Penalized maximum likelihood estimation of logit-based early warning systems

Penalized maximum likelihood estimation of logit-based early warning systems

Claudia Pigini[*]
Università Politecnica delle Marche (IT) and MoFiR

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, improves the predictive performance of EWS, both in simulation and out of sample, with respect to the pooled, random-effects and standard fixed-effects models.

KEYWORDS: BANKING CRISIS, BIAS REDUCTION, FIXED-EFFECTS LOGIT, PRECISION-RECALL, RARE EVENTS, SEPARATED DATA

JEL CLASSIFICATION: C23, C25, G17, G21

[*]*Corresponding Author.* Address: P.le Martelli 8, 60121, Ancona. Tel: +39 071 2207103. Email: c.pigini@univpm.it

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 shown 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–specific 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 faces 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 in 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 avoid 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 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

[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.

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).

I consider a dynamic logit model 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). In fact, a dynamic binary choice model in this context allows for the prediction of different crisis probabilities for a country at time t, according to whether the country was not in a crisis state at time $t-1$, that is the crisis *entry* rate, or it is facing a prolonged state of financial distress since a crisis already occurred in $t-1$, that is the crisis *persistence* rate. For this reason, and differently from Kunz et al. (2017) and Cook et al. (2018), I combine Firth's approach to overcome the complete separation problem with the analytical bias correction by Fernández-Val (2009), specifically derived for dynamic binary choice panel data models with the aim of reducing the bias due to the incidental parameters problem.

The performance of the PML estimator is first evaluated by means of a simulation study and then by its application to logit-based EWS estimated using an unbalanced panel dataset, which consists of 129 countries from 1983 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 compare the proposed approach with the dynamic pooled probit model proposed by Antunes et al. (2018) and with the pooled multinomial logit by Caggiano et al. (2016); these alternative approaches are outperformed both in simulation and in the empirical application.

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. However, with rare events the value of the AUROC alone can mislead to an overly optimistic view of the model accuracy: a high rate of cases correctly predicted may be given by the model ability to forecast non-crisis events and a small probability cut-off is needed to correctly predict crisis events, proucing in turn a relatively large number of false alarms. For this reason, I also consider predictions based on the Precision-Recall (PR) analysis, which it is argued to be more suitable for the forecasting of rare events (Davis and Goadrich, 2006), as it largely reduces the number of false alarms. The adoption of the PR analysis represents an additional contribution with respect to the related EWS literature that only considers forecasts based on the ROC.

The rest of the paper is organized as follows: Section 2 introduces the problem of separation with fixed-effects logit models, illustrates the dynamic logit model and its PML estimator, and 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 describe the binary panel logit model and the separation problem that originates with rare events and fixed effects; secondly, I recall the dynamic logit model; I then describe the PML approach derived using Firth's bias reduction technique and Fernández-Val's analytical bias correction; finally, I briefly review the ROC analysis for classification problems, illustrate the so-called "accuracy paradox", and describe the PR curve as an alternative tool in forecasting rare events.

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 $t = 1, \dots, T$.[3] 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 once), 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 $\alpha_i, i = 1, \dots, n$, 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

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

[3] In order to keep a clear notation, in the following the case of balanced panel data is considered. The extension to unbalanced panel datasets, such as the one used here for the empirical illustration, is straightforward.

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 represents a rare event, such as a financial crisis, 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 an inconsistent ML estimator of the model parameters if the unobserved heterogeneity specific to these countries were correlated with the model covariates, because of the omitted variable bias.

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, when correctly specified, 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 a function of the entire history of covariates, meaning that the conditional probability of $y_{it} = 1$ includes some function of $\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}$ in the set of covariates (see Mundlak, 1978; Chamberlain, 1980; Wooldridge, 2005). While this strategy can be effective in controlling for the unobserved heterogeneity when the goal is to identify the causal effects of predictors, it rules out the possibility of computing out-of-sample forecasts.

2.2 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

$$p(y_{it}|\boldsymbol{x}_{it}, y_{i,t-1}; \delta_i, \boldsymbol{\phi}, \gamma) = \frac{\exp\left[y_{it}\left(\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + y_{it-1}\gamma\right)\right]}{1 + \exp\left(\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + y_{it-1}\gamma\right)}, \tag{4}$$

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 dynamic logit model allows to predict the probability of facing a crisis at time t differently according to the status of the country at $t-1$. The probability that a crisis at time t occurs in country i given that the country was not in distress at time $t-1$, can be written as

$$\Pr(y_{it} = 1|\boldsymbol{x}_{it}, y_{i,t-1} = 0) = \frac{\exp\left(\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi}\right)}{1 + \exp\left(\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi}\right)}, \tag{5}$$

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

$$\Pr(y_{it}=1|\boldsymbol{x}_{it},y_{it-1}=1)=\frac{\exp\left(\delta_i+\boldsymbol{x}_{it}'\boldsymbol{\phi}+\gamma\right)}{1+\exp\left(\delta_i+\boldsymbol{x}_{it}'\boldsymbol{\phi}+\gamma\right)}, \tag{6}$$

and can be referred to as the *persistence* rate. From these definitions it becomes clear that the dynamic formulation introduces important heterogeneity in the computation of the crisis probability through the parameter γ, which can substantially improve the model explanatory power and predictive performance. The probability of a crisis can therefore be computed as the entry or persistence rate according to the value of the response variable in $t-1$. [4]

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.

2.3 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 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

$$B(\hat{\boldsymbol{\theta}})=\frac{B_1(\hat{\boldsymbol{\theta}})}{h}+\frac{B_2(\hat{\boldsymbol{\theta}})}{h^2}+\dots \tag{7}$$

Bias reduction techniques usually aim at estimating the leading term (or an approximation of it) in (7), 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

[4] Trivially $\Pr(y_{it}=1|\boldsymbol{x}_{it})=(1-y_{it-1})\Pr(y_{it}=1|\boldsymbol{x}_{it},y_{it-1}=0)+y_{it-1}\Pr(y_{it}=1|\boldsymbol{x}_{it},y_{it-1}=1)$.

of order $O(h^{-2})$. Then $\tilde{\boldsymbol{\theta}}$, which can be regarded as the PML estimator, is the solution to

$$U^*(\tilde{\boldsymbol{\theta}}) = U(\tilde{\boldsymbol{\theta}}) - I(\tilde{\boldsymbol{\theta}})\tilde{B}_1(\hat{\boldsymbol{\theta}}) = 0, \tag{8}$$

where $I(\boldsymbol{\theta})$ is the information matrix and $\tilde{B}_1(\hat{\boldsymbol{\theta}}) = \mathrm{E}\left(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\right)$, for which Firth relies on the approximation provided by Cox and Snell (1968). Expression (8) is the result of a first–order Taylor expansion of $U(\hat{\boldsymbol{\theta}})$ around $\tilde{\boldsymbol{\theta}}$.[5]

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. This is the approach adopted by Kunz et al. (2017) to evaluate hospital readmission reduction programs and by Cook et al. (2018) to forecast the occurrence of civil wars. But beacuse I estimate a dynamic logit model, differently from them I consider a formulation for $\tilde{B}_1(\hat{\boldsymbol{\theta}})$ that is specific for dynamic binary choice panel data models. In particular, I consider the analytical bias correction put forward by Fernández-Val (2009), which was derived to reduce the bias due to the incidental parameters problem.

Let $\boldsymbol{\theta}$ collect the parameters for model (4), that is $\boldsymbol{\theta} = (\boldsymbol{\phi}', \gamma, \delta_1, \dots, \delta_n)'$, and let $\boldsymbol{z}_{it} = \left(\boldsymbol{x}'_{it-1}, y_{it-1}, \boldsymbol{d}'_{it}\right)'$, where $\boldsymbol{d}_{it}$ is a vector of country dummies for the t-th time occasion. The analytical bias correction by Fernández-Val (2009) can be written as

$$B_1(\boldsymbol{\theta}) = I(\boldsymbol{\theta})^{-1} \sum_{i=1}^{n} b_i(\boldsymbol{\theta}),$$

so that the modified score function for subject i becomes

$$U_i^*(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left(y_{it} - F_{it}\right) \boldsymbol{z}'_{it} - b_i(\boldsymbol{\theta})', \tag{9}$$

where $F_{it} = \exp\left(\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + y_{it-1}\gamma\right) / \left[1 + \exp\left(\delta_i + \boldsymbol{x}'_{it}\boldsymbol{\phi} + y_{it-1}\gamma\right)\right]$, and clearly $U^*(\boldsymbol{\theta})$ in (8) is $\sum_{i=1}^{n} U_i^*(\boldsymbol{\theta})$. The expression for $b_i(\boldsymbol{\theta})$ is given by[6]

$$b_i(\boldsymbol{\theta}) = -\frac{c_i}{T}\sum_{t=1}^{T} f_{it}\boldsymbol{z}_{it} - \frac{\tau_i}{2T}\sum_{t=1}^{T} g_{it}\boldsymbol{z}_{it} - \frac{\tau_i}{T-1}\sum_{t=2}^{T} f_{it}(y_{it-1} - F_{it-1})\boldsymbol{z}_{it},$$

with $f_{it} = F_{it}(1 - F_{it})$, $g_{it} = f_{it}(1 - 2F_{it})$, and where $\tau_i = T / \sum_{t=1}^{T} f_{it}$ and

$$c_i = -\tau_i \left[\frac{1}{2T}\sum_{t=1}^{T} g_{it} + \frac{1}{T-1}\sum_{t=2}^{T} f_{it}(y_{it-1} - F_{it-1})\right].$$

The PML estimator is the solution to $U^*(\boldsymbol{\theta})$ being equal to zero and it can be obtained by Newton-

[5] 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.

[6] Every term in the following expressions should be wirtten as a function of the parameter vector $\boldsymbol{\theta}$, i.e. $c_i(\boldsymbol{\theta})$, which is suppressed to avoid abuse of notation.

Raphson with update for the $s+1$-th step

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

where the information matrix can be obtained as $I(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{t=1}^{T} f_{it}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'$. Panel robust standard errors can be derived in the ususal way based on the covariance matrix

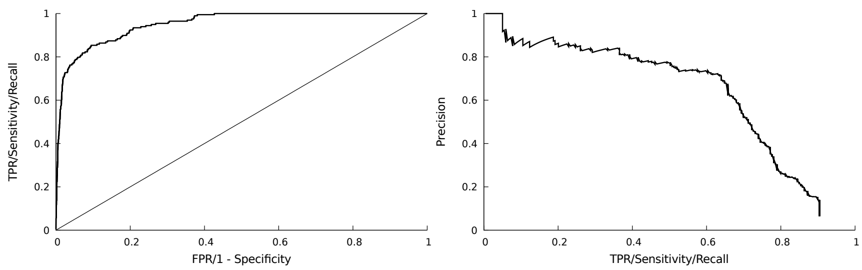
$$V(\tilde{\boldsymbol{\theta}}) = I(\tilde{\boldsymbol{\theta}})^{-1}S(\tilde{\boldsymbol{\theta}})I(\tilde{\boldsymbol{\theta}})^{-1}, \quad \text{with} \quad S(\tilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} U_i^{*}(\tilde{\boldsymbol{\theta}})U_i^{*}(\tilde{\boldsymbol{\theta}})'.$$

2.4 Evaluating model performance

In the early warning literature, the standard approach to evaluate the model prediction accuracy 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 or Recall) is the fraction of correctly classified crisis events and the False Positive Rate (FPR, also 1 - Specificity or 1 - True Negative Rate) 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 (see Figure 1, Panel (a)). The AUROC, that is

Figure 1: ROC and PR curves examples



(a) ROC curve (b) PR curve

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 corresponding to the maximum of the Youden's J statistic (Youden, 1950), 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 about 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 may, paradoxically, correspond to a large number of false alarms.

For this reason, I also present with the empirical results the forecast analysis based on the Precision-Recall (PR) curve, which is argued to be more suitable for the forecasting of rare events (Davis and Goadrich, 2006). The PR curve is the plot of the Precision, equal to the ratio between the true positives and the sum of true and false positives, against the TPR, for every predicted outcome used as cut-off (see Figure 1, Panel (b)). A measure of the model predictive performance is given by the area under the PR curve (AUPR): here a value equal to 0 refers to random classification, whereas a value equal to 1 represents perfect classification.

The optimal cut-off to compute the percentages (or numbers) of crises correctly predicted and false alarms is chosen to maximize the F-score (or F1-score), which is argued to be more suitable for the forecasting of rare events (Davis and Goadrich, 2006). The F-score is the armonic mean of Precision and Recall, computed as twice the ratio between Recall$\times$Precision and Recall$+$Precision. In this context, the selected cut-off for a models with a good predictive performance is related to a high TPR and to a high number of correctly predicted crises relatively to the overall predicted crises, which keeps the number of false alarms under control.

It is worth to clarify that the AUPR is not meant to be an alternative to the AUROC to perform selection between different modeling or estimation approaches *per sè*. In fact, Davis and Goadrich (2006) show that a curve dominates in the ROC space if and only if it dominates in the PR space, meaning that the model producing the highest value of the AUROC is also the one presenting the largest value of the AUPR. Rather I suggest that, once a classifier is selected, the F-score-based cut-off should be used to forecast crisis events because it largely reduces the number of false positives, as it will be shown in Section 5, thus avoiding to dismiss a very accurate model in terms of AUROC because of a high false alarm rate (based on the Youden's J). This recommendation represents an additional element of novelty with respect to the logit-based EWS literature that, so far, has only considered the ROC analysis.

3 Simulation study

In this section I report and discuss the results of a simulation study assessing the performance of the PML estimator for the fixed-effects logit-based EWS, compared to the other estimation strategies usually adopted in the applied early warning literature. I also compare the proposed approach with the performace of the dynamic models proposed by Antunes et al. (2018) and Caggiano et al. (2016).

3.1 Simulation design

For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, artificial data are generated from a dynamic logit model according to:

$$y_{it} = \mathbf{I}(c + \delta_i + x_{it}\phi + y_{it-1}\gamma + \varepsilon_{it} > 0),$$

where $\mathbf{I}(\cdot)$ is an indicator function, c is a common intercept, $\phi = 1$, and $\gamma = 0.5$. The individual effects δ_i are drawn from a standard normal distribution and the covariate x_{it} is assumed to be correlated with the unobserved heterogeneity and generated as a Gaussian copula of the type $x_{it} = \rho\delta_i + \sqrt{1-\rho^2} \cdot N(0, 1)$, with $\rho = 0.5$. Finally, the idiosyncratic error term follows a standard logistic distribution, that is $\varepsilon_{it} \sim \Lambda(0, \pi^2/3)$. At time $t = 0$, the process is initialized as $y_{i0} = \mathbf{I}(c + \delta_i + x_{i0} + \varepsilon_{i0}) > 0$.

The scenarios considered in this simulation study cover $n = 50, 100$ and $T = 10, 20$. I 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. I also consider scenarios where the model covariate is independent of the unobserved heterogeneity, by setting $\rho = 0$. The simulations for each scenario are based on 1000 Monte Carlo replications.

3.2 Main results

Figure 2 depicts the boxplots for the bias distribution of the estimators for ϕ and γ and Figure 3 displays the boxplots for the AUROC and AUPR in the benchmark scenarios with $c = -2$ and $\rho = 0.5$ 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 behavior of the bias, AUROC, and AUPR, Table 1 reports some descriptive statistics for the same scenarios.

Because the simulation design includes some unobserved heterogeneity via δ_i, the MLE Pooled is not consistent. Also, the MLE RE is expected to be biased since the individual effects are correlated with the model covariate through ρ. This is confirmed by the boxplots of their bias distributions for ϕ and γ not centered at zero in every panel of Figure 2, regardless of the sample size. The MLE FE is also 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. The proposed bias reduction technique seems to work quite nicely, in that the boxplot for the bias of ϕ is centered at zero and the bias for γ decreases in the values of n and T.

Looking at the distributions of the AUROC in Figure 3 for the five logit models, 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 boxplots 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 because 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. The distribution of the AUPR follows the same patterns as the AUROC, as they both give the same indications if used for comparisons between estimators, but shows that PMLE FE brings a greater degree of accuracy with respect to the MLE FE*, especially with a short time series.

Other than the expected reduction of the variability in the distributions of the bias and AUROC/AUPR, very similar patterns emerge for the scenarios when n and T increase. 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 Table A.1 and $c = -4$ in Table A.2 in Appendix A. Some bias for the PMLE FE shows up when the frequency of positive events is as low as about 3% ($c = -4$), but it reduces with $T = 20$.

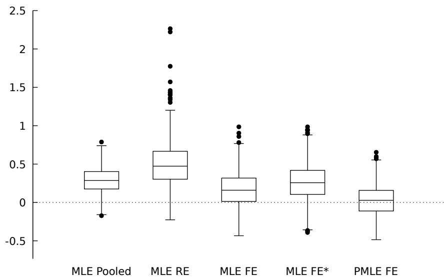
The results for the scenarios with $\rho = 0$ (only for $c = -2$) are presented in Appendix A in Table A.3. In this case, the MLE RE is a consistent estimator for ϕ and γ since the unobserved heterogeneity is generated as independent of the model covariate. It therefore shows a better performance in terms of bias than the PMLE FE, which only brings a bias reduction, yet under the rather restrictive assumption of independence between the model covariates and individual effects. The Appendix also contains the results for the scenarios with $c = -2$ and $\rho = 0.5$ where static logit models, that is assuming $\gamma = 0$, are estimated. Interestingly, omitting the lagged dependent variable from the set of the model explanatory variables does not seem to have an important effect on the bias and AUROC distribution, while it produces a slight decrease in the average value of the AUPR. It is worth mentioning, however, that this result could be specific of the data generating process, where the model covariate is independent of y_{it-1}.

3.3 Comparison with alternative models

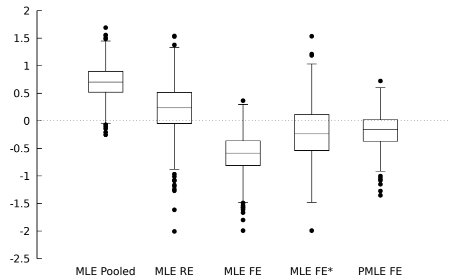
In this last part of the simulation study, I compare the finite sample performance of the proposed PML estimator for the dynamic logit model with the two approaches dealing with the so–called duration bias in the EWS literature for financial crises, namely the dynamic probit model by Antunes et al. (2018) and the multinomial logit model by Caggiano et al. (2016).

Antunes et al. (2018) specify the same dynamic binary choice model considered here, with the difference that the idiosyncratic component follows a standard normal instead of a standard logistic distribution. Moreover, they consider a pooled model, which assumes no country specific permanent heterogeneity. Similarly to expression (5), the probability of a crisis occurring in coun-

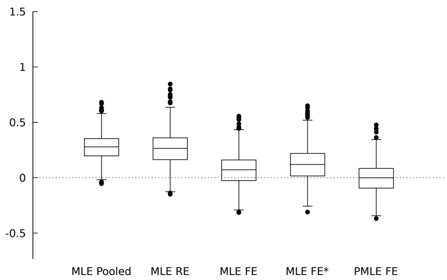
Figure 2: Simulation results: Bias distribution for ϕ and γ, $c = -2$, $\rho = 0.5$



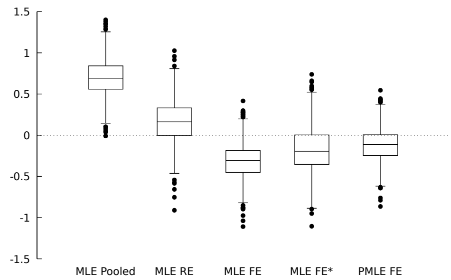
(a) Bias distribution for ϕ, $n = 50, T = 10$



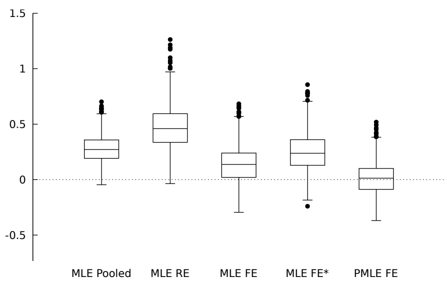
(b) Bias distribution for γ, $n = 50, T = 10$



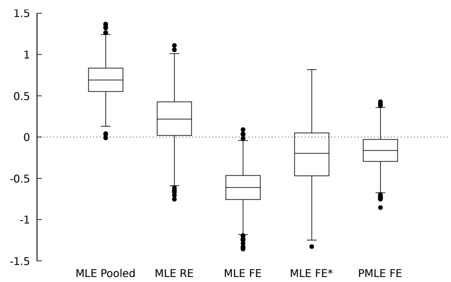
(c) Bias distribution for ϕ, $n = 50, T = 20$



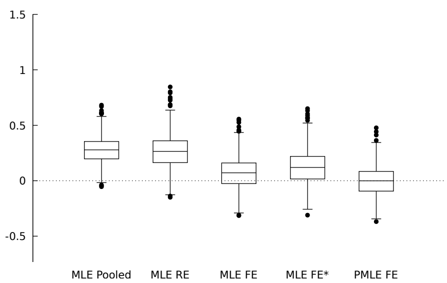
(d) Bias distribution for γ, $n = 50, T = 20$



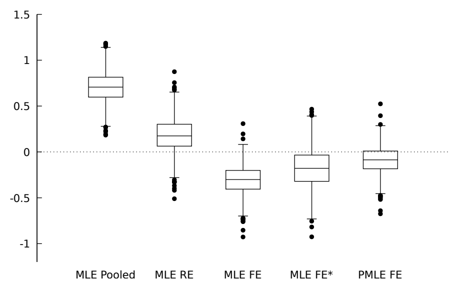
(e) Bias distribution for ϕ, $n = 100, T = 10$



(f) Bias distribution for γ, $n = 100, T = 10$

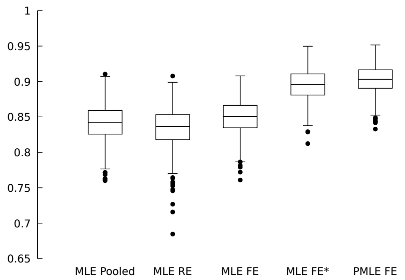


(g) Bias distribution for ϕ, $n = 100, T = 20$

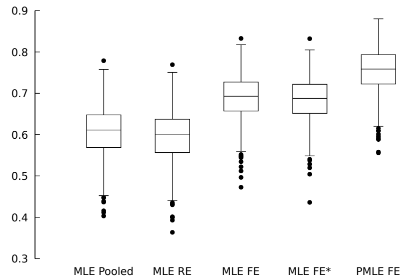


(h) Bias distribution for γ, $n = 100, T = 20$

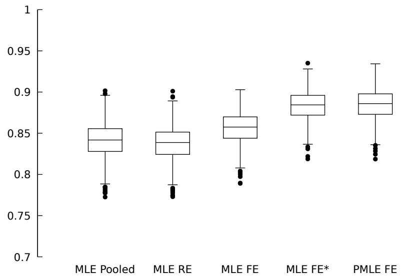
Figure 3: Simulation results: AUROC and AUPR distribution for $\Pr(y_{it} = 1)$, $c = -2$, $\rho = 0.5$



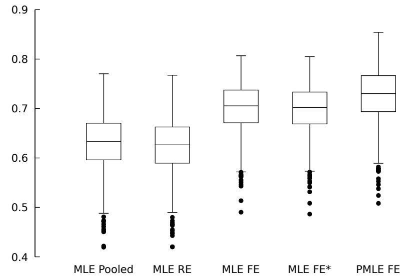
(a) AUROC distribution, $n = 50, T = 10$



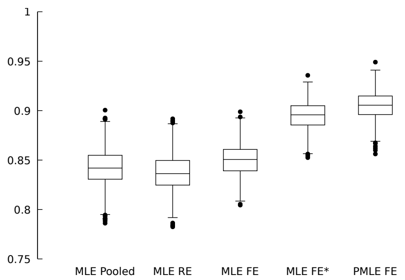
(b) AUPR distribution, $n = 50, T = 10$



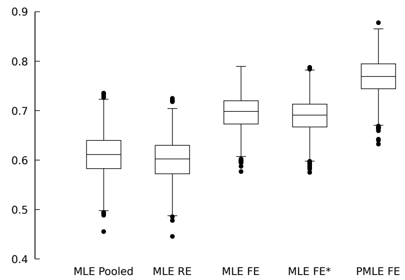
(c) AUROC distribution, $n = 50, T = 20$



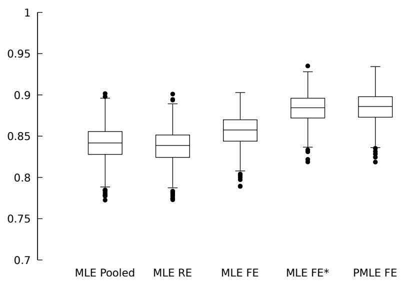
(d) AUPR distribution, $n = 50, T = 20$



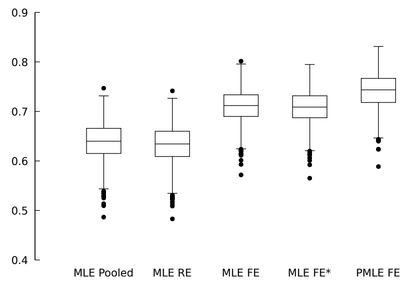
(e) AUROC distribution, $n = 100, T = 10$



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



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



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

Table 1: Simulation results: Mean and standard deviation of Bias, mean of AUROC and AUPR, $c = -2$, $\rho = 0.5$

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias ϕ	0.291	0.507	0.165	0.261	0.028
SD bias ϕ	0.166	0.337	0.223	0.226	0.191
Mean bias γ	0.712	0.212	-0.605	-0.221	-0.177
SD bias γ	0.294	0.477	0.336	0.474	0.294
Mean AUROC	0.842	0.835	0.850	0.895	0.903
Mean AUPR	0.608	0.595	0.690	0.686	0.756
$n = 50, T = 20$					
Mean bias ϕ	0.276	0.268	0.070	0.121	-0.001
SD bias ϕ	0.116	0.148	0.137	0.148	0.129
Mean bias γ	0.700	0.165	-0.313	-0.169	-0.120
SD bias γ	0.224	0.252	0.209	0.279	0.198
Mean AUROC	0.841	0.838	0.857	0.884	0.885
Mean AUPR	0.630	0.624	0.701	0.699	0.727
$n = 100, T = 10$					
Mean bias ϕ	0.280	0.474	0.138	0.244	0.014
SD bias ϕ	0.122	0.194	0.158	0.168	0.137
Mean bias γ	0.700	0.221	-0.610	-0.208	-0.162
SD bias γ	0.215	0.314	0.229	0.361	0.204
Mean AUROC	0.842	0.837	0.850	0.895	0.905
Mean AUPR	0.610	0.600	0.696	0.689	0.767
$n = 100, T = 20$					
Mean bias ϕ	0.279	0.273	0.073	0.121	0.008
SD bias ϕ	0.084	0.103	0.097	0.109	0.090
Mean bias γ	0.709	0.181	-0.304	-0.170	-0.089
SD bias γ	0.162	0.185	0.154	0.215	0.146
Mean AUROC	0.843	0.840	0.859	0.886	0.889
Mean AUPR	0.639	0.633	0.710	0.708	0.742

try i at time t given that the country was not in distress at time $t-1$ can, in general, be written as

$$\Pr(y_{it} = 1|\boldsymbol{x}_{it}, y_{i,t-1} = 0) = \Phi\left(\boldsymbol{x}_{it}'\boldsymbol{\phi}\right),$$

while, similarly to expression (6), the probability that country i at time t is still in distress given that $t-1$ was a crisis year is

$$\Pr(y_{it} = 1|\boldsymbol{x}_{it}, y_{i,t-1} = 0) = \Phi\left(\boldsymbol{x}_{it}'\boldsymbol{\phi} + y_{it-1}\gamma\right),$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Caggiano et al. (2016) adopt a different approach and consider a pooled multinomial logit model where the dependent variable takes values equal to: 0 if country i is not facing a crisis at time t nor at time $t-1$, $(y_{it} = 0, y_{it-1} = 0)$; 1 if the country is in state of financial distress in t but it was not in $t-1$, $(y_{it} = 1, y_{it-1} = 0)$; 2 if the country has faced a crisis in both $t-1$ and t, $(y_{it} = 1, y_{it-1} = 1)$. Notice that the conditioning with respect to the state in $t-1$ is here implicit in the definition of the response variable. For the resulting multinomial logit model, the probability of a crisis occurring in country i at time t given that the country was not in distress at time $t-1$ is

$$\Pr(y_{it} = 1|\boldsymbol{x}_{it}, y_{i,t-1} = 0) = \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\phi}_1)}{\sum_{l=0}^{2}\exp(\boldsymbol{x}_{it}'\boldsymbol{\phi}_l)},$$

while the probability that country i at time t is still in distress given that $t-1$ was a crisis year is

$$\Pr(y_{it} = 1|\boldsymbol{x}_{it}, y_{i,t-1} = 1) = \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\phi}_2)}{\sum_{l=0}^{2}\exp(\boldsymbol{x}_{it}'\boldsymbol{\phi}_l)}.$$

Each alternative is associated with a different parameter vector $\boldsymbol{\phi}_l$, $l = 0, 1, 2$, with $\boldsymbol{\phi}_0 = \boldsymbol{0}$ for identification. With respect to the dynamic logit and probit models, this approach introduces an additional sources of heterogeneity in the regression parameters, that are allowed to vary with the state of financial distress in $t-1$. [7]

Figure 4 and Table 2 report the simulation results for the probit model by Antunes et al. (2018), MLE Pooled Probit, the multinomial logit by Caggiano et al. (2016), and the proposed approach, here denoted as PMLE FE Logit, based on the simulation design described in Section 3.1 with $c = -2$ and $\rho = 0.5$. For brevity, the figure only reports the distribution of the AUROC and AUPR. In Table 2, the bias statistics are only reported for the MLE Pooled Probit and PMLE FE Logit, that have the same parametrization.

From Figure 4 it emerges that the model performance in forecasting positive events improves when subject–specific fixed effects are considered. The average values of the AUROC and AUPR

[7] It is straightforward to allow for such heterogeneity in dynamic binary choice models by using interaction terms between the model covariates and the lagged dependent variable. The resulting formulation would correspond to the transition probability model, put forward to study poverty dynamics by Cappellari and Jenkins (2002, 2004). Because I focus on the estimation approach and not on the model specification, this is not investigated further. Forecasts obtained with the PMLE FE transition probability model based on the empirical example in Section 5, available upon request, did not reveal a crucial role of this sort of heterogeneity.

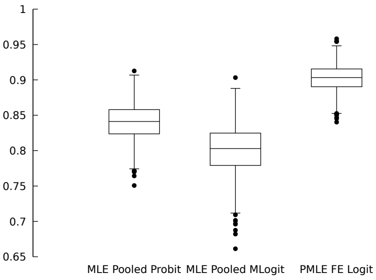
are the largest for the PMLE FE Logit, while those for the MLE Pooled Probit mirror closely the results for the MLE Pooled Logit model in Figure 3. On average, the AUROC and AUPR for the MLE Pooled Mlogit are about 10 to 20 percentage points smaller than those for the proposed approach in all the scenarios considered (see Table 2). The MLE Pooled Probit also exhibit a larger bias than the PMLE FE Logit in the estimate for ϕ and γ.[8] As for the MLE of the pooled logit model considered in Section 3.2, this is due to the presence of unobserved heterogeneity, which is neglected by a pooled model specification.

Table 2: Simulation results for alternative models: Mean and standard deviation of Bias, mean of AUROC and AUPR, $c = -2$, $\rho = 0.5$.

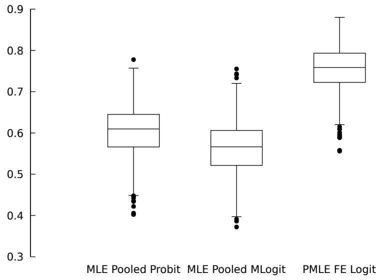
	MLE Pooled Probit (Antunes et al., 2018)	MLE Pooled MLogit (Caggiano et al., 2016)	PMLE FE Logit
$n = 50, T = 10$			
Mean bias ϕ	0.156		0.028
SD bias ϕ	0.144		0.191
Mean bias γ	0.583		-0.177
SD bias γ	0.279		0.294
Mean AUROC	0.841	0.801	0.903
Mean AUPR	0.603	0.547	0.756
$n = 50, T = 20$			
Mean bias ϕ	0.150		-0.001
SD bias ϕ	0.101		0.129
Mean bias γ	0.587		-0.120
SD bias γ	0.213		0.198
Mean AUROC	0.841	0.804	0.885
Mean AUPR	0.632	0.576	0.727
$n = 100, T = 10$			
Mean bias ϕ	0.149		0.014
SD bias ϕ	0.101		0.137
Mean bias γ	0.587		-0.162
SD bias γ	0.198		0.204
Mean AUROC	0.842	0.803	0.905
Mean AUPR	0.610	0.556	0.767
$n = 100, T = 20$			
Mean bias ϕ	0.153		0.008
SD bias ϕ	0.074		0.090
Mean bias γ	0.592		-0.089
SD bias γ	0.198		0.146
Mean AUROC	0.843	0.805	0.889
Mean AUPR	0.638	0.585	0.742

[8] Since the data are generated from a logit model, the estimates obtained from the probit model are multiplied by 1.6 in order to have a comparable order of magnitude with those of the logit model, following the rule of thumb by Amemiya (1981).

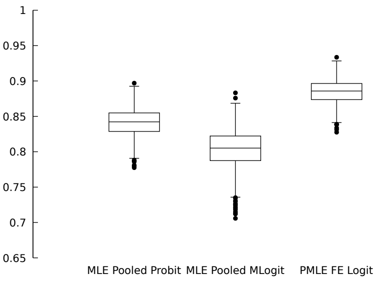
Figure 4: Simulation results for alternative models: AUROC and AUPR distribution for $\Pr(y_{it} = 1)$, $c = -2$, $\rho = 0.5$



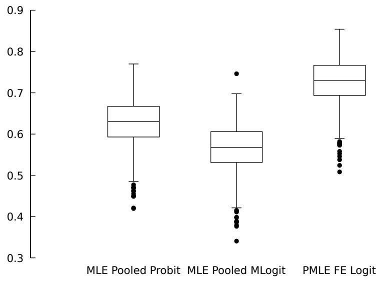
(a) AUROC distribution, $n = 50, T = 10$



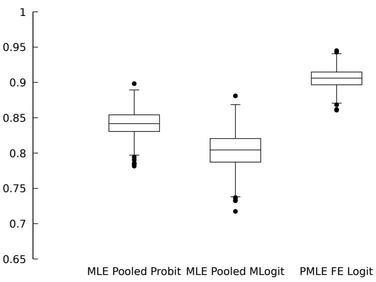
(b) AUPR distribution, $n = 50, T = 10$



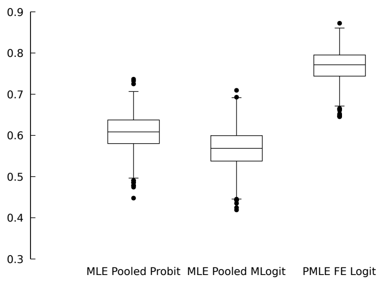
(c) AUROC distribution, $n = 50, T = 20$



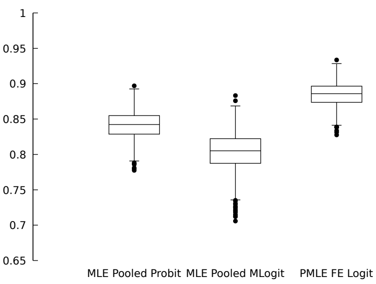
(d) AUPR distribution, $n = 50, T = 20$



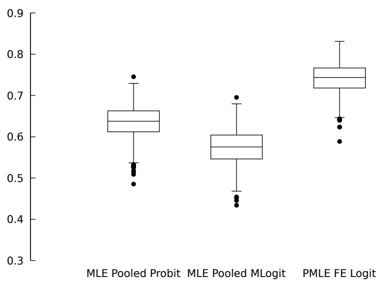
(e) AUROC distribution, $n = 100, T = 10$



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



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



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

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 (seeTable B.1 in Appendix B for the complete list), covering the period from 1983 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.[9] Based on this definition, I end up with a sample where 198 events of systemic banking crises are identified, 64 of which are new crises, and the average duration is 3 years. Table 3 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 4 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 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

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

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 3: 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	1998	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

5 Estimation results

In this section I report the estimation results for the proposed approach and alternative models/estimation techniques. Next I analyze in–sample predictions and then assess the ability of the discussed approaches to generate accurate out-of-sample forecasts.

5.1 Model estimation

Table 5 reports the results relative to the five ML estimators for the pooled, random- and fixed-effects models described in Section 3, along with the estimates for the alternative dynamic models in the EWS literature on banking crises, namely the pooled probit and pooled multinomial logit put forward by Antunes et al. (2018) and Caggiano et al. (2016), respectively.

Looking at the values of the McFadden R^2, it can be noticed that including country specific effects in the specification of EWS increases the model goodness of fit, except for the MLE FE,

Table 4: Variable sources and descriptive statistics

	Source	Mean	Median	SD	Min	Max
DEPENDENT VARIABLE						
Crisis	Laeven and Valencia (2018)	0.068	0.000	0.252	0	1
EXPLANATORY VARIABLES						
Real GDP growth(-1)	IFS	3.879	4.089	4.238	-13.82	18.18
Log per capita GDP(-1)	WDI	7.705	7.635	1.458	4.713	11.39
Inflation(-1)	WDI	26.50	5.940	324.7	-31.57	40.86
Real interest rate(-1)	IFS	2.384	2.125	11.23	-11.42	15.02
M2 to foreign exchange reserves(-1)	WDI	15.21	3.970	33.58	1.111	62.13
Growth of real domestic credit(-1)	WDI	16.25	12.64	32.86	-9.455	54.75
Growth of net foreign assets to GDP(-1)	WDI	0.019	0.015	0.231	-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.

since several countries do not contribute to the estimation sample (down to 64 from 129). The highest value is achieved by the proposed PML estimator of the fixed-effects dynamic logit model.

As for the estimated coefficients, it can be immediately noticed that the lag of the dependent variable has a strong explanatory power, as denoted by the t ratios. 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. This is a common feature of all the models and estimation approaches considered, with the exception of the pooled multinomial logit, in that the definition of the dependent variable already contains the information of the crisis status at $t-1$

As for the rest of the explanatory variables, all the related coefficients have the expected sign. In addition, M2 over foreign exchange reserves emerges as a strong predictor across all models. The fixed-effects models detect the real interest rate as the only other mildly significant early warning, whereas in the rest of the models other explanatory variables are found to have a statistically significant effect on the crisis probability. In particular: the real GDP growth rate in the pooled and random-effects logit, in the pooled probit, and in the multinomial logit model for the probability of a crisis occurring for one more year; the real interest rate also in the random-effects logit, pooled probit and multinomial logit, for the probability of a crisis onset; the growth rate of private credit in the random-effects logit, pooled probit and multinomial logit; the growth rate of net foreign assets to GDP in the random-effects logit and pooled multinomial logit model, relatively to the probability of a prolonged state of financial distress.

Discrepancies in the estimation results between models could be the consequence of a different treatment of the unobserved heterogeneity. In fixed-effects models, country–level intercepts may capture specific latent traits that, if correlated with the model covariates, may generate an omitted variable bias when such heterogeneity is considered independent of the explanatory variables or ignored.

Table 5: Estimation results: logit and alternative models

	MLE Pooled Logit	MLE RE Logit	MLE FE Logit	MLE FE* Logit	PMLE FE Logit	MLE Pooled Probit	MLE Pooled Mlogit $l = 1$	MLE Pooled Mlogit $l = 2^*$
# of observations	3045	3045	1683	3045	3045	3045	3045	
# of countries	129	129	64	129	129	129	129	
McFadden R^2	0.449	0.520	0.451	0.538	0.581	0.495	0.153	
Log-likelihood	-403.429	-383.530	-334.880	-338.241	-334.981	-403.220	-783.7836	
COEFFICIENTS								
Crisis(-1)	4.449	4.591	3.687	3.819	3.729	2.406		
	[21.42]	[20.24]	[20.42]	[16.79]	[21.65]	[22.11]		
Real GDP growth(-1)	-0.042	-0.040	-0.018	-0.021	-0.014	-0.019	-0.019	-0.160
	[-1.941]	[-1.668]	[-0.652]	[-0.829]	[-0.494]	[-1.720]	[-1.718]	[-9.180]
Log per capita GDP(-1)	0.032	0.015	-0.070	0.247	-0.044	0.007	0.007	-0.027
	[0.470]	[0.197]	[-0.316]	[1.391]	[-0.202]	[0.213]	[0.214]	[-0.456]
Inflation(-1)	-0.000	-0.000	0.000	-0.000	0.000	-0.000	-0.000	-0.000
	[-0.200]	[-0.537]	[0.190]	[-0.031]	[-0.317]	[-0.322]	[-0.482]	[-0.640]
Real interest rate(-1)	0.011	0.017	0.015	0.017	0.015	0.006	0.006	0.009
	[1.539]	[2.212]	[1.831]	[1.708]	[1.743]	[1.734]	[1.735]	[1.418]
M2 to foreign	0.007	0.009	0.028	0.031	0.027	0.004	0.004	0.007
exchange reserves(-1)	[3.104]	[3.159]	[2.888]	[4.794]	[2.812]	[0.178]	[3.176]	[3.606]
Growth of real	0.009	0.014	0.013	0.010	0.007	0.005	0.005	0.009
domestic credit(-1)	[1.554]	[2.322]	[2.112]	[1.376]	[1.063]	[1.684]	[1.679]	[1.865]
Growth of net foreign	-0.492	-1.110	0.008	-0.610	-0.221	-0.283	-0.282	-0.985
assets to GDP(-1)	[-0.822]	[-1.688]	[1.152]	[-0.780]	[-0.259]	[-0.951]	[-0.946]	[-2.035]

Panel robust t-tests are in square brackets. All models include an intercept term. The RE Logit model is based on the Gauss-Hermite quadrature method with 13 grid points, the estimated variance of the unobserved heterogeneity is $\hat{\sigma}_\delta^2 = 0.300$.
*: $l = 1$ corresponds to the alternative $y_{it} = 1, y_{it-1} = 0$; $l = 2$ corresponds to the alternative $y_{it} = 1, y_{it-1} = 1$.

5.2 In-sample forecasts

In the EWS literature, it is customary to assess the model performance by means of the ROC analysis. As argued in Section 2.4, evaluations based on the AUROC can mislead to an overly optimistic view of the model accuracy as it can, at the same time, generate a large number of false alarms. For this reason, the ROC analysis is here presented along with indicators based on the PR curve.

Table 6 reports the values of the AUROC and AUPR for all the models and estimation approaches considered. The corresponding plots can be found in Figures C.1–C.7 of Appendix C. Based on this values, cut-offs to compute crises predictions are derived based on the maximum of the Youden's J statistic and F-score for the AUROC and AUPR, respectively. Using these two thresholds, I then report the number of crises correctly predicted and the number of false positives, along with the accuracy, that is the percentage of cases correctly predicted, the true positive rate, the true negative rate, and the precision, which is the ratio of correctly predicted crises over all predicted crises.

The values of the AUROC and AUPR confirm the importance of dealing with the unobserved heterogeneity by including fixed effects, in terms of the model ability to detect crisis events. The ROC analysis however reveals that fixed-effects models yield very small probability cut-offs, which makes them correctly predict all the 198 crisis events in the sample, at the cost of over 2000 false positives. The number of false positives generated by the MLE FE logit is smaller only because its estimation is based on a smaller sample. Notice that, as a consequence, there is a high true positive rate, alongside with a low true negative rate (high rate of false alarms). A clear indication of the poor model performance in this respect is the value of the precision, which is rather small for all the models considered.

A different story is told by the forecasts obtained using the cut-off based on the F-score. Here the number of crises correctly predicted is highest for the pooled probit and the pooled and random-effects logit model, although the smallest number of false alarms is generated by the MLE FE*. It can be noticed that using the F-score cut-off dramatically reduces the number of false positives, at the expense of loosing a few correctly predicted crisis events. As a consequence, accuracy and precision increase by a sizable amount, compared to the indicators provided by the ROC analysis.

In light of these results, the MLE FE* dynamic logit seems to be the best performing approach. The out–of–sample forecast exercise will however show its limitations, its inability to predict first ever crisis events in particular. Yet this analysis suggests that, because of its ease of computation, can be used as a benchmark when performing preliminary model selection.

5.3 Out-of-sample forecasts

The results for out–of–sample forecast exercises for the years 2007–2014 are reported in Table 7, for the different models and estimation approaches considered so far. The method used to

Table 6: In-sample forecasts: logit and alternative models

	MLE Pooled Logit	MLE RE Logit	MLE FE Logit	MLE FE* Logit	PMLE FE Logit	MLE Pooled Probit	MLE Pooled Mlogit
# of crises: 198							
AUROC	0.886	0.896	0.908	0.949	0.952	0.891	0.877
Cut-off (Youden's J)	0.018	0.104	0.047	0.000	0.000	0.015	0.043
# of crises correctly predicted	171	195	173	198	198	190	103
# of false alarms	1047	2355	423	2305	2758	2046	101
Accuracy	0.647	0.226	0.734	0.243	0.094	0.325	0.936
True positive rate/Recall/Sensitivity	0.846	0.985	0.874	1.000	1.000	0.960	0.520
True negative rate/Specificity	0.632	0.173	0.715	0.190	0.031	0.281	0.965
Precision	0.140	0.076	0.290	0.079	0.063	0.085	0.505
AUPR	0.531	0.527	0.623	0.623	0.623	0.530	0.448
Cut-off (F-score)	0.201	0.465	0.460	0.467	0.486	0.237	0.045
# of crises correctly predicted	143	143	140	140	139	143	100
# of false alarms	75	75	56	20	54	75	70
Accuracy	0.957	0.957	0.932	0.963	0.963	0.957	0.945
True positive rate/Recall/Sensitivity	0.722	0.722	0.707	0.707	0.702	0.722	0.505
True negative rate/Specificity	0.974	0.974	0.962	0.981	0.981	0.974	0.975
Precision	0.656	0.656	0.714	0.718	0.720	0.656	0.588

compute these prediction is similar to the recursive technique adopted by Dawood et al. (2017). First the panel is restricted to the period 1983–2006 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 wave 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. Because there only few crisis events in each year, I only report the number of crises correctly predicted and false alarms, other than the AUROC, AUPR, and related thresholds.

From Table 7, it emerges that the proposed PMLE FE logit has the best performance out of sample. If the ROC analysis is considered, all the 34 crisis events in the years left out of the samples are correctly predicted, although with a high number of false alarms. If instead forecasts are derived on the basis of the F-score cut-off, the PMLE FE is able to correctly detect 31 crises against only 7 false alarms, which is the best performance among all the models considered. The pooled logit and probit models and the random effects logit model have a comparable performance within the ROC analysis, but the only predict 26 crises correctly with the F-score cut-offs, with the first two also generating 18 false alarms. This exercise also reveals the limitations of the standard MLE FE and of the MLE FE*, in that they are not able to forecast the occurrence of a first ever crisis episode.[10]

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

[10] There is also a computational problem with the evaluation of the AUROC for the MLE FE* in 2012.

Table 7: Out-of-sample forecasts: Logit and alternative models

Year # of crises	MLE Pooled Logit	MLE RE Logit	MLE FE Logit	MLE FE* Logit	PMLE FE Logit	MLE Pooled Probit	MLE Pooled Mlogit
2007 2							
AUROC	0.939	0.939	0.870	0.904	0.991	0.991	0.917
Cut-off (Youden's J)	0.023	0.035	0.068	0.001	0.000	0.023	0.017
# of crises correctly predicted	2	2	0	0	2	2	2
# of false alarms	32	23	18	20	111	63	15
AUPR	0.043	0.037	0.000	0.000	0.150	0.156	0.030
Cut-off (F-score)	0.034	0.049	0.044	0.000	0.121	0.058	0.020
# of crises correctly predicted	2	2	1	1	2	2	2
# of false alarms	8	8	30	22	1	1	10
2008 7							
AUROC	0.631	0.625	0.974	0.951	0.952	0.884	0.617
Cut-off (Youden's J)	0.016	0.025	0.000	0.000	0.000	0.011	0.012
# of crises correctly predicted	7	7	3	0	7	7	4
# of false alarms	93	94	82	51	80	85	62
AUPR	0.138	0.131	0.105	0.035	0.466	0.228	0.209
Cut-off (F-score)	0.096	0.149	0.181	0.159	0.061	0.063	0.045
# of crises correctly predicted	2	2	2	1	5	2	2
# of false alarms	0	0	0	0	2	0	0
2009 8							
AUROC	0.942	0.946	0.998	0.914	0.993	0.925	0.966
Cut-off (Youden's J)	0.022	0.032	0.004	0.000	0.027	0.014	0.015
# of crises correctly predicted	8	8	8	8	8	8	8
# of false alarms	73	60	46	23	27	82	40
AUPR	0.724	0.729	0.816	0.450	0.799	0.727	0.734
Cut-off (F-score)	0.077	0.114	0.162	0.064	0.190	0.055	0.039
# of crises correctly predicted	7	7	7	4	7	7	7
# of false alarms	0	0	1	23	0	0	0
2010 6							
AUROC	0.995	0.995	0.997	0.910	1.000	0.995	0.978
Cut-off (Youden's J)	0.062	0.094	0.000	0.000	0.000	0.010	0.018
# of crises correctly predicted	6	6	6	6	6	6	5
# of false alarms	11	8	106	24	67	102	14
AUPR	0.717	0.716	0.777	0.380	0.833	0.718	0.747
Cut-off (F-score)	0.091	0.125	0.760	0.283	0.617	0.578	0.119
# of crises correctly predicted	6	6	6	3	6	6	5
# of false alarms	2	2	1	0	0	1	0
2011 5							
AUROC	0.998	0.998	0.993	0.872	0.994	0.996	0.883
Cut-off (Youden's J)	0.052	0.080	0.011	0.174	0.000	0.031	0.011
# of crises correctly predicted	5	5	5	2	5	5	5
# of false alarms	8	8	47	1	52	11	75
AUPR	0.701	0.697	0.506	0.157	0.591	0.674	0.519
Cut-off (F-score)	0.069	0.098	0.160	0.174	0.165	0.059	0.035
# of crises correctly predicted	5	5	5	2	5	5	4
# of false alarms	1	1	1	1	1	1	0
2012 3							
AUROC	0.982	0.982	0.994	-	0.982	0.982	0.634
Cut-off (Youden's J)	0.019	0.027	0.000	-	0.000	0.019	0.017
# of crises correctly predicted	3	3	3	-	3	3	2
# of false alarms	76	84	84	-	65	52	20
AUPR	0.247	0.243	0.429	0.230	0.237	0.240	0.028
Cut-off (F-score)	0.074	0.102	0.844	0.226	0.186	0.071	0.039
# of crises correctly predicted	3	3	2	2	3	3	1
# of false alarms	2	2	0	0	2	2	2
2014 3							
AUROC	0.633	0.630	0.994	1.000	0.997	0.627	0.639
Cut-off (Youden's J)	0.017	0.024	0.000	0.001	0.000	0.010	0.033
# of crises correctly predicted	2	2	1	0	3	2	2
# of false alarms	95	92	89	75	81	100	87
AUPR	0.031	0.026	0.000	0.000	0.490	0.034	0.024
Cut-off (F-score)	0.042	0.064	0.085	0.073	0.067	0.038	0.019
# of crises correctly predicted	1	1	1	1	3	1	1
# of false alarms	5	5	2	0	1	3	2
Total 34							
AUROC							
# of crises correctly predicted	33	33	26	-	34	33	28
# of false alarms	388	369	472	-	483	495	313
AUPR							
# of crises correctly predicted	26	26	24	14	31	26	22
# of false alarms	18	18	35	56	7	8	14

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 in-sample performance. This strategy may represent an appealing alternative to perform preliminary analyses and model selection. 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.

Another insight provided in the paper concerns the way predictions of rare events are derived, in that assessing forecasting accuracy with standard analyses such as the ROC can lead to overstate the model ability to correctly predict crises, as it also entails 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. 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.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of economic literature*, 19(4):1483–1536.

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.

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.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1):39–54.

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

7 Appendix

A Additional simulation results

Table A.1: Simulation results: Mean and standard deviation of Bias, mean of AUROC and AUPR, $c = 0$, $\rho = 0.5$

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias ϕ	0.271	1.388	0.155	0.300	0.043
SD bias ϕ	0.149	1.987	0.198	0.190	0.172
Mean bias γ	0.681	0.350	-0.543	0.056	-0.092
SD bias γ	0.258	1.161	0.288	0.359	0.252
Mean AUROC	0.840	0.828	0.860	0.879	0.896
Mean AUPR	0.805	0.797	0.815	0.835	0.923
$n = 50, T = 20$					
Mean bias ϕ	0.257	0.405	0.070	0.144	0.019
SD bias ϕ	0.106	0.198	0.121	0.129	0.112
Mean bias γ	0.688	0.149	-0.263	-0.048	-0.033
SD bias γ	0.200	0.301	0.191	0.238	0.180
Mean AUROC	0.840	0.834	0.865	0.876	0.882
Mean AUPR	0.837	0.833	0.854	0.866	0.910
$n = 100, T = 10$					
Mean bias ϕ	0.260	0.844	0.142	0.296	0.029
SD bias ϕ	0.108	1.201	0.142	0.138	0.124
Mean bias γ	0.681	0.206	-0.548	0.081	-0.107
SD bias γ	0.185	0.603	0.189	0.245	0.170
Mean AUROC	0.841	0.831	0.860	0.879	0.898
Mean AUPR	0.806	0.799	0.816	0.836	0.926
$n = 100, T = 20$					
Mean bias ϕ	0.256	0.405	0.065	0.148	0.012
SD bias ϕ	0.074	0.145	0.086	0.090	0.080
Mean bias γ	0.691	0.157	-0.265	-0.033	-0.041
SD bias γ	0.139	0.210	0.131	0.167	0.124
Mean AUROC	0.841	0.835	0.865	0.876	0.883
Mean AUPR	0.841	0.837	0.856	0.869	0.914

Table A.2: Simulation results: Mean and standard deviation of Bias, mean of AUROC and AUPR, $c = -4$, $\rho = 0.5$

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias ϕ	0.384	0.537	0.216	0.173	-0.059
SD bias ϕ	0.265	2.132	0.407	0.463	0.327
Mean bias γ	0.744	0.354	-0.911	-0.853	-0.336
SD bias γ	0.617	3.010	0.647	0.683	0.558
Mean AUROC	0.848	0.845	0.820	0.924	0.934
Mean AUPR	0.327	0.317	0.491	0.466	0.440
$n = 50, T = 20$					
Mean bias ϕ	0.351	0.322	0.090	0.081	-0.086
SD bias ϕ	0.185	0.210	0.236	0.290	0.231
Mean bias γ	0.734	0.244	-0.500	-0.460	-0.217
SD bias γ	0.477	0.502	0.433	0.464	0.420
Mean AUROC	0.846	0.844	0.823	0.907	0.911
Mean AUPR	0.334	0.329	0.453	0.437	0.384
$n = 100, T = 10$					
Mean bias ϕ	0.363	0.433	0.171	0.002	-0.026
SD bias ϕ	0.186	0.230	0.267	0.449	0.226
Mean bias γ	0.756	0.297	-0.912	-0.852	-0.299
SD bias γ	0.430	0.547	0.449	0.487	0.385
Mean AUROC	0.848	0.846	0.819	0.905	0.939
Mean AUPR	0.329	0.321	0.499	0.454	0.489
$n = 100, T = 20$					
Mean bias ϕ	0.350	0.317	0.077	-0.009	-0.047
SD bias ϕ	0.129	0.147	0.163	0.318	0.156
Mean bias γ	0.772	0.250	-0.498	-0.492	-0.180
SD bias γ	0.315	0.345	0.295	0.337	0.284
Mean AUROC	0.848	0.847	0.827	0.904	0.918
Mean AUPR	0.349	0.341	0.470	0.446	0.440

Table A.3: Simulation results: Mean and standard deviation of Bias, mean of AUROC and AUPR, $c = -2$, $\rho = 0$

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias ϕ	-0.098	0.161	0.170	0.147	0.012
SD bias ϕ	0.135	0.224	0.196	0.191	0.164
Mean bias γ	0.601	-0.040	-0.608	-0.484	-0.182
SD bias γ	0.298	0.406	0.324	0.392	0.286
Mean AUROC	0.771	0.755	0.841	0.875	0.878
Mean AUPR	0.469	0.441	0.635	0.626	0.670
$n = 50, T = 20$					
Mean bias ϕ	-0.109	0.056	0.076	0.066	-0.010
SD bias ϕ	0.101	0.124	0.121	0.120	0.111
Mean bias γ	0.602	-0.002	-0.304	-0.244	-0.114
SD bias γ	0.225	0.232	0.210	0.226	0.197
Mean AUROC	0.771	0.759	0.839	0.853	0.853
Mean AUPR	0.486	0.464	0.623	0.617	0.630
$n = 100, T = 10$					
Mean bias ϕ	-0.110	0.137	0.152	0.133	0.013
SD bias ϕ	0.101	0.154	0.143	0.140	0.122
Mean bias γ	0.604	-0.036	-0.606	-0.505	-0.157
SD bias γ	0.215	0.288	0.232	0.288	0.204
Mean AUROC	0.771	0.755	0.841	0.878	0.881
Mean AUPR	0.469	0.442	0.639	0.633	0.683
$n = 100, T = 20$					
Mean bias ϕ	-0.115	0.055	0.074	0.066	0.001
SD bias ϕ	0.068	0.084	0.084	0.084	0.077
Mean bias γ	0.607	-0.006	-0.308	-0.266	-0.093
SD bias γ	0.161	0.162	0.148	0.158	0.139
Mean AUROC	0.771	0.758	0.840	0.855	0.857
Mean AUPR	0.489	0.466	0.630	0.626	0.647

Table A.4: Simulation results: Mean and standard deviation of Bias, mean of AUROC and AUPR, $c = -2$, $\rho = 0.5$, static logit model

	MLE Pooled	MLE RE	MLE FE	MLE FE*	PMLE FE
$n = 50, T = 10$					
Mean bias ϕ	0.344	0.467	0.136	0.246	0.002
SD bias ϕ	0.162	0.274	0.215	0.232	0.186
Mean AUROC	0.796	0.796	0.839	0.885	0.902
Mean AUPR	0.564	0.564	0.730	0.718	0.747
$n = 50, T = 20$					
Mean bias ϕ	0.340	0.221	0.046	0.105	-0.019
SD bias ϕ	0.127	0.152	0.140	0.155	0.130
Mean AUROC	0.796	0.796	0.849	0.875	0.881
Mean AUPR	0.571	0.571	0.716	0.710	0.716
$n = 100, T = 10$					
Mean bias ϕ	0.340	0.458	0.126	0.245	-0.001
SD bias ϕ	0.113	0.204	0.150	0.165	0.132
Mean AUROC	0.797	0.797	0.839	0.885	0.905
Mean AUPR	0.568	0.568	0.736	0.723	0.760
$n = 100, T = 20$					
Mean bias ϕ	0.338	0.214	0.039	0.094	-0.021
SD bias ϕ	0.085	0.105	0.097	0.112	0.091
Mean AUROC	0.797	0.797	0.851	0.879	0.885
Mean AUPR	0.573	0.573	0.724	0.719	0.728

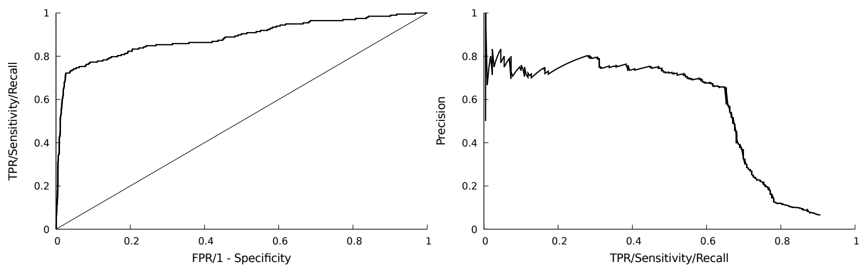
B Panel structure

Table B.1: List of countries and panel structure

Albania	1996-2017	Algeria	1983-2017
Angola	1997-2017	Argentina	1983-2017
Armenia	1996-2017	Australia	1983-2016
Azerbaijan	1999-2017	Bangladesh	1983-2017
Barbados	1983-2009	Belarus	1996-2017
Belize	1983-2017	Benin	2006-2017
Bhutan	1985-2016	Bolivia	1983-2017
Bosnia and Herzegovina	2006-2017	Brazil	1997-2017
Brunei Darussalam	2001-2017	Bulgaria	1995-2017
Burkina Faso	2006-2017	Burundi	1983-2017
Cabo Verde	2002-2017	Cambodia	1995-2017
Cameroon	1991-2017	Canada	1983-2008
Central African Republic	1991-2017	Canada	1983-2008
Chad	1991-2017	Chile	1986-2017
China	1983-2017	Colombia	1983-2017
Comoros	2000-2017	Costa Rica	1983-2017
Cote d'Ivoire	2006-2017	Croatia	1997-2014
Czech Republic	1996-2017	Democratic Republic of the Congo	2007-2017
Dominica	1983-2017	Dominican Republic	1992-2017
Ecuador	2008-2017	Egypt	1983-2017
El Salvador	1984-2000	Equatorial Guinea	1991-2017
Ethiopia	1986-2008	Fiji	1992-2017
Gabon	1991-2017	Gambia	1983-2014
Georgia	1998-2017	Ghana	1983-2016
Grenada	1983-2017	Guatemala	1998-2017
Guinea	1993-2016	Guinea-Bissau	2006-2017
Guyana	1983-2017	Haiti	1995-2017
Honduras	1983-2017	Hong Kong	1992-2016
Hungary	1996-2017	Iceland	1997-2017
India	1983-2017	Indonesia	1987-2017
Israel	1996-2017	Jamaica	1983-2016
Japan	1983-2016	Jordan	1983-2016
Kenya	1983-2017	Korea	1983-2017
Kuwait	1983-2017	Kyrgyz Republic	1997-2016
Lao P.D.R.	1991-2010	Lebanon	1991-2017
Lesotho	1983-2017	Liberia	2002-2016
Libya	1991-2003	Macedonia	2001-2016
Madagascar	1990-2017	Malaysia	1983-2017
Maldives	1983-2017	Mali	2006-2017
Mauritania	1983-2016	Mauritius	1983-2017
Mexico	1983-2017	Moldova	1997-2017
Mongolia	1993-2017	Morocco	1983-2017
Mozambique	1995-2017	Myanmar	2001-2017
Namibia	1993-2017	Nepal	1983-2016
New Zealand	1991-2015	Nicaragua	1989-2017
Niger	2006-2017	Nigeria	1991-2016
Pakistan	1995-2017	Panama	1987-2008
Papua New Guinea	1984-2017	Paraguay	1990-2017
Peru	1983-2016	Philippines	1983-2017
Poland	1996-2006	Republic of Congo	1991-2017
Rwanda	1996-2016	Sao Tome and Principe	2002-2017
Senegal	2006-2017	Serbia	2000-2015
Seychelles	1983-2017	Sierra Leone	1983-2017
Singapore	1983-2017	South Africa	1983-2017
Sri Lanka	2002-2016	St. Kitts and Nevis	1983-2017
Suriname	1991-2017	Swaziland	1983-2017
Sweden	1993-2006	Switzerland	1983-2016
Tajikistan	2000-2017	Tanzania	1989-2015
Thailand	1983-2016	Togo	2006-2017
Trinidad and Tobago	1984-2017	Turkey	1989-2017
Uganda	1983-2015	Ukraine	1994-2017
United Kingdom	1983-2014	United States	1983-2016
Uruguay	1983-2017	Venezuela	1985-2013
Vietnam	1997-2017	Yemen	1996-2013
Zambia	1983-2017	Zimbabwe	1991-2005

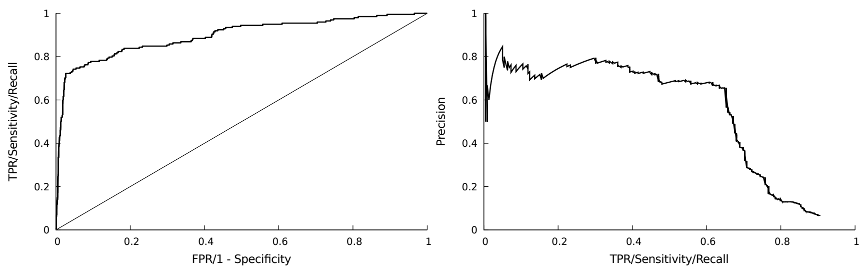
C ROC curves

Figure C.1: ROC and PR curves: MLE Pooled logit model



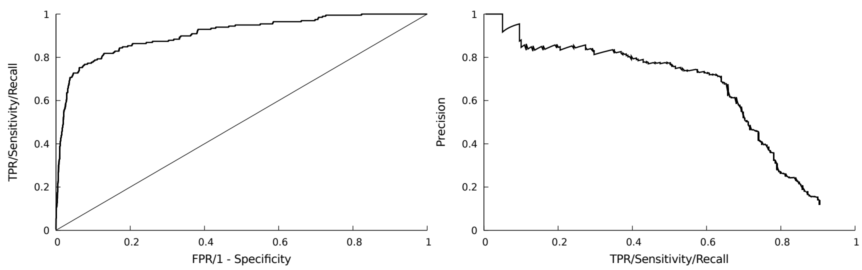
(a) ROC curve (b) PR curve

Figure C.2: ROC and PR curves: MLE RE logit model



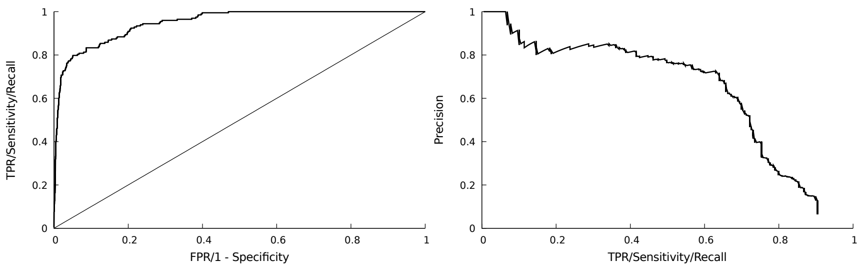
(a) ROC curve (b) PR curve

Figure C.3: ROC and PR curves: MLE FE logit model



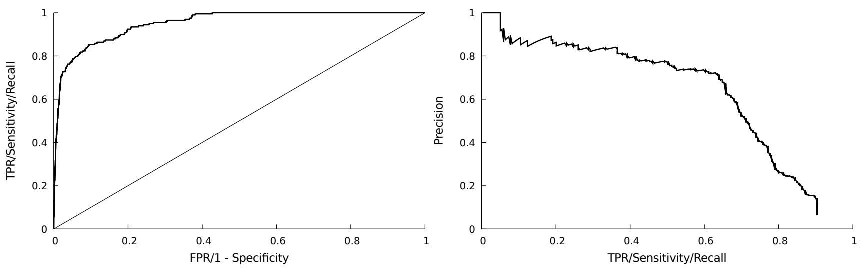
(a) ROC curve (b) PR curve

Figure C.4: ROC and PR curves: MLE FE* logit model



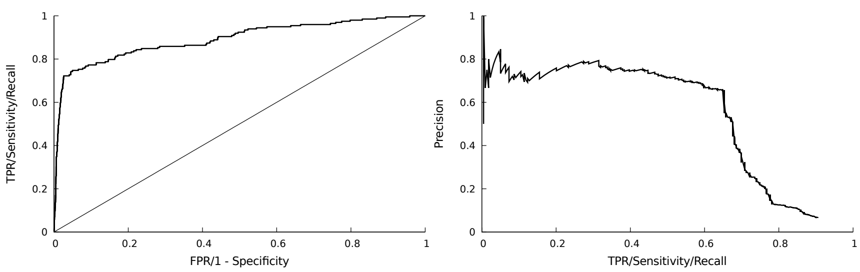
(a) ROC curve (b) PR curve

Figure C.5: ROC and PR curves: PMLE FE logit model



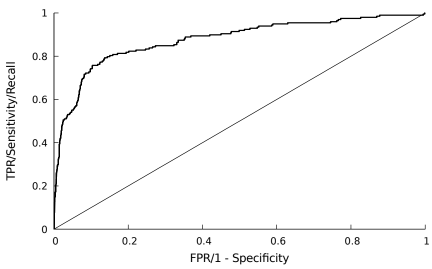
(a) ROC curve (b) PR curve

Figure C.6: ROC and PR curves: MLE Pooled probit model

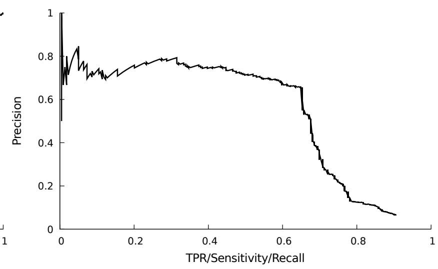


(a) ROC curve (b) PR curve

Figure C.7: ROC and PR curves: MLE Pooled multinomial logit model



(a) ROC curve



(b) PR curve