroeconomic uncertainty, monetary policy uncertainty and risk aversion. However, the relatively small Hausman statistic for PLOGIT(ng) versus PLOGIT at 4.12 supports the former specification. The joint restriction $H_0 : \alpha_1 = ... = \alpha_{T-1} = \alpha$ on the year dummies in the FTE model is rejected at the 10% level and several of the individual test statistics ($H_0 : \alpha_t = \alpha$) are significant at the 1% level. The Hausman statistic for FTE versus PLOGIT(ng) is clearly insignificant supporting the latter specification.[31]

The LR statistic for $H_0 : \beta_t = \beta$ (or $H_0 : \alpha_t = \alpha, \boldsymbol{\beta}_t = \boldsymbol{\beta}$) in the MCS model is insignificant. But this outcome may be an artefact of the huge number of restrictions being tested (above 200).[32] Indeed, the individual slope estimates suggest that there is marked time heterogeneity.

[29] Likewise, the Hausman test based on the Huber-White estimate of the sampling covariance matrix to account for unspecified latent heterogeneity is $\mathbf{31.16(0.01, 16)}$.

[30] The Hausman test statistic to compare the RE estimate with Chamberlain's FE estimate is very similar at 7.75 (0.96,16). Also, the Hausman test to compare the FE(ng) and RE(ng) gives **7.98(0.84, 13)**.

[31] The counterpart Hausman statistics based on the 'sandwich estimator' (Huber-White) are 12.19 (0.51,13) for PLOGIT versus PLOGIT(ng) and 10.84 (0.62,13) for FTE versus PLOGIT(ng).

[32] In these sequential cross-section regressions, 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 as well and the LR test may not be appropriate.

For instance, the coefficient on GDP growth and the volatility of GNP p.c. growth have ranges [-18.84, 18.16], [-29.21,27.89] and standard deviations of 10.12 and 13.55, respectively. For each of the $j = 1, ..., 13$ regressors, about 30% of $\hat{\beta}_{jt}, t = 1, ..., T$ are identified as outliers using $Z(\hat{\beta}_{jt})$.

The LR test for $H_0 : \rho_\alpha = \boldsymbol{\rho}_\beta = 0$ in the RC$^\beta$-AR model is highly significant and the individual tests for $\boldsymbol{\rho}_\beta$ are significant in 11/13 cases. The null $H_0 : \rho_\alpha = \boldsymbol{\rho}_\gamma = 0$ in the RC$^\gamma$-AR model cannot be rejected but 2/3 of the individual tests for $\boldsymbol{\rho}_\gamma$ are significant. These findings suggest that the influence of the domestic (and possibly the global) indicators $\mathbf{x}_{it}$ on the log-odds ratio of default varies both across countries and over time.

We now assess the importance of controlling for heterogeneity at a regional level in the intercept and slopes $(\alpha_j, \boldsymbol{\beta}_j)', j =$ I,...,IV. We use the same 13 regressors, $\mathbf{x}_{it}$, in each region and test for parameter stability across regions. The LR test indicates that the regional heterogeneity is significant even at the 1% level. This is borne out by the variation across the estimates of the four RLOGIT equations. For instance, the coefficient on debt/GDP, official/total debt and trade/GDP has a regional range of $[5.20, 23.03]$, $[4.19, 30.23]$ and $[-12.60, -3.29]$, respectively. The dispersion of the FE estimates $\hat{\alpha}_i$ within regions is $SD_{II} = 2.7$, $\mathrm{SD}_{III} = 2.9$ and $\mathrm{SD}_{IV} = 2.4$ (the FE is unfeasible for region I) which is lower than within all countries.[33] A Hausman test for homogeneity $(\alpha_i = \alpha)$ within regions gives smaller statistics than the world panel at 18.67(0.00), 31.89(0.00) and 7.00(0.43) for $j =$ II, III, IV, respectively.

Slope estimates: plausibility of the signs

The impact of the domestic indicators (on the probability of default) has a clearcut theoretical sign in 8 out of 13 cases.[34] This is denoted in parenthesis. Table 5 sets out the estimation results.[35]

[Table 5 around here]

[33] Likewise, the world FE model gives $\hat{\alpha}_i$ with $SD_I = 1.2, SD_{II} = 3.1, SD_{III} = 3.1$ and $SD_{IV} = 1.5$.

[34] Eaton and Gersowitz (1981) and Lee (1991) set out a theoretical framework where the default probability hinges on the 'willingness-to-pay'. The higher the volatility of export growth (and of GNP), the more an exclusion from the international capital markets is feared and so the more willing it is to honour its debt (–). Peter (2002) 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 (–).

[35] Let $y^* = \mathbf{x}'\beta + \varepsilon$, the marginal effect of x_j is $\partial p / \partial x_j = G(\boldsymbol{\beta}'\mathbf{x})[1 - G(\boldsymbol{\beta}'\mathbf{x})]\beta_j$. The sign of $\partial p / \partial x_j$ is that of β_j.

Three of the coefficients bear the correct sign and are significant in all models: external debt/GDP (+), official/total debt (+) and trade/GDP (–).[36] The credit to private sector/GDP is significantly negative in all models thus supporting the view that this ratio is a proxy for banking development which is linked with increased economic growth (Bekaert et al. 2002).[37]

The significantly negative effect of GNP per capita is correctly picked up by the FE, RC$^{\beta}$ and RC$^{\gamma}$ models. But 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 (also in the PLOGIT(ng) for Africa and EastEurope/MidEast/NorthAfrica; see Appendix D). For debt service/exports, only the RC$^{\beta}$ captures the expected positive sign although the coefficient is insignificant. For trade balance/GDP, the RC$^{\beta}$-AR model yields the expected negative signed coefficient although it is insignificant also. The AR estimates for RC$^{\beta}$-AR and RC$^{\gamma}$-AR are generally below 0.85 which is suggestive of stationary time-series dependence. Regarding the (unreported) coefficients of the global regressors, the US macro uncertainty has the expected (+) sign and is significant in all models (where included) except for the FE. The US monetary policy uncertainty has the correct (+) sign and is significant in the PLOGIT and RE models. The US risk aversion proxy has the correct (+) sign in the RC$^{\beta}$, RC$^{\beta}$-AR and RC$^{\gamma}$-AR. In contrast to the other models, the FE suggests that all three global indicators are insignificant.[38]

Although the individual CS estimates show massive instability as noted above, their average is plausible and comparable in magnitude to the PLOGIT estimates (Appendix D). This instability is reflected in large standard errors and six insignificant MCS coefficients. Large time series variation is expected in developing economy models due to measurement error. The CS regressions cannot allow for country heterogeneity and this provides a rationale for the instability of $\hat{\boldsymbol{\beta}}_t$. If the omitted factors (idiosyncratic shocks) responsible for the country heterogeneity change over time, this would

[36] An exception is *official/total debt* in the FE model (t-ratio=0.7). The *external debt/GDP* signals the ability to pay debt. A large *total trade/GDP* ratio signals openness and hence, the opportunity cost of default. 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.

[37] The FTE and PLOGIT slopes $\hat{\boldsymbol{\beta}}$ are very close so the former are not reported to preserve space. The t-ratios of the PLOGIT(ng), PLOGIT, R(S)LOGIT based on the asymptotic covariance matrix adjusted for unspecified latent heterogeneity (White's robust 'sandwich estimator') are qualitatively similar. But the t-tests must be interpreted with caution regarding autocorrelation in the residuals (see Fuertes and Kalotychou, 2004b).

[38] In the PLOGIT at least two global signals (macro and monetary policy uncertainty) are significant whereas in the RLOGIT equations none is significant. The latter may be an artefact of the smaller degrees of freedom.

induce different biases in $\hat{\beta}_t, t = 1, ..., T$ which may, however, cancel out when averaged.[39]

5.2 Out-of-sample Forecast Comparison

In sum, the information criteria suggest that models that allow for country heterogeneity (and possibly time effects also) fare better than models that control for heterogeneity at a broader, regional level. The (R)LOGIT and models that exclusively control for time effects such as the PLOGIT with globals, FTE and TCS fare relatively worse. The LR and Hausman tests tend to suggest that unobserved effects (of time, country and regional type) should not be overlooked in modelling the probability of default. However, the model comparison on the basis of the coefficient estimates (signs) is mixed.

We now compare the models' ability to predict outside of the estimation sample. Table 6 presents the overall losses over the entire holdout sample and the PDC subset. The model ranking according to the percentage of Type I and Type II errors is similar to that on the basis of MR for $\theta = 0.8$ and $\theta = 0.2$, respectively. For each loss function, the minimum-forecast-error (in bold) model is contrasted with all other models using the DM test. Asterisks denote a significant loss differential.

[Table 6 around here]

Interestingly, the RSLOGIT provides the best forecasts under the loss functions implicit in the QPS , LPS and EMR$_{0.8}$ criteria over the holdout sample and it significantly outperforms the RW-type naive benchmarks. The DM test suggests that the PLOGIT(ng), PLOGIT, TCS and FTE have similar forecast accuracy to the RSLOGIT for these loss functions. Under the MR loss, the R(S)LOGIT, TCS, FTE, RE(ng) and RC$^\gamma$ forecast significantly better than any other model but not the naive. Under EMR$_{0.2}$, the PLOGIT (ng), PLOGIT, FTE, RE and RC$^\gamma$ are the best models but again the naive generates equally satisfactory forecasts.

Over the PDC subset, the simple PLOGIT(ng) attains the minimum-loss according to the QPS and LPS criteria and it forecasts significantly better than the naive ($\hat{p}_{it}^{RW}$). Under QPS,

[39] In a logit, if the true DGP contains x_{1t} and x_{2t} but $y_{it}^* = \beta_1 x_{1t} + \varepsilon_{it}$ is specified, then plim$\hat{\beta}_1 = \delta_1\beta_1 + \delta_2\beta_2$ where δ_1 and δ_2 are complicated functions of the unknown parameters. If there are individual differences, the CS regression is $y_i^* = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + e_i$ where $e_i = (\alpha_i - \alpha) + \mathbf{x}_i'(\boldsymbol{\beta}_i - \boldsymbol{\beta}) + \varepsilon_i = \varsigma_i + \varepsilon_i$ and ς_i represents the factors responsible for the heterogeneous responses whereas ε_i are true innovations.

the PLOGIT, RSLOGIT and FTE models forecast equally well whereas all other models forecast significantly worse. Under LPS, the PLOGIT(ng) and PLOGIT forecast significantly better than any other model. The PLOGIT, TCS, RSLOGIT and RE models forecast better than any other model (including the naive $\hat{y}_{it}^{RW}$) for the MR loss. Under $\text{EMR}_{0.2}$, the best forecasts are obtained from the PLOGIT(ng), PLOGIT, FTE and $\hat{y}_{it}^{RW}$, all other forecasts are significantly worse. Finally, under $\text{EMR}_{0.8}$ the minimum-loss model is the RSLOGIT whereas the FE, RC^{β}, RC^{β}-AR, RC^{γ}-AR models and $\hat{y}_{it}^{RW}$ are clearly significantly worse even at the 1% level.

It is worth noting that under the $\text{EMR}_{0.2}$ loss, the best forecasting model cannot beat the RW naive. Given that our sample is dominated by 0s, the naive predictor $\hat{y}_{it}^{RW}$ is expected to perform well because this criteria unrealistically penalises more heavily the false alarms than the missed defaults. Nevertheless, the forecast accuracy of the PLOGIT(ng), PLOGIT and FTE is equivalent to that of $\hat{y}^{RW}$ over both the holdout and PDC sets.

We next turn to the predictor dependence (PT) and predictive performance (DoM) tests. The PT statistic is significant in all cases at better than the 1% level (except for the FE model) suggesting that there is positive dependence between predictions and realizations for all models. Unsurprisingly, the DoM test rejects less often since it is relatively more demanding. The hit rate of the naive predictor implicit in the DoM test ($HR^{DoM} = 0.59$) is significantly larger than that of the PT test (which varies with the model) but smaller than that of the RW naive ($HR^{\hat{y}^{RW}} = 0.80$) over the holdout sample.[40] Among the models which pass the DoM test, the FTE or RSLOGIT produce the largest statistics over the holdout sample and the PLOGIT gives the largest statistic over the PDC sample. The FE, RC^{γ}, RC^{γ}-AR, RC^{β} and RC^{γ}-AR models do not pass the DoM test.

In sum, the more general formulations such as FE, RE, RC^{γ}(-AR), RC^{β}(-AR) that allow for unobserved (random) heterogeneity across countries and possibly over time too are good descriptive models but do not predict well out of sample. In contrast, parsimonious regional logit regressions (RSLOGIT) forecast relatively well and outperform the naive benchmarks. The simple PLOGIT, TCS and FTE models that exclusively control for time-specific effects also work quite well as early warning devices.

[40] The benchmark predictor implicit in the DoM test coincides with $\hat{y}^{RW}$ over the PDC set with a hit rate $HR^{DoM} = 0.66$. Also, $HR^{PT} = HR^{y^{RW}} = 0.66$ over the PDC set whereas for all other models $HR^{PT} < 0.66$.

6 Conclusion

The empirical literature on sovereign default is quite vast but a systematic analysis of the importance of controlling for differences in behavior across countries and/or through time that are not captured by the included regressors is lacking. This paper seeks to fill this gap. We estimate different panel logit variants ranging from a simple pooled regression to a general random coefficients model where each country has its own coefficients that are specific to each time period also. The relative quality of the models is gauged on the basis of statistical hypotheses tests, model selection criteria, theoretical judgements on the plausibility of the estimates and forecast metrics.

The LR and Hausman type tests point towards the more general formulations. According to the AIC and SBC, simple models that exclusively control for time or regional heterogeneity are ranked last. The comparison is unclear, however, in terms of the plausibility of the coefficient estimates. Four domestic indicators — external debt to GDP, official to total debt, total trade to GDP and credit to private sector over GDP — emerge as robust default signals together with global macroeconomic uncertainty.

The loss function affects the forecast ranking. The overall picture is that panel logits that exclusively control for either time or regional heterogeneity in a simple manner provide more accurate sovereign default estimates than models that allow for random (country and time) effects. Moreover, the out-of-sample forecast ability of the selected models is superior to the naive benchmarks for most of the loss functions, particularly when a heavier penalty is attached to missed defaults (than to false alarms) or when the emphasis is in predicting entry into a debt crisis period. Therefore it seems reasonable to conclude that the panel model that best describes the data does not necessarily generate accurate sovereign default forecasts. More complexity in the models means, in effect, adding extra terms in the forecast error variance. These findings may have implications for the appropriate assessment of sovereign default risk, a task which is currently being pursued by rating agencies and international investors.

References

[1] Aylward, L., Thorne R., 1998. An econometric analysis of coutries' repayment performance to the IMF. IMF Working Paper.

[2] Arellano, M., Honore B., 2001. Panel data models: Some recent developments, in: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. 5. North-Holland, Amsterdam.

[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, B. H., 2002. Econometric Analysis of Panel Data, John Wiley, New York.

[5] Baltagi, B. H., Griffin, J. M., 1997. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. Journal of Econometrics 77, 303-327.

[6] Baltagi, B. H., Griffin, J. M., Xiong, W., 2000. To Pool or not to Pool: Homogeneous versus Heterogeneous Estimators Applied to Cigarette Demand. The Review of Economics and Statistics 82, 117-126.

[7] Baum, C.F., 2001 A Stata Module to Interpolate a Quarterly Flow Series from Annual Totals via the Proportional DENTON Method. Boston college, mimeo.

[8] Bekaert, A., Harvey, C.R., Lundblad C., 2002. Does financial liberalization spur growth? Duke University Working Paper, Durham.

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

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

[11] Cantor, R., Packer, F., 1996. Determinants and Impact of Sovereign Credit Ratings, Economic Policy Review, Federal Reserve Bank of New York.

[12] Chamberlain, G., 1980. Analysis of covariance with qualitative data. Review of Economic Studies 47, 225-238.

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

[14] Diebold, F.X, López, J.A., 1996. Forecast evaluation and combination, in: Maddala, G.S., Rao C.R. (Eds.), Statistical Methods in Finance, Handbook of Statistics Series, Vol. 14. North-Holland, Amsterdam, pp. 241-268.

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

[16] Donkers, B., Melenberg B., 2002. Testing predictive performance of binary choice models. Econometric Institute Working Paper.

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

[18] Feder, G. R., Just, R. E., Ross, K., 1985. Projecting debt-servicing capacity of developing countries. Journal of Financial and Quantitative Analysis 16, 651-669.

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

[20] Fuertes, A.-M., Kalotychou, E., 2004a. Towards the optimal design of an early warning system for debt crises. Cass Business School Working Paper. Presented at the 2004 SCE Conference on Computing in Economics and Finance, Amsterdam.

[21] Fuertes, A.-M., Kalotychou, E., 2004b. Inferences on the determinants of default. Does autocorrelation matter? Cass Business School Working Paper.

[22] Granger, C.W.J., Pesaran, M.H., 2000. Economic and statistical measures of forecast accuracy. Journal of Forecasting 19, 537-560.

[23] Greene, W.H., 2003. Econometric analysis, Prentice Hall, New Jersey.

[24] Hajivassiliou, V., 1987. The external debt repayment problems of LDS's: An econometric model based on panel data. Journal of Econometrics 36, 205-230.

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

[26] Kamin, S., Kleist, K., 1999. The evolution and determinants of emerging market credit spreads in the 1990s. Bank for International Settlements Working Paper 68, Basel.

[27] Kennedy, P., 2003. A Guide to Econometrics, Blackwell, Oxford.

[28] Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly testing equality and inequality restrictions. Econometrica 50, 1243-1248.

[29] Lee, S.H., 1991. Ability and willingness to service debt as explanation for commercial and official rescheduling cases. Journal of Banking and Finance 15, 5-27.

[30] Maddala, G.S., Srivastava, V.K., Li, H., 1994. Shrinkage estimators for the estimation of long-run parameters from panel data models. Ohio State University Working Paper.

[31] 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: Smith, G., Cuddington, J. (Eds.), International Debt and the Developing Countries. The World Bank, Washington, pp. 179-209.

[32] Oka, C., 2003. Anticipating arrears to the IMF: Early earning systems. IMF Working Paper.

[33] Oral, M., Kettani, O., Cosset, J.C., Daouas, M., 1992. An estimation model for country risk rating. International Journal of Forecasting 8, 583-593.

[34] Pesaran, M.H., Smith, R.P., 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79-113.

[35] Pesaran, M.H., Skouras, S., 2002. Decision-based methods for forecast evaluation, in: Clements, M.P., Hendry, D.F. (Eds.), A companion in economic forecasting. Blackwell, New York, pp. 241-267.

[36] Pesaran, M.H., Timmermann, A., 1992. A simple nonparametric test of predictive performance. Journal of Business and Economic Statistics 10, 461-465.

[37] 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.

[38] Shleifer, A., 2003. Will the sovereign debt market survive? Harvard Institure of Economic Research DP 2000.Sommerville, R.A., Taffler, R.J., 1995. Banker judgement versus formal forecasting models: The case of country risk assessment. Journal of Banking and Finance 19, 281-297.

[39] Staikouras, S.K., 2005. A chronicle of the banking and currency crises. *Applied Economics Letters*, forthcoming.

[40] Standard & Poor, 2001. Sovereign defaults decline through third-quarter 2000, in : S&P (Eds.), Ratings Performance 2000, default, transition, recovery, and spreads. New York, pp. 93-96.

[41] Yatchew, A., Griliches, Z., 1984. Specification error in probit models, Review of Economics and Statistics 66, 134-139.

Appendix A: Data Description

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 factors*
External credit exposure	External econ. activity	Domestic conditions	Global links	
Debt/GDP	Export growth[a]	Credit to private sector/GDP	Trade[g]/GDP	World interest rates[j]
Official debt/Total debt	Vol. of export growth[b]	GDP growth[a]	Net bond flow[f,h]	OECD GDP growth[k]
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]	FDI/GDP[i]	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. [i]Foreign direct investment. [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: Historical sovereign defaults per country 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: The cross-validation method

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 rate $\lambda = 0.5$ 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 $\check{X}$ 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

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

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: Cross-Section and regional estimates

	CS			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 1
Regressor set selected for world and regional PLOGIT models

World panel				Regional panels			
	mean			(I)	(II)	(III)	(IV)
N=96	y=0	y=1	t-stat	N=17	N=26	N=36	N=17
A) Country-specific indicators							
Total external debt/ GDP	0.3879	0.7032	23.76*	X	X	X	×
Official debt / Total debt	0.5633	0.5933	6.91*	X	X	X	X
Short term debt / Total debt	0.1244	0.1109	2.99*	X	×	X	X
Debt service / Exports	0.1492	0.1843	6.70*	×	X	X	×
IMF credit / Exports	0.0828	0.1529	9.68*	X	×	X	X
Volatility of export growth	0.1072	0.1391	6.14*	×	×	×	X
Trade balance / GDP	-0.0797	-0.0822	0.40	×	×	×	×
Credit to private sector/ GDP	0.2746	0.1785	13.41*	×	×	×	X
GDP growth	0.0398	0.0247	6.26*	×	×	×	×
GNP per capita	7.1236	6.4402	13.87*	X	×	X	X
Volatil. of GNP p.c. growth	0.0435	0.0529	6.35*	X	X	×	×
Real exchange rate	0.1273	0.1110	0.74	×	×	×	×
Trade / GDP	0.5427	0.4737	6.93*	X	X	X	X
B) Global indicators							
US macroeconomic uncertainty	0.2246	0.2281	0.81	×	×	×	×
US monetary policy uncertainty	0.2566	0.2572	0.18	×	×	×	×
US risk aversion	1.0453	1.0458	0.03	×	×	×	×

The variable selection is conducted over the in-sample [1984, 1995] period using the jacknife.
(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.
*denotes significant at the 1% level.

Table 2
Model comparison on the basis of information criteria

Model type	Controlled effects	MLL	s	AIC statistic	AIC ranking	BIC statistic	BIC ranking
PLOGIT(ng)	–	-606.3	14	620.3	12	656.5	10
PLOGIT	time	-601.4	17	618.4	11	662.4	12
MCS	time	-491.8	$14 \times T$	729.8	15	1007.5	15
FTE	time	-594.4	30	624.4	13	702.0	13
RLOGIT	regional	-500.1	$14 \times R$	556.1	10	660.0	11
RSLOGIT	regional	-646.7	$\sum_{j=1}^{R} p_j$	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$^{\beta}$ (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$^{\gamma}$	country, time	-457.8	21	478.8	6	533.1	5
RC$^{\gamma}$-AR	country, time	-455.7	25	480.7	7	545.4	6
RC$^{\beta}$	country, time	-408.6	31	439.6	4	519.8	2
RC$^{\beta}$-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.
$AIC = -\sum_{t=1}^{T} MLL_t + \sum_t s_t$ and $BIC = -\sum_t MLL_t + 0.5 s_t \sum_t \ln(N_t)$ for the TCS model, where N_t is the no. of available observations per cross-section.
$AIC = -\sum_{j=1}^{R} MLL_j + \sum_j s_j$ and $BIC = -\sum_j MLL_j + 0.5 s_j \sum_j \ln(NT_j)$ for the R(S)LOGIT, s_j are the number of coefficients, NT_j the data points per region (R=4).

Table 3
Statistical significance of country-specific effects

Tests	Model type					
	FE	RE	RC$^{\beta}$		RC$^{\gamma}$	
A) Likelihood ratio						
null hypothesis	$\alpha_i=\alpha$	$\sigma_\alpha=0$	$\sigma_\alpha=0$ $\sigma_{\beta_1}=...=\sigma_{\beta_k}=0$	$\sigma_{\beta_1}=...=\sigma_{\beta_k}=0$	$\sigma_\alpha=0$ $\sigma_{\gamma_1}=...=\sigma_{\gamma_3}=0$	$\sigma_{\gamma_1}=...=\sigma_{\gamma_3}=0$
	PLOGIT	PLOGIT	PLOGIT	RE	PLOGIT	RE
test statistic	141.1***	267.3***	385.7***	118.4***	287.3***	20.0***
	(0.00, 52)	(0.00, 1)	(0.00, 14)	(0.00, 13)	(0.00, 4)	(0.00, 3)
B) Hausman-type						
null hypothesis	CML=PLOGIT	RE=FE	RC$^{\beta}$=RE		RC$^{\gamma}$=RE	
test statistic	32.14***	7.559	292.2***		78.58***	
	(0.01, 16)	(0.96, 16)	(0.00, 16)		(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, FE) or by MSL using Halton draws (RE, RP). CML denotes the Chamberlain's conditional ML estimator of the FE slopes.
The MLL_{FE} and $\text{MLL}_{\text{PLOGIT}}$ for the LR test of $H_0: \alpha_i=\alpha$ are based on the largest possible sample (Ñ=53) for the former.
*, ** and *** denote significant at the 10% , 5% and 1% level.

Table 4
Statistical significance of time- and regional-specific effects

Tests	Time effects						Regional effects
	PLOGIT	FTE	MCS		RC$^{\beta}$-AR	RC$^{\gamma}$-AR	RLOGIT
A) Likelihood ratio							
null hypothesis	$\boldsymbol{\gamma}=0$	$\alpha_t=\alpha$	$\boldsymbol{\beta}_t=\boldsymbol{\beta}$	$\alpha_t=\alpha, \boldsymbol{\beta}_t=\boldsymbol{\beta}$	$\rho_\alpha=0, \boldsymbol{\rho}_\beta=0$	$\rho_\alpha=0, \boldsymbol{\rho}_\gamma=0$	$\alpha_j=\alpha, \boldsymbol{\beta}_j=\boldsymbol{\beta}$,
	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)	(0.09, 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 significance at the 10% , 5% and 1% level.

Table 5

Parameter estimates of default probability models 1984-2000

	latent effects											
		time		country	regional		country, time					
Variables	PLOGIT(ng)	PLOGIT	MCS	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 C for details).

Table 6
Out-of-sample forecast analysis

Model	Statistical loss: MR	QPS	LPS	PT	DoM	Economic loss: $EMR_{0.2}$	$EMR_{0.8}$
A: holdout sample							
naive	0.2014	0.4117^{*}	1.1588^{**}	5.00^{**}	3.73^{**}	0.0643	0.1372^{**}
PLOGIT(ng)	0.2700^{*}	0.3275	0.5130	4.41^{**}	1.90^{*}	0.0754	0.0860
PLOGIT	0.2562^{*}	0.3289	0.5168	4.51^{**}	2.20^{*}	0.0712	0.0927
MCS	0.2450	0.3435	0.5631	4.67^{**}	2.39^{**}	0.0973^{*}	0.0776
FTE	0.2354	0.3367	0.5353	4.53^{**}	2.87^{**}	0.0703	0.0971^{*}
RLOGIT	0.2314	0.3275	0.5904^{**}	4.79^{**}	2.72^{**}	0.1004^{*}	0.0835
RSLOGIT	0.2320	0.3060	0.5043	4.68^{**}	2.82^{*}	0.0960^{*}	0.0760
RE(ng)	0.2434	0.3520^{*}	0.6267^{*}	4.36^{**}	2.77^{**}	0.0938^{*}	0.0948
RE	0.2514^{*}	0.3530^{*}	0.6335^{*}	4.24^{**}	2.59^{**}	0.0897	0.0948
FE	0.3898^{**}	0.5251^{**}	0.8874^{**}	1.72^{*}	0.29	0.1725^{**}	0.1344
RC^{γ}	0.2492	0.3629^{*}	0.6672^{**}	4.51^{**}	1.37	0.0836	0.0943
RC^{γ}-AR	0.3068^{**}	0.5239^{**}	2.2647^{**}	3.61^{**}	1.46	0.1388^{**}	0.1236^{**}
RC^{β}	0.3220^{**}	0.6231^{**}	3.1384^{**}	3.06^{**}	1.37	0.1581^{**}	0.1588^{**}
RC^{β}-AR	0.3088^{**}	0.5752^{**}	2.6447^{**}	3.25^{**}	1.61	0.1456^{**}	0.1594^{**}
B: PDC sample							
naive	0.3354^{**}	0.4418^{**}	1.2197^{*}	0.00	0.00	0.0671	0.2683^{**}
PLOGIT(ng)	0.3292^{**}	0.3489	0.5469	3.38^{**}	0.07	0.0819	0.0970
PLOGIT	0.2400	0.3491	0.5508	3.79^{**}	2.92^{**}	0.0776	0.1023
MCS	0.2294	0.3713^{*}	0.6153^{**}	3.92^{**}	1.69^{*}	0.1055^{**}	0.0960
FTE	0.3228^{**}	0.3621	0.5788^{**}	3.16^{**}	0.16	0.0817	0.1047^{*}
RLOGIT	0.2716^{**}	0.3992^{**}	0.7290^{**}	3.71^{**}	1.57	0.1193^{**}	0.1037
RSLOGIT	0.2474	0.3685	0.5935^{*}	4.26^{**}	1.83^{*}	0.1233^{**}	0.0887
RE(ng)	0.2648^{**}	0.3776^{*}	0.6718^{**}	2.83^{**}	1.22	0.1064^{**}	0.0909
RE	0.2492	0.3743^{*}	0.6730^{**}	3.76^{**}	2.01^{*}	0.0996^{*}	0.0951
FE	0.3642^{**}	0.6321^{**}	1.0806^{**}	0.77	-1.12	0.2091^{**}	0.1691^{**}
RC^{γ}	0.2788^{**}	0.3916^{**}	0.7158^{**}	3.11^{**}	0.84	0.0900^{*}	0.0947
RC^{γ}-AR	0.3172^{**}	0.5483^{**}	2.1234^{**}	2.89^{**}	0.24	0.1523^{**}	0.1280^{**}
RC^{β}	0.3174^{**}	0.5994^{**}	2.9334^{**}	3.12^{**}	0.61	0.1782^{**}	0.1225^{**}
RC^{β}-AR	0.3096^{**}	0.5774^{**}	2.5995^{**}	2.16^{*}	0.30	0.1470^{**}	0.1658^{**}

Bold denotes the minimum loss. **, * under (E)MR, QPS or LPS indicates that the model's forecast ability is significantly worse than that of the minimum-loss model according to a Diebold-Mariano test at the 1% and 5% level, respectively. PT is the Pesaran-Timmerman test of predictor dependence and DoM is Donkers-Melenberg test of predictive performance. The naive prediction is $\hat{y}^{RW}$ under MR, $EMR_{0.2}$ and $EMR_{0.8}$. The naive predictor is $\hat{P}^{RW}$ under QPS and LPS.