



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



# PREDICTION OF EARLY WARNING CRISES BY A HIDDEN MARKOV MODEL WITH COVARIATES

---

LUCA BRUSA<sup>1</sup>, FULVIA PENNONI<sup>1</sup>,  
FRANCESCO BARTOLUCCI<sup>2</sup>, ROMINA PERUILH BAGOLINI<sup>2</sup>  
([luca.brusa@unimib.it](mailto:luca.brusa@unimib.it))

<sup>1</sup>University of Milano-Bicocca - Department of Statistics, and Quantitative Methods

<sup>2</sup>University of Perugia - Department of Economics

# Outline

- 1 Early warning systems
- 2 Hidden Markov models
- 3 Forecasting with the HM model
- 4 Application
- 5 References

# Early warning systems

- An **early warning system (EWS)** may be defined as any system of biological or technical nature that helps to assess, detect, and prevent hazards and failures in different fields
- Two basic **features** of EWS:
  - stable relationship between failures and a set failure drivers
  - these failure drivers can be identified in advance
- We propose a statistical model for EWS tailored to longitudinal data with missing values and time-varying covariates
- The proposal is related to a **hidden Markov (HM)** model to predict financial crises, with time-varying economic drivers and the lagged response variable

# Outline

- 1 Early warning systems
- 2 Hidden Markov models**
- 3 Forecasting with the HM model
- 4 Application
- 5 References

# Hidden Markov model: formulation for EWSs

- **Univariate binary response variables**  $\mathbf{Y}_i = (Y_i^{(1)}, \dots, Y_i^{(T)})$ , with

$$Y_i^{(t)} = \begin{cases} 1 & \text{if the financial crisis is observed at time } t \text{ for unit } i \\ 0 & \text{otherwise} \end{cases}$$

- **Time-varying covariates:**  $\mathbf{x}_i = (\mathbf{x}_i^{(1)}, \dots, \mathbf{x}_i^{(T)})$ , with  $\mathbf{x}_i^{(t)}$  representing the vector of observed individual covariates for unit  $i$  at time  $t$
- **Hidden process:**  $\mathbf{U}_i = (U_i^{(1)}, \dots, U_i^{(T)})$ , following a first-order Markov chain with state-space  $\{1, \dots, k\}$

# Model formulation

- **Measurement model:**  $p\left(y_i^{(t)} \mid u_i^{(t)}, \mathbf{x}_i^{(t)}, y_i^{(t-1)}\right)$ 
  - represents the conditional distribution of the response variable  $Y_i^{(t)}$  given the latent process  $U_i^{(t)}$ , with covariates  $\mathbf{x}_i^{(t)}$  and lagged response variables  $Y_i^{(t-1)}$
  - covariates directly influence the response variable
  - the lagged response among covariates allows for serial dependence between observed responses over time, thus relaxing the conditional independence of  $\mathbf{Y}$  and  $\mathbf{U}$  given  $\mathbf{x}$
- **Latent model:**  $p(\mathbf{u}_i)$ 
  - represents the distribution of the latent process
  - is not affected by covariates: the same latent model holds for all units
  - accounts for unobserved heterogeneity between individuals, which remains when observed covariates in the measurement model cannot fully explain the variability

# Model parameters

- 1 the **initial** and **transition probabilities**, denoted as  $\pi_u$  and  $\pi_{u|\bar{u}}$ , respectively
- 2 the **conditional response probabilities**, given the latent state and the covariate configuration, denoted as  $\phi_{ux}^{(t)}$ , such that:

$$\log \frac{\mathbb{P}\left(Y_i^{(t)} = 1 | u_i^{(t)}, \mathbf{x}_i^{(t)}, y_i^{(t-1)}\right)}{\mathbb{P}\left(Y_i^{(t)} = 0 | u_i^{(t)}, \mathbf{x}_i^{(t)}, y_i^{(t-1)}\right)} = \log \frac{\phi_{ux}^{(t)}}{1 - \phi_{ux}^{(t)}} = \mu + \alpha_u + \mathbf{x}_i' \boldsymbol{\beta} + \gamma y_i^{(t-1)}$$

- $\mu$ : intercept
- $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$ : support points corresponding to the latent states
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ : regression parameters for the covariates
- $\gamma$ : parameter for the lagged response variable

# Maximum likelihood estimation

- **Expectation-maximization** (EM) algorithm (Dempster et al., 1977) is usually employed to perform full **maximum likelihood estimation** (**MLE**) of discrete latent variable models
- It maximizes the observed-data log-likelihood function  $\ell(\boldsymbol{\theta})$  relying on the **complete data log-likelihood** function  $\ell^*(\boldsymbol{\theta})$
- It alternates the following steps until convergence:
  - **E-step: compute the conditional expected value** of  $\ell^*(\boldsymbol{\theta})$  given the value of the parameters at the previous step and the observed data; it relies on the posterior distribution  $q(\mathbf{u}_i | \mathbf{x}_i, \mathbf{y}_i)$
  - **M-step: update the model parameters** by maximizing the expected value of  $\ell^*(\boldsymbol{\theta})$

# Outline

- 1 Early warning systems
- 2 Hidden Markov models
- 3 Forecasting with the HM model
- 4 Application
- 5 References

# In-sample forecasting

	Units	Time occasions
Model estimation	$i = 1, \dots, n$	$t = 1, \dots, T$
Forecasting	$i = 1, \dots, n$	$t = 1, \dots, T$

## In-sample estimated crisis probability

The probability  $p_i^{(t)}$  of a crisis for unit  $i$  at time  $t$  is computed as

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

- $\hat{\phi}_{u\mathbf{x}}^{(t)}$ : estimated conditional probabilities at time  $t$
- $\hat{q}^{(t)}(u | \mathbf{x}_i, \mathbf{y}_i)$ : estimated posterior distribution of the latent variable  $U_i^{(t)}$  given the responses  $\mathbf{y}_i$  and the covariates  $\mathbf{x}_i$

# Out-of-sample forecasting

	Units	Time occasions
Model estimation	$i = 1, \dots, n$	$t = 1, \dots, t^* \ (t^* < T)$
Forecasting	$i = 1, \dots, n$	$t = t^* + 1$

## Out-of-sample estimated crisis probability

The probability  $p_i^{(t^*+1)}$  of a crisis for unit  $i$  at time  $t^* + 1$  is computed as

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

- $\hat{\phi}_{ux}^{(t^*+1)}$ : estimated conditional probabilities at time  $t^* + 1$
- $\hat{q}^{(t^*+1)}(u|\mathbf{x}_i, \mathbf{y}_i) = \sum_{\bar{u}=1}^k \hat{\pi}_{u|\bar{u}} \hat{q}^{(t^*)}(\bar{u}|\mathbf{x}_i, \mathbf{y}_i)$ : estimated posterior distribution of the latent variable  $U_i^{(t^*+1)}$  given the responses  $\mathbf{y}_i$  and the covariates  $\mathbf{x}_i$

# Choice of the cutoff

- In both cases, the choice of a suitable **cutoff**  $c \in [0, 1)$  to forecast the crisis is based:
  - either on the **Receiver Operating Characteristics** (ROC) curve, through the **Yuoden's J statistics**
  - or on the **precision-recall** (PR) curve, through the **F1 score**
- A crisis is finally predicted if

$$\hat{p}_i^{(t)} > c \quad \text{or} \quad \hat{p}_i^{(t^*+1)} > c$$

# Outline

- 1 Early warning systems
- 2 Hidden Markov models
- 3 Forecasting with the HM model
- 4 Application**
- 5 References

# Financial crisis of countries

- Data refer to **129 countries** and cover the period **from 1983 to 2017**
- For each country-year observation, a binary variable indicates whether or not the country has experienced a **financial crisis** (overall we observe 227 crises over 4,415 records)
- Time-varying covariates are lagged by one period and belong to these categories:
  - **macroeconomic variables**: real GDP growth rate, logarithm of the per-capita GDP, inflation and real interest rate
  - **monetary variables**: broad money over foreign exchange reserves, and growth of private credit
  - **financial variables**: growth rate of net foreign assets to GDP

# Feature of the data

- Data are **unbalanced**, with the number of available observations considerably varying across years
- Partially **missing outcomes** at a given time are considered under the missing-at-random assumption
- Partially **missing values on the covariates** are set to 0 and are handled by dummy variables serving as missing indicators. This allows us to evaluate the informativeness of the missing observations

# Estimation settings

	Training		Test
In-sample	$t = 1983, \dots, 2016$	$\rightarrow$	$t = 1983, \dots, 2016$
Out-of-sample	$t = 1983, \dots, 2006$	$\rightarrow$	$t = 2007$
	$t = 1983, \dots, 2007$	$\rightarrow$	$t = 2008$
	$\dots$		$\dots$
	$t = 1983, \dots, 2016$	$\rightarrow$	$t = 2017$

- Each HM model is estimated considering a number of latent components  **$k$  ranging from 1 to 4**
- The estimation of each HM model is repeated 25 times, employing both deterministic and random initialization methods

# In-sample forecasting

Model selection in terms of AIC criterion 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
<i>ROC curve</i>				
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
<i>PR curve</i>				
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

# Out-of-sample forecasting

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
<b>Total</b>	<b>43</b>	<b>32 (74.42)</b>	<b>8</b>

# Conclusions

- The proposed HM model constitutes a **simple and interpretable alternative** for early warning systems to machine learning methods (higher predictive performance but results difficult to interpret)
- The application reveals that the HM model with covariates and  $k = 4$  latent components yields the most **accurate in-sample forecasts**, effectively **predicting all banking crises**, with **no false alarms**
- **Out-of-sample forecasting** provides a high level of accuracy, correctly **predicting approximately three-quarters (32 out of 43) of banking crises** occurring between 2007 and 2017, with minimal false alarms

# Outline

- 1 Early warning systems
- 2 Hidden Markov models
- 3 Forecasting with the HM model
- 4 Application
- 5 References

# References I

- BARTOLUCCI, F. AND FARCOMENI, A. (2009). 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.
- BARTOLUCCI, F., FARCOMENI, A., AND PENNONI, F. (2013). *Latent Markov models for longitudinal data*. Chapman & Hall/CRC, Boca Raton.
- BRADLEY, A. P. (1997). The use of the area under the curve in the evaluation of machine learning algorithms. *Pattern recognit.*, **30**, 1145–1159.
- DEMPSTER, A., LAIRD, N., AND RUBIN, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Stat. Soc. Series B Stat. Methodol.*, **39**, 1–38.
- LAEVEN, L. AND VALENCIA, F. (2018). *Systemic Banking Crises Revisited*. IMF Working Papers, International Monetary Fund.

## References II

- LEVINSON, S. E., RABINER, L. R., AND SONDHI, M. M. (1983). 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.
- PIGINI, C. (2021). Penalized maximum likelihood estimation of logit-based early warning systems. *Int. J. Forecast.*, **37**, 1156–1172.
- SAMBRIDGE, M. (2005). *Hsiao, C.* Cambridge University Press, New York.
- WELCH, L. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Inform. Theory Soc. Newsl.*, **50**, 10–13.

# Acknowledgments

**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.

# Estimated model parameters

Covariate	Estimate	Standard Error	<i>p</i> -value
Crisis	11.574	2.787	< 0.0001
Growth of real GDP	-0.006	0.056	0.9185
Log per capita GDP	-0.228	0.164	0.1638
Real interest rate	0.066	0.026	0.0100
M2 to foreign exchange reserves	0.040	0.021	0.0572
Inflation	-0.002	0.002	0.2623
Growth of real domestic credit	0.054	0.020	0.0085
Growth of net foreign assets to GDP	-3.845	2.028	0.0580
Auxiliary binary variable 1	-3.384	1.449	0.0196
Auxiliary binary variable 2	1.693	0.785	0.0310
Support point latent state 1	-7.004		
Support point latent state 2	3.838		
Support point latent state 3	14.715		
Support point latent state 4	15.127		

# Estimated conditional crisis probability

