Prediction of early warning crises by a hidden Markov model with covariates

Luca Brusa[1], Fulvia Pennoni[1], Francesco Bartolucci[2], and Romina Peruilh B.[2]

[1] University of Milano-Bicocca, Milano 20126, Italy,
luca.brusa@unimib.it, fulvia.pennoni@unimib.it
[2] University of Perugia, Perugia 06123, Italy,
francesco.bartolucci@unipg.it, romina.peruilh@unipg.it

Abstract. We propose an early warning system for financial crisis prediction tailored to longitudinal data with missing values and time-varying covariates. The proposed system is based on a hidden Markov model, which includes selected time-varying economic drivers and the lagged response variable, thus relaxing the local independence assumption. Partially missing outcomes at a given time are considered under the missing-at-random assumption, and partially missing values on the covariates are accounted for by dummy indicators. We study in-sample and out-of-sample model performance in terms of forecasting, considering an application related to country-level financial crises.

Keywords: discrete latent variable models, financial crises, maximum likelihood estimation, missing values

1 Early warning systems

Early warning systems are nowadays relevant in many research areas. These systems are aimed at timely classifying units that are at risk of certain types of failure. Since the pioneering work of [6], hidden Markov (HM) models [2] have been proposed to predict future observations. Considering heterogeneous dynamics between subjects, these models provide learning, decoding, and prediction. Motivated by macroeconomic data referred to countries that experienced financial crises as rare events, we illustrate an HM model to study the influence of key indicators in the measurement (sub-)model and the lagged response variables. We also account for missing indicators through dummies to avoid changes in the forecast horizon. We explore the capability of the HM model to interpret signals of crises, to identify important predictors of crises, and to produce in-sample and out-of-sample forecasts.

The remainder of this paper is structured as follows. Section 2 describes the model and outlines the forecasting methodology. Section 3 shows the results of the model applied to the financial crises dataset, mainly focusing on the predictive performance evaluation. Section 4 reports main conclusions.

2 Hidden Markov model with covariates and lagged dependence

Dealing with longitudinal data, let $\boldsymbol{Y}_i = (Y_i^{(1)}, \ldots, Y_i^{(T)})$ denote the univariate binary response variables, where $Y_i^{(t)} = 1$ if the event of interest is observed at time t for unit i and $Y_i^{(t)} = 0$ otherwise. Additionally, let $\boldsymbol{x}_i = (\boldsymbol{x}_i^{(1)}, \ldots, \boldsymbol{x}_i^{(T)})$, with $\boldsymbol{x}_i^{(t)}$ representing the vector of individual covariates at time t. The HM model is based on a hidden process $\boldsymbol{U}_i = (U_i^{(1)}, \ldots, U_i^{(T)})$, following a first-order Markov chain with state-space $\{1, \ldots, k\}$. The response variables are generally assumed to be conditionally independent given $\boldsymbol{U}_i$.

The model is characterized by (i) the measurement model $p_{\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{X}_i}(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{x})$, representing the distribution of the response vector $\boldsymbol{Y}_i$ given the latent process $\boldsymbol{U}_i$, and (ii) the latent model $p_{\boldsymbol{U}_i}(\boldsymbol{u})$. As clear from the two formulas above, covariates are here assumed to affect solely the measurement model, while the same latent model holds for all units, so as to account for unobserved heterogeneity between these units. In such a way, we postulate that there may be an effect of unobservable covariates, which may follow a specific dynamic over time. Considering data where the binary response variable indicate if a country had a financial crisis, through this model formulation we study possible effects of unobservable drivers on the crises. Moreover, we include in each vector of covariates $\boldsymbol{x}_i^{(t)}$ the lagged response variable, so as to allow for serial dependence between observed responses over time. The model parameters are:

1. the initial and transition probabilities, denoted as π_u and $\pi_{u|\bar{u}}$, $\bar{u}, u = 1, \ldots, k$, respectively;

2. the conditional response probabilities, given the latent state and the covariate configuration, denoted as $\phi_{u\boldsymbol{x}}^{(t)}$.

Following [1], a logistic parameterization is employed for the measurement model:

$$\log \frac{\mathbb{P}(Y_i^{(t)} = 1 | U_i^{(t)} = u, \boldsymbol{X}_i^{(t)} = \boldsymbol{x})}{\mathbb{P}(Y_i^{(t)} = 0 | U_i^{(t)} = u, \boldsymbol{X}_i^{(t)} = \boldsymbol{x})} = \log \frac{\phi_{u\boldsymbol{x}}^{(t)}}{1 - \phi_{u\boldsymbol{x}}^{(t)}} = \mu + \alpha_u + \boldsymbol{x}'\boldsymbol{\beta}, \qquad (1)$$

for $t = 1, \ldots, T$ and $u = 1, \ldots, k$. In this way the model extends the dynamic logit model proposed in [4], since μ is the intercept, $\alpha_1, \ldots, \alpha_k$ are specific support points corresponding to the latent states, and $\boldsymbol{\beta}$ is the vector of regression parameters for the covariates.

In the following, we refer to the manifest distribution of the response variables given the covariates as

$$p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{y}|\boldsymbol{x}) = \sum_{\boldsymbol{u}} p_{\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{X}_i}(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{x}) p_{\boldsymbol{U}_i}(\boldsymbol{u}),$$

and to the posterior distribution of the latent variables given the responses and the covariates as

$$q_{\boldsymbol{U}_i|\boldsymbol{X}_i,\boldsymbol{Y}_i}(\boldsymbol{u}|\boldsymbol{x},\boldsymbol{y}) = \frac{p_{\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{X}_i}(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{x}) p_{\boldsymbol{U}_i}(\boldsymbol{u})}{p_{\boldsymbol{Y}_i|\boldsymbol{X}_i}(\boldsymbol{y}|\boldsymbol{x})}.$$

Maximum likelihood estimation of the model parameters, collected in the vector $\boldsymbol{\theta}$, is performed through the expectation-maximization (EM) algorithm [3] that relies on the complete-data log-likelihood function

$$\ell^*(\boldsymbol{\theta}) = \log p_{\boldsymbol{Y}_i,\boldsymbol{U}_i|\boldsymbol{X}_i}(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{x}) = \log p_{\boldsymbol{Y}_i|\boldsymbol{U}_i,\boldsymbol{X}_i}(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{x}) + \log p_{\boldsymbol{U}_i}(\boldsymbol{u}).$$

After a proper initialization of the model parameters, the EM algorithm alternates the following two steps until convergence: (i) expectation step, which computes the conditional expected value of $\ell^*(\boldsymbol{\theta})$ given the observed data and the estimates of the parameters at the previous step, and (ii) maximization step, which maximizes the expectation of $\ell^*(\boldsymbol{\theta})$ and so updates the value of the parameters. The expectation step, which is based on the posterior distribution $p_{\boldsymbol{U}_i|\boldsymbol{Y}_i}(\boldsymbol{u}|\boldsymbol{y})$, is computationally unfeasible and requires suitable recursions; see [8] for more details.

2.1 Forecasting

Given the estimated model, and relying on the parameterization expressed in Equation (1), the estimated conditional response probabilities $\phi_{u\boldsymbol{x}}^{(t)}$ are computed as

$$\hat{\phi}_{u\boldsymbol{x}}^{(t)} = \frac{\exp(\hat{\mu} + \hat{\alpha}_u + \boldsymbol{x}'\hat{\boldsymbol{\beta}})}{1 + \exp(\hat{\mu} + \hat{\alpha}_u + \boldsymbol{x}'\hat{\boldsymbol{\beta}})},$$

for time $t = 1, \ldots, T$, latent state $u = 1, \ldots, k$, and any possible covariate configuration $\boldsymbol{x}$.

In-sample forecasting To perform in-sample prediction, we estimate the HM model using all the available data and subsequently forecast the occurrence of the event of interest for every unit $i = 1, \ldots, n$ and time $t = 1, \ldots, T$. To this aim, the probability $p_i^{(t)}$ of a crisis is estimated as a weighted average of the conditional probabilities $\hat{\phi}_{u\boldsymbol{x}}^{(t)}$ with weights equal to the estimated posterior distribution $\hat{q}(u|\boldsymbol{x}_i, \boldsymbol{y}_i)$ of the latent variable $U_i^{(t)}$ given the responses $\boldsymbol{y}_i$ and the covariates $\boldsymbol{x}_i$ of subject i:

$$\hat{p}_i^{(t)} = \sum_{u=1}^{k} \hat{q}^{(t)}(u|\boldsymbol{x}_i, \boldsymbol{y}_i)\ \hat{\phi}_{u\boldsymbol{x}}^{(t)}.$$

The choice of a suitable threshold $c \in [0, 1)$ to forecast the crisis is often based on the receiver operating characteristics (ROC) curve, through the Yuoden's J statistics, or on the precision-recall (PR) curve, through the so called F1 score. We recall that a crisis is predicted if $\hat{p}_i^{(t)} > c$.

Out-of-sample forecasting To perform out-of-sample forecasts, data are restricted to a temporal period spanning from $t = 1$ to $t = t^*$, with $t^* < T$. We estimate the proposed model on this restricted dataset and forecast the occurrence of a crisis for each unit at time $t^* + 1$. In this case, the probability $p_i^{(t^*+1)}$ is estimated, similarly to the previous case, as

$$\hat{p}_i^{(t^*+1)} = \sum_{u=1}^{k} \hat{q}^{(t^*+1)}(u|\boldsymbol{x}_i, \boldsymbol{y}_i)\ \hat{\phi}_{u\boldsymbol{x}}^{(t^*+1)},$$

once c is chosen according to the measures illustrated above. Here we have that

$$\hat{q}^{(t^*+1)}(u|\boldsymbol{x}_i, \boldsymbol{y}_i) = \sum_{\bar{u}=1}^{k} \hat{\pi}_{u|\bar{u}}\ \hat{q}^{(t^*)}(\bar{u}|\boldsymbol{x}_i, \boldsymbol{y}_i).$$

3 Application to country-level data on financial crises

Data motiving the applicative example are analyzed in [7] and refer to annual macroeconomic (real GDP growth rate, logarithm of the per-capita GDP, inflation and real interest rate), monetary (broad money over foreign exchange reserves, and growth of private credit), and financial (growth rate of net foreign assets to GDP) measurements for 129 developed countries between 1983 and 2017. For each country-year observation, data contain a binary variable indicating whether or not the country suffered from a financial crisis in a particular year. Overall, in the dataset we observe 230 crises over 4,415 records. See [5] for more details on the definition of the type of financial crisis.

Data are unbalanced, with the number of available observations considerably varying across years. Missing values of the key indicators are set to 0 and are handled by dummy variables serving as missing indicators; in this way, we

avoid excluding of observations and we can evaluate the informativeness of the covariates with missing observations.

Due to space constraints, in this article only the results of the models evaluations in forecasting terms are given below.

3.1 In-sample forecasting

We estimated the HM model on the whole data considering a number of latent components k ranging from 1 to 4. To mitigate the risk of convergence to local maxima, we repeated 25 times the estimation of each model, employing both deterministic and random initialization methods. Inference is subsequently based on the solution corresponding to the highest likelihood value at convergence.

Table 1. Model selection in terms of Akaike and Bayesian information criteria of the HM models for k ranging from 1 to 4. Crisis prediction and false alarms for the threshold based on the Youden's J statistics and the F1 score

	$k = 1$	$k = 2$	$k = 3$	$k = 4$
AIC	1060.77	1059.69	1045.64	1041.06
BIC	1095.09	1102.59	1102.83	1118.28
			ROC curve	
Threshold (c)	0.03	0.06	0.08	0.38
Youden's J	0.65	0.98	0.99	1.00
Predicted crises	170	227	227	227
False alarms	403	76	35	0
			PR curve	
Threshold (c)	0.21	0.17	0.19	0.38
F1 score	0.66	0.91	0.95	1.00
Predicted crises	150	203	208	227
False alarms	81	14	2	0

Table 1 reports values of the Akaike and Bayesian information criteria for the models estimated with each value of k, as well as the thresholds computed according to both Youden's J statistic and F1 score, and the obtained results in terms of in-sample crisis prediction and false alarms for both thresholds.

Looking at Table 1, we first observe that the best performance in terms of forecast is obtained with $k = 4$, which is also the model suggested by the Akaike information criterion. In this case, the two approaches based on the ROC and PR curves provide the same threshold $c = 0.38$. The corresponding model correctly forecasts all banking crises, and no false alarms are provided.

3.2 Out-of-sample forecast

Following the approach proposed in [7], the dataset is first restricted to the period from 1983 to 2006 and used for in-sample model selection and estimation, with k ranging from 1 to 4; the estimated parameters are then used to forecast the probability of crises for the year 2007. Subsequently, the data are augmented with the observations from 2007, the model is estimated on the resulting data, and the estimates are used to forecast the crisis probabilities one year ahead. The whole procedure is repeated for each subsequent year up to 2017.

Table 2. Number of correctly predicted crises and false alarms obtained with the out-of-sample forecast procedure for the years from 2007 to 2017

Year	Total crises	Predicted (%)	False alarms
2007	2	0 (0.00)	0
2008	7	2 (28.57)	0
2009	8	7 (87.50)	0
2010	6	6 (100.00)	2
2011	5	5 (100.00)	1
2012	3	3 (100.00)	2
2013	0	0 (-)	3
2014	3	0 (0.00)	0
2015	3	3 (100.00)	0
2016	3	3 (100.00)	0
2017	3	3 (100.00)	0
Total	43	32 (74.42)	8

Table 2 reports the number of correctly predicted crises and false alarms for each year. The proposed approach demonstrates a high level of accuracy. In fact, it forecasts approximately three-quarters of banking crises occurring between 2007 and 2017 (32 out of 43). The number of false alarms is exceedingly low. In particular, the proposed approach achieves a perfect prediction rate for crises that were already present in the previous time period. On the contrary, it encounters difficulties in forecasting first-ever crisis events. Interestingly, the same behavior persists even when excluding the lagged crisis variable from the set of predictors.

4 Conclusions

The proposed HM model may constitute a simple and interpretable alternative for early warning systems to the most common machine learning methods, which generally guarantee higher predictive performance but their results are often difficult to interpret. Additional research will explore possible non-linearities and interactions among key indicators and penalized likelihood methods which can allow gains in predictive accuracy.

Acknowledgment. The authors acknowledge the financial support from the grant "Hidden Markov Models for Early Warning Systems" of Ministero dell'Università e della Ricerca (PRIN 2022TZEXKF) funded by European Union - Next Generation EU, Mission 4, Component 2, CUP J53D23004990006.

References

1. Bartolucci, F., Farcomeni, A.: A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. J. Am. Stat. Assoc. 104, 816-831 (2009)
2. Bartolucci, F., Farcomeni, A., Pennoni, F.: Latent Markov Models for Longitudinal Data. Chapman and Hall/CRC, Boca Raton, FL (2013)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. R. Stat. Soc. Ser. B 39, 1-38 (1977)
4. Hsiao, C.: Analysis of panel data. Cambridge University Press, New York (2005)
5. Laeven, L., Valencia, F.: Systemic Banking Crises Revisited. IMF Working Papers, International Monetary Fund (2018)
6. Levinson, S. E., Rabiner, L. R., Sondhi, M. M.: An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. Bell Labs Tech. J. 62, 1035-1074 (1983)
7. Pigini, C.: Penalized maximum likelihood estimation of logit-based early warning systems. Int. J. Forecast. 37, 1156-1172 (2021)
8. Welch, L.R.: Hidden Markov models and the Baum-Welch algorithm. IEEE Inform. Theory Soc. Newsl. 53, 10-13 (2003)