BANK OF GREECE
EUROSYSTEM

Working Paper

A ternary-state early warning system for the European Union

Savas Papadopoulos
Pantelis Stavroulias
Thomas Sager
Etti Baranoff

222

APRIL 2017

BANK OF GREECE
Economic Analysis and Research Department – Special Studies Division
21, E. Venizelos Avenue
GR-102 50 Athens
Tel: +30210-320 3610
Fax: +30210-320 2432

www.bankofgreece.gr

Printed in Athens, Greece
at the Bank of Greece Printing Works.

ISSN 1109-6691

A TERNARY-STATE EARLY WARNING SYSTEM FOR THE EUROPEAN UNION

Savas Papadopoulos
Bank of Greece

Pantelis Stavroulias
Democritus University of Thrace

Thomas Sager
University of Texas

Etti Baranoff
Virginia Commonwealth University

ABSTRACT

The global financial crisis of 2007-2008 focused the attention of financial authorities on developing methods to forecast and avoid future financial crises of similar magnitude. We contribute to the literature on crisis prediction in several important ways. First, we develop an early warning system (EWS) that provides 7-12 quarters advance warning with high accuracy in out-of-sample testing. Second, the EWS applies region-wide to the leading economies in the European Union. Third, the methodology is transparent – utilizing only publicly available macro-level data and standard statistical classification methodology (multinomial logistic regression, discriminant analysis, and neural networks). Fourth, we employ two relatively novel methodological innovations in EWS modeling: ternary state classification to guarantee a minimum advance warning period, and a fitting and evaluation criterion (the total harmonic mean) that prioritizes avoiding classification errors for the relatively infrequent events of most interest. As a consequence, a policymaker who uses these methods will enjoy a high probability that future crises will be signaled well in advance and that warnings of crisis will not be false alarms.

Keywords: Banking crisis; financial stability; macroprudential policy; classification methods; goodness-of-fit measures.

JEL-classifications: C53; E58; G28

Acknowledgement: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Correspondence:
Savas Papadopoulos
Bank of Greece
Department of Financial Stability
10250 Athens, Greece
Tel.:0030-210-3205106
Email: sapapa@bankofgreece.gr

1. Introduction

Financial crises are historically associated with the "four deadly D's": downturns, deficits, debt, and downgrades. Sharp economic downturns follow banking crises; fiscal deficits mount as enterprise revenues decline; rising debt follows deficits; rating downgrades ensue as debt accumulates. For the least fortunate enterprises, the crisis leads to a fifth and deadliest "D": default. The global financial crisis of 2007–08 played a significant role in the failure of key businesses, declines in consumer wealth estimated in trillions of U.S. dollars, and a downturn in economic activity leading to the 2008–2012 global recession and contributing to the European sovereign-debt crisis. Many economists consider the 2007-08 crisis to have been the worst since the Great Depression of the 1930s. The 2007-08 crisis almost brought down the world's financial system. The collapse of most large financial institutions was prevented by government bailouts, but stock markets still plunged worldwide. In many areas, the housing market suffered severely, fed by substantial declines in home values and prolonged unemployment, and resulting in evictions and foreclosures.

Following the Financial Crisis of 2007-08, the G-20 (the 20 developed nations with the largest economies) charged the Financial Stability Board (FSB) with developing a regulatory system that would avoid similar systemic crises in the future. Out of this developed an emphasis on understanding systemic risk and identifying the institutions that may contribute most strongly to it. The FSB defined systemic risk (BIS, FSB and IMF, 2009) as "*The risk of disruption to the flow of financial services that is (i) caused by an impairment of all or parts of the financial system; and (ii) has the potential to have serious negative consequences for the real economy.*" It is evident that this directive intends to ensure financial stability in terms of the whole financial system as well as the overall economies. The FSB also defined systemically important financial institutions (SIFIs) as: "*Firms whose disorderly failure, because of their size, complexity, and systemic interconnectedness, would cause significant disruption to the wider financial system and economic activity.*" [1] The FSB designated 29 banks as SIFIs in 2012 and followed with 9 insurers in 2013. These institutions would be subject to additional supervision and scrutiny.

[1] IMF, BIS and FSB (2009) *Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations*. Available at http://www.bis.org/publ/othp07.pdf

The need for macro-level supervision was also recognized. In a speech at the Federal Reserve Bank of Chicago Conference on Bank Structure and Competition on May 7, 2009, Federal Reserve Chairman Ben Bernanke concluded that the best way to avoid future crises is to improve supervision of the financial sector. A more macroprudential approach to supervision – one that supplements the supervision of individual institutions with mitigation of risks to the financial system as a whole – could help to enhance overall financial stability. The regulatory system must include the capacity to monitor, assess, and, if necessary, address potential systemic risks within the financial system.

Despite the potential importance of individual institutions like the SIFIs for triggering future crises, the literature on EWS focuses on prediction of macro-level crisis for an entire economy, rather than on prediction of micro-level crisis for individual institutions. A separate body of literature, often known as insolvency prediction, focuses on the latter. However, many of the methodologies of the two literatures overlap. Statistical methods in common to the two fields include logistic regression, discriminant analysis, neural networks, and other machine learning methodologies. In this paper, we follow convention and develop an EWS for a severe macro-level crisis for the entire economy.

Although excruciatingly painful, the crisis of 2007-08 provides a natural experiment for testing crisis prediction and avoidance methodologies. The development of early warning systems (EWS) for the prediction and avoidance of future crises on the scale of 2007-08 is clearly of critical importance. Despite the proliferation of studies on such systems since 2008, the area remains an open and fertile field for research. The academic literature reflects a variety of approaches and foci. For a historical review of EWS and for a classification of crisis generations see Kaminsky (2006). For a comparison of EWS methods before 2008 see Davis and Karim (2008). The following comments briefly summarize some of the literature that is most relevant for this article and that has appeared since the financial crisis of 2008. Misina and Tkacz (2009) consider the predictive ability of credit and asset for EWS and test the results in Canada, Japan and USA, including the 2008 FC. Männasoo and Mayes (2009) constructed an EWS for banks from Eastern Europe for the period 1995-2004. Borio and Drehmann (2009) applied existing research to the 2007-2008 crisis and found that it could be predicted by credit and asset price growth. They also found that property prices could provide an early signal for a banking crisis. Dungey et al. (2010) use stock and bond returns to predict financial crisis in Russia, Brazil, Argentina and USA. Büyükkarabacak et al. (2010) show that private credit is a strong predictor and enterprise credit is a weaker predictor than household but statistically and economically significant for crises before 2007. Demyanyk and Hasan (2010), summarize the methodologies and empirical results of various models “that attempt to explain, predict, or suggest remedies for financial crises or banking defaults”. Reinhart and Rogoff (2011) test the hypothesis that banking crises follow external debt inflation and give possible signals for sovereign debt crises. Davis, Karim and Liadze (2011) show that multivariate EWS for banking crises might not pool samples across regions. Yu Cao (2012) created aggregated multiple classification results by using Choquet integral for financial distress early warning. An updated database for systemic banking crises is presented by Laeven and Valencia (2012). Schularick and Taylor (2012) also show that credit growth is a strong early signal and comment on the impact of monetary policy on the financial cost of crises. Sarlin and Peltonen (2013) created detailed and compelling maps that show the state of financial stability. Babecký et al. (2013) studied EU and OECD countries, during the 1970-2010 period using panel VAR and model averaging techniques, and suggest that house and share prices, domestic and global variables warn of impending crises. A recent study by Behn et al. (2013) analyzed 23-EU countries between 1982 and 2012 and built a model with macro, market and bank variables at the local and global levels. Lo Duca and Peltonen (2013) also test a model with domestic and public predictors and construct a financial stress index that identifies the beginning of crises. Sarlin and Peltonen (2013) set up a visual EWS that classifies multi-categories of financial stability. Betz et al. (2014) construct an EWS for European banks within 2000-2013 with variables at the bank and country levels. Caggiano et al. (2014) follows a multinomial logit approach to early warning systems and systemic banking crises in low income countries. Lainà, Nyholm and Sarlin (2015) tested predictors for systemic bank crises for 11 EU countries within 1980-2013 and identified loans over deposits and house price growth as early warning quantities. Papadopoulos, Stavroulias and Sager (2016) develop a methodology that predict a crisis several quarters in advance for the same region and time window but preprocessing the data by moving averages and filtering and utilizing a binary dependent variable.

In this study, we provide a layered early warning model for the overall macro-economic system. Departing from the existing literature, we construct an Early Warning System (EWS) with three pre-crisis states instead of the customary two. We have an Early Warning Period

(EWP), extending from 7-12 quarters before the crisis erupts; a Late Warning Period (LWP), 1-6 quarters before the crisis; and a Normal Period (NP) for all other preceding pre-crisis quarters. Our implementation of a three-state classification has several advantages. Correctly identifying the current quarter as an EWP has two implications: (1) A crisis is definitely coming; but (2) the crisis lies at least 6 quarters in the future – so there should be time to take remedial measures. Correctly identifying the current quarter as an NP means there may or may not be a crisis in the future, but if there is, it is at least three years off. In conventional binary state EWS, the pre-crisis period immediately precedes the crisis. In that case, the crisis may occur as soon as the next time period – which may allow insufficient time for remedial action. Our LWP state corresponds most nearly to the event state in a binary EWS. But the LWP state is still valuable. If a policy maker misses predicting the early warning time window, he has a second chance to predict the later state: "Better late than never." Furthermore, prediction of a late warning period following prediction of an early warning period would reinforce confidence in the coming storm.

We employ three standard statistical methodologies in our predictive models: multinomial logistic regression, multiple discriminant analysis (MDA), and neural networks (NN). However, we adopt the innovative metric of total harmonic mean (THM) of misclassification rates for fitting and evaluating models. THM enjoys advantages in our case over more classical metrics, such as sensitivity and specificity. THM emphasizes avoidance of errors in classifying relatively less numerous state-events such as crises or pre-crisis periods.

Since we target national economies and their financial sectors as a whole, rather than individual banks, insurers or other financial institutions, our models utilize macro-level predictor variables that reflect overall conditions and that are not unique to one industry. All the predictive variables are quarterly, transformed into year-on-year growth rates to remove trend and seasonality. Since these variables are publicly available, our models are transparent and easily reproducible.

Our results in predicting future crises that may be similar to the crisis of 2007-08 are very strong. The NN models have the best in-sample accuracy (about 90%), but MDA gives the best out-of-sample accuracy (about 85%) by our classification metrics. MDA is also more consistent between in and out-of-sample classification success. MDA drops only from 90% in-sample to

85% out-of-sample, in contrast to NN 94% to 81%, respectively. Multinomial logistic regression is similar to NN.

2. Data

Our predictive models are regional – for the EU-15, less Luxembourg[2]; Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the United Kingdom – which we henceforth call the EU-14. We use quarterly data from the first quarter of 2001 through the last quarter preceding the start of the crisis in each country. There is a substantial literature on defining the start of a financial crisis, for example Babecky et al. (2013), Laeven and Valencia (2008, 2012), Laina, Nyholm, Sarlin (2015). For identifying the start of the crisis, we follow the guidance of Kaminsky and Reinhart (1999), Demirguc¸ Kunt and Detragiache (1998) and Eichengreen and Arteta (2000). For the UK, we set the start of the crisis as first quarter of 2007; for the rest of the countries, we set it as the first quarter of 2008.

2.1 Explanatory variables

Theory provides ample guidance for selection of crisis prediction variables for EWS models. In an effort to make our methodology and forecasts as transparent and reproducible as possible, we limited our predictive variables to data that are freely and publicly available. Our variables are obtained from Eurostat, the World Bank, the Organization for Economic Cooperation and Development, and the Bank for International Transactions. We selected 12 relevant macro-level variables suggested by the literature (Kaminsky and Reinhart, 1999; Borio and Lowe, 2002, 2004, and Alessi and Detken, 2011) and organized into four categories: macroeconomic, banking, market, and government and banking. Each variable is an aggregate national level quarterly time series, beginning with the first quarter of 2001 and extending through the last quarter preceding the crisis for each country:

Macroeconomic variables.

- GDP - Gross domestic product (€ billions)

[2] Luxemburg was excluded due to missing values of variables important for the study, such as 10-year bonds yields.

- UNE - unemployment rate
- CPI - consumer price index (index, 2005=100)
- CST - production in construction (index, 2010=100)
- LBC - real unit labor cost (index, 2005=100)

Banking variables.

- CRE - credit to the private sector (€ billions)

Market variables.

- EQU – equities/ stock market quarterly mean (points)
- BON - quarterly mean of 10-year bond yields (units)
- EXR - real effective exchange rate (index, 2005=100)

General government variables.

- GDE - general government consolidated gross debt (as percent of GDP)
- GRE - total general government revenue (as percent of GDP)
- GTE - total general government expenditure (as percent of GDP).

2.2 Pre-Processing the Data

The data require pre-processing before analysis can begin. Spurious predictive power can result from using predictors like population or consumer prices that tend to increase naturally over time. The highest levels of such variables would tend to occur immediately prior to the end of the time series, whether a crisis were to follow or not. Many time series also exhibit other regularities like seasonal cycles. The presaging of crisis may lie in obscure disturbances in the trend and seasonal regularities. Therefore, the time series must be adjusted for trend and seasonal regularities.

The data pre-processing consists of eliminating seasonal cycles and linear trend by calculating a 4-quarter (year-on-year) growth rate for each predictive variable. For a quarterly frequency variable $X_{i,t}$, in which *i* is the country and *t* is the quarter number, the annual growth rate (AGR) equals

$$AGR = \frac{X_{i,t} - X_{i,t-4}}{X_{i,t-4}}$$

Differencing ($X_{i,t} - X_{i,t-4}$) filters a linear trend and annual cycles. In particular, since most of the EU-14 financial crises occurred in 2008Q1, if there is a temporal trend in $X_{i,t}$, then spurious prediction of crisis at the end of the time series would result from use of unfiltered $X_{i,t}$ in a forecasting model. Dividing by $X_{i,t} - X_{i,t-4}$ scales the predictive variables for comparability.

As an illustration, Figure 1 shows time plots of AGR for the consumer price index for each of the EU-14 from 2001Q1 extended past the crisis through 2013Q1. The red broken vertical line marks the start of the crisis in each country. Up to the start of the crisis, many of the countries show an average level of about 0.02 for the AGR of CPI – that is an average inflation rate of about 2%. But each country also shows an acceleration in the AGR a few quarters before the crisis, followed by a substantial decline after the crisis. Our prediction models look to subtleties in the behavior of inflation, in combination with the AGR of the other 11 predictors, to presage the coming of crisis. The models use no data after the start of the crisis. Table 1 reports the summary statistics for the 12 explanatory variables.

2.3 Coding the three states

The three early warning states are coded by their proximity to the start of the crisis in each country. For all countries but the UK, the three states are coded as follows:

- LWP (2) - the late warning period: 2006Q3-2007Q4;
- EWP (1) - the early warning period: 2005Q1-2006Q2;
- NP (0) - the normal period: 2001Q1-2004Q4.

For the UK, LWP and EWP are shifted back in time by one year. However, the first 4 quarters of data for each country will be missing because of the AGR transformation of the predictor variables to year-on-year (4-quarter) growth rates. Altogether, there are 13x24 + 1x20 = 332 quarters to be classified, consisting of 84 EWPs, 84 LWPs and 164 NPs.

Many EWS try to predict the time periods when a crisis will occur. Others try to identify the pre-crisis period (e.g, Lo Duca and Peltonen, 2013). In order for a forecast to have operational value to a policy maker, the forecast must occur sufficiently in advance of the crisis to permit remedial or ameliorative actions. If an EWS merely classifies the current time period as crisis or not, a crisis warning will be of limited value, as there will be no advance time for preventive action. Better is an EWS that correctly identifies that the current period is within some advance time, like 6 quarters, of a crisis. However, even then there would be uncertainty about whether the crisis is in the immediate offing or still six quarters in advance. Very different policy options might be called for, depending upon the prospective immediacy of the crisis.

In this study as mentioned above, we adopt a ternary classification scheme that ensures at least six quarters early warning on correct calls. The early warning period (EWP) extends from 7-12 months prior to a crisis. If our EWS correctly calls the current quarter as an EWP, then the policy maker will have 6-12 months advance notice to deploy remediation. The period 1-6 quarters preceding a crisis is the late warning period (LWP). Periods prior to EWP are normal periods. Rather than trying to identify whether a period will be a crisis period, we shift the focus to identifying whether the current period occurs in the run-up to a crisis. Of course, our classification will be correct only if the crisis actually follows within 7-12 quarters. This time horizon comports with the announcement period of twelve months specified in the EU CRD IV (article 126(6)), as well as for an adequate time lag required to implement the necessary policies. Additionally, most publications of the Bank for International Settlements (BIS) specify that the necessary prediction period should be between one and five years before the onset of the actual crisis. Correctly calling a crisis in the LWP will still be valuable – just not as valuable as calling the crisis in the EWP. Although we have data through 2014, our models are developed only with pre-crisis data. We do not include any crisis or post-crisis quarters in our model development.

We hope that the NP, EWP, and LWP quarters will be economically and financially sufficiently distinctive from each other to correctly identify them. Typically, the further ahead in time the event, the more difficult it is to predict. The economic and financial disturbances that augur a crisis are more subtle the more distant the crisis. So there will be a trade-off between confidence in the accuracy of the prediction and allowance of policy time for prophylaxis. We find that for the 2008 crisis (and presumably for similar future crises) the coming of crisis can be forecast 7-12 quarters in advance with high accuracy. This should be sufficient advance warning for effective policy action.

3. Methodology and assessment

3.1 Evaluation of model signals

The general methodology for our EWS is multinomial statistical classification using a training data set of 332 observations (country-quarters). The specific statistical methodologies that we employ include multinomial logistic regression, multiple linear discriminant analysis, and neural networks. Each methodology will produce a probability function $P_{i,t}^{(k)}$ that estimates the probability that a quarter t for country i is in state k (NP, EWP, LWP). It is natural to classify a quarter as being in state k if the probability of state k is sufficiently high. Therefore, we seek threshold values τ_1 and τ_2 to define cutoffs for classifying quarters as EWP (1) or LWP (2), respectively. We classify quarter t for country i to be in state 1 (EWP) when $P_{i,t}^{(1)}$ is greater or equal to the threshold value τ_1 and at the same time $P_{i,t}^{(2)}$ is smaller than the threshold value τ_2, to be in state 2 (LWP) when $P_{i,t}^{(2)}$ is greater or equal to the threshold value τ_2, and to be in state 0 (NP) otherwise. We select optimal thresholds τ_1 and τ_2 by adjusting the approach of Alessi and Detken (2011) for multiclass threshold creation. The resulting classifications are denoted by $\hat{Y}_{i,t}$ as follows:

$$\hat{Y}_{i,t} = \begin{cases} 0 & \text{otherwise} \\ 1 & P_{i,t}^{(1)} \geq \tau_1 \text{ and } P_{i,t}^{(2)} < \tau_2 \\ 2 & P_{i,t}^{(2)} \geq \tau_2 \end{cases}$$

These estimated states may be compared with the actual known states $Y_{i,t}$ of the training dataset. After classifying all quarters in the training dataset, we combine the three states of $Y_{i,t}$ in a table with each of the three states of $\hat{Y}_{i,t}$ to form the nine frequency counts of the classical classification (or "confusion") table:

$$F_{l,m} = \sum_{i,t}[Y_{i,t} = l,\ \hat{Y}_{i,t} = m] \text{ for } l, m = 0,1,2,$$

in which [ ] is the Iverson bracket function[3] that takes the value 1 if the condition in square brackets is satisfied, and 0 otherwise. The confusion table is a basic tool for displaying and evaluating the performance of a classification algorithm. General notation for the confusion table is given in Table 2.

[3] Iverson bracket, named after Kenneth E. Iverson, is a notation that generalises the Kronecker delta. It converts any logical proposition into a number that is 1 if the proposition is satisfied, and 0 otherwise, and is generally written by putting the proposition inside square brackets.

Clearly, the larger the frequency counts in the main diagonal (F_{00}, F_{11}, F_{22}) relative to the off-diagonal frequencies, the more accurate the classifier. In addition to the usual accuracy metrics, such as sensitivity and specificity, for evaluating a confusion table, we introduce some lesser known metrics, related to the harmonic mean, that are especially suitable for our context. These metrics are labeled for display in the margins of Table 2 and include the row harmonic mean (RHM), the column harmonic mean (CHM) and the total harmonic mean (THM). The definitions are as follows:

$$RHM = \frac{1}{1+\frac{1}{3}\left(\frac{F_{01}+F_{02}}{F_{00}}+\frac{F_{10}+F_{12}}{F_{11}}+\frac{F_{20}+F_{21}}{F_{22}}\right)}$$

$$CHM = \frac{1}{1+\frac{1}{3}\left(\frac{F_{10}+F_{20}}{F_{00}}+\frac{F_{01}+F_{21}}{F_{11}}+\frac{F_{02}+F_{12}}{F_{22}}\right)}$$

$$THM = \frac{1}{1+\frac{1}{6}\left(\frac{F_{10}+F_{20}+F_{01}+F_{02}}{F_{00}}+\frac{F_{10}+F_{12}+F_{01}+F_{21}}{F_{11}}+\frac{F_{20}+F_{21}+F_{02}+F_{12}}{F_{22}}\right)}$$

The harmonic mean of positive numbers $x_1, x_2, ..., x_n$ is the inverse of the mean of their inverses given by

$\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$, alternatively $\frac{n \cdot \prod_{j=1}^{n} x_j}{\sum_{i=1}^{n}\frac{\prod_{j=1}^{n} x_j}{x_i}}$.

To understand the origin of the terminology for our metrics, rewrite:

$$\text{RHM} = \frac{3}{\frac{F_{01}+F_{02}+F_{00}}{F_{00}}+\frac{F_{10}+F_{12}+F_{11}}{F_{11}}+\frac{F_{20}+F_{21}+F_{22}}{F_{22}}} =$$

$$\frac{3}{\frac{1}{\frac{F_{00}}{F_{01}+F_{02}+F_{00}}}+\frac{1}{\frac{F_{11}}{F_{10}+F_{12}+F_{11}}}+\frac{1}{\frac{F_{22}}{F_{20}+F_{21}+F_{22}}}}$$

This is the harmonic mean of the sensitivities of the three states – the proportions correct in each row of the confusion table are equal to the proportions of actual events correctly called. Similarly, CHM is the harmonic mean of the true positive rates – the proportions correct in each column of the confusion table are equal to the proportions of called events correctly called. In addition, THM is the harmonic mean of RHM and CHM.

One major reason for our preference for THM as a fitting and evaluation criterion is the large weight that THM gives to cells with relatively low counts. Typically, studies of financial crisis have relatively few crisis events. We have 84 EWP quarters, against 164 NPs and 84 LWPs that must be distinguished. Errors in classifying EWPs are weighted relatively highly in THM. For example, consider the first of the three ratios in the denominator of RHM:

$$\frac{F_{10}+F_{12}}{F_{11}}$$

Errors F_{10} and F_{12} in classifying EWPs will be magnified by virtue of the relatively small F_{11} in the denominator. Consequently, maximizing RHM places a premium on keeping the EWP error counts small and discourages fitting strategies that tolerate high error counts among low-frequency states in order to bulk up on high-frequency states. Similarly, overclassifying quarters as EWPs will be penalized in CHM by the relatively small F_{11} in the denominator of $\frac{F_{01}+F_{21}}{F_{11}}$.

The weighting algorithm in THM thus realistically reflects policy maker interest in getting early crisis calls correct. Another reason for our preference is that THM takes into account all nine individual cells in the confusion table – unlike some other performance criteria for multiple-state classifications.

As a numerical illustration, consider the following hypothetical confusion table for 100 quarters:

		Predicted		
		NP	EWP	LWP
Observed	NP	50	6	4
	EWP	10	5	0
	LWP	0	5	20

Overall, this table shows apparently good conformance between observed states and predicted states with 75% of the quarters correctly classified. However, only 5 of the 15 actual EWPs are correctly called, and only 5 of 16 of the predicted EWPs actually are EWPs. THM would penalize this solution because of the large number (10) of actual EWPs called as other states relative to the correctly called EWPs (5), and because of the large number (11) of predicted EWPs that were not, relative to the correctly called EWPs (5).

In our work, THM plays a double role. First, it guides the optimal selection of cutoff thresholds for the probability function $P_{i,t}^{(k)}$, which is estimated using the classification methodologies multinomial logit regression (MLR), discriminant analysis (MDA) and neural networks (NN), as explained in the next sections. The thresholds τ_1 and τ_2 are chosen such that the THM performance criterion is maximized. Second, the THM is used to evaluate and compare the classification methodologies. In maximizing THM, we maximize a combination of the specificities and true positive rates for the whole table – which is intuitively appealing.

3.2 Classification methods

In this section, we briefly describe the general approach to modeling $P_{i,t}^{(k)}$ for each prediction method employed. Empirical estimates are found in Section 4.

3.2.1 Multinomial logistic regression

Multinomial logistic regression is an extension of binary logistic regression to the case of more than two classification states. The multinomial logit has been employed, for example, by Bussiere and Fratzscher (2006) and Comeli (2014) in the context of currency crises, by Ciarlone and Trebeschi (2006) for debt crises and Caggiano et al. (2014) in the context of banking crises in low income countries.

In our paper, the model assesses the likelihood that the current quarter for a given country is in early or late pre-crisis and thereby signals both whether palliative action is needed, and also how much time is available for remedial measures. More formally, our three-state multinomial logistic regression model specifies that each quarter (t) for each country (i), can be in one of three states k = 0, 1, 2, corresponding respectively to NP, EWP and LWP, as explained previously, with the form of the probability function given by:

$$P_{i,t}^{(k)} = \frac{e^{\boldsymbol{\beta}_k' \cdot \boldsymbol{x}}}{1+e^{\boldsymbol{\beta}_1' \cdot \boldsymbol{x}}+e^{\boldsymbol{\beta}_2' \cdot \boldsymbol{x}}} \quad , k\text{=1,2} \qquad \text{and} \qquad P_{i,t}^{(0)} = 1 - P_{i,t}^{(1)} - P_{i,t}^{(2)},$$

where x is a vector of predictor variables and β_k is a state-specific vector of coefficients to be estimated by maximum likelihood. The mathematical form of $P_{i,t}^{(k)}$ ensures that the probabilities are nonnegative and lie between 0 and 1 and sum to 1. We set the “normal period” (NP) as the base state in order to make the multinomial logit model identifiable. The vector of parameters β_k

reflects the sensitivity of the state probability to changes in the independent variables. Note that the coefficient vector β_k is neither country-specific nor time-specific. That is, we assume that our model applies across all EU14 countries for all quarters. Thus, the subscripts i and t on $P_{i,t}^{(k)}$ signify merely that the probability for country i and quarter t varies because the values of the predictor variables for country i and quarter t are different from other countries and other quarters.

3.2.2 Multi-state linear discriminant analysis

Linear discriminant analysis is inherently multi-state and, to our knowledge, it is has not been used for multi-level financial crises prediction. The form of the probability classification function $P_{i,t}^{(k)}$ for our three-state discriminant analysis is:

$$P_{i,t}{}^{(k)} = \frac{f_k(\mathbf{x})\,p_{i,t}(k)}{f_0(\mathbf{x})\,p_{i,t}(0) + f_1(\mathbf{x})\,p_{i,t}(1) + f_2(\mathbf{x})\,p_{i,t}(2)}\,,\ k = 0,1,2$$

Here, $P_{i,t}^{(0)}$, $P_{i,t}^{(1)}$, $P_{i,t}^{(2)}$ represent prior probabilities of the three states, and $f_k^{(x)}$ is the probability density for data x, given state k. As in multinomial logistic regression, the form of $P_{i,t}^{(k)}$ ensures nonnegative probabilities that lie between 0 and 1 and sum to 1. Again, as in our multinomial logistic regression model, we assume that the probability classification function $P_{i,t}^{(k)}$ applies across all countries and through time. Similarly, the form of the data density function $f_k^{(x)}$ depends upon the state k but not the country or quarter. In standard linear discriminant analysis $f_k^{(x)}$ is taken to be multivariate normal with a common covariance matrix for all states. We follow the standard practice in our model.

3.2.3 Neural networks

Neural network (NN) models have developed from the fields of artificial intelligence and brain modeling. The NN models have been utilized for example by Boyacioglu et al. (2009) and Aydin et al. (2015) for predicting bank financial failures in Turkey and by Yu et al. (2010) for multiscale financial crisis forecasting.

NN have mathematical and algorithmic features intended to mimic the biological neural networks of the human nervous system. The NN method posits an interrelated group of artificial neurons that processes information using a so-called connectionist approach, in which network units communicate by flows of information. During a learning phase with a training data set, the structure of the model adapts, based on the external or internal information that flows through the network. Compared to other statistical methods, NN have two advantages. First, the models make no assumptions about the statistical distribution or properties of the data, and therefore tend to be useful in practical situations common to finance where the assumptions of common statistical models are not met. Second, NN models use nonlinear functions that permit considerable flexibility to adapt to a training data set. The result is often highly successful classification rates for the training data. However, the flexibility of NN models makes them prone to overfitting, unless care is taken. Our implementation of NN modeling can be illustrated schematically in Figure 2.

Predictor variables are shown on the left. They connect by information flows to a single hidden middle layer, represented by three neurons in our model.[4] Each neuron processes the input data through a hyperbolic tangent "activation function" of the form

$$H_j = \frac{e^{\gamma_j \mathbf{x}} - 1}{e^{\gamma_j \mathbf{x}} + 1},\ j = 1, 2, 3$$

where x is a vector of predictor variables and γ_j is a neuron-specific vector of coefficients to be estimated by the NN methodology. The three neural signals H_1, H_2, H_3 are sent to the output unit on the right in the diagram for final processing into classification states. This processing takes the form of the aggregation of the neural signals into state-specific linear functions:

$$\theta_k = \beta_{k0} + \beta_{k1} H_1 + \beta_{k2} H_2 + \beta_{k3} H_3,\ k = 1, 2,$$

with the coefficients estimated by the NN methodology, and then for calculation of the NN model probability function:

$$P_{i,t}^{(k)} = \frac{e^{\theta_k}}{1 + e^{\theta_1} + e^{\theta_2}},\ k = 1, 2 \qquad \text{and} \qquad P_{i,t}^{(0)} = 1 - P_{i,t}^{(1)} - P_{i,t}^{(2)}$$

As with the other models, the form of $P_{i,t}^{(k)}$ ensures the probabilities are nonnegative, between 0 and 1 and sum to 1. Since the state probabilities sum to 1, it is unnecessary to compute θ_0; θ_1 and θ_2 are sufficient.

[4] Although there are three neurons, the neurons do not correspond to the three states, NP, EWP, and LWP.

4. Empirical results

Models are built *ex post* but used *ex ante*, and hindsight provides 20-20 vision. To assess model performance, it is necessary not only that the model fit historical data well, but also predict future data well. There is always a danger that a model has been overfit to historical data and will not replicate its success with future data. Since future data are never conveniently available, statistical schemes are used to simulate the availability of future data. The most common of these schemes is cross-validation. The basic idea of cross-validation is to divide the available data into a training dataset and a holdout sample. The model is estimated on the training dataset and then applied to the "new" data in the holdout sample. If performance does not decline substantially between the training and holdout samples, then one may infer that the model will apply as well with future data.

To implement cross-validation with our multinomial logistic regression, discriminant analysis, and neural network models, we performed three-fold cross-validation. That is, the dataset of quarters was divided randomly into thirds. Using two of the thirds as the training dataset and the remaining third as the holdout sample, we ran the three classification models on the training dataset and then classified the holdout sample into our three states of NP, EWP, and LWP for each methodology. Then we repeated this procedure with one of the other thirds as holdout sample and remaining two thirds as training dataset. Finally, the procedure was repeated a third time so that all of the quarters had been used once in a holdout sample and twice in a training dataset. Classification successes and failures were tallied, evaluation metrics like THM were applied and averaged across the three runs.

As noted in Section 3, the models estimate the probability functions $P_{i,t}^{(k)}$ of the states NP (*k*=0), EWP (*k* =1), and LWP (*k* =2), as well as the threshold probabilities τ_1 and τ_2 that trigger classification of quarters into these states. Figures 3, 4 and 5 graphically display these estimates for the three classification methods and all EU-14 countries. Each time plot is divided into the three states NP, EWP, and LWP, from left to right, by the two broken red vertical lines placed at 2005Q1 and 2006Q3 (except for the U.K.). The solid lines are the estimates of $P_{i,t}^{(k)}$, *k*=0,1,2. The broken red horizontal lines mark the thresholds τ_1 and τ_2. When the plot of $P_{i,t}^{(1)}$ lies above τ_1, the state EWP is predicted; when the plot of $P_{i,t}^{(2)}$ lies aboveτ_2, the state LWP is predicted; otherwise, the state NP is predicted.

Even a cursory glance at Figures 3,4,5 shows that each model correctly classifies most quarters. The probability function $P_{i,t}^{(1)}$ tends to rise above τ_1 in the true EWP period and remain below it otherwise; $P_{i,t}^{(2)}$ tends to rise above τ_2 in the true LWP period and remain below it otherwise; $P_{i,t}^{(0)}$ is high in the true NP period. Thus correct classifications are triggered in the appropriate states. The neural network model does especially well with $P_{i,t}^{(k)}$ being very close to either 0 or 1. Neural networks enjoy great flexibility and are therefore prone to overfitting. They tend to be very optimistic during in sample analysis. Our NN models did less well in the holdout samples.

The probability functions $P_{i,t}^{(k)}$ can be calculated from the estimates of the models given in Section 3.2. Tables 3, 4, 5 present estimates of the coefficients of those models.

Results of the cross-validation analysis are shown in Table 6. For each of the 9 cells ($\mathbf{F_{00}}$ – $\mathbf{F_{22}}$) of the confusion table, Table 6 reports the mean (of the three-fold cross-validations) cell count as a percent of the total cell count. For example, on average across the three cross-validations, the neural networks model classified 47.3% of the 332 quarters as NP and they actually were NP ($\mathbf{F_{00}}$.) ACC is the mean gross accuracy rate = 3-fold average of:

$$\frac{F_{00}+F_{11}+F_{22}}{F_{00}+F_{01}+F_{02}+F_{10}+F_{11}+F_{12}+F_{20}+F_{21}+F_{22}}$$

Each method is estimated by maximizing THM.

All three methods work very well. The most consistent method is discriminant analysis, which declines by only about 5% from the training sample to the holdout sample. Multinomial logistic regression and neural networks perform very similarly, with a very high success rate in the training sample, but much poorer performance in the holdout sample.

To get an idea of what these results might mean for a policymaker, let us take the holdout sample results as indicative of future performance. Out of 332 quarters, there are 84 EWPs. Using discriminant analysis, 19.6% x 332 = 65 are correctly classified as EWPs. 1.8% x 332 = 6 are incorrectly classified as NPs. And 3.6% x 332 = 12 are incorrectly classified as LWPs.[5] Thus,

[5] The total, 65 + 6 + 12 = 83, is one short due to averaging and roundoff.

a policymaker could expect to be warned of the coming of a severe financial crisis in 65/84 = 77.4% of the early quarters that precede the crisis (and probably warned multiple times during that early period). For those quarters, the method correctly calls both the crisis and its timing. In addition, the method would correctly call the crisis, but call it too early, in about 12/84 = 14.3% of the early quarters. In those quarters, the method would suggest there is less time for remedial action than there actually would be. Only in 1.8% x 332 = 6 of 84 or 7.1% of early quarters does the model incorrectly signal a crisis coming. Thus, the *sensitivity* of the method is high: It tends to call correctly a high proportion of actual early pre-crisis quarters.

However, one must also examine the *true positive rate* of the method. That is, what proportion of the called EWPs actually are EWPs? From Table 6, the holdout discriminant analysis method calls 3.6% x 332 = 12 quarters as EWPs that are actually NPs; 19.6% x 332 = 65 quarters as EWPs that actually are EWPs; and 0% x 332 = 0 quarters as EWPs that are actually LWPs. Thus, 12 + 65 + 0 = 77 quarters are actually classified as EWPs. Of these calls, 65/77 = 84.4% are correct both as to the fact of the coming of the crisis and as to its timing. 12/77 = 15.6% are actually NPs, so no crisis is coming. Thus, when the method calls a quarter as EWP, the method is correct a high proportion of the time.

A policymaker wants to avoid missing a severe crisis. Suppose he adopts the discriminant analysis method. If a crisis is really coming 6-12 quarters in the future, the probability of getting a correct early warning in the current quarter is about 0.774. But there is also about 0.143 probability of getting a premature warning (LWP). Such a warning is correct as to the fact of a coming crisis, but the timing is off. The policymaker's probability is only about 0.071 of being told that no crisis is coming (NP). And that is just for the current quarter. The policymaker still has several more quarters to go in which warnings could be given before the crisis.

A policymaker also wants to avoid falsely calling a severe crisis. Suppose he adopts the discriminant analysis method. If an EWP warning is given in the current quarter, the probability that the warning is correct both as to the fact of the coming crisis and as to its timing is about 0.844. There is about a 0.156 probability that the early warning is a false alarm and there is really no crisis in the foreseeable future. Before taking action, the policymaker could await a confirming warning in a subsequent quarter, and have a lower probability of a false alarm. But we have not investigated the statistics of decision rules based upon the signals of multiple quarters.

5. Conclusion

In this study, we have developed an EWS for forecasting a financial crisis of the magnitude of the 2007-08 crisis for the European Union (the EU-14). The EWS shifts focus from forecasting the crisis itself to classifying pre-crisis quarters by the amount of time remaining until the start of the crisis. The EWS has the dual objective of forecasting both the fact of a crisis and the time remaining until its arrival. We achieve this double goal by adopting a ternary state classification system for pre-crisis quarters as EWP (7-12 quarters in advance), LWP (1-6 quarters in advance), and NP (more than 12 quarters in advance). Thus, the correct classification of a quarter as EWP means both that a severe crisis is coming and that there remain at least 6 quarters to take measures to prevent it or reduce its impact should it occur.

The EWS is transparent. It utilizes only publicly available macro-level data and standard statistical classification methodology. The data include 12 standard macroeconomic, market, banking and government variables. There are no proprietary data. The statistical methodologies are multinomial logistic regression, discriminant analysis, and neural networks.

Since EWP quarters are relatively less numerous than NP and LWP quarters, we introduce a fitting and evaluation metric, the total harmonic mean, that places a premium on avoiding errors in classifying and calling EWP quarters.

The EWS enjoys high accuracy. Cross-validation provides a means to test classification accuracy with simulated future data in the form of holdout samples. In our tests, discriminant analysis did best with an accuracy of about 85% in 3-fold holdout samples. The neural network model was best in the training dataset with about 94% accuracy. Further analysis shows that a policymaker who uses these methods will enjoy a high probability that future crises will be signaled well in advance and that warnings of crisis will not be false alarms. For example, when a crisis is coming in 7-12 quarters from the current quarter, the EWS will correctly warn both of the crisis and of the correct timing with 0.774 probability, and will warn prematurely of the crisis with probability 0.143, but will fail to warn either of the crisis or its timing with probability only

0.071. Moreover, if the EWS does issue an EWP warning, then the probability is about 0.844 that the warning is correct. Since the EWS issues a signal for every quarter, failure to correctly warn in the current quarter may still be remedied in a subsequent quarter in a timely manner.

References

Alessi, L., Detken, C. (2011). *Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity*. European Journal of Political Economy, 27(1), 520–533.

Aydin, A. D., Cavdar, S. C. (2015). *Prediction of Financial Crisis with Artificial Neural Network: An Empirical Analysis on Turkey*. International Journal of Financial Research, 6(4), 36.

Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K., Vašíček B. (2013). *Leading indicators of crisis incidence: Evidence from developed countries*. Journal of International Money and Finance, 35, 1–19.

Behn, M., Detken, C., Peltonen, T.A., Schudel, W. (2013). *Setting countercyclical capital buffers based on early warning models: Would it work?* ECB WP no. 1604.

Betz, F., Oprică, S., Peltonen, T.A., Sarlin, P. (2014). *Predicting distress in European banks*. Journal of Banking & Finance, 45, 225–241.

BIS. (2009). 79th Annual Report.

Borio, C., Lowe, P. (2002). *Asset Prices, Financial and Monetary Stability: Exploring the Nexus*. BIS Working Papers, No. 114.

Borio, C., Lowe, P. (2004). *Securing Sustainable Price Stability: Should Credit Come Back from the Wilderness?* BIS Working Papers, No. 157.

Borio, C., Drehmann, M. (2009). *Assessing the risk of banking crises – revisited*. BIS Quarterly Review, (March), 29–46.

Boyacioglu, M. A., Kara, Y., Baykan, Ö. K. (2009). *Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey*. Expert Systems with Applications, 36(2), 3355-3366.

Büyükkarabacak, B., & Valev, N.T. (2010). *The role of household and business credit in banking crises*. Journal of Banking & Finance, 34(6), 1247–1256.

Caggiano, G., Calice, P., Leonida, L. (2014). *Early warning systems and systemic banking crises in low income countries: A multinomial logit approach*. Journal of Banking & Finance, 47, 258-269.

Cao, Y. (2012). *Aggregating multiple classification results using Choquet integral for financial distress early warning*. Expert Systems With Applications, 39(2), 1830-1836. Chicago.

Ciarlone, A., Trebeschi, G. (2006). *A Multinomial Approach to Early Warning Systems for Debt Crises*. Bank of Italy, Economic Research and International Relations Area. (No. 588).

Comelli, M. F. (2014). *Comparing the performance of logit and probit early warning systems for currency crises in emerging market economies*. International Monetary Fund. (No. 14-65)

Davis, E.P., Karim, D. (2008). *Comparing early warning systems for banking crisis*. Journal of Financial Stability 4 (2), 89–120.

Davis, E.P., Karim, D., Liadze, I. (2011). *Should multivariate early warning systems for banking crises pool across regions?* Review of World Economics, 147, 693–716.

Demyanyk, Y., Iftekhar, H. (2010). *Financial crises and bank failures: A review of prediction methods*. Omega, Volume 38, Issue 5, Pages 315–324.

Dungey, M., Fry, R., Martin, V., Tang, C., Gonzalez-Hermosillo, B. (2010). *Are Financial Crises Alike?* IMF Working Paper, WP/10/14.

FSB. (2009). *Guidance to assess the systemic importance of financial institutions, markets and instruments: initial considerations*. Report to G20 finance ministers and governors.

IMF. (2009). How Linkages Fuel the Fire: The Transmission of Financial Stress from Advanced to Emerging Economies. World Economic Outlook. World Economic and Financial Surveys.

Kaminsky, G. (2006). *Currency crises: Are they all the same?* Journal of International Money and Finance, 25(3), 503–527.

Kaminsky, G., Reinhart, C.M. (1999). *The twin crises: the causes of banking and balance of payments problems*. American Economic Review, 89 (3), 473– 500.

Laeven, L., Valencia, F. (2008). *Systemic Banking Crises: A New Database*. Working Paper WP/08/224, International Monetary Fund.

Laeven, L., Valencia, F. (2012). *Systemic Banking Crises Database: An Update*. Working Paper WP/12/163, International Monetary Fund.

Lo Duca, M., Peltonen, T.A. (2013). *Assessing systemic risks and predicting systemic events*. Journal of Banking & Finance, 37(7), 2183–2195.

Männasoo, K., Mayes, D.G. (2009). *Explaining bank distress in Eastern European transition economies*. Journal of Banking & Finance, 33(2), 244–253.

Misina, M., Tkacz, G. (2009). *Credit, asset prices, and financial stress*. International Journal of Central Banking, 5 (4), 95–122.

Papadopoulos, S., Stavroulias, P., Sager, T. (2016). *Systemic early warning systems for EU15 based on the 2008 crisis*. Bank of Greece Eurosystem WP, No. 202.

Reinhart, C.M., Rogoff, K.S. (2011). *From financial crash to debt crisis*. American Economic Review, 101(5), 1676–1706.

Sarlin, P., Peltonen, T.A. (2013). *Mapping the state of financial stability*. Journal of International Financial Markets Institutions and Money, 26, 46–76.

Schularick, M., Taylor, A. (2012). *Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870–2008*. American Economic Review, 102(2), 1029–1061.

Tables and figures

Table 1. Summary statistics for the 12 AGR[1] predictor variables. Figures are for the transformed AGR variables for all quarters, across all countries.

Variable	Variable Explanation	MIN	Q1	Mean	Median	Q3	MAX	STD
GDP	Gross Domestic Product	-0.046	0.032	0.047	0.046	0.063	0.132	0.025
UNE	Unemployment Rate	-0.229	-0.067	0.008	0	0.067	0.478	0.114
CPI	Consumer Price Index	-0.001	0.015	0.021	0.020	0.027	0.050	0.008
CST	Production in Construction	-0.511	-0.018	0.020	0.018	0.056	0.447	0.093
LBC	Real Unit Labor Cost	-0.063	-0.014	-0.003	-0.005	0.004	0.090	0.018
CRE	Credit to the Private Sector	-0.061	0.042	0.083	0.076	0.118	0.230	0.057
EQU	Equities	-0.477	-0.036	0.094	0.146	0.240	0.605	0.216
BON	10y Bonds Yield	-0.312	-0.120	-0.014	-0.019	0.112	0.251	0.136
EXR	Real Effective Exchange Rate	-0.092	-0.003	0.013	0.009	0.030	0.118	0.029
GDE	Gen. Gov. Consolidated Gross Debt	-0.232	-0.046	-0.015	-0.014	0.021	0.112	0.052
GRE	Gen. Gov. Revenue	-0.161	-0.016	0.001	0.001	0.020	0.202	0.035
GTE	Gen. Gov. Expenditure	-0.070	-0.013	0.007	0.005	0.024	0.155	0.031

[1] AGR stands for annual growth rate. See p.8 for details.

Table 2. A general 3x3 confusion table[1]

		Predicted			
		0 – NP	1 - EWP	2 - LWP	
Observed	0 – NP	F_{00}	F_{01}	F_{02}	RHM
	1 – EWP	F_{10}	F_{11}	F_{12}	
	2 – LWP	F_{20}	F_{21}	F_{22}	
		CHM			THM

[1] where F stands for frequency.

Table 3. Multinomial logistic regression: Estimates of coefficients of $P_{i,t}^{(1)}$ and $P_{i,t}^{(2)}$. State 0 (NP) is the baseline case.

	State 1 (EWP)		State 2 (LWP)	
Variable	Coefficient	P-value	Coefficient	P-value
CPI	233.08	0.001	-7.78	0.847
GDP	91.67	0.003	2.66	0.904
UNE	-14.57	0.001	-10.32	0.004
CST	-20.36	0.000	-4.44	0.147
GTE	-61.92	0.001	-33.68	0.003
LBC	118.92	0.000	33.33	0.069
GRE	37.41	0.001	8.19	0.318
GDE	-12.71	0.101	-9.44	0.158
EXR	-334.25	0.000	-1.41	0.952
CRE	-6.57	0.449	10.51	0.160
BON	-13.67	0.000	24.43	0.000
EQU	12.40	0.000	2.29	0.308
const.	-10.36	0.000	-3.68	0.001

Table 4. Discriminant analysis: Estimates of coefficients of $P_{i,t}^{(1)}$ and $P_{i,t}^{(2)}$. State 0 (NP) is the baseline case.

	State 1 (EWP)	State 2 (LWP)
Variable	Coefficient	Coefficient
CPI	-0.24	-0.03
GDP	-0.44	-0.06
UNE	0.18	0.47
CST	0.22	0.24
GTE	0.02	0.32
LBC	-0.17	-0.21
GRE	-0.15	-0.04
GDE	-0.10	0.13
EXR	0.99	0.18
CRE	0.12	-0.17
BON	0.89	-0.51
EQU	-0.24	-0.21

Table 5. (a) Neural network: Estimates of coefficients in the activation functions H_1, H_2, H_3 of the hidden layer.

	γ_1 (H_1)	γ_2 (H_2)	γ_2 (H_3)
Variable	Coefficient	Coefficient	Coefficient
CPI	27.1	145.7	-39.7
GDP	-40.9	-4.7	-110.8
UNE	-7.1	-4.7	-3.3
CST	3.6	-5.6	-26.8
GTE	-10.8	10.9	10.9
LBC	2.8	-10.9	-126.8
GRE	2.6	-3.3	-34.6
GDE	-1.1	-12.0	-17.1
EXR	-9.0	-210.3	28.0
CRE	7.2	-4.1	20.1
BON	20.8	-11.0	27.7
EQU	-5.0	0.3	-19.9
const.	-1.5	1.4	1.7

(b) Neural network: Estimates of information feeds from hidden middle layer to output categorizer.

$$\Theta_1 = -24.8 - 145.5 \cdot H_1 + 32.2 \cdot H_2 + 76.3 \cdot H_3$$

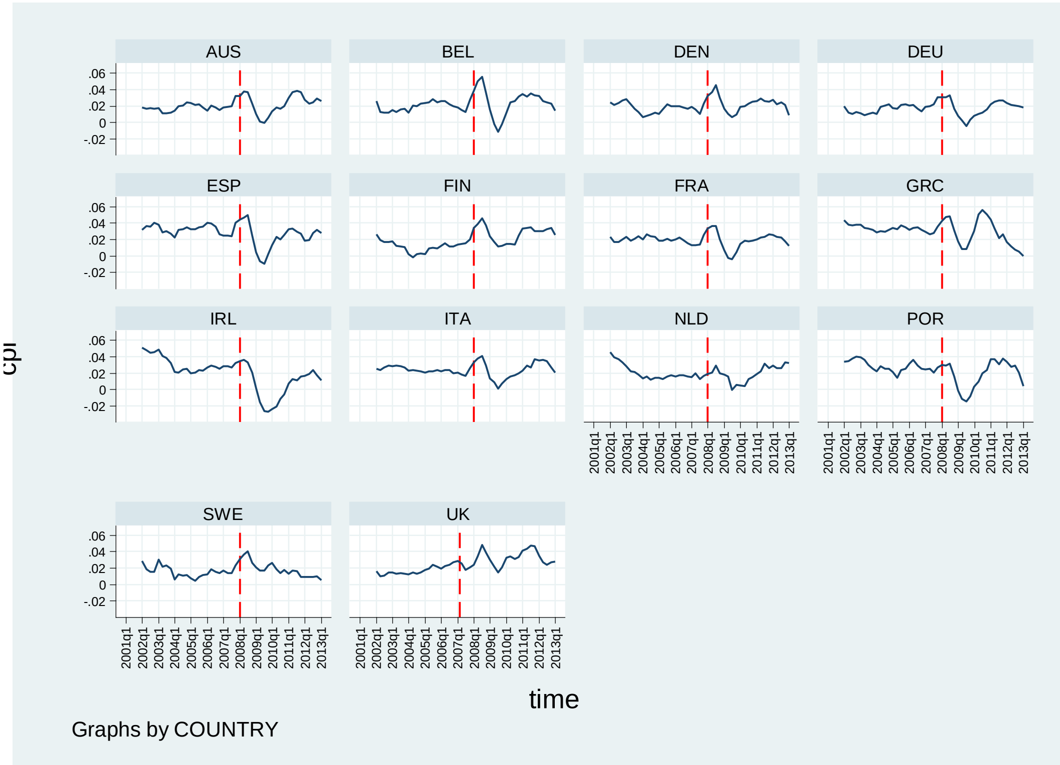
$$\Theta_2 = -136.1 + 54.5 \cdot H_1 + 185.1 \cdot H_2 + 126.4 \cdot H_3$$

Only Θ_1 and Θ_2 are necessary to calculate all $P_{i,t}^{(k)}$, k=0,1,2. See Section 3.2 for details.

Table 6. Cross-validation results for three classification methods: Means of three-fold cross-validation cell count percentages and ACC, THM evaluation metrics.

Method	F_{00}	F_{01}	F_{02}	F_{10}	F_{11}	F_{12}	F_{20}	F_{21}	F_{22}	ACC	THM
Training Sample											
Neural Network	47.3	0.9	1.8	1.3	22.3	1.3	0.4	0	24.6	94.2	93.7
Multinomial Logistic Regression	48.2	1.3	0.4	0.9	23.2	0.9	0.9	1.3	22.8	94.2	93.0
Discriminant Analysis	48.2	1.3	0.4	0.9	20.5	3.1	2.2	0.4	22.3	91.1	90.0
Holdout Sample											
Neural Network	41.1	7.1	1.8	2.7	19.6	2.7	2.7	0.9	21.4	82.1	81.0
Multinomial Logistic Regression	42.0	6.3	1.8	2.7	18.8	3.6	1.8	1.8	21.4	82.1	80.3
Discriminant Analysis	43.8	3.6	1.8	1.8	19.6	3.6	2.7	0	22.3	85.7	85.2

Figure 1: Time plots of AGR[1] for CPI (consumer price index)



[1] AGR stands for annual growth rate. See p.8 for details.

Figure 2. Schematic illustration of the neural network model

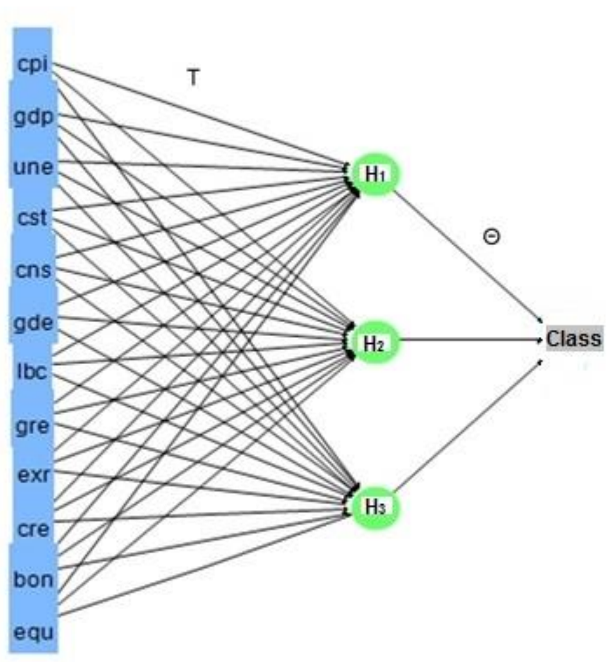


Figure 3. The multinomial logistic regression model: Estimates of the probability functions $P_{i,t}^{(k)}$ and thresholds τ_1 = 0.306 and τ_2 = 0.612.

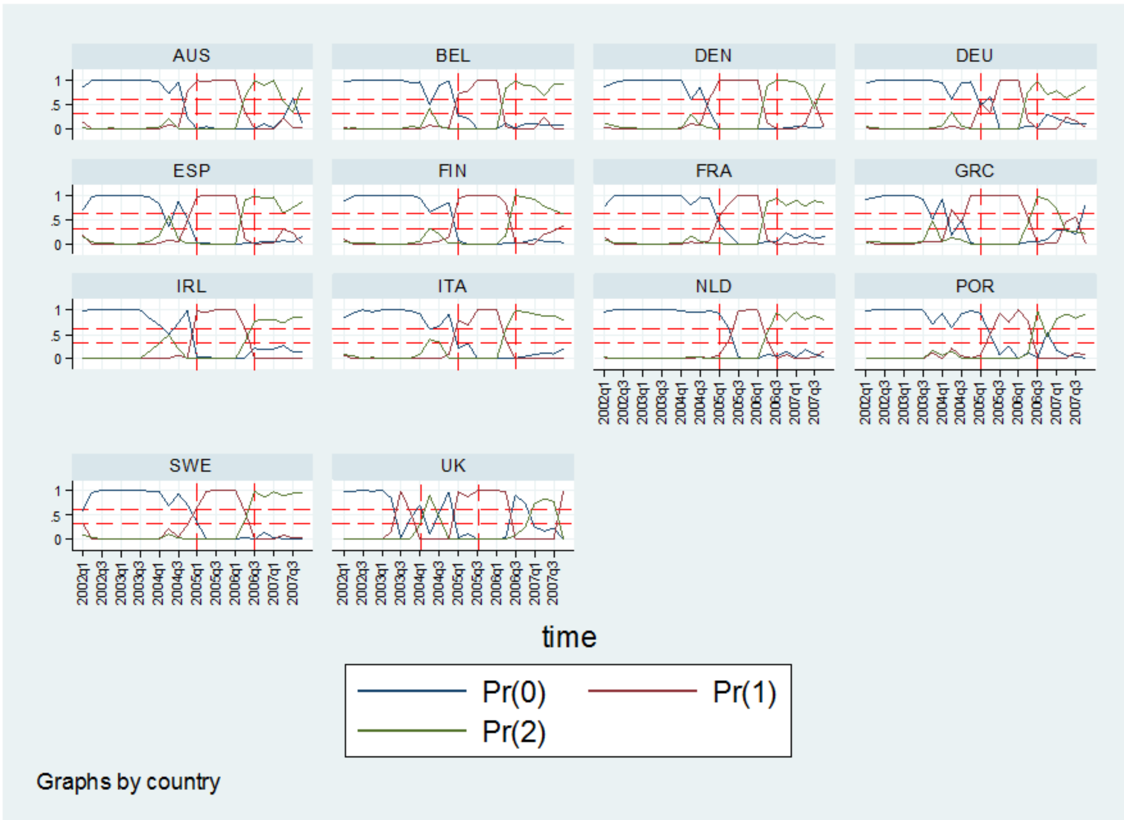


Figure 4. The discriminant analysis model: Estimates of the probability functions $P_{i,t}^{(k)}$ and thresholds τ_1 = 0.825 and τ_2 = 0.879.

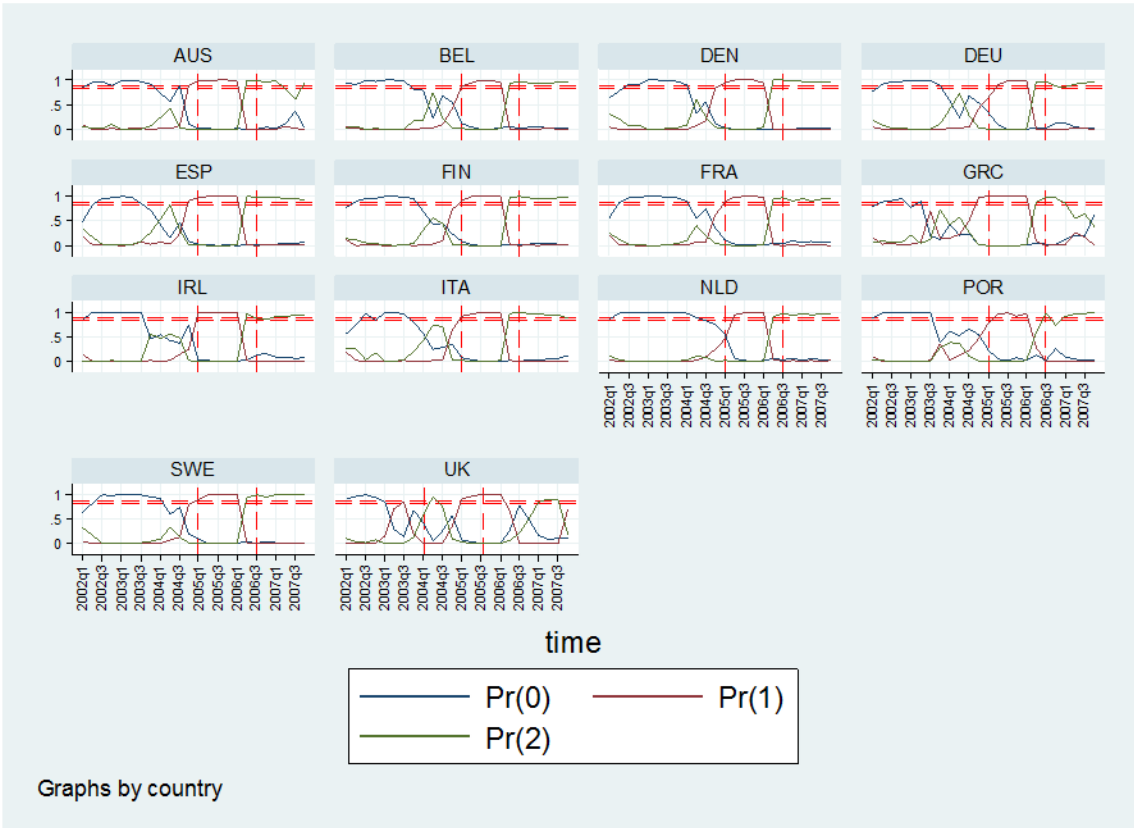
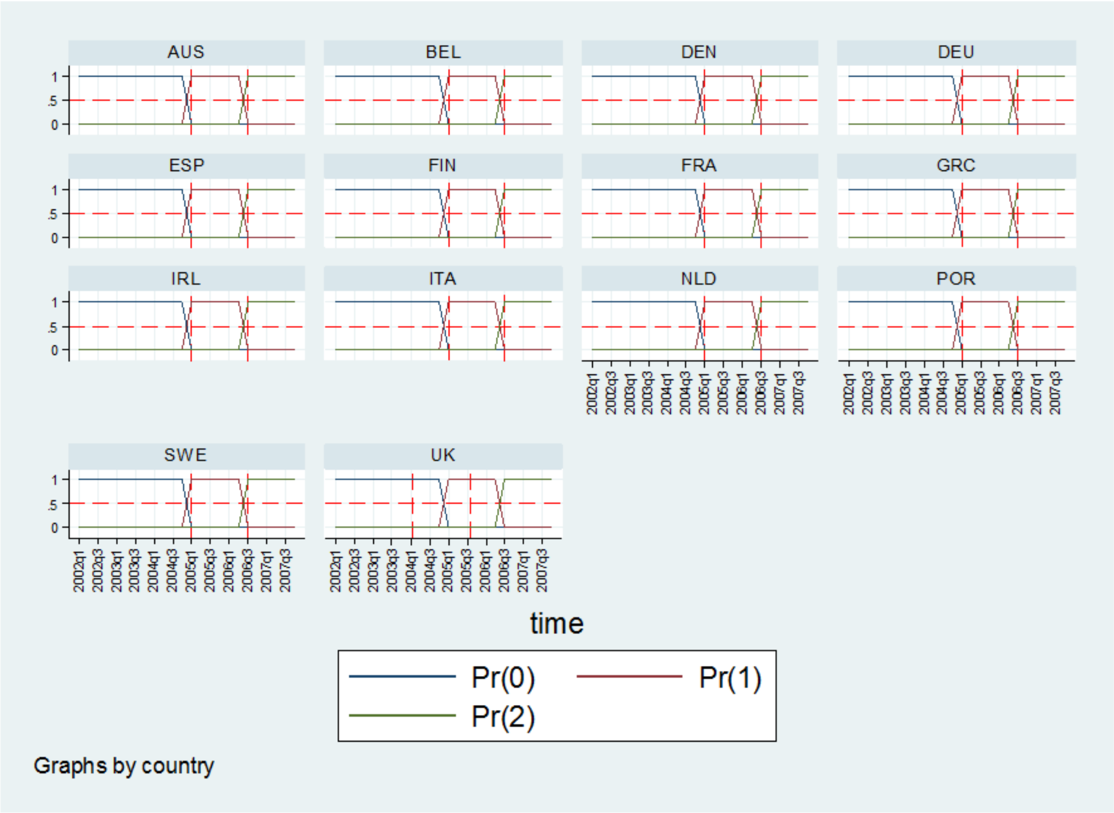


Figure 5. The neural network model: Estimates of the probability functions $P_{i,t}^{(k)}$ and thresholds τ_1 and τ_2[1].



[1] For the neural network model, the thresholds are not uniquely determined. The neural network model was nearly perfect in its training set classifications. That is why its probability functions are nearly 0 or 1. In that case, the thresholds may take any value between 0.05 and 0.95 (say) without affecting the results. Figure 5 shows them at 0.5 for illustration.

BANK OF GREECE WORKING PAPERS

202. Papadopoulos, S., P. Stavroulias and T. Sager, “Systemic Early Warning Systems for EU15 Based on the 2008 Crisis”, January 2016.

203. Papadopoulos, G., S. Papadopoulos and T. Sager, “Credit Risk Stress Testing for EU15 Banks: a Model Combination Approach”, January 2016.

204. Louzis, P. D., “Macroeconomic Forecasting and Structural Changes in Steady States”, March 2016.

205. Christodoulakis, N. and C. Axioglou, “Underinvestment and Unemployment: the Double Hazard in the Euro Area”, April 2016.

206. Bardaka, C. I., “Structural and Cyclical Factors of Greece’s Current Account Balances: A Note”, May 2016.

207. Lothian, J. R., and G. S. Tavlas, “How Friedman and Schwartz Became Monetarists”, May 2016.

208. Balfoussia, H. and H. D. Gibson, “Firm Investment and Financial Conditions in the Euro Area: Evidence from Firm-Level Data”, June 2016.

209. Samantas, G.I., “On the Optimality of Bank Competition Policy”, July 2016.

210. Malliaropulos, D., and P.M. Migiakis, “The Re-Pricing of Sovereign Risks Following the Global Financial Crisis”, July 2016.

211. Asimakopoulos, I., P.K. Avramidis, D. Malliaropulos, N. G. Travlos, “Moral Hazard and Strategic Default: Evidence from Greek Corporate Loans”, July 2016.

212. Tavlas, S.G., “New Perspectives on the Great Depression: a Review Essay”, September 2016.

213. Papapetrou, E. and P. Tsalaporta, “Inter-Industry Wage Differentials in Greece: Rent-Sharing and Unobserved Heterogeneity Hypotheses”, November 2016.

214. Gibson, D.H, S.G. Hall, and G.S. Tavlas, “Self-Fulfilling Dynamics: the Interactions of Sovereign Spreads, Sovereign Ratings and Bank Ratings during the Euro Financial Crisis”, November 2016.

215. Baumann, U. and M. Vasardani, “The Slowdown in US Productivity Growth - What Explains it and Will it Persist?”, November 2016.

216. Tsionas. G.M., “Alternative Bayesian Compression in Vector Autoregressions and Related Models”, November 2016.

217. Tsionas. G.M., “Alternatives to Large VAR, VARMA and Multivariate Stochastic Volatility Models”, December 2016.

218. Balfoussia, H. and D. Papageorgiou, “Insights on the Greek Economy from the 3D Macro Model”, December 2016.

219. Anastasiou, D., H. Louri and M. Tsionas, “Non-Performing Loans in the Euro Area: are Core-Periphery Banking Markets Fragmented?,” December 2016.

220. Charalambakis, E., Y. Dendramis, and E. Tzavalis, “On the Determinants of NPLS: Lessons from Greece”, March 2017.

221. Christou, G. and P. Chronis, Markups and fiscal policy: analytical framework and an empirical investigation, April 2017.