BANK OF GREECE
EUROSYSTEM

Working Paper

Systemic early warning systems for EU15 based on the 2008 crisis

Savas Papadopoulos
Pantelis Stavroulias
Thomas Sager

202

JANUARY 2016

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

SYSTEMIC EARLY WARNING SYSTEMS FOR EU15 BASED ON THE 2008 CRISIS

Savas Papadopoulos
Bank of Greece

Pantelis Stavroulias
Democritus University of Thrace

Thomas Sager
University of Texas

Abstract

Reliable forecasts of an economic crisis well in advance of its onset could permit effective preventative measures to mitigate its consequences. Using the EU15 crisis of 2008 as a template, we develop methodology that can accurately predict the crisis several quarters in advance in each country. The data for our predictions are standard, publicly available macroeconomic and market variables that are preprocessed by moving averages and filtering. The prediction models then utilize the filtered data to distinguish pre-crisis from normal quarters through standard statistical classification methodology plus a proposed new combined method, enhanced by an innovative threshold selection and goodness-of-fit measure. Empirical results are very satisfactory: Country-stratified 14-fold cross validation achieves 92.1% correct classification and 85.7% for both true positive rate and positive predictive value for the EU15 crisis of 2008. Results will be of use to policy makers, investors, and researchers who are interested in estimating the probability of a crisis as much as one and a half years in advance in order to deploy prudential policies.

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

JEL-classifications: C53; E58; G28

Acknowledgments: This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

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

1. Introduction

"Achieving financial stability is perhaps the most urgent task facing the world economy at the present time. If the international financial system cannot be made to operate in a more stable way, the prospects for an open and liberal approach to trade and capital flows are poor … the fundamental goals of development and poverty alleviation will be set back." Andrew Crockett (1998). Although research oriented toward developing Early Warning System (EWS) models of financial crisis was underway in the mid 90s, a strong stimulus was given in recent years following the global financial crisis, which started in the United States in 2007 and spread through the rest of the world in the following years.

The burgeoning of EWS was to be expected since the direct and indirect cost of the global financial crisis was vast and had a significant impact on the health and soundness of the entire global financial system. Subsequently, policy makers started to reconsider the existing early warning models, the predictors of financial instability and their policy tools in order to build new ones with the objective of preventing or at least reducing the intensity of future financial crises, using all the knowledge so far.

Since the global crisis of 2008, the literature on EWS has proliferated. 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. 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 growth. They also found that property prices could provide an early signal for a banking crisis. 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 financial 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. 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. 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. 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. Lainà, Nyholm and Sarlin (2015) tested predictors for systemic bank crises for 11 EU countries within 1980-2013 and identify loans over deposits and house price growth as early warning quantities.

This paper builds upon the existing literature and extends it in many ways. The goal of this research is to construct an EWS for a future financial crisis in the EU15 countries, based on public data from the recent global financial crisis. Further, the intent is that the EWS should provide a useful tool for macroprudential policy. That is, the EWS should inform financial regulation aiming to mitigate the entire financial risk of a system, called the systemic risk. In order to achieve this goal, the EWS should predict a crisis sufficiently in advance to permit timely mitigation. The EWS should also provide reliable predictions that are correct most of the time, taking into account the relative costs of incorrect predictions. This research proposes an EWS that achieves these desiderata. The predictive power of the proposed macroprudential model benefits from its focus on the relatively financially and economically homogeneous EU15. Additionally, the use of public, nonproprietary data makes the model attractive to a wide range of policy makers and analysts.

The main contribution of this study is to potentially enhance the future financial stability of the EU15 economic/financial system. Financial stability is threatened by high systemic risk which can lead to systemic events with financial turmoil and economic losses. The financial crisis of 2008, which provides the benchmark for this study, is the most severe systemic event since the Great Depression. Regulators would like to have reliable early prediction of a systemic event in order to adopt preventive measures like bolstering bank capital buffers. Capital buffers help eliminate the procyclicality of lending by creating countercyclical buffers to curb credit risk. But buffers require time to implement. A reasonable timeframe for implementing a capital buffer regulation is 7 to 12 quarters advance warning. If draconian requirements like capital buffers are to be imposed on the basis of an EWS, then high predictive accuracy for systemic events should be demanded of the EWS several quarters in advance.

To best serve the need of adequate advance warning, this research creates a binary indicator variable having the value 1 for quarters 7 - 12 before the onset of crisis (pre-crisis quarters) and the value 0 for all quarters 16 or more before the crisis (normal quarters). Prospectively, then, classification of the current quarter as a 1 (pre-crisis) means that the EWS predicts the occurrence of crisis within 7 - 12 quarters following the present quarter. Thus, the EWS methodology shifts the methodological focus from classification of crisis quarters to classification of pre-crisis quarters. For this approach to be feasible, the crisis must be foreshadowed in the data of 7 - 12 quarters before the crisis. The pre-crisis quarters must already have become distinguished from normal quarters, but not yet transformed into crisis quarters. The analysis finds this to be true for each country in the EU15.

The macroprudential predictive modeling applied in this paper includes panel data models with random effects and discriminant analysis. Each model has the binary response variable described above for pre-crisis and normal quarters. Each model statistically classifies each quarter as 1 or 0, for every country. In the statistical classification the independent variables are selected country - specific time series that represent the real, the fiscal, the financial, the market and the external sectors for 2001-2008. These data series are all publicly available quarterly data from the EU15 countries. The raw time series are smoothed for randomness and are seasonally adjusted by applying a moving average of order 4. The trend is filtered by the one-sided Hodrick-Prescott technique executed by Kalman filtering. Applying the statistical classification methodology to the processed data results in a score for each quarter for each country. Quarters with scores that exceed an innovatively estimated threshold are classified as pre-crisis quarters. An error weighted harmonic mean is used to estimate the threshold and to assess goodness of fit. Moreover, scores can be easily converted into probabilities. Therefore, applied researchers will be able to estimate the probability of systemic risk by processing and filtering 10 values of independent systemic variables, then substituting them into one linear equation. Additionally, a new to the literature, as far as we know, combined classification method is proposed. This method takes into account the classification results of various predictive models and optimizes the crisis and non-crisis period identification.

All extracted models have good predictive accuracy. Country stratified k-fold cross-validation was used to assess out-of-sample predictive success. Discriminant analysis and panel linear regression with random effects and then the proposed combined method, predict better than standard non-linear models.

Section 2 describes the data, the classification variable, the independent variables and their transformation. Section 3 presents the four estimation methods for parameters and thresholds, provides the used notation and the proposed goodness of fit. In Section 4 we illustrate the probability of default for the classification methods, the extracted thresholds and the evaluation using confusion tables. Section 5 summarizes the main conclusions.

2. Data

In this study, as we mentioned earlier, we focus on one geographic sector. The country sample consists of the EU15 minus Luxemburg. Luxemburg was excluded due to missing values of variables important for the study, such as 10 years bonds. The 14 countries are Austria (AUS), Belgium (BEL), Denmark (DEN), Germany (DEU), Spain (ESP), Finland (FIN), France (FRA), Greece (GRC), Ireland (IRL), Italy (ITA), Holland (NLD), Portugal (POR), Sweden (SWE) and United Kingdom (UK). Since the EU is relatively homogeneous in relation to the world's economies, we are hopeful that a single predictive model may suffice for its members. Although EWS studies that use global samples offer more data, they may lose predictive power because crises often erupt in regionally distinctive manners. It may therefore be advantageous to model regions separately (Davis, et al., 2011). Since we focus on the EU, our predictive model may not apply to other regions. However, the EU is the world's largest economy if member GDPs are summed, so there is ample justification for a special focus on a crisis in the EU.

Additionally we focus on one time period: the European Union (EU) and the first quarter of 2001 through the first quarter of 2014. This time period covers the run-up to the Great Recession – the period of greatest financial distress since the Great Depression of the 1930s. Therefore, our predictive model may not apply to other periods of financial distress. However, because of the depth and breadth of the distress of the Great Recession, there has been much attention paid to understanding it and trying to prevent its recurrence. Therefore, there is ample justification for a special focus on the prediction of similar crises in the future. It should be stated that we do not intend to predict the crisis itself but the early warning stage, 7 - 12 quarters prior to the actual systemic crisis event. This time horizon accounts for the announcement period of 12 months specified in the article 126(6) of the EU CRD IV regulation, as well as for a time lag to implement the necessary policies.

2.1 Systemic crisis identification

There are important issues involved in defining "crisis" – the dependent variable in EWS models. The most commonly cited problem is inconsistency in the definition of a banking crisis, which is necessarily defined with a degree of subjectivity (Kaminsky and Reinhart, 1999, Demirguc¸ Kunt and Detragiache, 1998, Eichengreen and Arteta, 2000). There is no unique quantitative variable for banking crisis. The problem lies in the fact that a banking crisis is an event, so proxies for banking crises would not necessarily be perfectly correlated with banking crises themselves. For instance, if we were to use a measure for banking insolvency such as aggregate banking capital, we would need to define a lower bound threshold for a crisis event. However, government intervention or deposit insurance could prevent crisis and the threshold could still be violated.

The bank crisis data we use comes from the literature deployed in this field. Laeven and Valencia (2012) are updating their own (2008, 2010) widely used banking crisis database including all systemic banking, currency and sovereign debt crisis during the period 1970-2011. A banking crisis is defined as systemic if two conditions are met. Significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations) and significant banking policy intervention measures in response to significant losses in the banking system. The authors consider the first year that both criteria are met to be the year when the crisis became systemic. The data show some striking differences in policy responses between advanced and emerging economies as well as many similarities between past and ongoing crisis.

The European System of Central Banks' (ESCB) Macroproduential Research Network (Babecky et. al., 2012) has constructed a quarterly database of the occurrence of banking, debt, and currency crises (or, alternatively, balance of payment crises) for a panel of 40 developed countries over 1970 – 2010. To minimize subjective judgment in defining crisis episodes, they consider various available sources, including both published studies and country experts' opinions based on their survey. The data demonstrate that there is substantial variation in the definition of crises across the published studies. Importantly, one can observe greater discrepancy in the determination of crisis endpoints compared to crisis onsets. To cross-check for the timing of crisis periods, they conduct a comprehensive survey among country experts (mostly from central banks) from all the sample countries.

In a more recent research Laina, Nyholm, Sarlin (2015) define a systemic banking crisis as the occurrence of simultaneous failures in the banking sector, which significantly impairs the capital of the banking system as a whole, and accordingly a crisis mostly results in large economic effects and government intervention. Their database includes banking, currency and debt crisis events for a global set of advanced economies from 1970 to 2012. The database is a compilation of crisis events from a large number of influential papers, which have been cross-checked and complemented by ESCB Heads of Research. They further cross-check and complement the crisis database using events in Laeven and Valencia (2012), Kindleberger and Aliber (2011), Freystatter and Mattila (2011), IMF (2010), Reinhartand Rogoff (2009), Laeven and Valencia (2008), Caprio et al. (2005), Caprio and Klingebiel (2003), and Kaminsky and Reinhart (1999). Using the above sources, they have tried to find consensus in the literature when choosing the crisis periods, particularly from the viewpoint of systemic stress in the banking sector.

Periods of financial distress are characterized by a rapid collapse, following a gradual ascent to a prosperous climax. The collapse is faster than the ascent. The aftermath of a crisis is a regime of financial distress, in which financial authorities deploy ameliorative measures to reverse the collapse. Eventually, recovery ensues. Many EWSs focus on forecasting only the onset of crisis. Others include the immediate post-onset as a period of distress to be predicted. Still others try to forecast the pre-crisis period – before crisis onset when warning signs may become visible (e.g., Lo Duca and Peltonen, 2013). There are valid arguments for all three approaches. The successful forecast of pre-crisis gives policy-makers time to deploy avoidance measures. However, a pre-crisis period is certainly economically distinct from a crisis period, and the economic conditions characteristic of pre-crisis need not always precipitate crisis. Moreover, the intent of early warning can be realized by onset forecasters through sufficiently advancing the time of forecast.

In the present study we will follow the third approach, predicting the pre-crisis period and defining a vulnerable state of the economy from which a banking crisis could emerge (given a suitable trigger). From that perspective our dependent variable is binary, taking the value of 1 between twelve and seven quarters before the onset of a banking crisis. Shorter time horizons are less relevant for policy makers because the potential for effective pre-emptive actions is lower. Most publications of BIS, also, consider a time horizon between 1 and 5 years. In our case we use a time window of 7 to 12 quarters as the most useful compromise between timely prediction and better accuracy. The dependent variable takes the value of 0 for all other quarters in the dataset. Furthermore, some authors observe that the post-crisis period is different from the onset period precisely because of remedial interventions and natural healing processes during the aftermath that are absent during onset (e.g., Bussière and Fratzscher, 2006). They observe that biased forecasts may result from including recovery periods within the periods to forecast, thus all country quarter observations which are in the crisis period, or six quarters after the end of the crisis are omitted from the analysis.

For the 13 EU countries of our sample, we identify the onset of the banking crisis in the first quarter of 2008, while for UK in the first quarter of 2007. We code the binary crisis variable accordingly. All data after the quarters coded as 1, were dropped because as mentioned above the subsequent quarters of a crisis or the tranquil times shorter than 6 quarters after a crisis, could cause post crisis bias. For all countries the dataset starts from the first quarter of 2001. Altogether, there are 13x22

+ 1x18 = 304 quarterly values to be predicted, consisting of 84 ones (vulnerable state or Early Warning Period - EWP) and 220 zeros (tranquil state or Normal Period - NP).

2.2 Explanatory variables

Theory provides ample guidance for selection of crisis prediction variables for EWS models. Following the literature (Kaminsky and Reinhart, 1999; Borio and Lowe, 2002, 2004, and Alessi and Detken, 2011) we survey 14 national level aggregate macroeconomic, market, government and banking time series for each country. In this regard we focus on 10 year bond yields and equities (quarterly mean and standard deviation for these two variables), consumer price index (index 2005=100), gross domestic product (€ billions), production in construction (index 2010=100), real unit labor cost (index 2005=100), unemployment rate (rate), purchase or build home (balance), general government consolidated gross debt (% GDP), general government final consumption expenditure (€ millions), total general government revenue (% GDP), total general government expenditure (% GDP), net current account (€ millions), credit to the private sector (€ billions).

One main problem, pointed out by Peduzzi et. al. (1996) is the number of events (crises in our case) per variable when performing a logistic regression analysis. They indicate that for events per variable values less than 10 the regression coefficients are biased in both positive and negative directions and the large sample variance estimates from the logistic model both overestimated and underestimated the sample variance of the regression coefficients. Rule of thumbs require at least 6 events per variable. Considering these, our final model includes 10 variables from the original 14 variables, excluding total general government expenditure, general government final consumption expenditure, total general government revenue and credit to the private sector. In our case, since the number of events (1's) is 84 the number of 10 variables is acceptable in order to get asymptotically unbiased estimators and valid asymptotic standard errors.

We should underline that all the data collected and used in this paper are publicly and freely available data gathered from Eurostat, World Bank, Organization for Economic Cooperation and Development and the Bank for International

Transactions. Although most of the literature suggests models that are difficult or even impossible to reproduce, our model is transparent. Moreover, the model is very easily and readily applicable for researchers and policy makers to draw direct conclusions about the possibility or not of the future occurrence of systemic banking crises.

Following the literature we pre-processed the data through filters in order to enhance resolution, to make patterns in the data more interpretable and help to meet the assumptions of inferential statistics (Sokal and Rohlf, 1995). Since the crisis for each country comes in the final quarter of a time series, the crisis would be easy to predict successfully if it were legitimate to forecast on the basis of quarter number, population, or any other variable that grows naturally over time: Just classify the quarter of the highest value to be a crisis quarter. In modeling crisis prediction, one must therefore exercise care to avoid spurious correlation between the crisis variable and the predictive criteria. Moreover, as time series, many of our predictive criteria exhibit seasonality and secular trends. It is likely that foreshadowings of crisis may be registered as perturbations that are superimposed upon seasonal and long-term trends. It is necessary to separate the crisis-predictive components from the others. Therefore, we apply pre-processing to all of our predictor time series to enhance the resolution of those perturbations.

First, we apply a four-quarter moving average (MA4) to each predictor time series for each country eliminating seasonality (quarterly data) and random noise. Then in order to eliminate cyclicality, we apply to each of the moving averages the one (left) sided Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997) and not the standard two-sided HP. The standard two-sided HP filter is non-causal as it is not purely backward looking. Hence, it should not be used when predicting future eventsestimating dynamic stochastic general equilibrium models based on recursive state-space representations. The reason is that the HP filter uses observations at $t + i, i > 0$ to construct the current time point t, while the recursive setting assumes that only current and past states influence the current observation. The method used in this study though implements the procedure described on p. 301 of Stock and Watson (1999). The one-sided HP trend estimate is constructed as the Kalman filter estimate of τ_t in the model:

$$y_t = \tau_t + e_t$$

$$(1 - L)^2 \cdot \tau_t = n_t$$

where y_t is the logarithm of the data series, τ_t is the unobserved trend component and e_t and n_t are the mutually uncorrelated white noise sequences with relative variance

$q = var(n_t)/\mathrm{var}(e_t)$.

Subsequently, we take the percentage differences between the moving average and the one sided HP filter (residuals). We look to these residuals to point to the onset of vulnerable state. The HP filter is designed to suppress seasonality and enhance long-term trend. However, we preceded the HP filter by a four-quarter moving average because we found that the HP filter, by itself, with the usual smoothing parameter λ=*1600*, insufficiently damps the seasonality of some of our series. The idea is that the initial four-quarter moving average suppresses seasonality and short-term random noise that is without significance for a crisis. The HP filter then removes long-duration momentum from the moving average, leaving those intermediate-term effects of substance that act to brake or accelerate the economic engine.

As an illustration of the Moving Average and the one sided HP transformation, Figure 1 displays the original and processed time series for quarterly unemployment in Greece.

Summarizing, we apply the following transformations:

- Moving Average of order 4 (MA4)
- One-sided HP
- Percentage differences off the trend

Table 1 reports the summary statistics and Table 2 the unit roots, the t-test for mean differences and (Shapiro – Wilk) normality tests for each variable for the EU 14 for the quarters 2001Q1 to 2014Q1 after doing the above transformations.

Finally, we transform the data to comparable scales using the typical standardization procedure in order to equalize the range and the data variability.

3. Methodology and assessment

The methodology presented in this section consists of two main blocks. First a framework for evaluating signals of early warning models and second the estimation and prediction methods.

3.1 Evaluation of model signals

In this paper we consider statistical binary classification identifying to which of two categories a new observation belongs using several predictor variables. In our case, the two categories are the "*Normal Period*" (*NP*) and the "*Early Warning Period*" (*EWP*), denoted by 0 and 1 respectively. As a statistical hypothesis test the null and the alternative hypotheses are stated as follows:

$$H0: NP = Normal\ Period = (2001Q1, 2004Q4)$$

$$H1: EWP = Early\ Warning\ Period = (2005Q1, 2006Q2)$$

The EWP only for U.K. is one year earlier, that is (2004Q1, 2005Q2). Let the vector $\boldsymbol{x}_{it}$ contain K predictors, known also as independent variables, from the *i*-th country and the *t*-th period (year and quarter). The linear-binary-classifier algorithms produce a linear predictor function, $\beta' \cdot \boldsymbol{x}_{it}$, that predicts the category 1 when it is greater than a threshold value, τ, and 0 otherwise. Once we have the estimation results, we use the approach by Alessi and Detken (2011) to evaluate whether the policy maker can extract useful signals from them. Thus we find the threshold τ that maximizes the performance evaluation criterion using the appropriate Goodness of Fit (GoF) for each model. The best model is the one that achieves the highest score for the performance evaluation criterion.

To get a simplified notation, we use the Iverson-bracket that takes the value 1 if the condition in square brackets is satisfied, and 0 otherwise. Then, the binary dependent variable is denoted as $Y_{i,t}$=[EWP] and is predicted by the variable,

$$Y_{i,t} = \beta' \cdot \boldsymbol{x}_{i,t} > \tau$$

In the hypothesis test, the predictor function is the relevant *test statistic* and the area above the threshold is the *critical region*. Also, the counts of the confusion (classification, contingency, error, or matching) table are denoted as

$n_{k,m} = \quad {}_{i,t}[Y_{i,t} = k,\ Y_{i,t} = m]$ for k,m=0,1.

The notation of the confusion table is given in Table 3 and allows us to visualize the performance of the algorithms.

The confusion table reports the true and false positives and negatives (*TP*, *FP*, *TN*, *and FN*). *TPR* (true positive rate or sensitivity or recall) and *TNR* (true negative rate or specificity) are the percentages of positives (*EWP*, 1's) and negatives (*NP*, 0's), respectively, that are correctly classified ($\%CC_1$ & $\%CC_0$ respectively). Also, *PPV* (positive predictive value or precision) and *NPV* (negative predictive value) are the proportions of positive and negative signals that are correct predicted, $\%CP_1$ & $\%CP_0$, respectively. The proportion of total correct classified *(%CC*) is given by the accuracy measure (*ACC*). The *F* measure is the harmonic mean (HM) of *TPR* and *PPV*. *Type I error* equals to *1 – TPR* and Type II error equals to *1 – TNR.*

The parameter vector β is estimated by the standard linear classifiers for panel data: linear discriminant analysis and linear regression with random effects, panel logistic regression with random effects, and probit regression with random effects regression. The threshold τ is chosen such that the proposed performance criterion EWHM is maximized and it is explained in the following subsection.

A new performance measure and threshold criterion (EWHM)

In practice very often we need just one overall measure of model performance. The F measure does not take into account the value of *TN*. The *ACC* measure just takes into account only the sum of errors, *FP* and *FN*, and not their individual values. Therefore, we propose a new performance measure and threshold extraction criterion, called Error Weighted Harmonic Mean (*EWHM*) of *TPR*, *TNR*, *PPV,* and *NPV* with weights the errors *1-TPR*, *1-TNR*, *1-PPV*, and *1-NPV*, respectively. The *EWHM* is given as

$$EWHM \;\; TPR, TNR, PPV, NPV \;\; = \tag{1}$$

$$= \frac{4 - TPR - TNR - PPV - NPV}{\frac{1 - TPR}{TPR} + \frac{1 - TNR}{TNR} + \frac{1 - PPV}{PPV} + \frac{1 - NPV}{NPV}} =$$

$$= \frac{1 - AM(TPR, TNR, PPV, NPV)}{1 - HM(TPR, TNR, PPV, NPV)} \cdot HM(TPR, TNR, PPV, NPV) =$$

$$= \frac{1 - AM(TPR, TNR, PPV, NPV)}{AM \quad \frac{1}{TPR}, \frac{1}{TNR}, \frac{1}{PPV}, \frac{1}{NPV} \quad - 1}$$

Recall that the Weighted Harmonic Mean *(WHM)* is always smaller than the Weighted Arithmetic Mean (*WAM*) and the error weights make the *WHM* even smaller (see the third part of Equation (1)). In other words, the *WHM* is closer to the smaller values of *TPR*, *TNR*, *PPV*, and *NPV* than to the larger values of them, and the *EWHM* is smaller than the *WHM.*

Table 4 presents the indicative performance of *TPR*, *PPV*, *TNR*, *NPV*, *ACC*, *F* measure, *T1+T2 errors* and *EWHM* for various possible combinations of *FP* and *FN* and constant *TP* and *TN*. The total of *TP*, *FP*, *FN*, *TN* is 304 as the number of the observation of our study presented in Section 4. Also in the application, the sum of *TP* and *FN* equals to 84 as the number of *EWP* and the sum of *TN* and *FP* equals to 220 as the number of *NP*. The values of *FP* and *FN* change from 0 to 10 such that their summation remains the same, equal to 10, with $TP = 84 - FN$ and $TN = 220 - FP$.

As we notice from Table 4, the measure ACC remains exactly the same and the arithmetic mean of *TP*, *FP*, *FN*, *TN* does not change significantly, under these cases. The *F* and *Type I error* plus *Type II error* (T1+T2) measures indicate the case of (*FP*, *FN*) = (10, 0). In the contrary, *EWHM* changes significantly from 90% to 95% giving the maximum value when the difference between *FP* and *FN* is minimized. In this example, *FP* and *FN* become equal to 5. Therefore, the standard deviation of *TPR*, *PPV*, *TNR*, *NPV* is minimized when the *EWHM* takes its maximum value.

3.2 Classification methods

There are many methods described in literature for constructing a prediction model. In our study we use logistic and probit regression which are the most common, but also discriminant analysis and linear panel regression with random effects, which are not widespread. In addition, we propose a new combined method.

Logistic Regression

The logistic regression model assumes that the log-odds ratio is a linear combination of the predictor variables $\boldsymbol{x}$ for the *i*th quarter and the *t*th country, plus error:

$$log \frac{\pi}{1-\pi} = \beta' \boldsymbol{x} + \varepsilon$$

where π is $P_{i,t}^{(1)}$ or else the probability that the (i, t) quarter is in category 1 (e.g., a EWP quarter). Parameters are estimated by maximum likelihood, based upon the Bernoulli likelihood:

$$L \;\; \beta \;\; \boldsymbol{\chi} \;\; = \pi^{y_1+\cdots+y_2}(1-\pi)^{n-(y_1+\cdots+y_n)}$$

$$= (\frac{e^{\beta'\chi}}{1+e^{\beta'\chi}})^{y_1+\cdots+y_n} \cdot (\frac{e^{\beta'\chi}}{1+e^{\beta'\chi}})^{n-(y_1+\cdots+y_n)}$$

and

$$P_{i,t}^{(1)} = \frac{e^{\beta' \cdot x}}{1+e^{\beta' \cdot x}}$$

Probit regression

The probit regression model assumes that the inverse of the standard normal distribution function is a linear combination of the predictor variables $\boldsymbol{x}$ plus error:

$$\Phi^{-1} \;\; \pi \;\; = \beta' \boldsymbol{x} + \varepsilon$$

where π is $P_{i,t}^{(1)}$, i.e., the probability that the observation is in the category 1 (e.g., a EWP quarter). Parameters are estimated by maximum likelihood, based upon the Bernoulli likelihood:

$$L \;\; \beta \;\; \boldsymbol{x} \;\; = \pi^{y_1+\cdots+y_2}(1-\pi)^{n-(y_1+\cdots+y_n)}$$

$$= \Phi(\beta' \boldsymbol{x})^{y_1+\cdots+y_n} \cdot (1 - \Phi \;\; \beta' \boldsymbol{x} \;\;)^{n-(y_1+\cdots+y_n)}$$

and

$$P_{i,t}^{(1)} = \Phi(\beta' \boldsymbol{x})$$

Linear discriminant analysis

The discriminant analysis model supposes that the *p*-variate predictor variables $\boldsymbol{x}$ are multivariate normal, given the classification *y* (0 or 1):

$$f\left(\boldsymbol{x}|y\right) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_y)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_y)}$$

where the covariance matrix $\boldsymbol{\Sigma}$ is assumed to be the same for the two categories, but not necessarily the mean vectors, $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$. The vector $\boldsymbol{x}$ is classified as 1 if the logarithm of the likelihood ratio is below a threshold:

$$log\frac{f\left(\boldsymbol{x}|y=0\right)}{f\left(\boldsymbol{x}|y=1\right)} < c$$

This is equivalent to the linear combination $\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\right) > l$ for some l. Again, the decision rule is based on a score $\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\right) \cdot \boldsymbol{x}$ computed from the predictor variables. The probability (posterior probability) that a quarter t for country i is classified in the category $k = 0,1$ (NP, EWP) is

$$P_{i,t}^{(k|\boldsymbol{x})} = \frac{f_k(\boldsymbol{x})\cdot P_{i,t}^{(k)}}{f_0\left(\boldsymbol{x}\right)\cdot P_{i,t}^{(0)} + f_1\left(\boldsymbol{x}\right)\cdot P_{i,t}^{(1)}}, k = 0,1$$

where $P_{i,t}^{(0)}$ and $P_{i,t}^{(1)}$ are the prior probabilities of each category $k = 0,1$ (NP, EWP) computed as

$$P_{i,t}^{(k)} = \frac{\text{\# of initial observations in } k}{\text{Total \# of observations}}$$

and $f_k(\boldsymbol{x})$ is the probability density for data $\boldsymbol{x}$ if $\boldsymbol{x}$ comes from group k.

Linear panel regression RE

The linear panel regression with random effects model assumes that the binary classification variable *Y* is a linear combination of the predictor variables x plus error:

$$Y_{i,t} = \beta'\boldsymbol{x}_{i,t} + \varepsilon_{i,t}$$

where the error ε satisfies the assumptions of the standard linear regression model. It is logically impossible for ε to meet those assumptions. However, the linear combination $\beta'\boldsymbol{x}_{i,t}$ may still be a successful classifier despite the lack of theoretical purity, as long as the rule $I(\beta'\boldsymbol{x}_{i,t}) > c$ agrees with the data sufficiently often, where I(*condition*) = 1 if *condition* is true and = 0 if *condition* is false. Again

$$P_{i,t}^{(1)} = \beta'\boldsymbol{x}_{i,t}$$

For some of our classification methods (logit, probit, discriminant), the score can be interpreted as a probability of crisis. For linear panel regression, the score is just a number. Whether the scores are probabilities or just numbers, the relevant assessment of the prediction is classification success or failure, for which we provide a variety of informative metrics.

It should be noted that each of the four classification methods has underlying theoretical assumptions that are problematic. Problematic modeling assumptions also affect most of the other literature on EWS. As we note here, that is not of concern for the parts of the analysis devoted to classification success. But it is of concern for the parts devoted to statistical significance of parameters. Use of robust standard errors and/or other corrective measures is warranted. For example, panel logit, probit, and linear regression assume that the crisis indicators are conditionally independent given the predictors. Although that is more likely after the moving average/HP processing than before, time dependence may still be present. Discriminant analysis assumes that the crisis quarters have the same multivariate normal distribution as the non-crisis quarters. That is unlikely, although how important the assumption is in this case is open to question. Satisfying the theoretical specifications of these models would be of concern if one of the objectives of our study were to estimate parameters and probabilities of crisis. It is not. We wish to classify correctly. For classification purposes, we need a rule that successfully classifies. The test is whether the rule classifies correctly. To that end we look to the classical "confusion" table and metrics of success based upon that table (Table 3).

The coefficients of the standardized variables for the four estimation methods are given in Table 5. The size of the coefficients in absolute value reflect their contribution to classification with the following order *Equities (mean), Unemployment Rate, 10y Bond Yield (std), Equities (std), 10y Bond Yield (mean), Gross Domestic Product, Production in Construction, General Government Consolidated Gross Debt, Real Unit Labor Cost* and *Consumer Price Index*. All the signs of the statistically significant coefficients are reversed, except for *10y Bond Yield (std),* from the signs of the t-tests of Table 2 indicating the higher the values of the variables the lower the probability of having *EWP*, for the variables with negative signs *Unemployment Rate, 10y Bond Yield (mean),* and *Production in Construction.* Increase of *Equities (mean)* and *Gross Domestic Product* with unchanged the other variables increase the probability of having *EWP*. It is worth mentioning that *Credit to the Private Sector* that most of the literature highlights as one of the most crucial variables in financial crisis prediction, in our case did not contribute to the proposed classification models.

Combined method

In addition to the other classification methods we propose in this study, a combined classification method for predicting systemic banking crises that has not previously appeared in the literature, to our knowledge. This method is called the Combined Method, as it combines the results of classification of the other used methods and it gives very good prediction results. To apply the combined method, first classify all quarters by each of the other classification methods (logistic regression, probit regression, linear discriminant analysis and linear regression in our case). The combined method then assigns a 1 (*EWP*) to a quarter if two or more of the other classification models assign a 1; otherwise, the combined method assigns a 0. The combined method with two or more 1's performs better than the same method applied with one, three, or four 1's. If m are the classification methods and l, p, r, d stand for panel logistic regression, probit regression, linear regression and discriminant analysis respectively, then the combined method is defined by

$$Y_{i,t}^{m} = Y_{i,t}^{l} + Y_{i,t}^{p} + Y_{i,t}^{r} + Y_{i,t}^{d} > 2 \ .$$

4. Empirical results

All uses of historical data run the danger of over-tuning the model to the past, at the cost of poor prediction in the future. Therefore, in order to assess our model performance, we evaluate its success not only in fitting the in-sample historical data, but also in fitting out of sample data. To do this we follow one version of the Jackknife technique. The jackknife, or "leave one out", technique of cross-validation has a long history of acceptance in the statistical literature for data that are not numerous enough to support a completely separate hold-out sample (e.g. European Central Bank, Financial Stability Review, p. 118, November 2013). Originally, since the observational unit is the quarter, a jackknife technique would leave out one quarter. We apply 14-fold stratified cross validation leaving out all data from one country. We build the model with the data from the remaining 13 countries and successes and failures for the omitted country are tallied. Then we return the omitted country to the dataset and remove a different country. We repeat the process until each country has been rotated out of the data and its quarters classified by the other 13 countries. We then evaluate the set of successes and failures for the combined set of omitted countries as an out-of-sample dataset.

Figures 2, 3, 4 and 5 show the Probability of Default for each country and each quarter, along with the optimal threshold as a horizontal dashed line. The vertical axis can be interpreted as the probability of crisis. Each quarter above the dashed line is classified as *EWP*, each quarter below the line is classified as *NP*.

The extracted thresholds of the classification methods of panel logistic, probit, linear regression and discriminant analysis are 0.643, 0.599, 0.492 and 0.726, respectively.

Table 6 presents the confusion tables with performance measures of the five methods, for in-sample and 14-fold stratified cross validation techniques. As a threshold criterion we used the *EWHM* measure. All five methods give very satisfactory and similar results. The out-of-sample results do not deteriorate substantially from the in-sample ones. The discriminant analysis and the panel linear regression with random effects perform a bit better than the other three methods, showing smaller deviations between in and out-of-sample results. The discriminant and the panel linear methods give the same in and out-of-sample results scoring 85.7% for *TPR* and *PPV* and 94.5% for *TNR* and *NPV* in the out-of-sample case. The *ACC*, *F* and *EWHM* measures are 92.1%, 85.7% and 88.0%, respectively. The *EWHM* is between the *ACC* and *F* measure and smaller than the mean of *TPR*, *PPV*, *TNR* and *NPV* as it was expected and explained above.

Table 7 presents the performance for out of sample analysis (14 - fold stratified cross validation) using *EWHM* and three standard threshold criteria (*ACC*, *F* measure, Sum of *Type I* and *Type II* error) for all methods.

Interpreting Table 7 we conclude once again that the linear panel regression and the discriminant analysis models score the best *EWHM* performance when using the *EWHM* and other standard in literature threshold extraction criteria. Furthermore, 2 - combined method is rated again in the third best place.

The out-of-sample value of TPR is 85.7% implying that within a six-quarter time window period, researchers and policy makers might require as many as five positive signals in order to take prudential measures.

5. Conclusion

The analysis in this paper shows that the global crisis of 2008 for the EU15 countries could be predicted at least 7 quarters in advance by standard macroeconomic variables such as GDP and unemployment rate, by standard market variables such as stock and sovereign bonds (prices, spreads and volatilities) and by the construction and government debt variable. Note that most of the literature finds that different variations of credit variables are the best predictors for systemic banking crisis. A model with such broad variables might predict satisfactorily a future crisis not necessarily banking or loan oriented. Also, financial stability measures could be constructed by computing the probability of default of the utilized models.

The paper proposes an overall goodness of fit (*EWHM*) for classification methods which takes into account all four standard measures (*TPR, TNR, PPV, NPV*) giving more weight to the smaller measures. The *EWHM* is also proposed as a criterion for threshold extraction demonstrating very satisfactory performance.

We apply four classification methods (panel logistic, probit, linear with random effects regression and discriminant analysis). All contribute very adequate results, while most researchers disregard discriminant analysis and linear regression with the excuse of theoretical assumptions, such as normality. Additionally, we could have a combination of the four original methods for classification purposes.

We refine the above methods by applying the one-sided Hodrick Prescott filter that is more realistic than the two-sided HP for predicting future crisis.

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 (3), 520–533.

Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K., and Vašíček, B. (2012). Banking, Debt, and Currency Crises: Early Warning Indicators for Developed Countries. Czech National Bank, mimeo.

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.

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.

Bussire, M., Fratzscher, M. (2006). Towards a new early warning system of financial crises. Journal of International Money and Finance 25 (6), 953–973.

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

Crockett, A. (1998). Progress towards Greater International Financial Stability, Speech at the General Manager of the Bank for International Settlements, at the 'Reforming the Architecture of Global Economic Institutions'.

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

Davis, E. Philip, Karim, Dilruba, & Liadze, Iana (2011). Should multivariate early warning systems for banking crises pool across regions? Review of World Economics, 147, 693–716.

Demirgóc-Kunt, A., Detragiache, E. (1998). The determinants of banking crises in developed and developing countries. IMF Staff Paper 45 (1), 81–109.

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

ECB (2009a). Global index for financial turbulence. Financial Stability Review (December), 21–23 (Box 1).

ECB (2009b). The concept of systemic risk. Financial Stability Review (December), 134–142 (Special Feature B).

Eichengreen, B., Arteta, C. (2000). Banking Crisis in Emerging Markets: Risks and Red Herrings, in Blejer, Mario and Marko Škreb, (Eds.), Financial Policies in Emerging Markets. MIT Press, Cambridge, pp. 47-94.

Freystatter, H., Mattila, V. (2011). Finanssikriisin vaikutuksista Suomen talouteen. BoF Online 1/2011.

Hodrick, R.J., Prescott, E.C. (1997). Post-war US business cycles: An empirical investigation. Journal of Money, Credit, and Banking 29, 1–16

IMF, 2010: "The Financial Stress Index for Advanced Economies" in World Economic Outlook, October 2008, index updated in January 2010.

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

Laeven, L., Valencia, F. (2012). Systemic Banking Crises Database: An Update. Working Paper WP/12/163, International Monetary Fund and the ugly. IMF working paper no. 10/146.

Patrizio, L.,Nyholm, J., Sarlin, P. (2015). Leading indicators of systemic banking crises: Finland in a panel of EU countries. Review of Financial Economics 24 : 18-35.

Lo Duca, Marco, & Peltonen, Tuomas 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.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology, 49(12), 1373-1379.

Reinhart, C.M., Rogoff, K.S. (2009). The aftermath of financial crises. American Economic Review 99 (2), 466–472.

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

Rohlf, F. J., Sokal, R. R. (1995). Statistical tables. Macmillan.

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.

Sokal, R., Rohlf F. (1995). The principles and practice of statistics in biological research. New York: Edition 3. Freeman and Company 887 pp.

Stock, J.H., Watson M. W. (1999). Forecasting inflation. Journal of Monetary Economics, vol. 44(2), pages 293-335, October.

Table 1. Variables list and summary statistics

Variable	MIN	Q1	Mean	Median	Q3	MAX	STD
Gross Domestic Product	-2.45	-0.24	0.13	0	0.57	2.45	0.66
Unemployment Rate	-18.37	-1.42	-0.16	0.08	1.55	8.28	3.65
Consumer Price Index	-0.83	-0.11	-0.02	0	0.16	0.65	0.29
Real Unit Labor Cost	-3.08	-0.9	-0.41	-0.37	0	2	0.79
Production in Construction	-44.76	-1.75	-1.46	-0.38	0.43	8.05	5.7
General Government Consolidated Gross Debt	-13.14	-0.39	0.05	0	0.6	4.39	1.6
Total General Government Expenditure	-4.85	-1.38	-0.68	-0.39	0.01	3.87	1.29
General Government Final Consumption Expenditure	-3.03	-1.37	-0.53	-0.39	0	6.55	1.23
Total General Government Revenue	-5.24	-1.17	-0.22	0	0.64	4.11	1.41
Credit to the Private Sector	-5.2	-0.07	0.68	0.25	1.38	6.14	1.56
Equities (mean)	-9.59	-0.39	8.07	5.64	17.13	25.04	9.67
10y Bond Yield (mean)	-14.23	-3.89	-1.56	-0.87	0	6.38	3.32
Equities (std)	-55.84	-6.33	10.19	3.41	25.63	105.5	28.31
10y Bond Yield (std)	-57.46	-19.45	-11.88	-8.55	-0.41	12.3	13.87

Table 2. Unit roots, mean differences and normality test

Variable	Unit Root Test	t-test for two groups	SW Test for Normality	
			NP	EWP
Gross Domestic Product	-1.79	-6.88	0.94	0.98
Unemployment Rate	-0.15	10.58	0.95	0.87
Consumer Price Index	-0.96	1.4	0.91	0.95
Real Unit Labor Cost	-0.52	-0.34	0.96	0.97
Production in Construction	-3.26	1.38	0.76	0.48
General Government Consolidated Gross Debt	0.48	3.26	0.86	0.8
Total General Government Expenditure	-1.51	4.31	0.91	0.94
General Government Final Consumption Expenditure	-2.36	1.73	0.95	0.7
Total General Government Revenue	-1.49	-4.75	0.99	0.98
Credit to the Private Sector	-0.9	-8.36	0.87	0.99
Equities (mean)	-4.86	-14.26	0.83	0.97
10y Bond Yield (mean)	-11.71	2.05	0.92	0.97
Equities (std)	-0.19	-13.05	0.86	0.97
10y Bond Yield (std)	0.11	1.13	0.83	0.98

Table 3. Confusion table

		Predicted		
	$n = n_{1,\cdot} + n_{0,\cdot}$	***EWP*=1** $n_{\cdot,1} = n_{1,1} + n_{0,1}$	***NP=0*** $n_{\cdot,0} = n_{1,0} + n_{0,0}$	
Observed	***EWP*=1** $n_{1,\cdot} = n_{1,1} + n_{1,0}$	$TP = n_{1,1}$	$FN = n_{1,0}$	*%CC₁* $TPR = n_{1,1}/n_{1,\cdot}$
	NP=0 $n_{0,\cdot} = n_{0,1} + n_{0,0}$	$FP = n_{0,1}$	$TN = n_{0,0}$	*%CC₀* $TNR = n_{0,0}/n_{0,\cdot}$
		%CP₁ $PPV = n_{1,1}/n_{\cdot,1}$	*%CP₀* $NPV = n_{0,0}/n_{\cdot,0}$	$ACC = (n_{1,1} + n_{0,0})/n$ $F = HM(TPR, PPV)$ *EWHM (*given below*)*

Table 4. Confusion table with performance measures

TP	**FP**	**FN**	**TN**	**TPR**	**PPV**	**TNR**	**NPV**	**ACC**	**F**	**T1+T2**[1]	**MEAN**[2]	**EWHM**
74	0	10	220	88.1	100.0	100.0	95.7	96.7	93.7	11.9	95.9	90.0
75	1	9	219	89.3	98.7	99.5	96.1	96.7	93.8	11.2	95.9	91.8
76	2	8	218	90.5	97.4	99.1	96.5	96.7	93.8	10.4	95.9	93.2
77	3	7	217	91.7	96.3	98.6	96.9	96.7	93.9	9.7	95.9	94.2
78	4	6	216	92.9	95.1	98.2	97.3	96.7	94.0	9.0	95.9	94.8
79	5	5	215	94.0	94.0	97.7	97.7	96.7	94.0	8.2	95.9	95.0
80	6	4	214	95.2	93.0	97.3	98.2	96.7	94.1	7.5	95.9	94.9
81	7	3	213	96.4	92.0	96.8	98.6	96.7	94.2	6.8	96.0	94.5
82	8	2	212	97.6	91.1	96.4	99.1	96.7	94.3	6.0	96.0	93.7
83	9	1	211	98.8	90.2	95.9	99.5	96.7	94.3	5.3	96.1	92.5
84	10	0	210	100.0	89.4	95.5	100.0	96.7	94.4	4.5	96.2	91.1

[1] *Type I error + Type II error*
[2] The arithmetic mean of *TP*, *FP*, *FN*, *TN*

Table 5. Estimation results

Variable	Discriminant	Linear	Logit	Probit
Gross Domestic Product	0.39***	0.09***	0.85*	0.47*
Unemployment Rate	-0.59***	-0.14***	-2.94***	-1.63***
Consumer Price Index	0.18	0.04*	0.19	0.1
Real Unit Labor Cost	0.2	0.04*	0.44	0.21
Production in Construction	-0.24***	-0.05**	-0.72	-0.37
General Government Consolidated Gross Debt	0.21***	0.043*	0.6	0.34
Total General Government Expenditure	0.87***	0.23***	3.28***	1.83***
General Government Final Consumption Expenditure	-0.30***	-0.06***	-0.89*	-0.47*
Total General Government Revenue	0.47***	0.12***	0.90**	0.51**
Credit to the Private Sector	0.47	0.1***	1.54**	0.90**
Constant	-	0.28	-2.87***	-1.58***

Table 6. Evaluation for classification by confusion tables

Method	TP	FP	FN	TN	TPR	PPV	TNR	NPV	ACC	F	EWHM
In Sample											
Logit	75	9	6	214	92.6	89.3	96.0	97.3	95.1	90.9	92.1
Probit	75	9	7	213	91.5	89.3	95.9	96.8	94.7	90.4	91.8
Discriminant	75	9	10	210	88.2	89.3	95.9	95.5	93.8	88.8	90.6
Linear	75	9	10	210	88.2	89.3	95.9	95.5	93.8	88.8	90.6
2-Combined	77	7	10	210	88.5	91.7	96.8	95.5	94.4	90.1	91.5
Out of Sample (14 - fold stratified cross validation)											
Logit	71	13	16	204	81.6	84.5	94.0	92.7	90.5	83.0	85.6
Probit	72	12	16	204	81.8	85.7	94.4	92.7	90.8	83.7	86.1
Discriminant	72	12	12	208	85.7	85.7	94.5	94.5	92.1	85.7	88.0
Linear	72	12	12	208	85.7	85.7	94.5	94.5	92.1	85.7	88.0
2-Combined	74	10	16	204	82.2	88.1	95.3	92.7	91.4	85.1	86.9

Table 7. GoF (*EWHM*) for various threshold criteria for all methods

(out-of-sample, 14-fold stratified cross validation)

	Logit Reg	Probit Reg	Discriminant	Linear Reg	2-Combined
EWHM	85.6	86.1	88.0	88.0	86.9
ACC	84.5	85.1	86.3	87.1	86.1
***F* measure**	84.5	85.1	87.1	87.1	86.1
***Type I+II* errors**	83.5	83.2	84.7	84.7	81.6

Figure 1. Effect of time series processing. Left panel: Greek quarterly unemployment rate; Right panel: percentage differences between 4-quarter moving average and HP filter

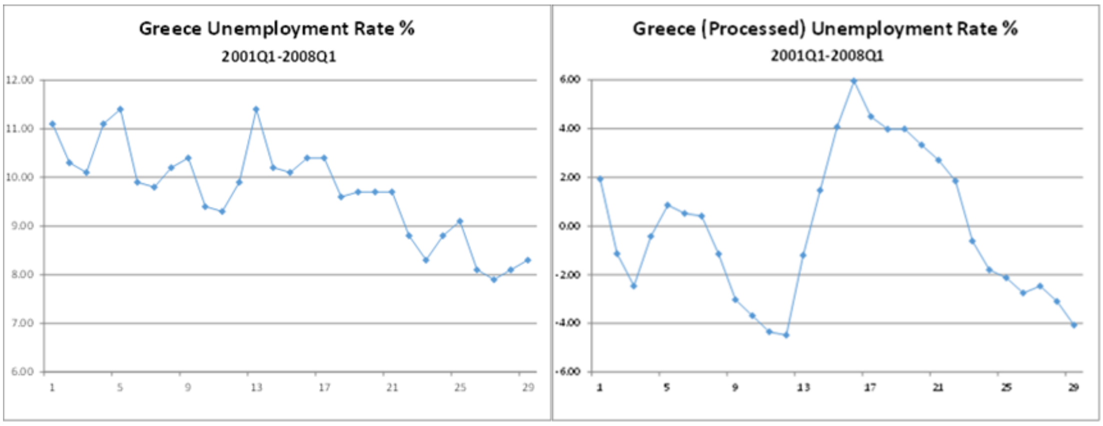


Figure 2. Score and threshold for logistic regression model

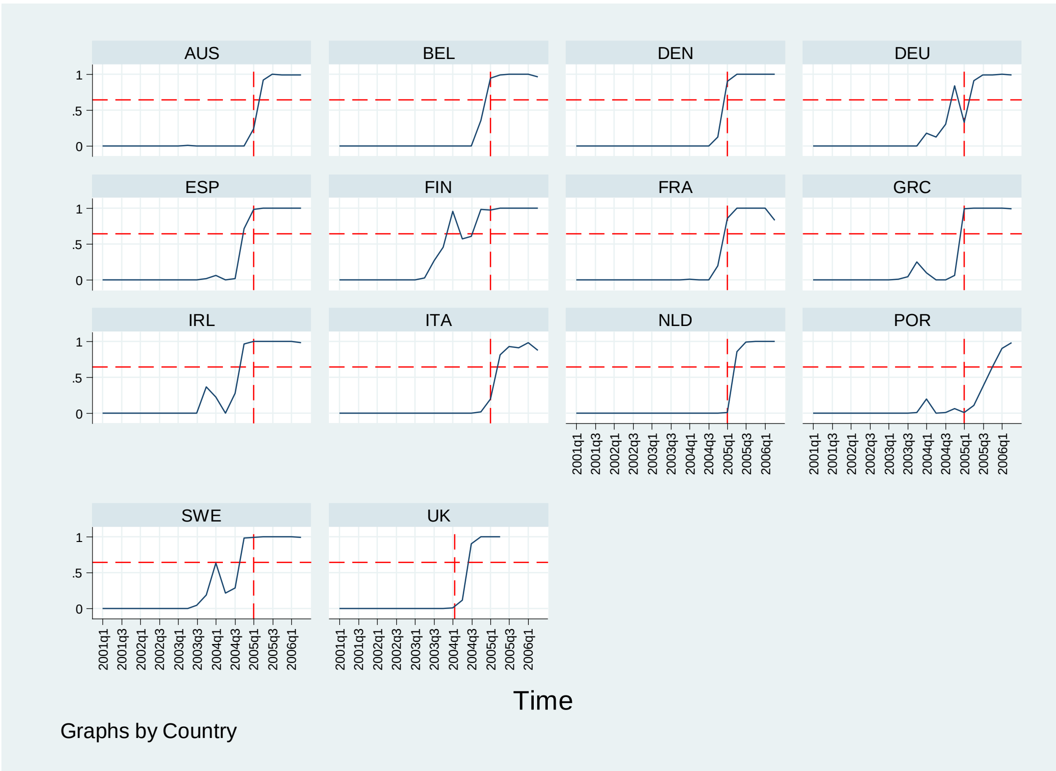


Figure 3. Score and threshold for probit regression model

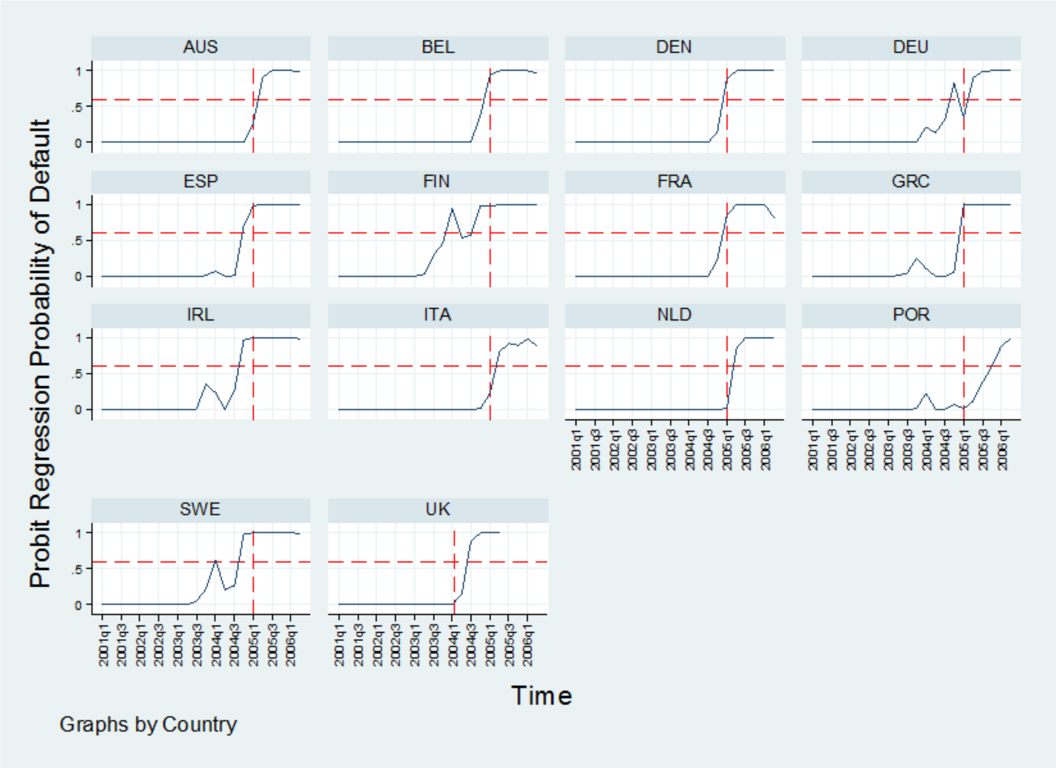


Figure 4. Score and threshold for discriminant analysis model

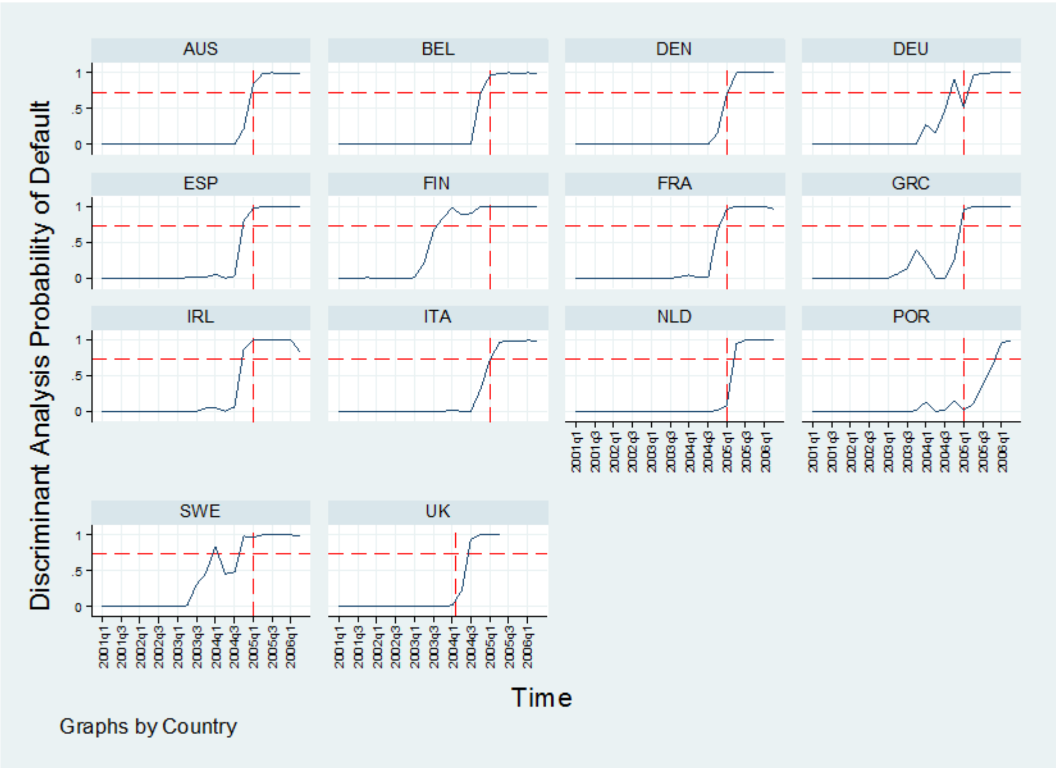
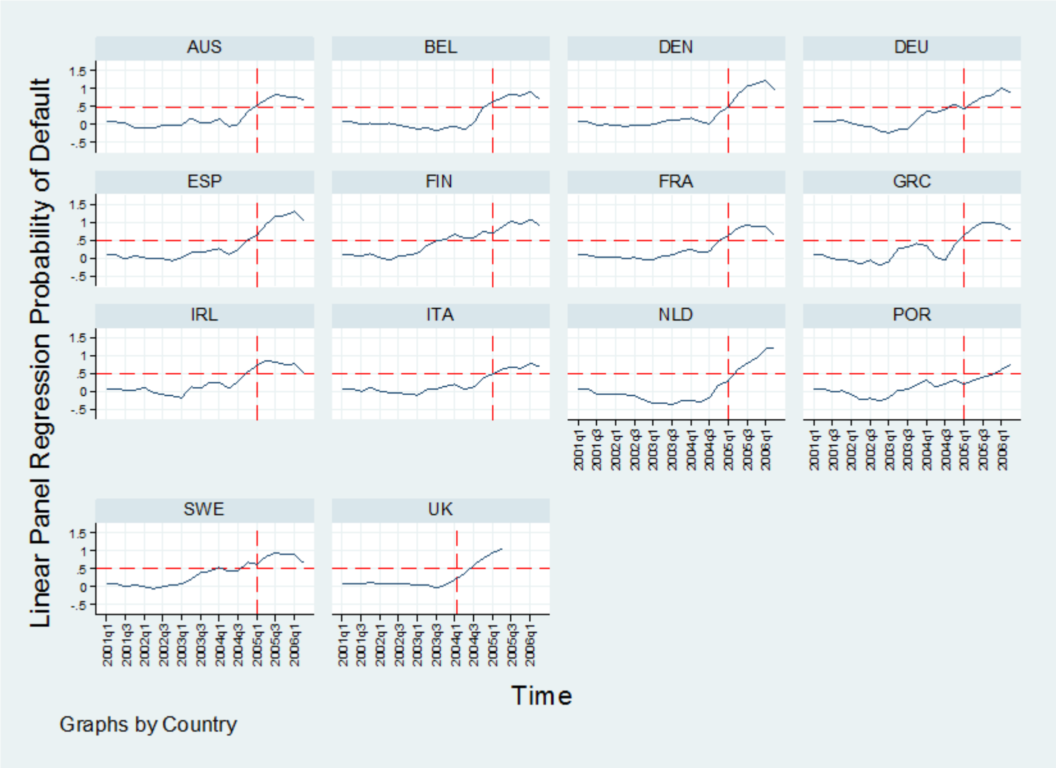


Figure 5. Score and threshold for linear panel regression RE model



BANK OF GREECE WORKING PAPERS

185. Adam, A., and T., Moutos, “Industry-Level Labour Demand Elasticities Across the Eurozone: Will There Be Any Gain After the Pain of Internal Devaluation?” July, 2014.
186. Tagkalakis, O.A., “Fiscal Policy, Net Exports, and the Sectoral Composition of Output in Greece”, September 2014.
187. Hondroyiannis, G. and D., Papaoikonomou, “When Does it Pay To Tax? Evidence from State-Dependent Fiscal Multipliers in the Euro Area”, October 2014.
188. Charalambakis, C. E., “On Corporate Financial Distress Prediction: What Can we Learn From Private Firms in a Small Open Economy?, November 2014.
189. Pagratis, S., E., Karakatsani and E. Louri, “Bank Leverage and Return on Equity Targeting: Intrinsic Procyclicality of Short-Term Choices”, November 2014.
190. Evgenidis, A. and C., Siriopoulos, “What are the International Channels Through Which a US Policy Shock is Transmitted to the World Economies? Evidence from a Time Varying Favar, January 2015.
191. Louzis, D. P., and A.T., Vouldis, “Profitability in the Greek Banking System: a Dual Investigation of Net Interest and Non-Interest Income”, February 2015.
192. Papaspyrou, S.T, “EMU 2.0 - Drawing Lessons From the Crisis - a New Framework For Stability and Growth”, March 2014.
193. Litina, A and T, Palivos, “Corruption and Tax Evasion: Reflections on Greek Tragedy”, June 2015.
194. Balfoussia, H. and H.D. Gibson, “Financial Conditions and Economic Activity: The Potential Impact of the Targeted Longer-Term Refinancing Operations (TLTROS)”, July 2015.
195. Louzis, P. D., “Steady-State Priors and Bayesian Variable Selection in VAR Forecasting”, July 2015.
196. Zografakis, S. and A., Sarris, “The Distributional Consequences of the Stabilization and Adjustment Policies in Greece During the Crisis, with the Use of A Multisectoral Computable General Equilibrium Model”, August 2015.
197. Papageorgiou, D. and E., Vourvachaki, “The Macroeconomic Impact of Structural Reforms in Product and Labour Markets: Trade-Offs and Complementarities”, October 2015.
198. Louri, H., and P., M., Migiakis, “Determinants of Euro-Area Bank Lending Margins: Financial Fragmentation and ECB Policies”, October 2015.
199. Gibson, D. H, S.G. Hall, and G. S. Tavlas, “The effectiveness of the ECB’s asset purchase programs of 2009 to 2012”, November 2015.
200. Balfoussia, H and D, Malliaropulos, “Credit-less recoveries: the role of investment-savings imbalances”, November 2015.
201. Kalyvitis, S., “Who Exports High-Quality Products? Some Empirical Regularities from Greek Exporting Firms”, December 2015.