UNIVERSITÀ POLITECNICA DELLE MARCHE
Dipartimento di Scienze Economiche e Sociali

PENALIZED MAXIMUM LIKELIHOOD ESTIMATION OF LOGIT-BASED EARLY WARNING SYSTEMS

CLAUDIA PIGINI

QUADERNI DI RICERCA n. 441

ISSN: 2279–9575

Novembre 2019

Abstract

Panel logit models have proved to be simple and effective tools to build Early Warning Systems (EWS) for financial crises. But because crises are rare events, the estimation of EWS does not usually account for country fixed effects, so as to avoid losing all the information relative to countries that never face a crisis. I propose using a penalized maximum likelihood estimator for fixed-effects logit-based EWS where all the observations are retained. I show that including country effects, while preserving the entire sample, greatly improves the predictive power of EWS with respect to the pooled, random-effects and standard fixed-effects models.

JEL Class.: C23, C25, G17, G21
Keywords: Keywords: Banking Crisis, Bias Reduction, Fixed-Effects Logit, Separated Data

Penalized maximum likelihood estimation of logit-based Early Warning Systems

Claudia Pigini

1 Introduction

Logit models have proved to be simple and effective tools to build Early Warning Systems (EWS) for financial crises. Their predictive power is employed to generate accurate out of sample warning signals and their specification as binary choice models offers a clear interpretation of the drivers of financial, especially banking, crises. Building on the seminal works by Demirgüç-Kunt and Detragiache (1998, 2005), cross-country logit-based EWS have been proved to outperform country-specific signal extraction EWS as concerns in and out of sample predictions (Davis and Karim, 2008). These models have also been recently extended in order to exploit the information about the crisis duration, when it lasts more than one year (Caggiano et al., 2016; Antunes et al., 2018).[1]

When logit-based EWS are built on panel data, permanent country-specific unobserved heterogeneity could be accounted for by including fixed effects, which would supposedly improve the model predictive power. But because crises are rare events, the estimation of EWS does not usually account for country fixed effects, so as to avoid losing a sizable number of countries in the dataset. This is due to the complete separation problem, because of which the Maximum Likelihood (ML) estimate of the intercept of a country that never experiences a crisis is $-\infty$ and thereby prevents it from contributing to the estimation sample. Unfortunately, retaining the whole sample of subjects and excluding only the intercepts for those countries with all zeros (or ones) in the dependent variable will lead to a biased ML estimator (Heinze and Schemper, 2002). Moreover, the ML estimator of a fixed-effects binary choice model has the additional bias caused by the incidental parameters problem, if the panel time-series is short (Neyman and Scott, 1948; Lancaster, 2000).[2]

The common practice when estimating logit-based EWS is therefore to neglect fixed effects and rely on a pooled logit model, while retaining those countries that never face a banking crisis in the period considered, so as to be able to predict the occurrence of a first crisis out of sample. Besides, doing so makes it possible to investigate the factors that help avoiding the occurrence of a crisis, rather than just identifying those that signal it (Eberhardt and Presbitero, 2018). One exception is the empirical analysis conducted by Schularick and Taylor (2012), whose dataset covers the period 1870-2008 and where

[1] The empirical literature on applications of EWS for banking crises is extensive and not presented in this paper, where I only focus on methodological issues. For a recent review, see (Antunes et al., 2018). Recently, EWS have also been used to identify the drivers of sovereign debt crises. See, for instance, Dawood et al. (2017) and Bassanetti et al. (2018) and references therein.

[2] A way to overcome the incidental parameters problem is to estimate a fixed-effects logit model by Conditional Maximum Likelihood (Andersen, 1970; Chamberlain, 1980), which removes the unobserved heterogeneity by conditioning on suitable sufficient statistics for the individual intercepts. This way, however, country-specific effects are not estimated and, therefore, not included in the computation of the predicted probability.

every country faces at least one year of crisis; their baseline specification is a fixed-effects logit and the inclusion of country-specific intercepts is shown to greatly improve the model predictive power upon the pooled model. Alternatively, Eberhardt and Presbitero (2018) and Dawood et al. (2017) consider random-effects binary choice models, that allow for some country unobserved heterogeneity to be accounted for although subject to strong parametric restrictions, namely the normality assumption with constant variance across countries.

In this paper, I propose estimating logit-based EWS by a Penalized Maximum Likelihood (PML) approach, which allows for the inclusion of fixed effects while retaining the whole sample of countries. The bias of the ML estimator for the intercepts relative to countries that never face a crisis is dealt with by applying the bias reduction technique put forward by Firth (1993) and Kosmidis and Firth (2009), who defined the PML estimator as the solution to a modified score function. Under the assumption that every country will experience a crisis as $T \to \infty$, the bias of the ML estimator is reduced from $O(T^{-1})$ to $O(T^{-2})$, where T is the number of time occasions. This bias reduction technique was first adapted to the separation problem in the binary logit model by Heinze and Schemper (2002) and Heinze (2006); recently, in the context of fixed-effects logit models, it has been applied to the evaluation of hospital readmission reduction programs by Kunz et al. (2017) and to the forecasting of civil wars by Cook et al. (2018). Moreover, Firth's approach reduces, by the same order, the bias due to the incidental parameters problem (Hahn and Newey, 2004), which affects the estimates of the slope coefficients and the remaining country effects.

The performance of the PML estimator is first evaluated by means of a brief simulation study and then by its application to logit-based EWS estimated using an unbalanced panel dataset, which consists of 129 countries from 1982 to 2017 and where systemic banking crisis events are defined as in Laeven and Valencia (2018). I show that including country effects, while preserving the entire sample, greatly improves the predictive power of EWS with respect to the pooled, random-effects and standard fixed-effects logit models. I also consider a dynamic formulation of EWS by including the lagged dependent variable among the set of covariates, which has been shown to substantially enhance the model predictive performance (Antunes et al., 2018). Adding to this framework and similarly to Ghulam and Derber (2018), I show that a dynamic binary choice model in this context allows for the prediction of two separate probabilities which can be of interest to policymakers: the crisis *entry* rate, that is the probability of a crisis a time t given that the country was not in a crisis state at time $t-1$, and the crisis *persistence* rate, which is the probability of that a country faces a prolonged state of financial distress at time t conditional on that country already having faced a crisis at time $t-1$

I compare both in and out of sample forecasts for EWS estimated by different methods and evaluate their performance in classifying crisis events by plotting Receiver Operating Characteristics (ROC) curves and comparing the related Areas Under ROC (AUROC), as customary in this literature. I also consider predictions based on the F-score, which it is argued to be more suitable for the forecasting of rare events (Davis and Goadrich, 2006).

The rest of the paper is organized as follows: Section 2 introduces the problem of separation with fixed-effects logit models, illustrates the PML estimator and briefly describes the tools for evaluating model performance; Section 3 reports the results of the simulation study; Section 4 describes the panel dataset and reports descriptive statistics; Section 5 illustrates the estimation results, and finally Section 6 concludes.

2 Methodology

In this section I illustrate the proposed approach to the estimation of logit-based EWS. I first recall the binary panel logit model and the separation problem that originates with rare events and fixed effects; secondly, I describe the PML derived using Firth's bias reduction technique; I then recall the dynamic logit model and illustrate the definition of crisis *persistence* and *entry* rates; finally, I briefly describe the techniques to evaluate the model predictive power.

2.1 The fixed-effects logit model

Consider a sample of countries indexed by i, for $i = 1, \dots, n$, observed in year t, for $t = 1, \dots, T$. In a logit-based EWS, the probability formulation is

$$p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) = \frac{\exp\left[y_{it}\left(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}\right)\right]}{1 + \exp\left(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}\right)}, \tag{1}$$

where y_{it} is a binary variable equal to 1 if a crisis occurs in country i at time t and 0 otherwise, $\boldsymbol{x}_{it}$ is a vector of relevant predictors (usually lagged), and $\boldsymbol{\beta}$ is a vector of parameters representing the early warnings. The country-specific intercept α_i collects permanent traits that are unobserved to the researcher such as, for instance, unmeasured cultural factors.

Relying on a fixed-effects approach means treating the country-specific effects α_i, $i = 1, \dots, n$, as parameters to be estimated. The ML estimator of $\boldsymbol{\beta}$ is obtained by concentrating out the α_i as the solution to

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \log p(y_{it}|\boldsymbol{x}_{it}; \hat{\alpha}_i(\boldsymbol{\beta}), \boldsymbol{\beta}), \tag{2}$$

where

$$\hat{\alpha}_i(\boldsymbol{\beta}) = \underset{\alpha_i}{\operatorname{argmax}} \sum_{t=1}^{T} \log p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}). \tag{3}$$

If the occurrence of $y_{it} = 1$ is a rare event, the ML estimate of α_i might not be finite for some i. To see this, rewrite the log-likelihood in (3) as

$$\sum_{t=1}^{T} \log p(y_{it}|\boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) = \sum_{t=1}^{T} y_{it}(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}) - \sum_{t=1}^{T} \log\left[1 + \exp\left(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}\right)\right].$$

The above function is always decreasing in α_i if $y_{i1} = \ldots = y_{iT} = 0$, that is country i never faces a crisis in the period considered. Therefore, the ML estimator $\hat{\alpha}_i(\boldsymbol{\beta}) = -\infty$. This is known as the *separation* problem, as characterized by Albert and Anderson (1984) for logistic regression models, which can occur every time responses are perfectly predicted by a single binary variable. If the sample contains subjects such that the dependent variable is always 0 (or 1), then the ML estimator of $\boldsymbol{\beta}$ is obtained only using those subjects for which $0 < \sum_{t=1}^{T} y_{it} < T$ and consequently only the related individual intercepts will be estimated.

This means that if the response variable indeed represents a rare event, fixed-effects estimation entails losing a sizable portion of the estimation sample, due to the fact that all countries that never face a crisis in the period considered have to be discarded. Other than the obvious loss of efficiency, this limits the potential of EWS, which could be used for forecasting financial crises out of sample, even for those countries that have never seen any. One solution could be to keep the whole sample of countries while not estimating the intercepts for those that never face a crisis. This strategy would give rise, however, to a biased ML estimator of the model parameters if the unobserved heterogeneity specific to these countries were correlated with the model covariates.

In addition, the ML estimator of the parameters of a fixed-effects binary choice model is generally inconsistent because of the incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000). From equation (3) it can be noticed that $\hat{\alpha}_i(\boldsymbol{\beta})$ depends on the data only through $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})$. As a consequence, the ML estimator $\hat{\alpha}_i(\boldsymbol{\beta})$ from (3) is consistent only if $T \to \infty$, meaning that the ML estimator of $\boldsymbol{\beta}$ obtained from (2) is also not consistent unless $T \to \infty$.

The incidental parameters problem together with the exclusion of a potentially high number of countries from the estimation sample has led practitioners to rely on pooled models for the estimation of logit-based EWS. It must be stressed, however, that the ML estimator of a pooled binary choice model is consistent if there is no time-invariant unobserved heterogeneity, meaning that country-specific permanent effects can be safely neglected. Clearly this is a rather strong assumption, which is why random effects logit-based EWS have been recently considered (Dawood et al., 2017; Eberhardt and Presbitero, 2018). Yet, this approach also presents some limitations, as it requires the unobserved heterogeneity to be independent of the model covariates and to have constant variance across countries. The first can be partially overcome by a correlated random-effects approach (Mundlak, 1978) applied to logit-based EWS (Eberhardt and Presbitero, 2018), although the dependence between the model covariates and the unobserved heterogeneity is modeled parametrically and consistency of the ML estimator relies on the strict exogeneity assumption.[3] On the contrary, a fixed-effects approach allows the unobserved heterogeneity to be correlated with the model covariates in a nonparametric way and the ML estimator does not require the strict exogeneity assumption. In addition, the presence of country–specific fixed effects is bound to yield a better predictive performance.

2.2 Penalized maximum likelihood estimator

Heinze and Schemper (2002) addressed the complete separation problem in logistic regressions by adopting the bias reduction technique put forward by Firth (1993), which removes the leading term of the small sample bias of the ML estimator. In fact, the bias introduced by the complete separation problem can be viewed as a small sample bias due to fixed T under the assumption that, for instance, every country will eventually experience a crisis, which we could see if only we observed all of its history. The bias caused by the incidental parameters problem will be reduced as well, as it vanishes as $T \to \infty$ (Hahn and Newey, 2004; Fernández-Val, 2009).

In order to illustrate Firth's bias reduction, let me consider a regular problem with the parameter

[3]Following the expression given by Chamberlain (1982) for panel data binary choice models, the strict exogeneity assumption requires that, conditional on the unobserved heterogeneity, a crisis (or no crisis) event at time t has no effect the value of the explanatory variables at time $t+1, t+2, \ldots$.

vector $\boldsymbol{\theta}_0$. The ML estimator is derived as the solution to the score equation $U(\hat{\boldsymbol{\theta}}) = 0$. If $U(\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}}$ is unbiased, $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}_0$. However, even if the score function is unbiased, bias in $\hat{\boldsymbol{\theta}}$ can arise from the curvature in the score function. This bias shows up in finite samples and vanishes as the number of observations goes to infinity. Given h the number of observations in a sample, an approximation of the the bias in $\hat{\boldsymbol{\theta}}$ can be written as

$$b(\hat{\boldsymbol{\theta}}) = \frac{b_1(\hat{\boldsymbol{\theta}})}{h} + \frac{b_2(\hat{\boldsymbol{\theta}})}{h^2} + \dots \tag{4}$$

Bias reduction techniques usually aim at estimating the leading term (or an approximation of it) in (4), so as to obtain the bias corrected estimator $\hat{\boldsymbol{\theta}}_{BC} = \hat{\boldsymbol{\theta}} - \frac{b_1(\hat{\boldsymbol{\theta}})}{h}$, whose order of bias has been reduced from $O(h^{-1})$ to $O(h^{-2})$. By contrast, Firth's proposal can be viewed as a *preventive* rather than a *corrective* strategy, as the above, in that he proposes reducing the first order bias by introducing a bias in the score function $U(\hat{\boldsymbol{\theta}})$. Let me denote as $\tilde{\boldsymbol{\theta}}$ the ML estimator that has bias of order $O(h^{-2})$. Then $\tilde{\boldsymbol{\theta}}$, which can be regarded as the PML estimator, is the solution to

$$U^*(\tilde{\boldsymbol{\theta}}) = U(\tilde{\boldsymbol{\theta}}) - I(\tilde{\boldsymbol{\theta}})\tilde{b}_1(\hat{\boldsymbol{\theta}}) = 0, \tag{5}$$

where $I(\boldsymbol{\theta})$ is the information matrix and $\tilde{b}_1(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}$, for which Firth relies on the approximation provided by Cox and Snell (1968). Expression (5) is the result of a first–order Taylor expansion of $U(\hat{\boldsymbol{\theta}})$ around $\tilde{\boldsymbol{\theta}}$.[4]

Based on these premises, Heinze and Schemper (2002) provided the expression of the modified score function for the logistic regression, which can be easily adapted to the fixed-effects logit model. Let $\boldsymbol{\theta}$ collect the parameters for model (1), that is $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha_1, \dots, \alpha_n)'$, then the modified score function becomes

$$U^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{t=1}^{T} \left[y_{it} - \pi_{it} + m_{it}(0.5 - \pi_{it}) \right] \boldsymbol{z}'_{it}, \tag{6}$$

where $\pi_{it} = \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}) / [1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]$, $\boldsymbol{z}_{it} = (\boldsymbol{x}'_{it}, \boldsymbol{d}'_{it})'$, with $\boldsymbol{d}_{it}$ being a vector of country dummies for the t-th time occasion, and m_{it} is the diagonal element of the matrix $\boldsymbol{M} = \boldsymbol{W}^{1/2}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{W}^{1/2}$, where $\boldsymbol{W} = \mathrm{diag}[\boldsymbol{\pi}(1-\boldsymbol{\pi})]$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n)'$, with $\boldsymbol{Z}_i = (\boldsymbol{z}'_{i1}, \dots, \boldsymbol{z}'_{iT})'$. The PML estimator is the solution to (6) being equal to zero and it can be obtained by Newton-Raphson with update for the $s+1$-th step

$$\tilde{\boldsymbol{\theta}}^{s+1} = \tilde{\boldsymbol{\theta}}^{s} + I\left(\tilde{\boldsymbol{\theta}}^{s}\right)^{-1} U^*\left(\tilde{\boldsymbol{\theta}}^{s}\right).$$

Heinze and Schemper (2002) also show that a good approximation of the standard errors can be based on the Information Matrix.

[4]Firth also shows that an equivalent approach is based on the penalized likelihood $L^*(\boldsymbol{\theta}) = L(\boldsymbol{\theta})|I(\boldsymbol{\theta})|^{1/2}$, where the penalty function $|I(\boldsymbol{\theta})|^{1/2}$ is Jeffreys invariant prior.

2.3 Dynamic logit model

Clearly, once they occur, financial crises can be persistent over time, meaning that there might be a certain likelihood of observing crisis periods for a number of subsequent years due to a prolonged state of financial distress. This poses a methodological issue that Caggiano et al. (2016) denoted as *duration bias*, arguing that not accounting for a period of adjustment after the outbreak of a crisis might reduce the performance of the EWS.

The common practice in the empirical literature is setting to zero the binary outcome variable for the years right after the occurrence of the crisis. Differently, in order to improve the predictive performance of the EWS, Caggiano et al. (2016) rely on a multinomial logit specification, where three different categories of the outcome variable represent a tranquil year, the onset of a crisis, and an additional year of financial distress, respectively.

Within the binary logit-based EWS, persistence can be accounted for by specifying a dynamic model (Hsiao, 2015), i.e. the lagged dependent variable is included in the set of regressors. The probability of a crisis can be written as

$$p(y_{it}|\boldsymbol{x}_{it}, y_{i,t-1}; \delta_i, \boldsymbol{\phi}, \gamma) = \frac{\exp\left[y_{it}\left(\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi} + \gamma y_{i,t-1}\right)\right]}{1 + \exp\left(\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi} + \gamma y_{i,t-1}\right)}, \tag{7}$$

where the additional parameter γ identifies the *true* state dependence, that is the effect for country i at time t of having experienced a crisis in $t-1$ on the probability of a crisis occurring again, separately from the propensity to be financially distressed at all times (Heckman, 1981). This strategy is also considered by Antunes et al. (2018), who rely on a dynamic pooled probit model and show that it greatly outperforms its static counterpart in terms of both in and out of sample forecasts. The PML estimator proposed here can be directly tailored to the fixed-effects dynamic logit model.

The dynamic logit model allows to predict both the marginal probability of facing a crisis and its conditional probability given the country status at $t-1$. Therefore, one can compute the probability that a crisis at time t occurs in country i given that the country was not in distress at time $t-1$, which is the crisis *entry* rate and can be written as

$$e_{it} = \Pr(y_{it} = 1|y_{it-1} = 0) = \frac{\exp\left(\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi}\right)}{1 + \exp\left(\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi}\right)}. \tag{8}$$

Similarly, the probability that country i at time t is still in distress given that $t-1$ was a crisis year is the *persistence* rate, that is

$$p_{it} = \Pr(y_{it} = 1|y_{it-1} = 1) = \frac{\exp\left(\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi} + \gamma\right)}{1 + \exp\left(\delta_i + \boldsymbol{x}_{it}'\boldsymbol{\phi} + \gamma\right)}. \tag{9}$$

These definitions are borrowed from the empirical literature on poverty traps, where poverty dynamics are modeled as a Markov process (Cappellari and Jenkins, 2002, 2004) and have the same rationale of the approach adopted by Ghulam and Derber (2018) in the context of duration models for sovereign defaults.

2.4 Evaluating model performance

In the early warning literature, the standard approach to evaluate the accuracy of the model prediction is based on the correct classification of crisis and non-crisis events, which is often assessed by means of ROC curves and the related AUROC (Hanley and McNeil, 1982; Hsieh et al., 1996).

With logit-based EWS, a crisis for country i at time t is predicted if the estimated response probability is greater than some cut-off value. Based on this cut-off, the True Positive Rate (TPR, or sensitivity) is the fraction of correctly classified crisis events and the False Positive Rate (FPR, also 1 - specificity) is the number of misclassified crisis events over the total non-crisis events. The plot of the TPR against the FPR, computed for every predicted outcome used as cut-off, is the ROC curve. The AUROC, that is the area under the ROC curve, is then a measure of the model performance, where a value of the AUROC equal to 0.5 refers to random classification, whereas a value equal to 1 represents perfect classification. Alongside the AUROC, the percentages of crises correctly predicted and false alarms for an optimal cut-off is usually presented as an immediate and straightforward signal on the model ability to predict financial crises. The cut-off is often chosen (as, for instance, by Dawood et al., 2017) as the value that maximizes the Youden's index (Youden, 1950) or J statistic , which is defined as the distance between any point in the ROC curve and the $45°$ line.

Even though computing statistics based on the ROC curve and AUROC is customary in the EWS literature, the rare occurrence of financial crises may give rise to the so-called *accuracy paradox* (Valverde-Albacete and Peláez-Moreno, 2014). As it happens with the sample used for the empirical application, crises represent only 7% of the events; a percentage of cases correctly predicted of, say, 90% could easily mean that the model is able to forecast 90% of non-crisis events, but it has no ability to predict the occurrence of crises. Therefore, the value of the AUROC alone can mislead to an overly optimistic view of the model accuracy in presence of rare events. This is because a rather small threshold is needed in order to have a high rate of crisis events correctly predicted that will, paradoxically, correspond to a high rate of false alarms. For this reason, I also present with the empirical results the percentages of crises correctly predicted and false alarms based on the F-score (or F1-score), which is argued to be more suitable for the forecasting of rare events (Davis and Goadrich, 2006). Is is computed as twice the ratio between TPR$\times$PR and TPR$+$PR, where PR is called Precision and is equal to the ratio between the true positives and the sum of true positives and false negatives.

3 Simulation study

In this section I show the results of a simple simulation study aimed at assessing the performance of the PML estimator for the fixed-effects logit-based EWS, compared to the other models usually adopted in the applied early warning literature.

For $i = 1, \dots, n$ and $t = 1, \dots, T$, artificial data are drawn from a static logit model according to:

$$y_{it} = \mathbf{I}(c + \alpha_i + \beta x_{it} + \varepsilon_{it} > 0),$$

where $\mathbf{I}(\cdot)$ is an indicator function, c is a common intercept, $\beta = 1$ and $x_{it} \sim N(0, 1)$, α_i is drawn from a standard normal distribution, and the idiosyncratic error term follows a standard logistic distribution, that is $\varepsilon_{it} \sim \Lambda(0, \pi^2/3)$. The scenarios considered in this simulation study cover $n = 50, 100$ and $T = 10, 20$. I also control how rare the events $y_{it} = 1$ are by tuning c, which is set to $c = 0, -2, -4$, corresponding to a frequency of events in the sample of about 50%, 15%,and 3%, respectively. The simulations for each scenario are based on 1000 Monte Carlo replications.

Figure 1 depicts the boxplots of the bias of the estimator of β and the model AUROC [5] in the scenarios with $c = -2$ for the ML estimators of five different logit models: the pooled model (MLE Pooled), the random-effects model (MLE RE), the fixed-effects model (MLE FE), a fixed-effects model where all observations are retained but the intercepts for the countries never facing a crisis are dropped (MLE FE*), and the proposed PMLE of the fixed-effects model (PMLE FE). In order to have a firmer grasp of the values of the bias and AUROC, Table 1 reports their mean and standard deviations for the scenarios with $c = -2$. The results for the scenarios with $c = 0$ and $c = -4$ are presented in Appendix A in Figures A.1-A.2 and Tables A.1-A.2.

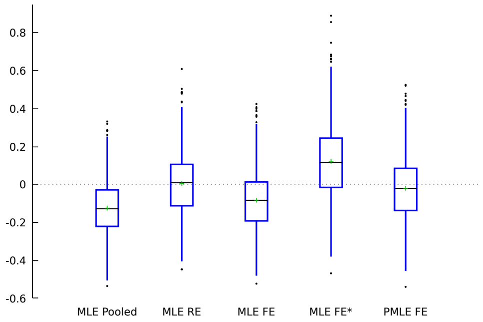
Because the simulation design includes some unobserved heterogeneity via α_i, the MLE Pooled is expected not to be consistent, which is confirmed by its boxplot not centered at zero in Panel (a) of Figure 1. On the contrary, the MLE RE produces a consistent estimator of β since the unobserved heterogeneity is generated as independent of the model covariate, which makes this model a benchmark for the bias distribution. It is worth noticing that the MLE FE is not consistent due to the incidental parameters problem as is MLE FE*, with the additional bias due to relevant omitted variables, , i.e. the intercepts for countries never facing a crisis. Finally, the proposed bias reduction technique seems to work quite nicely, in that the boxplot for the PMLE FE looks much like the one for the MLE RE.

Looking at the distribution of the AUROC for the five logit models in Panel (b), it is clear that retaining the whole sample while including country specific intercepts in the model specification greatly improves the model predictive performance. This emerges directly from the positions of the AUROC boxplot for the MLE FE* and PMLE FE that, at their average, almost reach 90%. By contrast, the AUROC of the ML estimators for the other three models only reach, about 72%. This is becasue the MLE Pooled does not account for any unobserved heterogeneity, theMLE RE only allows for a random intercept, and the MLE FE, even though includes fixed effects, is estimated using a limited number of subjects.

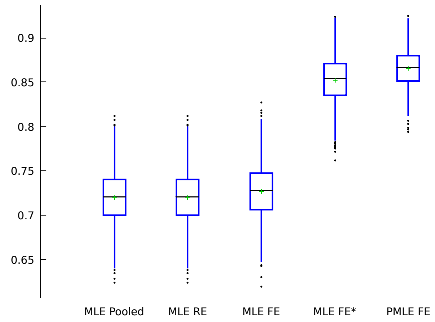
Other than the expected reduction of the variability in the distributions of the bias and AUROC, very similar patterns emerge for the scenarios with $n = 100$ and $T = 10, 20$ (see Panels c – h). All in all, the proposed PML estimator for the fixed-effects logit based on the whole sample has the best performance in terms of forecast against a negligible bias. This is still the case if we look at the results with $c = 0$ in Appendix (see Figure A.1 and Table A.1). Some bias for the PMLE FE shows up when the frequency of positive events is as low as about 3% ($c = -4$), but it quickly vanishes with $T = 20$ (see Figure A.2 and Table A.2).

[5]Even though the AUROC is subject to some limitation coming from the accuracy paradox discussed in Section 2.4, it still represents a viable tool to assess the models ability to predict the baseline at least (i.e., the non crisis events) and allows us to compare the performance of the different estimators in this respect.

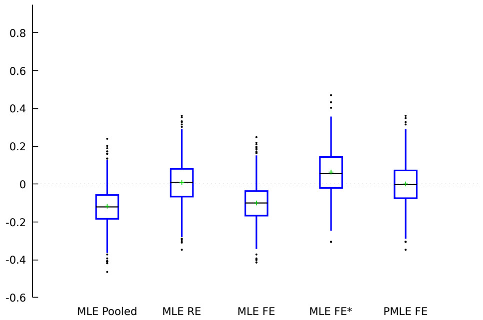
Figure 1: Simulation results: Bias and AUROC distributions, $c = -2$



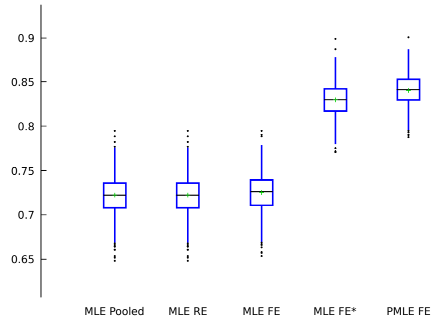
(a) Bias distribution, $n = 50, T = 10$



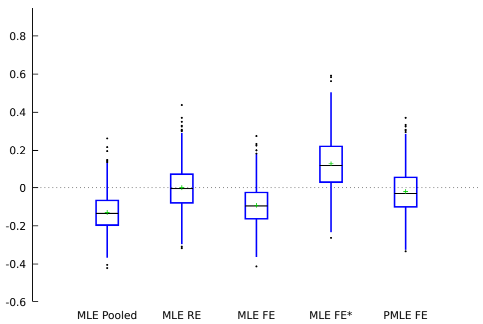
(b) AUROC distribution, $n = 50, T = 10$



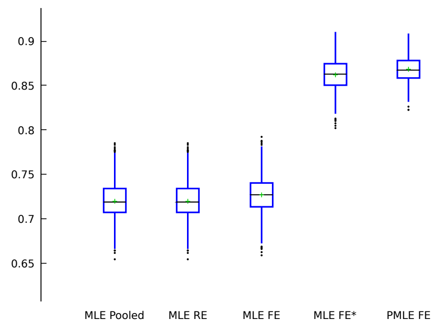
(c) Bias distribution, $n = 50, T = 20$



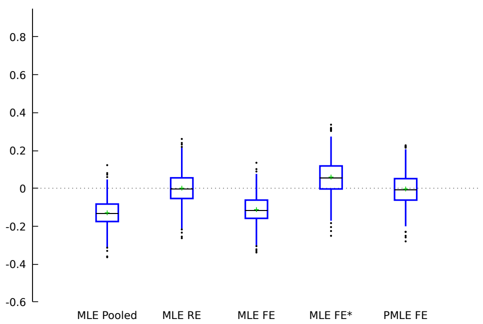
(d) AUROC distribution, $n = 50, T = 20$



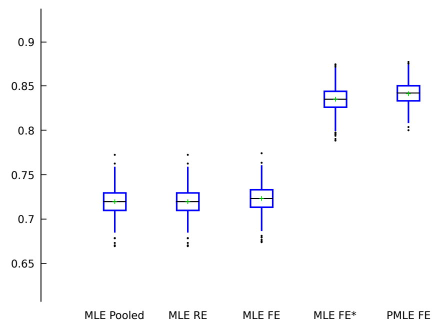
(e) Bias distribution, $n = 100, T = 10$



(f) AUROC distribution, $n = 100, T = 10$



(g) Bias distribution, $n = 100, T = 20$



(h) AUROC distribution, $n = 100, T = 20$

Table 1: Simulation results: Mean and standard deviation of Bias and AUROC, $c = -2$

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias	-0.124	0.007	-0.083	0.121	-0.018
SD bias	0.142	0.162	0.152	0.194	0.162
Mean AUROC	0.720	0.720	0.727	0.852	0.865
SD AUROC	0.030	0.030	0.031	0.026	0.021
$n = 50, T = 20$					
Mean bias	-0.118	0.011	-0.100	0.063	-0.001
SD bias	0.101	0.110	0.103	0.119	0.110
Mean AUROC	0.722	0.722	0.725	0.830	0.841
SD AUROC	0.021	0.021	0.021	0.019	0.017
$n = 100, T = 10$					
Mean bias	-0.128	0.002	-0.090	0.127	-0.018
SD bias	0.097	0.112	0.103	0.136	0.114
Mean AUROC	0.720	0.720	0.727	0.862	0.868
SD AUROC	0.021	0.021	0.021	0.017	0.014
$n = 100, T = 20$					
Mean bias	-0.130	0.002	-0.110	0.059	-0.004
SD bias	0.069	0.077	0.071	0.085	0.078
Mean AUROC	0.720	0.720	0.724	0.835	0.842
SD AUROC	0.015	0.015	0.015	0.014	0.012

4 Data

The performance of the proposed PML estimator of the fixed-effects logit-based EWS is also evaluated by means of a real data application. I consider an unbalanced panel dataset of 129 countries (see Appendix B for the complete list), covering the prediod from 1982 to 2017.

The dependent variable describes systemic banking crises and its definition is taken after Laeven and Valencia (2018), according to whom this event occurs if, in a given year, there are signs of financial distress in the banking system, such as bank runs, losses in the banking system, and bank liquidations, and there have been policy interventions as a consequences of significant losses in the banking system.[6] Based on this definition, I end up with a sample where 237 events of systemic banking crises are identified, 64 of which are new crises, and the average duration is 3 years. Table 2 reports the detailed list of crisis episodes identified by Laeven and Valencia (2018)

The list of explanatory variables is compiled following the EWS literature, especially the contributions by Demirgüç-Kunt and Detragiache (2005), Davis and Karim (2008), and Caggiano et al. (2016). Some descriptive statistics for the dependent and explanatory variables are reported in Table 3 along with their sources. Data on the latter are publicly available as International Financial Statistics (IFS), issued by the International Monetary Fund, or as World Development Indicators (WDI), from the World Bank. As the regression parameters in the logit model represent the early warnings, the associated covariates are lagged by one period.

According to the relevant stream of literature, early warnings are signaled by three main groups of

[6] The dataset is available at `https://www.imf.org/en/Publications/WP/Issues/2018/09/14/Systemic-Banking-Crises-Revisited-46232`

variables. The first one contains fundamental macroeconomic information, such as the real GDP growth rate, the logarithm of per capita GDP, inflation, and the real interest rate. All these variables, with the exception of the last one, are expected to negatively affect the probability of a systemic banking crisis. A second group is identified by monetary variables, specifically broad money (M2) over foreign exchange reserves and the growth rate of private credit. The former represents the (in)ability to face a negative shock to capital inflows and it is therefore expected to positively affect the probability of a crisis occurring. Similarly, the latter is supposed to exert a positive effect on the crisis probability, which can be expected to be increasing with over–indebtedness and deterioration of banks asset quality. The final group of covariates usually comprises financial information, which is here represented by the growth rate of net foreign assets to GDP and it is expected to negatively affect the likelihood of a systemic banking crisis.

It is worth to clarify that the dataset and the chosen set of explanatory variables do not represent a state–of–the–art basis for the investigation of early warnings in applied research, in that recent contributions have considered a richer set of covariates, different crisis definitions, models with a more sophisticated dynamic specification for the explanatory variables, and sometimes specific subsets of countries. Yet this work is intended to provide insights on the advantages of using a fixed-effects logit-based EWS, whose well–known drawbacks can be effectively overcome by the proposed methodology. To this aim, I believe that it may be sufficient to consider a sample and model specification that resemble those of policy oriented applications in terms of size and complexity, without accounting for specializations driven by specific operational choices.

Table 2: List of crises episodes

Algeria	1990-1994	Argentina	1982, 1991, 1995, 2001-2003
Bangladesh	1987	Belarus	1995
Bolivia	1994	Brazil	1996-1998
Bulgaria	1996-1997	Burundi	1994-1998
Cameroon	1990-1997	Central African Republic	1995-1997
Chad	1992-1996	Chile	1985
China		Colombia	1982, 1998-2000
Costa Rica	1987-1991, 1994	Croatia	1998-1999
Czech Republic	1996-2000	Dominican Republic	2003-2004
El Salvador	1989-1990	Ghana	1982-1983
Guinea	1993	Guinea Bisau	2014 - 2017
Guyana	1993	Haiti	1994-1998
Hungary	1995-2012	Iceland	2008-2012
India	1993	Indonesia	1997-2001
Jamaica	1996-1998	Japan	1997-2001
Jordan	1989-1991	Kenya	1985, 1992-1994
Korea	1997-1998	Kuwait	1982-1985
Kyrgyz Republic	1996-1999	Lebanon	1990-1993
Malaysia	1997-1999	Mauritania	1984
Mexico	1982-1985, 1994-1996	Moldova	2014-2017
Mongolia	2008-2009	Morocco	1982-1984
Nepal	1988	Nicaragua	1990-1993, 2000-2001
Nigeria	1991-1995, 2009-2012	Panama	1988-1989
Paraguay	1995	Peru	1983
Philippines	1983-1986, 1997-2001	Republic of Congo	1992-1994
Sierra Leone	1990-1994	Swaziland	1995-1999
Sweden	1992-1995	Switzerland	2008-2009
Tanzania	1988	Thailand	1983, 1997-2000
Turkey	2000-2001	Uganda	1994
Ukraine	1998-1999	Ukraine	1998-1999, 2008-2010, 2014-2017
United Kingdom	2007-2011	United States	1988, 2007-2011
Uruguay	1982-1985, 2002-2005	Venezuela	1994-1998
Vietnam	1997	Yemen	1996
Zambia	1995	Zimbabwe	1995-1999

Table 3: Variable sources and descriptive statistics

	Source	Mean	Median	SD	Min	Max
DEPENDENT VARIABLE						
Crisis	Laeven and Valencia (2018)	0.070	0.000	0.256	0	1
EXPLANATORY VARIABLES						
Real GDP growth(-1)	IFS	3.856	4.046	4.208	-3.185	9.961
Log per capita GDP(-1)	WDI	7.719	7.649	1.459	5.524	10.37
Inflation(-1)	WDI	29.64	5.873	381.6	-1.274	40.86
Real interest rate(-1)	IFS	2.315	2.047	11.25	-11.42	15.02
M2 to foreign exchange reserves(-1)	WDI	15.23	3.946	33.73	1.111	62.13
Growth of real domestic credit(-1)	WDI	15.86	12.51	33.92	-9.455	54.75
Growth of net foreign assets to GDP(-1)	WDI	0.000	0.013	0.386	-0.387	0.335

IFS: International Financial Statistics (International Monetary Fund). WDI: World Development Indicators (World Bank). Inflation is the growth rate of the GDP deflator.

5 Estimation results

Table 4 reports the results relative to the static logit model for the five ML estimators for the pooled, random- and fixed-effects models described in Section 3. The regression coefficients all have the expected sign and are statistically significant across all the models considered, with the exception of the real interest rate. Looking at the values of the McFadden R^2, it can be noticed that including fixed-effects increases the goodness of fit by a sizable amount, except for the MLE FE, because several countries do not contribute to the estimation sample (down to 67 from 129).

The models performance in terms of in–sample forecasts is assessed by means of an AUROC analysis. The results for the five models are presented in Table 5, where I report the AUROC along with its standard error and the percentages of correct crises and false alarms at the cut-offs computed on the basis of the Youden's index and F-score, as discussed in Section 2.4. The plots of ROC curves for the five models are reported in Figure C.1 in Appendix C. The best performance in terms of AUROC is provided by the proposed PMLE FE.

As concerns the percentages of crises correctly predicted and false alarms, looking at those based on the Youden's index, the proposed estimator for the fixed-effects logit provides 64.1% of crises correctly predicted against the lowest rate of false alarms among the models considered, that is 11.9%. As discussed in Section 2.4, a high percentage of crises correctly predicted alongside a rather large rate of false alarms can be interpreted as an accuracy paradox, which may arise with rare events. This seems to occur, for instance, in the case of the MLE FE*, where the model is apparently very performing (the AUROC is 0.87) in predicting non-crisis events. In fact, with an extremely low cut-off, 100% of the crises are predicted against 66% of false alarms. For this reason, I also report the percentages of crises correctly predicted and false alarms based on the F-score and it can be noticed that the cut-off values are always larger than those based on the Youden's index. As a result, the percentage of false alarms decreases for all models. Here it is worth noticing that the proposed PMLE FE provides the second best performance in terms of crises correctly predicted against a 9.4% of false alarms.

I also report the results of some out–of–sample forecast exercises in Table 6, only for the models

Table 4: Estimation results: logit models, crisis probability

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
# of observations	3365	3365	2407	3365	3365
# of countries	129	129	67	129	129
McFadden R^2	0.059	0.123	0.144	0.260	0.245
Log-likelihood	-806.898	-700.465	-614.657	-634.323	-647.251
COEFFICIENTS					
Real GDP growth(-1)	-0.120	-0.111	-0.104	-0.107	-0.095
	[-5.925]	[-4.180]	[-3.939]	[-4.093]	[-4.506]
Log per capita GDP(-1)	-0.024	-0.225	-0.482	-0.141	-0.462
	[-0.332]	[-1.560]	[-2.283]	[-0.446]	[-2.894]
Inflation(-1)	-0.039	-0.056	-0.056	-0.059	-0.037
	[-3.408]	[-2.989]	[-1.985]	[-2.425]	[-2.806]
Real interest rate(-1)	0.010	0.009	0.007	0.007	0.006
	[1.650]	[1.132]	[0.710]	[0.804]	[0.774]
M2 to foreign	0.005	0.012	0.015	0.019	0.013
exchange reserves(-1)	[2.029]	[2.650]	[2.913]	[3.548]	[3.455]
Growth of real	0.008	0.009	0.008	0.010	0.007
domestic credit(-1)	[2.534]	[2.317]	[1.652]	[2.037]	[1.829]
Growth of net foreign	-0.651	-0.784	-0.776	-0.882	-0.620
assets to GDP(-1)	[-3.578]	[-2.859]	[1.807]	[-2.481]	[-2.521]

MLE Pooled: ML estimator of the pooled logit model; MLE RE: random-effects logit model; MLE FE: ML estimator of the fixed-effects model; MLE FE*: ML estimator a fixed-effects logit model where all observations are retained and the intercepts for the countries never facing a crisis are dropped; PMLE FE: PMLE of the fixed-effects logit model. Panel robust t-tests are in square brackets. All models include an intercept term. The random effect logit model is based on the Gauss-Hermite quadrature method with 12 grid points, the estimated variance of the unobserved heterogeneity is $\hat{\sigma}_{\alpha}^2 = 1.525$.

Table 5: In-sample forecasts: logit models, crisis probability

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
AUROC	0.694	0.673	0.781	0.870	0.877
(st. err.)	(0.019)	(0.019)	(0.016)	(0.010)	(0.010)
Cut-off (Youden index)	0.050	0.105	0.127	0.002	0.171
% Correct crises	0.810	0.772	0.755	1.000	0.641
% False alarms	0.634	0.593	0.307	0.658	0.119
Cut-off (F-score)	0.103	0.160	0.173	0.226	0.188
% Correct crises	0.350	0.409	0.603	0.426	0.582
% False alarms	0.107	0.130	0.177	0.054	0.094

See notes to Table 4.

adopted more frequently in the EWS literature, namely MLE Pooled and MLE RE, and for the proposed PMLE. The method used to compute these forecasts is similar to the dynamic recursive stochastic technique adopted by Dawood et al. (2017). First the sample is restricted to the period 1982–2006 and used to estimate the model in sample. The estimated parameters are then used to predict the crisis probability for the year 2007. In the next step, the sample is updated with the 2007 cross-section and used to obtain the necessary estimates to compute the crisis probability for 2008, and so forth. The year 2013 is missing from the table because no crisis occurred, but it is included in the sample used to predict the out–of–sample crisis probabilities in 2014 and 2015.

From Table 6, it emerges that the proposed model does not always exhibit the best performance in terms of AUROC, compared to the MLE Pooled and MLE RE, and the results seem overall rather volatile, probably due to the very low incidence of crisis events in the samples, that ranges from 0.1% in 2007 and 2015 to 0.4% in 2008 and 2009. Yet the proposed PMLE is able to catch 1 of the 2 crises in 2007 and of the 8 in 2008, 6 of the 8 crisis episodes in 2009 and 4 out of 6 in 2010, 4 out of 5 in 2011, all 3 crises in 2012, and 1 in 3 crises in 2014 and 2015. This exercise also confirms that basing predictions on the F-score cut-off sensibly reduces the rate of false alarms.

The same set of results is presented for the dynamic logit model and the related estimated coefficients are reported in Table 7. It can be immediately noticed that the lag of the dependent variable has a strong explanatory power, as denoted by the t ratios and reflected by the values of the McFadden R^2, which now takes similar values across all the models considered. The associated coefficients are positive meaning that, as expected, the occurrence of a crisis in a given year significantly increases the likelihood of the crisis lasting the year after that. As for the rest of the explanatory variables, results are coherent across the models and similar to those obtained by estimating the static logit specifications, with the exception of loosing the GDP related variables as relevant predictors.

Table 8 reports the statistics related to the in–sample forecast exercises for the dynamic logit model. Here the analysis is split in two parts considering the crisis *entry* rate, i.e. the probability that in a country a crisis occurs at time t given that the country did not face a crisis at time $t-1$ as per expression (8) (also called crisis onset in the table), and the crisis *persistence* rate, that is the probability that a country still faces a crisis at time t given that the crisis already occurred at time $t-1$ as per expression (9) (also denoted as crisis return). The table reports the AUROC for both measures, followed by the percentages of crises (onsets or returns) correctly predicted and false alarms based on both the Youden's index and F-score. The plots of the related ROC curves are reported in Figures C.2 and C.3 in Appendix C.

In terms of AUROC, the proposed PMLE FE shows the best and second to best performance in forecasting crisis onsets and returns, respectively. The percentage of crises correctly predicted is much higher when considering the persistence rate than the entry rate. This can be expected since crisis outbreaks even rarer than a general crisis event. It is also confirmed that basing the cut-off on the F-score greatly reduces the rate of predicted false alarms. With the events of crisis onsets and returns being even more sporadic than crises themselves, out–of–sample predictions seem to be very unreliable (see the results in Tables 9 and 10). Even so, the results of these exercises seem to show the same patterns of those for the static logit model.

Overall, these results confirm the insights that emerged from the simulation study: the proposed PML estimator and, in general, including fixed effects in the model specification while retaining the whole

Table 6: Out-of-sample forecasts: logit models, crisis probability

	2007	2008	2009	2010	2011	2012	2014	2015
# of observations	1826	1941	2057	2171	2285	2397	2624	2737
# of crises	2	7	8	6	5	3	3	3
MLE Pooled								
AUROC	0.907	0.501	0.629	0.810	0.721	0.524	0.639	0.763
(st. err.)	(0.034)	(0.139)	(0.115)	(0.152)	(0.181)	(0.106)	(0.302)	(0.237)
Youden index cut-off								
% Correct crises	1.000	0.429	0.750	0.833	0.200	1.000	0.667	0.667
% False alarms	0.531	0.395	0.708	0.472	0.066	0.863	0.964	0.542
F-score cut-off								
% Correct crises	1.000	0.286	0.250	0.500	0.600	1.000	0.667	0.667
% False alarms	0.115	0.092	0.057	0.094	0.093	0.600	0.082	0.000
MLE RE								
AUROC	0.920	0.515	0.667	0.793	0.729	0.558	0.633	0.738
(st. err.)	(0.030)	(0.136)	(0.114)	(0.149)	(0.183)	(0.108)	(0.300)	(0.226)
Youden index cut-off								
% Correct crises	1.000	0.286	0.750	1.000	0.800	1.000	0.667	0.667
% False alarms	0.177	0.183	0.717	0.972	0.785	0.773	0.345	0.710
F-score cut-off								
% Correct crises	1.000	0.286	0.750	0.500	0.600	1.000	0.333	0.333
% False alarms	0.097	0.092	0.283	0.009	0.103	0.600	0.027	0.000
PMLE FE								
AUROC	0.646	0.469	0.759	0.948	0.961	0.967	0.576	0.850
(st. err.)	(0.302)	(0.115)	(0.072)	(0.026)	(0.020)	(0.017)	(0.241)	(0.133)
Youden index cut-off								
% Correct crises	0.500	1.000	0.750	1.000	1.000	1.000	0.667	0.667
% False alarms	0.071	0.872	0.311	0.278	0.636	0.154	0.809	0.045
F-score cut-off								
% Correct crises	0.500	0.143	0.750	0.667	0.800	1.000	0.333	0.333
% False alarms	0.053	0.000	0.264	0.037	0.047	0.045	0.073	0.000

See notes to Table 4.

Table 7: Estimation results: dynamic logit models, crisis probability

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
# of observations	3236	3236	2407	3236	3236
# of countries	129	129	67	129	129
McFadden R^2	0.383	0.427	0.426	0.480	0.473
Log-likelihood	-529.308	-457.471	-385.871	-421.235	-420.828
COEFFICIENTS					
Crisis(-1)	4.053	4.301	3.531	3.565	3.186
	[22.37]	[23.23]	[18.69]	[17.48]	[20.48]
Real GDP growth(-1)	-0.050	-0.037	-0.022	-0.025	-0.021
	[-2.719]	[-1.620]	[-0.793]	[-1.165]	[-1.061]
Log per capita GDP(-1)	0.040	-0.002	-0.129	-0.029	-0.134
	[0.671]	[-0.040]	[-0.617]	[-0.210]	[-1.019]
Inflation(-1)	-0.033	-0.030	-0.082	-0.058	-0.036
	[-1.252]	[-1.380]	[-1.377]	[-1.443]	[-1.940]
Real interest rate(-1)	0.011	0.009	0.015	0.014	0.016
	[1.732]	[1.650]	[1.893]	[1.724]	[2.221]
M2 to foreign	0.005	0.006	0.025	0.017	0.021
exchange reserves(-1)	[1.952]	[2.660]	[3.325]	[4.689]	[4.597]
Growth of real	0.009	0.009	0.013	0.011	0.009
domestic credit(-1)	[2.761]	[2.350]	[2.112]	[2.633]	[2.859]
Growth of net foreign	-0.620	-0.644	-1.164	-0.893	-0.814
assets to GDP(-1)	[-2.590]	[-2.410]	[-2.224]	[-2.991]	[-3.354]

See notes to Table 4. Panel robust t-tests are in square brackets. All models include an intercept term. The random effect logit model is based on the Gauss-Hermite quadrature method with 12 grid points, the estimated variance of the unobserved heterogeneity is $\hat{\sigma}_\alpha^2 = 0.001$.

Table 8: In-sample forecasts: dynamic logit models, crisis entry and persistence rates

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
Entry rate					
AUROC	0.609	0.641	0.672	0.766	0.783
(st. err.)	(0.036)	(0.036)	(0.035)	(0.025)	(0.026)
Cut-off (Youden index)	0.017	0.248	0.042	0.021	0.098
% Correct onsets	0.984	0.016	0.609	0.781	0.266
% False alarms	0.962	0.000	0.382	0.432	0.068
Cut-off (F-score)	0.048	0.042	0.100	0.077	0.087
% Correct onsets	0.125	0.141	0.219	0.281	0.375
% False alarms	0.037	0.045	0.065	0.064	0.087
Persistence rate					
AUROC	0.720	0.714	0.766	0.869	0.850
(st. err.)	(0.020)	(0.020)	(0.018)	(0.011)	(0.013)
Cut-off (Youden index)	0.475	0.587	0.578	0.088	0.464
% Correct crisis returns	0.994	0.821	0.827	1.000	0.931
% False alarms	0.976	0.575	0.395	0.785	0.416
Cut-off (F-score)	0.677	0.660	0.668	0.693	0.646
% Correct crisis returns	0.393	0.439	0.618	0.543	0.613
% False alarms	0.108	0.134	0.209	0.088	0.122

See notes to Table 7.

Table 9: Out-of-sample forecasts: logit models, crisis entry rate

	2007	2008	2009	2014
# of observations	1826	1940	2055	2618
# of onsets	2	5	1	2
MLE Pooled				
AUROC	0.938	0.532	0.647	0.633
(st. err.)	(0.024)	(0.130)	(0.107)	(0.299)
Youden index cut-off				
% Correct onsets	1.000	0.200	1.000	1.000
% False alarms	0.619	0.486	0.549	0.982
F-score cut-off				
% Correct onsets	1.000	1.000	1.000	0.500
% False alarms	0.071	0.829	0.504	0.036
MLE RE				
AUROC	0.929	0.460	0.546	0.633
(st. err.)	(0.021)	(0.132)	(0.111)	(0.295)
Youden index cut-off				
% Correct onsets	0.000	0.000	0.000	0.000
% False alarms	0.000	0.000	0.000	0.000
F-score cut-off				
% Correct onsets	1.000	1.000	1.000	0.500
% False alarms	0.071	0.836	0.241	0.027
PMLE FE				
AUROC	0.633	0.450	0.733	0.703
(st. err.)	(0.315)	(0.117)	(0.079)	(0.200)
Youden index cut-off				
% Correct onsets	0.500	0.000	0.000	1.000
% False alarms	0.133	0.171	0.062	0.838
F-score cut-off				
% Correct onsets	0.500	1.000	1.000	0.500
% False alarms	0.053	0.937	0.336	0.198

See notes to Table 7.

Table 10: Out-of-sample forecasts: logit models, crisis persistence rate

	2008	2009	2010	2011	2012	2014	2015
# of observations	1940	2055	2168	2281	2392	2618	2730
# of crisis returns	2	7	6	5	3	1	3
MLE Pooled							
AUROC	0.929	0.667	0.755	0.710	0.482	0.893	0.729
(st. err.)	(0.025)	(0.121)	(0.127)	(0.172)	(0.129)	(0.029)	(0.262)
Youden index cut-off							
% Correct crisis returns	1.000	0.571	0.833	0.800	0.667	1.000	0.667
% False alarms	0.175	0.206	0.287	0.542	0.636	0.982	0.748
F-score cut-off							
% Correct crisis returns	1.000	0.571	0.333	0.400	1.000	1.000	0.667
% False alarms	0.079	0.206	0.046	0.075	0.700	0.107	0.019
MLE RE							
AUROC	0.921	0.508	0.685	0.657	0.529	0.874	0.679
(st. err.)	(0.023)	(0.123)	(0.116)	(0.209)	(0.110)	(0.028)	(0.289)
Youden index cut-off							
% Correct crisis returns	1.000	0.000	1.000	0.000	0.000	1.000	0.667
% False alarms	0.351	0.000	0.833	0.056	0.055	0.252	0.708
F-score cut-off							
% Correct crisis returns	1.000	0.667	0.400	0.750	1.000	1.000	0.333
% False alarms	0.079	0.411	0.139	0.196	0.624	0.126	0.000
PMLE FE							
AUROC	0.785	0.740	0.907	0.907	0.492	0.982	0.959
(st. err.)	(0.199)	(0.090)	(0.070)	(0.070)	(0.124)	(0.028)	(0.031)
Youden index cut-off							
% Correct crisis returns	0.500	0.714	0.833	0.800	1.000	1.000	1.000
% False alarms	0.018	0.308	0.426	0.168	0.491	0.393	0.495
F-score cut-off							
% Correct crisis returns	0.500	0.429	0.667	0.600	1.000	1.000	0.333
% False alarms	0.018	0.112	0.018	0.019	0.000	0.019	0.000

See notes to Table 7.

sample (that is also MLE FE*), strongly improves the model predictive performance with respect to not accounting for unobserved heterogeneity or treating it as an independent random variables, as is done in the MLE Pooled and MLE RE models, respectively. The performance of the MLE FE falls somewhat in between pooled and random-effects models on one side and the other fixed-effects models on the other. It is worth recalling, though, that the MLE FE, as well as the MLE FE*, are both biased due to the incidental parameters problem, which can mislead the identification of the relevant early warnings, and do not allow for out–of–sample forecasts for those countries that never face a crisis.

6 Concluding remarks

The applied literature has made large use of logit–based EWS to forecast financial crises. To this aim, the model predictive performance is crucial for forecast accuracy and yet, despite the availability of panel data, country–specific effects are always neglected so as to avoid a substantial sample reduction. This limitation can be overcome by the approach proposed in this paper, which allows to include fixed-effects using the whole sample of countries when estimating logit–based EWS. Country–specific effects largely improve the model forecasting performance and retaining the entire sample enables the possibility of computing out–of–sample predictions of crisis events in countries that never had any before.

Both the simulation study and empirical application show that the proposed PML estimator of the fixed-effects EWS outperforms the pooled, random-effects and standard fixed-effects counterparts. Yet these exercise show that including fixed effects only for those countries that faced at least one crisis and retaining the whole sample gives a comparable performance. This strategy may represent an appealing alternative to the practitioner who can, this way, avoid using non–standard ML software routines and just select the relevant country dummies to include in the model specification beforehand. The practitioner, however, must be warned that proceeding this way does not allow to compute forecasts of first–ever crises and the incidental parameters problem, along with omitting country specific intercepts, can misguide the identification of the relevant early warnings.

There are also two secondary results provided in the paper. First, and as already shown by recent contributions, the inclusion of the lagged dependent variable largely improves the predictive performance of logit–based EWS. I add that the use of a dynamic binary choice model in this context allows for the prediction not only of the probability of a crisis event, but also of entry and persistence rates, which can be of interest to policymakers. Secondly, as crises are rare events, assessing forecasting accuracy with standard analyses such as the AUROC can lead to overstate the model ability to correctly predict crises, against a high rate of false alarms. In such cases, alternative measures can be adopted to select the cut-offs used to compute these statistics. As an example, I show that thresholds selected on the basis of the F-score help reducing the rate of false alarms, but give rise to a lower true positive rate. Given the relevance of EWS in operational research, I believe that further investigation in this direction is warranted.

References

Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71:1–10.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B*, 32:283–301.

Antunes, A., Bonfim, D., Monteiro, N., and Rodrigues, P. M. (2018). Forecasting banking crises with dynamic panel probit models. *International Journal of Forecasting*, 34:249–275.

Bassanetti, A., Cottarelli, C., and Presbitero, A. F. (2018). Lost and found: Market access and public debt dynamics. *Oxford Economic Papers*, 71:445–471.

Caggiano, G., Calice, P., Leonida, L., and Kapetanios, G. (2016). Comparing logit-based early warning systems: Does the duration of systemic banking crises matter? *Journal of Empirical Finance*, 37:104–116.

Cappellari, L. and Jenkins, S. P. (2002). Who stays poor? Who becomes poor? Evidence from the British household panel survey. *The Economic Journal*, 112:60–67.

Cappellari, L. and Jenkins, S. P. (2004). Modelling low income transitions. *Journal of Applied Econometrics*, 19:593–610.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47:225–238.

Chamberlain, G. (1982). The general equivalence of Granger and Sims causality. *Econometrica*, 50:569–581.

Cook, S. J., Hays, J. C., and Franzese, R. J. (2018). Fixed effects in rare events data: A penalized maximum likelihood solution. *Political Science Research and Methods*, 2018:1–14.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30:248–265.

Davis, E. P. and Karim, D. (2008). Comparing early warning systems for banking crises. *Journal of Financial stability*, 4:89–120.

Davis, J. and Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on machine learning*, pages 233–240. ACM.

Dawood, M., Horsewood, N., and Strobel, F. (2017). Predicting sovereign debt crises: An early warning system approach. *Journal of Financial Stability*, 28:16–28.

Demirgüç-Kunt, A. and Detragiache, E. (1998). The determinants of banking crises in developing and developed countries. *IMF Staff Papers*, 45:81–109.

Demirgüç-Kunt, A. and Detragiache, E. (2005). Cross-country empirical studies of systemic bank distress: A survey. *IMF Working Paper: Cross-Country Empirical Studies of Systemic Bank Distress-A Survey*, 5(96):1–33.

Eberhardt, M. and Presbitero, A. (2018). Commodity price movements and banking crises. IMF Working Paper 18/153, International Monetary Fund.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150:71–85.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80:27–38.

Ghulam, Y. and Derber, J. (2018). Determinants of sovereign defaults. *The Quarterly Review of Economics and Finance*, 69:43–55.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the Area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, 143:29–36.

Heckman, J. J. (1981). Heterogeneity and state dependence. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press: Cambridge MA.

Heinze, G. (2006). A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in Medicine*, 25:4216–4226.

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21:2409–2419.

Hsiao, C. (2015). *Analysis of Panel Data*. Cambridge University Press, New York, 3rd edition.

Hsieh, F., Turnbull, B. W., et al. (1996). Nonparametric and semiparametric estimation of the Receiver Operating Characteristic curve. *The Annals of Statistics*, 24:25–40.

Kosmidis, I. and Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika*, 96:793–804.

Kunz, J. S., Staub, K. E., and Winkelmann, R. (2017). Estimating fixed effects: Perfect prediction and bias in binary response panel models, with an application to the hospital readmissions reduction program. IZA Discussion Paper 11182, IZA.

Laeven, L. and Valencia, F. (2018). Systemic banking crises revisited. IMF Working Paper 18/206, International Monetary Fund.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95:391–413.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32.

Schularick, M. and Taylor, A. M. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review*, 102:1029–61.

Valverde-Albacete, F. J. and Peláez-Moreno, C. (2014). 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PloS one*, 9:e84217.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3:32–35.

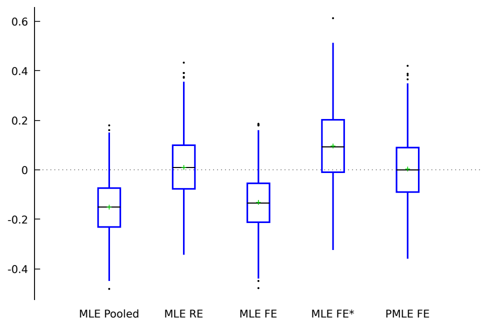
Table A.1: Simulation results: Mean and standard deviation of Bias and AUROC, $c = 0$

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias	-0.150	0.011	-0.130	0.098	0.003
SD bias	0.113	0.128	0.114	0.148	0.128
Mean AUROC	0.710	0.710	0.714	0.825	0.840
SD AUROC	0.024	0.024	0.024	0.019	0.019
$n = 50, T = 20$					
Mean bias	-0.152	0.010	-0.148	0.057	0.003
SD bias	0.081	0.089	0.081	0.095	0.088
Mean AUROC	0.710	0.710	0.711	0.818	0.821
SD AUROC	0.017	0.017	0.016	0.015	0.015
$n = 100, T = 10$					
Mean bias	-0.159	0.002	-0.139	0.088	-0.000
SD bias	0.078	0.088	0.079	0.103	0.090
Mean AUROC	0.709	0.709	0.713	0.825	0.842
SD AUROC	0.016	0.016	0.017	0.013	0.013
$n = 100, T = 20$					
Mean bias	-0.158	0.005	-0.153	0.056	0.003
SD bias	0.058	0.064	0.058	0.069	0.064
Mean AUROC	0.709	0.709	0.710	0.820	0.822
SD AUROC	0.012	0.012	0.012	0.011	0.011

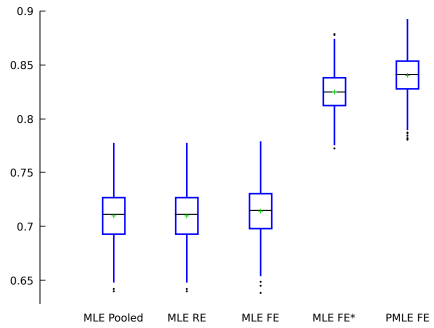
Appendix

A Additional simulation results

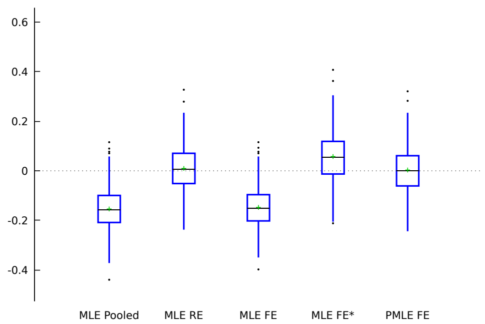
Figure A.1: Simulation results: Bias and AUROC distributions, $c = 0$



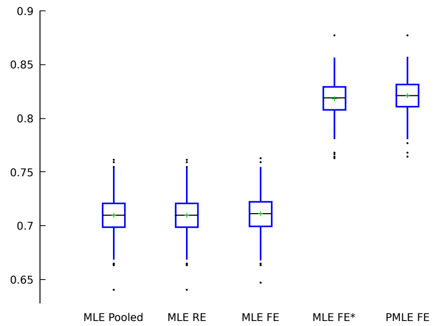
(a) Bias distribution, $n = 50, T = 10$



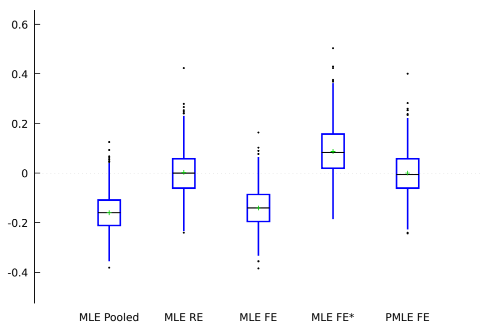
(b) AUROC distribution, $n = 50, T = 10$



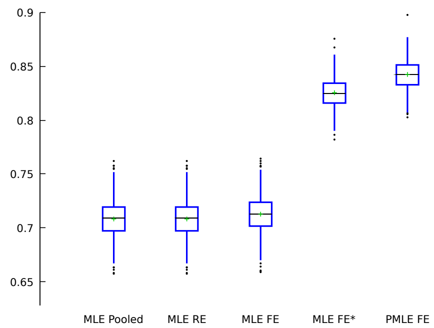
(c) Bias distribution, $n = 50, T = 20$



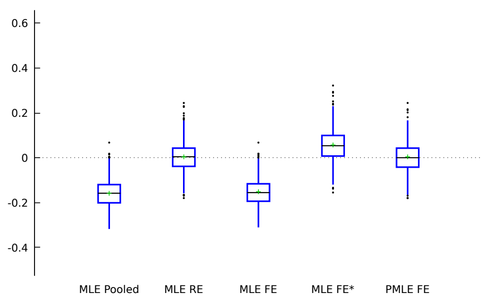
(d) AUROC distribution, $n = 50, T = 20$



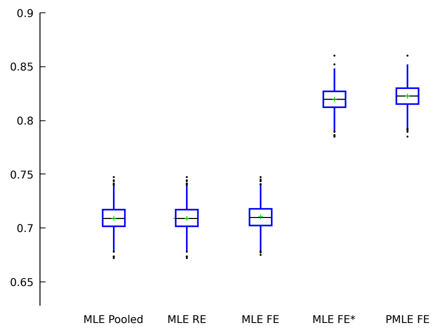
(e) Bias distribution, $n = 100, T = 10$



(f) AUROC distribution, $n = 100, T = 10$

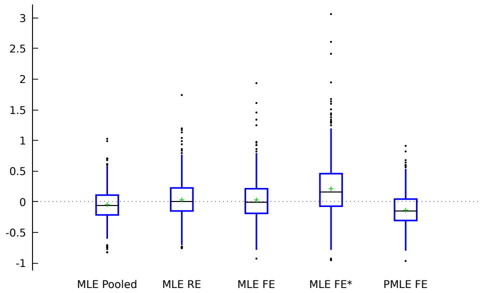


(g) Bias distribution, $n = 100, T = 20$

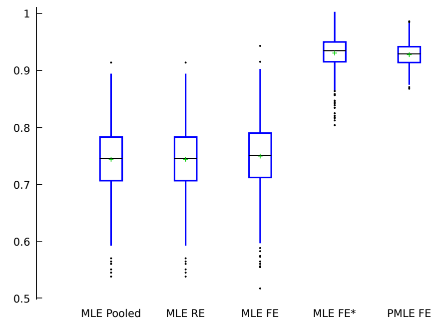


(h) AUROC distribution, $n = 100, T = 20$

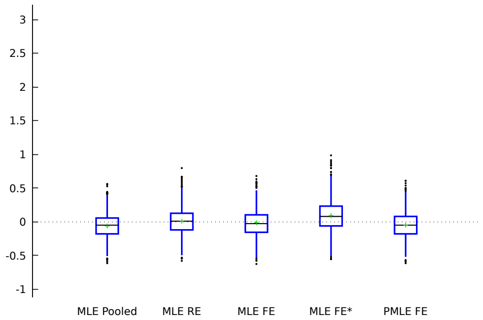
Figure A.2: Simulation results: Bias and AUROC distributions, $c = -4$



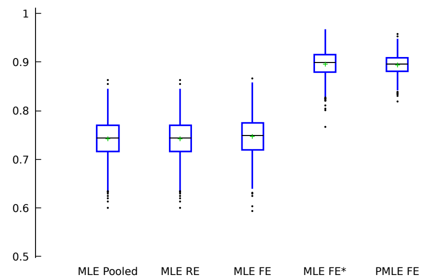
(a) Bias distribution, $n = 50, T = 10$



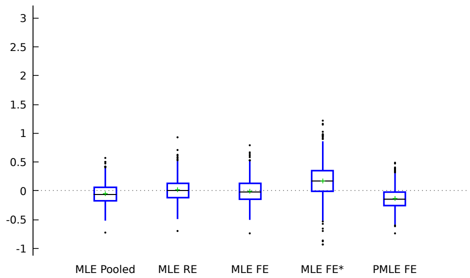
(b) AUROC distribution, $n = 50, T = 10$



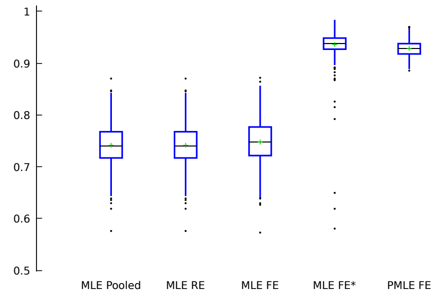
(c) Bias distribution, $n = 50, T = 20$



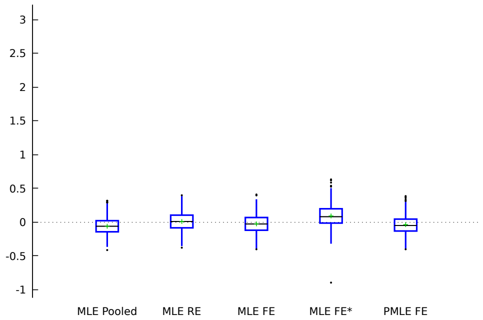
(d) AUROC distribution, $n = 50, T = 20$



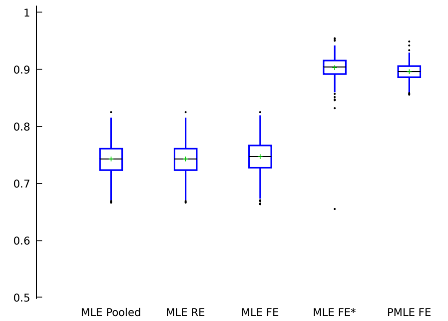
(e) Bias distribution, $n = 100, T = 10$



(f) AUROC distribution, $n = 100, T = 10$



(g) Bias distribution, $n = 100, T = 20$



(h) AUROC distribution, $n = 100, T = 20$

Table A.2: Simulation results: Mean and standard deviation of Bias and AUROC, $c = -4$

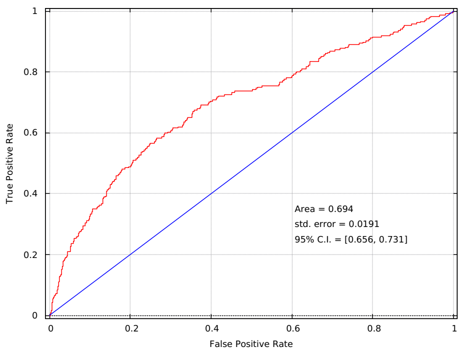
	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias	-0.047	0.035	0.027	0.212	-0.133
SD bias	0.244	0.285	0.313	0.417	0.254
Mean AUROC	0.744	0.744	0.750	0.931	0.929
SD AUROC	0.055	0.055	0.059	0.027	0.020
$n = 50, T = 20$					
Mean bias	-0.058	0.009	-0.022	0.090	-0.049
SD bias	0.182	0.199	0.203	0.233	0.190
Mean AUROC	0.743	0.743	0.747	0.896	0.895
SD AUROC	0.040	0.040	0.042	0.027	0.021
$n = 100, T = 10$					
Mean bias	-0.056	0.013	-0.007	0.177	-0.135
SD bias	0.177	0.198	0.206	0.279	0.178
Mean AUROC	0.742	0.742	0.748	0.936	0.928
SD AUROC	0.039	0.039	0.042	0.025	0.014
$n = 100, T = 20$					
Mean bias	-0.061	0.005	-0.025	0.088	-0.044
SD bias	0.124	0.135	0.135	0.162	0.131
Mean AUROC	0.742	0.742	0.747	0.904	0.896
SD AUROC	0.028	0.028	0.028	0.019	0.014

B List of countries

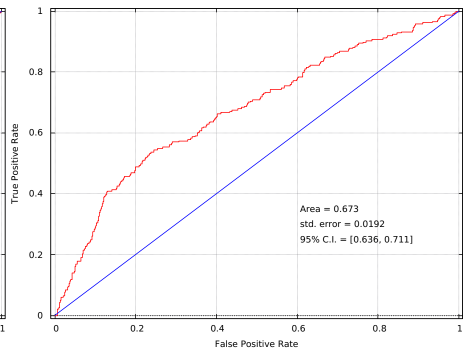
Albania, Algeria, Angola, Argentina, Armenia, Australia, Azerbaijan, Bangladesh, Barbados, Belarus, Belize, Benin, Bhutan, Bolivia, Bosnia and Herzegovina, Brazil, Brunei, Darussalam, Bulgaria, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Canada, Central African Republic, Chad, Chile, China, Colombia, Comoros, Costa Rica, Croatia, Czech Republic, Côte d'Ivoire, Democratic Republic of the Cong, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Macedonia, Fiji, Gabon, Georgia, Ghana, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong, SAR, Hungary, Iceland, India, Indonesia, Israel, Jamaica, Japan, Jordan, Kenya, Korea, Kuwait, Kyrgyz Republic, Laos, Lebanon, Lesotho, Liberia, Libya, Madagascar, Malaysia, Maldives, Mali, Mauritania, Mauritius, Mexico, Moldova, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nepal, New Zealand, Nicaragua, Niger, Nigeria, Pakistan, Panama, Papua, New, Guinea, Paraguay, Peru, Philippines, Poland, Republic of Congo, Rwanda, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, South Africa, Sri Lanka, St. Kitts and Nevis, Suriname, Swaziland, Sweden, Switzerland, São Tomé and Príncipe, Tajikistan, Tanzania, Thailand, Gambia, Togo, Trinidad and Tobago, Turkey, Uganda, Ukraine, United Kingdom, United States, Uruguay, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

C ROC curves

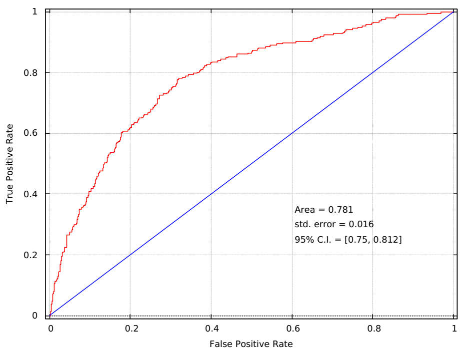
Figure C.1: ROC curves: logit models for crisis probability



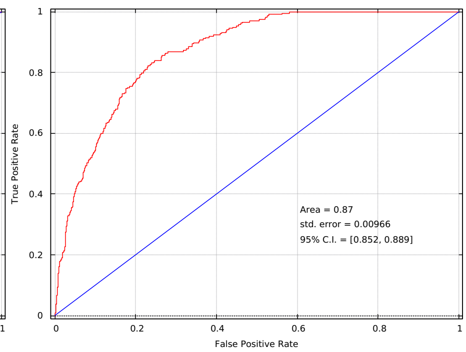
(a) MLE Pooled



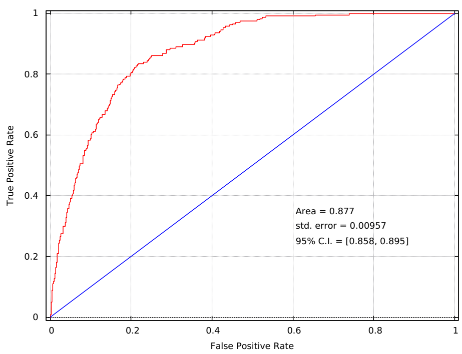
(b) MLE RE



(c) MLE FE

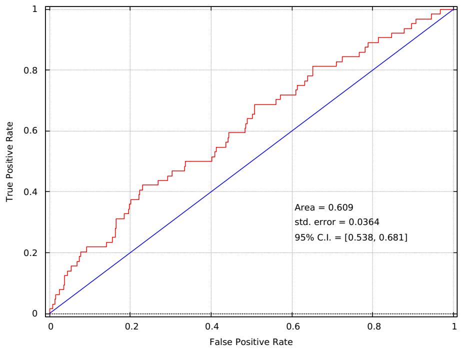


(d) MLE FE*

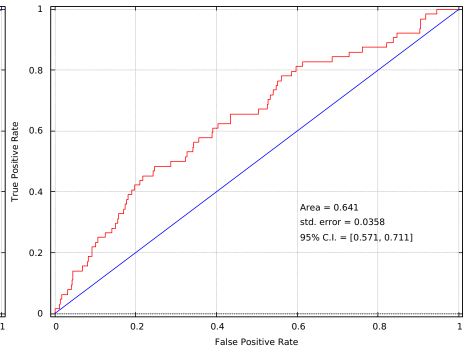


(e) PMLE FE

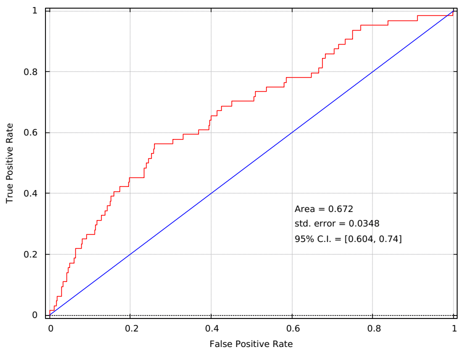
Figure C.2: ROC curves: dynamic logit models, entry rate



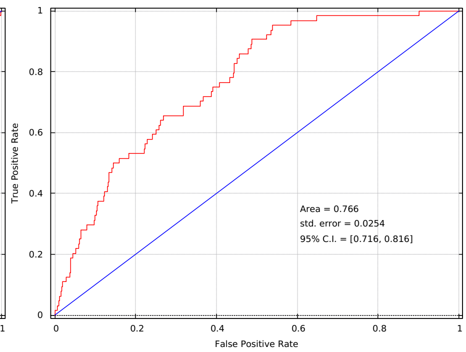
(a) MLE Pooled



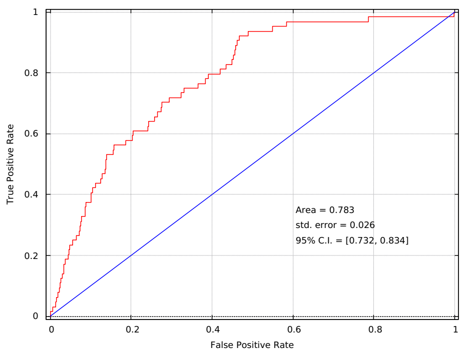
(b) MLE RE



(c) MLE FE

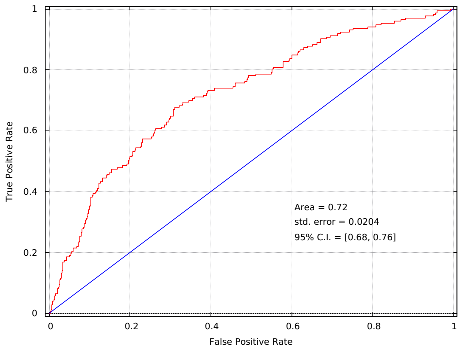


(d) MLE FE*

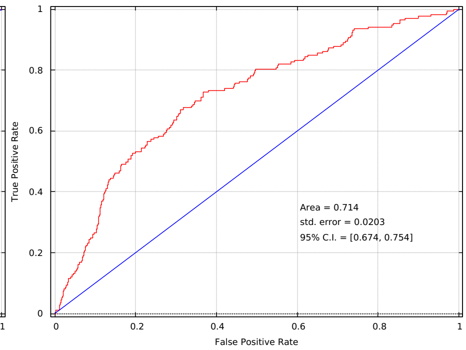


(e) PMLE FE

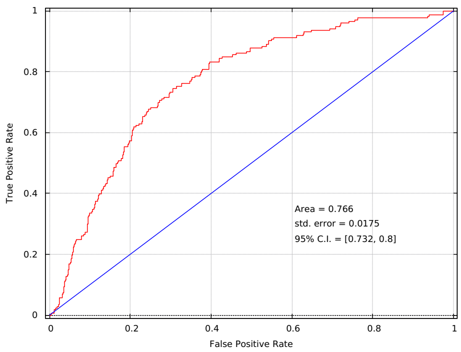
Figure C.3: ROC curves: dynamic logit models, persistence rate



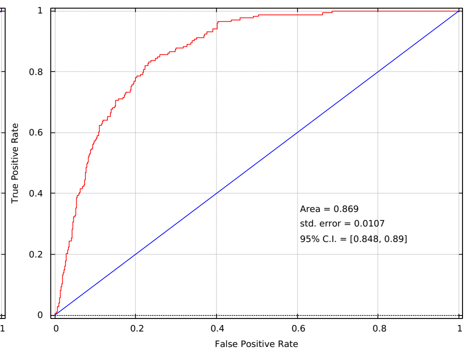
(a) MLE Pooled



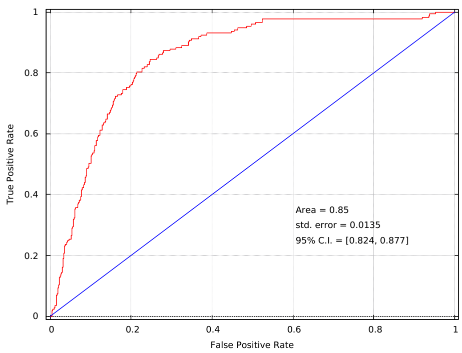
(b) MLE RE



(c) MLE FE



(d) MLE FE*



(e) PMLE FE