VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro baigiamasis darbas

Fiskalinio streso indekso vertinimas Europos Sąjungos šalims

Measuring fiscal stress index in European Union countries

Rosita Karietaitė

VILNIUS 2020

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
EKONOMETRINĖS ANALIZĖS KATEDRA

Darbo vadovas Dr. Dmitrij Celov

Darbas apgintas 2020-01-13

Registravimo NR. ______________

Fiskalinio streso indekso vertinimas Europos Sąjungos šalims

Santrauka

Pasaulinė finansinė krizė atkreipė paryškino valstybės skolos tvarumo problemą. Pradėta nagrinėti kaip būtų galima stebėti valdžios sektoriaus finansų sukeliamus svyravimus ir numatyti fiskalinio streso epizodus. Šiame baigiamąjame darbe vertinamas fiskalinio streso lygis 25 Europos Sąjungos šalims. Vertinama naudojant signalizavimo ir panelinės logistinės regresijos metodus. Darbe parodoma, kad į makroekonominių ir fiskalinių kintamųjų rinkinį būtų tikslinga įtraukti ir socialinės tematikos kintamuosius. Taip pat, parodoma šalių heterogeniškumo svarba.

Raktiniai žodžiai: Fiskalinis stresas, panelinė logistinė regresija, valstybės skola, signalizavimo metodas.

Measuring fiscal stress index in European Union countries

Abstract

Global financial crisis increased attention to public debt sustainability. Various methodologies, early warning systems were developed seeking to identify fiscal imbalances and stress episodes. In this thesis fiscal stress level was measured in 25 European Union countries. Index was constructed based on signalling approach and panel logistic regression. Evidence in this analysis shows the importance of social environment for fiscal stress level and suggests inclusion of social oriented variables into predictors' list. Moreover, significance of cross-country heterogeneity has been shown.

Key words: Fiscal stress, panel logistic regression, government debt, signalling approach.

Contents

[1] Chapter based on Fawcett (2006) and Somers (1969)

1. Introduction

The consequences of the global financial crisis, which began in 2008, brought attention to various methodologies, early warning systems (EWS), that could help to identify the collapse before it emerges. Policy-makers prioritized the fiscal sustainability that could be clearly noticed from the analytic work undergone by the European Commission and the International Monatary Fund (IMF). In 2011 the IMF proposed an indicator (so-called S0), with an updated definition of fiscal stress (Baldacci et al., 2011).

Historically, according to various studies, fiscal stress was observed when hyperinflation, IMF programme, government debt default or restructuring took place. However, some episodes may be severe enough to alter the stability of country economics, but not end up as default or near-default. Then Baldacci et al. (2011) proposed to adjust the definition and include a sharp deterioration in market access as criteria of stress episode. Moreover, the authors suggested a set of fiscal variables for building a composite early warning indicator. Later, the list was expanded by other studies, for instance, European Commission suggested to include financial competitiveness variables (Berti et al., 2012). In this thesis, we propose to add income inequality indicators to the predictors' list such as Gini and S80-S20 – income quintile share ratio, which appears to be significant when modelling stress episodes.

In this research, we use quarterly data to assess the risk of fiscal stress in 25 European Union countries. Using higher frequency data helps monitoring the changes in the economic situation more precisely and detect the change points sooner than working with the annual data. Period covered is from 2000Q1 to 2018Q4 (76 quarters). Moreover, the usage of quarterly data results in significantly more observations, which directly impacts model stability and the quality of statistical inference.

Two methodologies are analysed in this thesis. First, signalling approach based on an extended list of predictors. Second, indicator based on logit model. Furthermore, we take into consideration the idea proposed by Manasse et al. (2003) about heterogeneity across countries. With higher frequency data available, thresholds, when fiscal stress signal is sent, can be obtained on country level. The variation of optimal values of threshold suggests that common critical value is not efficient, because for highly developed economies such as Germany or Sweden stress would never be observed, on the contrary, emerging economies like Romania would always send a signal of stress. Moreover, classification precision increased significantly for both methods when cross-country differences were taken into account.

This thesis is organised as follows: section 2 shortly presents previous work in the field. Section 3 describes data and different methodologies. Section 4 provides results of the analysis and performance of models. Finally, conclusions and summary provided in Section 5.

2. Literature review

Various methodologies have been proposed to asses fiscal stress. It can be divided into two main categories. Early Warning System (EWS) models and multivariate regression.

The literature on early warning indicators differs on variables chosen, and composition of the countries are taken into consideration. But the main idea and algorithm to obtain stress remain the same for all reviewed studies. Berti et al. (2012), ECB (2014) and Hernández de Cos et al. (2014) used fiscal and financial competitiveness variables to construct a country-specific composite indicator (so-called S0 indicator) that determines fiscal risk. A set of variables had been selected based on the theoretical background of previous works, the behaviour of variables (anomalous or not), performance in terms of the applied methodology. Optimal threshold, when stress is observed, is chosen by maximizing signaling power based on the analysis of historical data. Maximization of signalling power is done by minimizing type I and type II errors. This technique is employed in this research (see 3.4.1). Then the index is constructed for each country. It is weighted interaction of variables sending fiscal stress signals (reached their optimal threshold) with the weights given by the signalling power of individual variables. Berti et al. (2012) state that financial competitiveness variables perform better than fiscal indicators, in the detection of fiscal stress. Hernández de Cos et al. (2014) show that their proposed indicator would have predicted the fiscal stress in 2007 for the next year with no False Positive and False Negative signals.

The literature on the regression approach is more empirical. It suggests a number of methods that could help to explain how fiscal stress could be triggered. Literature can be split into two main streams: logit/probit models to assess probability to be in stress and Fiscal Reaction Function to measure the relationship proposed by Bohn (1998) that relates primary balance and level of debt as a basis of calculations of fiscal space.

Recent literature on probit/logit models to measure fiscal stress level is very limited. As stated by Baldacci et al. (2011) economy is changing and even the definition of stress was adjusted. So, investigating papers written long ago and predicting the current economic environment based on them would not be reasonable from the methodological point of view. Composition of the explanatory variables – information relevant to carry the early warning signals,– depends on the nature of the stress in the economy and is time dependent. Furthermore, due to the lack of higher than annual frequency data, panel models are widely used to achieve stability of a model behaviour.

In Manasse et al. (2003) study, a different definition of stress is used. Crisis definition includes those cases in which near-default was avoided through the provision of large-scale official financing by the IMF. Aiming to detect fundamentals in affecting the risk of sovereign default and a debt crisis logit model was chosen with a robust variance estimator (Huber White sandwich estimator) with country-specific variances. Also, Classification and Regression Tree (CART) analysis were performed to identify relations between variables that can help to predict whether the country experiencing a crisis. After a comparison of two methods, Manasse et al. (2003) concluded that CART method predicts better but with the cost of sending more false-alarms than logit.

Sumner and Berti (2017) proposed a panel binary response model as a complementary tool to S0 indicator to monitor fiscal stress. The paper concluded the important list of variables to obtain stress. Such as a change in gross public debt. Moreover, the country's effect on classification precision was tested and a model with no cross-country impact was performing better compared to the Fixed effects regression and Signalling approach indicator. Authors suggest that a combination of methods should be used for monitoring, building on the respective strengths of the two approaches, while compensating for their limitations.

To sum up, reviewed works pay a lot of attention to the selection of variables. Many authors suggest the importance of not only fiscal but in addition macroeconomic variables such as real growth or inflation. Also, there is no consensus on if the country-effect should be included.

3. Methodology

3.1. Definition of Fiscal Stress Episodes

In related literature fiscal stress is usually defined as the country's debt crisis. Various indicators can be used to investigate the periods of government funding difficulties. In this thesis definition proposed by Baldacci et al. (2011) is used to indicate stress episodes. Fiscal stress episode has been observed in the country, if any of the below-mentioned conditions hold.

1. Public debt is in default or restructuring[1].
2. The IMF supported programme is present.
3. Hyperinflation – high inflation rate above 35%. Rarely observed after year 2000 in EU. Only Romania experienced hyperinflation from 2000Q1 to 2001Q2.
4. Extreme financing constraint of the sovereign – sovereign bond yield spreads greater than 2 standard deviations from the country average. The latter were defined comparing German six months bonds interest rate used as risk-free rate with country-specific 10 years government bonds interest rates.

3.2. Signalling window

The core of the EWS methodology is to obtain a crisis before its actual emergence. Otherwise, the indicator is not useful. Subsequently, stress episodes were transformed.

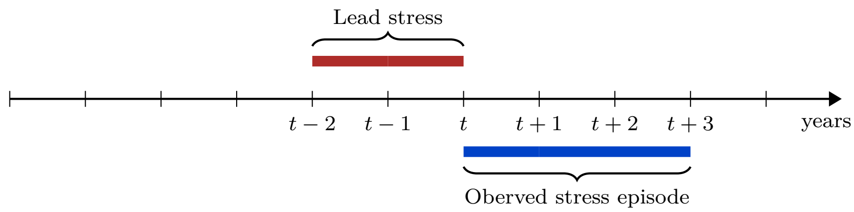


Fig. 3.1.: Stress episodes scenarios

[1]Due to data unavailability only default data from S&P was used. This will result in index sensitivity for debt adjustments.

Taking into account, that new decisions by policy-makers can be effective in no earlier than half a year, two quarters before the beginning of the actual episode is considered as lead stress (see fig. 3.1). Furthermore, following Drehmann and Juselius (2014) data from a period of stress were omitted, relying on the assumption that the economy reacts to stress and behaves systematically different. That implies crisis period data could cause bias in the final indicator. Similar methodology was applied by Banbula and Pietrzak (2017) when analysing the banking crisis.

In this thesis, lead and observed scenarios of stress episodes, corresponding to transformed vector of values and actual observed data, were analysed and compared.

3.3. The data set

Fiscal stress can be caused by a wide range of factors. It can take weak fundamentals in the fiscal part of the economy. In this case, stress reflects in government debt levels or deficits. But it can also be caused by the shocks from both domestic and external sectors of the economy. As 2008 crisis demonstrated the stress in the private sector of one country can easily grow into global economic crisis. Moreover, social and living conditions in the country may be a crucial indicator, signalling internal vulnerabilities and sensitivity to the fiscal stress. For example, financial competitiveness data may look impeccable at first glance, with a cost of inequality or people at risk of poverty peaking in the country. Consequently, it is critical to take a wide range of data into account while estimating fiscal stress.

In this thesis financial competitiveness data along with fiscal and social conditions variables will be used to measure budgetary situation in 25 European Union countries sampled from 2000Q1 to 2018Q4 (76 quarters). Luxembourg, Croatia, and Hungary had to be omitted due to very limited data availability, resulting in the final dataset consists of 25 EU countries. Data set was chosen based on the results of previously mentioned researches, the behaviour of variable[2] and performance in terms of the method chosen. Eurostat, the ECB and the IMF databases are the main sources of data. All variables listed in appendix E.

3.3.1. Data preparation process

Before modelling, data tidying has to be performed due to comparability and irregularity issues, such as: measurement errors, missing observations, time series of different lengths (usually, lack of history). All of these and similar issues can be very influential to results and inferences if handled inadequately. Therefore, before application of any methodology, data related issues have to be resolved.

[2]Excluding slowly varying variables, variables that have no significant or no at all effect on the quality of early warning signal.

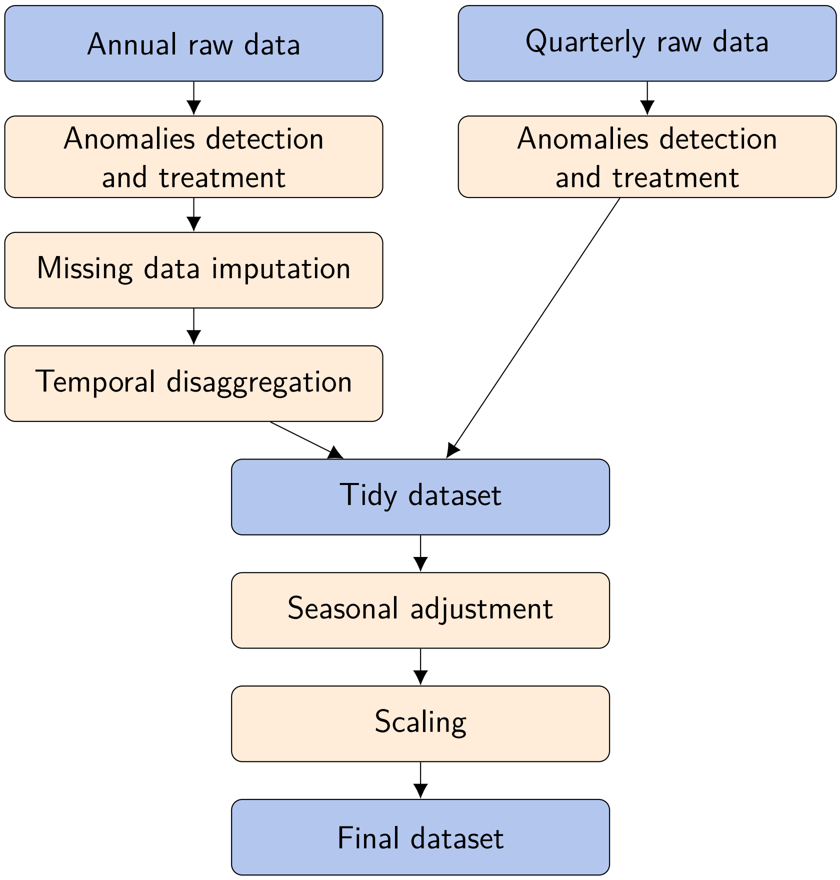


Fig. 3.2.: Data pre-processing

The database used in this thesis is a mixed frequency. The aim is to have a list of quarterly pre-processed (see Figure 3.2) variables starting the year 2000. The starting point was chosen based on the most frequent start point of time series, consequent long enough time series (76 quarters). But not all desired data is available in quarters. Some variables, inequality measurements, for example, observed only annually. A few variables have a lack of history issues for a higher frequency but contain full-time series since the year 2000 as annual. For those two groups, lower frequency data were interpolated using the Denton-Cholette method from **tempdisagg** library in R, to quarterly time series (Sax and Steiner, 2016). This method primarily concerned with movement preservation, generating a series that is similar to the indicator, whether or not the indicator is correlated with the low-frequency series (Eurostat, 2018). In this case, the trend is a crucial part to recreate, as signals will be used to construct a fiscal stress index. After preprocessing and disaggregation annual and quarterly variables were merged into the final dataset.

Before disaggregation, the standard procedure of data preparation was performed (see Figure 3.2). Different frequency data were treated separately. To begin with, outliers for both, annual and quarterly datasets, the presence of abnormal one-off deviations (additive outliers) from other values can reduce the overall quality of temporal disaggregation exercise so they should be corrected before running the interpolation procedure (Eurostat, 2018). Only additive outliers were considered following the best Eurostat and OECD practise.

The approach, initially proposed by Chen and Liu (1993) uses an iterative two-step procedure:

- Simultaneously detects additive outliers upon a chosen ARIMA model;
- Chooses/refits the ARIMA model including the additive outliers detected in the previous step and removes those that are not significant in the new fit.

The data then is adjusted for the detected outliers and the stages listed above are repeated until no more outliers are detected or until the maximum number of iterations (in this thesis, 10 is used as maximum) is reached.

The next step in data preparation is missing values imputation. In many cases, data are only available for a limited number of countries or only for certain variables. Seeking to have a complete time series dataset, missing values have to be replaced. This step was completed for the annual dataset only. Lower frequency time series contains significantly less missing values, so in this way, data was kept closer to reality as the time series models prediction accuracy deteriorates with the longer prediction horizons. Based on the same reason, variables with more than 5 years of data missing were not considered.

The strategy followed to develop a database of long time-series is based on predictions from univariate unobserved components models. Most of the remaining missing values in the raw data were either at the beginning or the end of the sample. Hence, this problem is equivalent to backcasting and forecasting. Exponential smoothing (ETS) and Seasonal autoregressive integrated moving average (SARIMA) methods were used as a solution. The choice between method based on out-of-sample validation. Predicting and validation methodology based on Hyndman and Athanasopoulos (2014). The sample was split into training (2005Q1-2018Q4) and test (2000Q1-2004Q4) sets. Accuracy of backcast results from both methods measured using MASE (mean absolute scaled error) which is units-free and has correct specification for the variables with both positive and negative values. Imputation strategy was chosen for every time series individually considering MASE values.

After missing values imputation and temporal disaggregation, different frequency data were merged resulting in one tidy dataset (see Figure 3.2).

A further stage is seasonality and calendar adjustments. In a few cases when seasonally adjusted data was unavailable, corrections have to be done. Seasonal fluctuations and calendar effects can mask short and long-term movements in time series and impede a clear understanding of underlying phenomena (Eurostat, 2015). In this thesis, new trends and turning points are the area of interest as fiscal stress is not a periodically repeated event. For seasonally unadjusted data own seasonal adjustment calculations were done applying

X13ARIMA-SEATS seasonal adjustment functions following Eurostat recommendations (Eurostat, 2015). X13ARIMA-SEATS is the US Census Bureau's latest program that implements both the X11 (Shiskin et al., 1967) and SEATS (Gómez and Maravall, 1996) methods plus some additional diagnostic and model selection tools.

To make variables comparable scaling has been performed. Fiscal and macroeconomic literature suggests expressing macroeconomic variables as a percent of nominal GDP. Scaling allows focussing on structural changes and intensity of socio-economic fluctuations purifying specific imbalances.

3.4. Early Warning System Approach

The signalling approach is based on the idea that economies behave in a systematically different way in the period preceding fiscal stress (Berti et al., 2012). The method allows using a set of explanatory variables to observe the stress level in the country and detect structural changes in the behaviour of the economy. This method is widely used in literature (Baldacci et al., 2011; Berti et al., 2012; Hernández de Cos et al., 2014).

An early warning system is composed of four components. The first important step in conducting any data-driven analysis is to specify the dataset and pre-process variables properly. A detailed description of this step is provided in 3.3.1. Second, fiscal stress episodes have to be specified following the definition provided in subsection 3.1. Then, thresholds are determined for each selected variable. This step is the core of the method optimizing the strength of the signal detecting the economies in fiscal stress. After examining data by variable, the composite indicator has to be built. Having one composite indicator instead of a long list makes surveillance of risk easier and more comprehensive. The idea is that the inference regarding the fiscal stress is more reliable when the same signal is supported by the most of indicators. The latter three steps are discussed in more depth below.

3.4.1. Critical thresholds calculation

A crucial part of the signalling system is to identify optimal thresholds when a variable sends a signal of stress. Ideally, the chosen value always predicts stress or no stress episodes correctly. But in practice, it is very unlikely that one characteristic can predict the behaviour of a complex economy without errors. Subsequently, the composite indicator is a proposed solution to this problem.

Each variable is signalling an increasing risk of funding difficulties when taking value below or above, depending on the indicator of interest, optimal limit. To find the correct threshold, a number of misclassified episodes have to be minimized. Prior to this, the probability of having type I – variable indicated stress, but no stress episode was observed,–

	Fiscal stress episode (FS)	No-fiscal stress episode (NFS)
Predicted fiscal stress episode	True Positive signal	True Positive signal (FP), Type I error
Predicted no-fiscal stress episode	False Negative signal (FN), Type II error	True Negative signal

Table 3.1.: Possible classification cases

or type II – variable did not indicate stress, but stress occurred,– errors. Four possible combinations of classification reported in table 3.1 in the form of confusion matrix.

For each variable i the optimal threshold t_i^* will be calculated by minimising sum of type I and type II errors. To achieve the latter total misclassification error (TME) was used.

$$t_i^* = \arg\min_{t_i \in T_i}(TME_i(t_i)) = \arg\min_{t_i \in T_i}\left(\frac{FN_i(t_i)}{Fs} + \frac{FP_i(t_i)}{Nfs}\right), i = 1, ..., n,$$

where T_i – the set of all values taken by variable i across all countries and time, n – number of variables, *FN* – total number of false negative signals, *FP* – total number of false positive signals, *Fs* – sum of fiscal stress episodes, *Nfs* – sum of fiscal not-stress episodes.

Every variable predicts with different precision, so weights need to be considered when building composite indicator. Interpretation of weights is straightforward – the bigger value of weight is, the higher "signalling power" (z_i) the variable has:

$$z_i = 1 - \left(\frac{FN_i(t_i)}{Fs} + \frac{FP_i(t_i)}{Nfs}\right).$$

3.4.2. Composite indicator

Early warning indicator *S0* is built following the methodology mentioned in most of the literature of early warning indicators e.g. Baldacci et al. (2011), Berti et al. (2012), Hernández de Cos et al. (2014) . The index combines earlier described "signalling power" and threshold to observe stress level by country, using various indicators:

$$S0_{jt} = \sum_{i=1}^{n} w_i d_{jt}^i = \sum_{i=1}^{n} \frac{z_i}{\sum\limits_{k=1}^{n} z_k} d_{jt}^i,$$

where w – weight and d – dummy variable, which shows whether value of indicator i for country j at time period t sends a signal of stress.

Moreover, methodology of identifying optimal threshold values (see section 3.4.1) applied to *S0* by country and for a panel. As quarterly data is used in this thesis, the country-specific threshold is more stable compared to annual, proposed by Hernández de Cos et al. (2014). As a result, the country level stress benchmark was obtained. As economies in European union are very heterogeneous, it would be imprecise to use a common value to decide if a country faces fiscal tension.

3.5. Binary response panel data regression

Regression analysis was also used to estimate fiscal distress in the country, which is more formally described as:

$$p\left(stress\ episode_{t,i}\right) = \phi\left(\Delta debt_{(t-4,t-1)}; Controls_{i,t})\right). \tag{3.1}$$

Two types of panel logistic regression were estimated. First, without taking country effect into account (Pooled). Second, the same model was assessed with country dummy variables (Fixed Effects), to confirm the significance of country level heterogeneity in the EU. List of controls was chosen as a subset of variables previously used to obtain *S0* indicator applying elastic net (EN) regularization technique proposed by Zou and Hastie (2005).

Sumner and Berti (2017) also include lagged debt ($debt_{t-1}$) in the model equivalent to equation 3.1. Various lags and transformations of debt were considered in this thesis, yet only annual change of debt was chosen as a significant explanatory variable by EN.

The model was estimated using a penalized least squares estimator. EN approach does both continuous shrinkage and automatic variable selection simultaneously and it can select groups of correlated variables. Moreover, often outperforms the another widespread sparse least absolute shrinkage and selection operator (LASSO) approach in terms of prediction accuracy Zou and Hastie (2005). Estimate from EN defined in equation 3.2:

$$\hat{\beta}_{enet} \equiv \underset{\beta}{\operatorname{argmin}}(\|y - X\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1), \tag{3.2}$$

where $X - t \times n$ predictor matrix of standardized variables, $y-$ response vector, $\lambda_1 \geq 0, \lambda_2 \geq 0$, $\|\beta\|_1 = \sum_{i=1}^{p} |\beta_i|, \|\beta\|^2 = \sum_{i=1}^{p} \beta_i^2$.

Seeking to assess the performance of each model, reported some goodness-of-fit characteristics, such as AUROC, Somers' Delta (see appendix D), overall predictive accuracy. Considering that the aim is to predict crisis episodes, it can be concluded that the most important measure, in this case, should be the percentage of crisis episodes correctly classified, balancing the I and type II errors in the same way as for signalling approach. A probability cut-off of the crisis appearance involves either pooled or country-specific thresholds derived from the regression analysis. The probability prediction and the choice of thresholds are two separate steps. Therefore, country-specific thresholds are deemed to result in higher detection rate of the stress episodes in out-of-sample (10-fold) validation, whilst the pooled threshold is expected to perform better for the comparison of more homogeneous groups of countries (clusters).

4. Results

4.1. Signalling approach

As discussed before, the signalling approach depends on critical thresholds. If variable exceeds or is lower, depending on the variable of interest, than critical cut-off value, a signal of fiscal stress is sent. Overview of critical values and signalling powers along with other characteristics are reported accordingly in tables 4.1 and 4.2 for observed stress episodes (Scenario 1) and with lead stress transformation (Scenario 2). Results generalize 25 EU countries in the time period from 2000Q1 until 2018Q4.

Comparing results reported in two tables can be concluded that the transformation of stress episodes timing makes a significant impact on threshold values and their characteristics. In the case, when stress episodes are observed (Scenario 1), income distribution measures Gini and S80-S20 have high signalling power, 38% and 44% accordingly. Fiscal variables perform well as expected. Balance is the second best after S80-S20 measure and has 43% signalling power, also primary balance's signalling power is 35%, but the latter has a high missed crisis ratio.

After making stress episodes earlier and removing crisis periods' data, threshold results change. Social inequality indicators show lower signalling power, though perform still quite well. Hence, the hypothesis that social situation impacts the fiscal stress level should not be rejected. Fiscal indicators such as balance and real GDP growth takes the leading positions with the signalling power of 36% and 33% correspondingly.

Next, $S0$ indicator is calculated based on single-variable thresholds for both scenarios. Pooled and country-specific thresholds for composite indicator were also computed together with signalling power (see 4.1 and 4.2). Cut-off's by country reported in appendix A (see tables A.1 and A.2). A visible variation between country-specific threshold values proves the importance of countries heterogeneity in both scenarios of fiscal stress episodes.

For the early warning indicator, $S0$ optimal common threshold is 0.57. Above this value signal of stress is sent by the indicator. Index with this critical value missed 18% of crisis episodes and sent false alarm 24% of the time. Overall signalling power is 0.61. Performance of $S0$ is improved by 10% when a critical value is country-specific. To highlight the heterogeneity across countries results: lowest critical value 0.16 for Austria and Portugal reaches the limit of 0.75. Lithuania has a value of 0.68. Compared to other countries,the Portuguese and Lithuanian cases indicate a higher overall fiscal stress level than in most countries.

Variable	Threshold		Signalling power	False alarms ratio	Missed crisis ratio
Gross debt, % GDP	108.99	>	0.18	0.05	0.77
Yearly change in gross debt, % GDP	1.78	>	0.26	0.15	0.49
Short term debt, % GDP	0.00	>	0.04	0.96	0.00
Interest rate growth	1.34	>	0.15	0.27	0.58
Change in expenditure GG, % GDP	1.63	>	0.12	0.14	0.75
Change in final consumption GG, % GDP	1.10	>	0.12	0.05	0.83
Nominal interest, % gross debt	0.86	>	0.22	0.65	0.13
Debt, non-financial corporations, %GDP	127.69	>	0.11	0.17	0.73
Short term debt, non-financial corporations, % GDP	7.48	>	0.07	0.89	0.05
S80-S20	4.45	>	0.44	0.47	0.10
GINI	28.73	>	0.38	0.51	0.11
Expenditure on social protection, % GDP	34.49	>	0.06	0.00	0.94
Expenditure on pensions, % GDP	15.18	>	0.14	0.04	0.82
Unemployment	6.60	>	0.29	0.64	0.07
Balance, % GDP	-3.38	<	0.43	0.29	0.28
Primary balance, % GDP	-1.64	<	0.35	0.25	0.40
Current account, 1Y MA, % GDP	-1.09	<	0.35	0.41	0.24
Nominal unit labour cost	-0.30	<	0.25	0.16	0.58
Part of non-employed	42.72	<	0.21	0.15	0.64
Fertility rate	1.51	<	0.23	0.50	0.28
Real GDP growth	-0.42	<	0.24	0.11	0.66
S0 (panel threshold)	0.57	>	0.61	0.18	0.21
S0 (country-specific threshold)	A.1	>	0.71	0.15	0.14

Table 4.1.: Thresholds' overview for observed stress

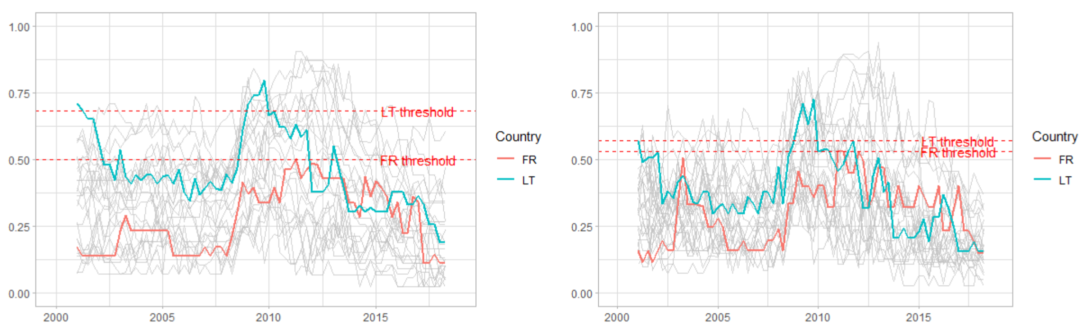
As expected, performance of the index became worse when the stress period is transformed. For the whole panel of countries, the optimal threshold is lower than before transformation: 0.32. Lower value results in a high – 0.39% false alarms ratio. Signalling power became 0.58 and only 2% of missed crisis.

Comparing all analysed scenarios, the first conclusion is country importance when assessing stress. All EU economies are different and what could be considered as stress for developed economies such as Germany or France, may be usual instabilities for others, such as Lithuania or Romania. To illustrate the statement see figure 4.1. France did not have any stress episodes, it can be seen in lower values of $S0$ index. For that reason, a common threshold is less efficient when predicting the crisis, especially for developed economies.

Variable	Threshold		Signalling power	False alarms ratio	Missed crisis ratio
Gross debt, % GDP	115.94	>	0.13	0.06	0.81
Yearly change in gross debt, % GDP	1.36	>	0.31	0.21	0.48
Short term debt, % GDP	3.44	>	0.17	0.51	0.33
Interest rate growth	5.03	>	0.16	0.13	0.71
Change in expenditure GG, % GDP	0.75	>	0.14	0.34	0.52
Change in final consumption GG, % GDP	1.29	>	0.10	0.04	0.86
Nominal interest, % gross debt	0.85	>	0.18	0.66	0.16
Debt, non-financial corporations, %GDP	88.73	>	0.08	0.44	0.48
Short term debt, non-financial corporations, %GDP	5.63	>	0.05	0.95	0.00
S80-S20	4.50	>	0.26	0.46	0.28
GINI	30.53	>	0.23	0.41	0.36
Expenditure on social , % GDP	24.92	>	0.12	0.42	0.47
Expenditure on pensions, % GDP	13.86	>	0.15	0.12	0.72
Unemployment	12.10	>	0.07	0.15	0.78
Balance, % GDP	-3.38	<	0.33	0.31	0.36
Primary balance, % GDP	-1.05	<	0.32	0.35	0.33
Current account, % GDP	-1.50	<	0.35	0.39	0.26
Nominal unit labour cost	-1.40	<	0.07	0.09	0.84
Part of non-employed	42.95	<	0.16	0.17	0.67
Fertility rate	1.66	<	0.05	0.68	0.28
Real GDP growth	-0.41	<	0.36	0.12	0.52
S0 (panel threshold)	0.32	>	0.58	0.39	0.02
S0 (country-specific threshold)	A.2	>	0.68	0.16	0.16

Table 4.2.: Thresholds' overview for transformed stress

Fig. 4.1.: S0 values in time



(a) Observed stress episodes

(b) Transformed stress episodes

Finally, the main difference between a lead and observed stress indicator is the missed crisis ratio. If obtain the crisis is more preferable than miss one – lead stress should be used.

4.2. Binary response panel model

Four regression models presented below for each fiscal stress episode scenario together with the goodness-of-fit measures.

One particular aspect of research requires attention prior to further analysis of the estimation results. Usually, standard errors are provided together with regression coefficients. However, as penalized regression were estimated in this thesis, we deliberately do not provide standard errors, because standard errors are not very meaningful for strongly biased estimates such as arise from penalized estimation methods. Penalized estimation is a procedure that reduces the variance of estimators by introducing substantial bias (Goeman, 2010).

For all models, control variables were selected applying the Elastic Net algorithm. For potential explanatory variables lags up to 5 time periods were tested, but only change of gross debt and primary balance was selected by the algorithm. In some cases both current and lagged value was significant with opposite signs of coefficients. In such cases, change term was included in the model to reduce the number of variables.

The first model is pooled panel (Model 1 – observed stress, Model 3 – lead stress) logistic regression. The second (Model 2 – observed stress, Model 4 – lead stress) is panel fixed effects model. Countries' effects were added as dummy variables to the panel. Germany was chosen as a base variable corresponding to the common intercept and was not included in the model directly. Thus, estimated parameters for dummies show if the deviation from Germany on average is statistically significant. Results are presented in table 4.3.

Model 1 and 2 include significant variables such as a change in gross debt and primary balance. Signs of coefficients are in line with other authors (Sumner and Berti, 2017; Manasse et al., 2003). Income inequality indicators also appear in the list of selected variables. Intercept in the case of Model 2 is lower. It means that the start point to stress for Germany is lower than for the country pool. The coefficient of real GDP growth in Fixed Effects case increases significantly compared to Model 1, also short term debt of General Government appears in Model 2 with a highly negative coefficient. This result suggests that the level of short-term debt is only meaningful mainly to obtain a scope of stress in the country and represents features of the country that are less observable.

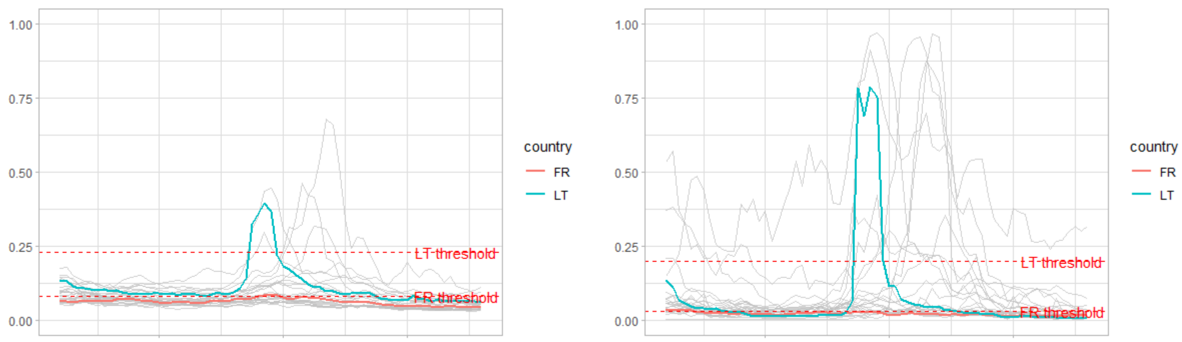
Models 3 and 4 with transformed stress episodes include a smaller number of significant explanatory variables. Primary balance, real GDP growth, interest rate and change in government debt can be concluded as robustly selected indicators. They were selected by

	Scenario 1		Scenario 2	
	Model 1	Model 2	Model 3	Model 4
(Intercept)	-3.97	-5.08	-3.68	-4.36
Δ Gross debt $GG_{(t-4,t-1)}$	1.28	3.16	2.21	6.60
Short term debt, GG, % GDP		-9.79		
Final consumption GG, % GDP	-0.71	1.22		
Change consumption GG, % GDP				-0.46
Nominal unit labour cost		-0.03	0.01	0.02
Current account, 1Y MA			-0.01	-0.02
Debt, non-fin corp, % GDP		0.67		
S80-S20	0.11	0.03		
GINI	0.02			
Unemplouyment	0.002			
Interest rate	0.14	0.37	0.05	0.05
Real GDP growth	-0.27	-9.99	-8.74	-31.25
Primary balanse$_{t-1}$, % GDP	-5.30	-5.17	-0.76	-2.63
Gross debt GG_{t-1}, % GDP		1.68		
Country effect	No	Yes	No	Yes

Table 4.3.: Logit regression results

algorithm in all four situations. Real GDP growth shows the same trend in both Scenario 1 and Scenario 2. The coefficient is significantly larger in the case of fixed effects. This indicates the importance of cross-country heterogeneity of GDP growth patterns when assessing the fiscal stress.

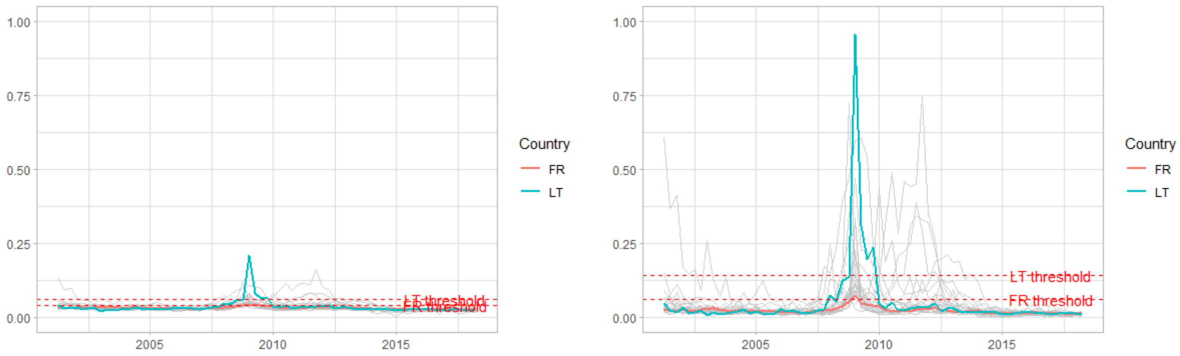
Fig. 4.2.: Model 1 and Model 2 values in time



(a) No country effect (b) With country effect

Similar to in the Signalling approach, values of country-specific optimal thresholds were obtained minimizing total misclassification error. Examples in figures 4.2 and 4.3, and accuracy measures (see table 4.4, ROC plots depicted in appendix C) demonstrate the importance of county effect for classification precision. In the case of pooled model (Model 1 and 3) probabilities are close to zero for all countries. It implies, that model only obtains stress when the level is actually peaking. This suggests that the model does not explain

Fig. 4.3.: Model 3 and Model 4 values in time



(a) No country effect

(b) With country effect

the case of stress properly and should not be used for predictions.

It can be seen from example plots the variety of trajectories. With evidence of accuracy measures can be concluded that country effect is crucial in order to have a stable model.

	AUROC	Somers' Delta	Country effects
Model 1	0.81	0.62	No
Model 2	0.92	0.83	Yes
Model 3	0.74	0.49	No
Model 4	0.80	0.59	Yes

Table 4.4.: Accuracy measures

While AUROC and Somers' Delta values relatively high for all models it can be clearly seen that models with observed stress episodes (Model 1 and 2) solve the classification task better. Model 2 outperforms all according to AUROC and Somers' Delta.

5. Conclusions

In this thesis fiscal stress level was measured by eight different approaches. Overview of methods reported in table 5.1.

From a classification precision point of view, Fixed Effects regression with observed stress has the highest signaling power, compared to other methods. However, the literature suggests removing crisis period data when modelling the probability of being in stress Drehmann and Juselius (2014). Models with lead stress have a higher false alarms ratio. For the banking sector has been concluded that type I errors (FN) are more costly than type II errors ESRB (2017). There is no literature or evidence yet that concludes the same for the government sector crisis. Therefore, could be an interesting point for analysis.

Method	Stress	Country effect	Signalling power	False alarms ratio	Missed crisis ratio
Signalling	Observed	No	0.61	0.18	0.21
Signalling	Observed	Yes	0.71	0.15	0.14
Signalling	Lead	No	0.58	0.39	0.02
Signalling	Lead	Yes	0.68	0.16	0.16
Regression	Observed	No	0.71	0.18	0.11
Regression	Observed	Yes	0.81	0.14	0.05
Regression	Lead	No	0.63	0.23	0.14
Regression	Lead	Yes	0.66	0.21	0.12

Table 5.1.: Methods overview

Equally important conclusion is cross country heterogeneity importance for model performance. For all tested scenarios, the model with country effects has significantly higher signalling power. To support the evidence from signalling power, in the regression case, ROC curves (see appendix C) together with other accuracy measures illustrate the result visually.

Last but not least, the social environment in a country should be considered when assessing the fiscal stress level, next to fiscal and macroeconomic measures. Only in regression with lead stress income inequality variables were not significant. In the rest of the models, these variables suggest that with an increase in inequality country risk to be in fiscal stress increases.

Bibliography

(2014). Early warning indicators for fiscal stress in European budgetary surveillance. Monthly bulletin, European Central Bank.

Baldacci, E., Dobrescu, G., Petrova, I., and Belhocine, N. (2011). Assessing Fiscal Stress. IMF Working Papers 11/100, International Monetary Fund.

Banbula, P. and Pietrzak, M. (2017). Early warning models of banking crises applicable to non-crisis countries. NBP Working Papers 257, Narodowy Bank Polski, Economic Research Department.

Berti, K., Salto, M., and Lequien, M. (2012). An early-detection index of fiscal stress for EU countries. European Economy - Economic Papers 2008 - 2015 475, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Bohn, H. (1998). The behavior of u. s. public debt and deficits. *The Quarterly Journal of Economics*, 113(3):949–963.

Chen, C. and Liu, L.-M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421):284–297.

Drehmann, M. and Juselius, J. (2014). Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting*, 30(3):759–780.

ESRB (2017). RECOMMENDATION OF THE EUROPEAN SYSTEMIC RISK BOARD of 18 June 2014 on guidance for setting countercyclical buffer rates. Technical report, EUROPEAN SYSTEMIC RISK BOARD.

Eurostat (2015). ESS guidelines on seasonal adjustment.

Eurostat (2018). European Statistical System (ESS) guidelines on temporal disaggregation, benchmarking and reconciliation.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Gómez, V. and Maravall, A. (1996). Programs tramo and seats, instruction for user (beta version: september 1996). Working papers, Banco de España.

Goeman, J. J. (2010). L1 penalized estimation in the cox proportional hazards model. *Biometrical Journal*, (52):–14.

Hernández de Cos, P., Nickel, C., Koester, G., and Moral-Benito, E. (2014). Signalling fiscal stress in the euro area - a country-specific early warning system. Working Paper Series 1712, European Central Bank.

Hyndman, R. J. and Athanasopoulos, G. (2014). *Forecasting : principles and practice / Rob J Hyndman and George Athanasopoulos*. OTexts.com [Heathmont, Victoria], print edition edition.

Manasse, P., Schimmelpfennig, A., and Roubini, N. (2003). Predicting sovereign debt crises. *IMF Working Papers*, 03.

Sax, C. and Steiner, P. (2016). *tempdisagg: Methods for Temporal Disaggregation and Interpolation of Time Series*. R package version 0.25.0.

Shiskin, J., Young, A., and Musgrave, J. (1967). *The X-11 Variant of the Census Method II Seasonal Adjustment Program*. Technical paper. U.S. Department of Commerce.

Somers, R. (1969). *A New Asymmetric Measure of Association for Ordinal Variables*. Bobbs-Merrill reprint series in the social sciences. Bobbs-Merrill, College Division.

Sumner, S. P. and Berti, K. (2017). A Complementary Tool to Monitor Fiscal Stress in European Economies. European Economy - Discussion Papers 2015 - 049, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320.

A. Thresholds by country based on S0

	Threshold	signalling power	missed episodes	false alarm
BE	0.32	0.93	0.00	7.35
BG	0.41	0.66	0.00	34.48
CZ	0.44	0.85	0.00	14.71
DK	0.29	0.96	0.00	4.48
DE	0.38	0.79	0.00	20.90
EE	0.38	0.70	0.00	29.69
IE	0.55	0.93	0.00	7.02
EL	0.75	0.85	0.00	15.38
ES	0.81	0.98	0.00	1.52
FR	0.50	1.00	0.00	0.00
IT	0.62	0.92	0.00	7.58
CY	0.81	1.00	0.00	0.00
LV	0.57	1.00	0.00	0.00
LT	0.68	0.98	0.00	1.52
MT	0.52	1.00	0.00	0.00
NL	0.19	0.58	0.00	42.42
AT	0.16	0.61	0.00	38.81
PL	0.52	0.43	0.00	57.14
PT	0.75	0.62	38.46	0.00
RO	0.54	0.33	48.48	18.92
SI	0.45	1.00	0.00	0.00
SK	0.54	1.00	0.00	0.00
FI	0.19	0.37	33.33	29.85
SE	0.29	1.00	0.00	0.00
UK	0.64	0.99	0.00	1.45

Table A.1.: Threshold values by country for observed stress episodes

	Threshold	signalling power	missed_episodes	false_alarm
BE	0.36	0.91	0.00	0.07
BG	0.23	0.16	0.00	0.34
CZ	0.43	0.93	0.00	0.15
DK	0.24	0.72	0.00	0.04
DE	0.17	0.25	0.00	0.21
EE	0.47	0.67	0.00	0.30
IE	0.38	0.74	0.00	0.07
EL	0.72	0.80	0.00	0.15
ES	0.72	0.92	0.00	0.02
FR	0.53	1.00	0.00	0.00
IT	0.63	1.00	0.00	0.08
CY	0.81	1.00	0.00	0.00
LV	0.45	0.95	0.00	0.00
LT	0.57	1.00	0.00	0.02
MT	0.55	1.00	0.00	0.00
NL	0.28	0.74	0.00	0.42
AT	0.38	0.93	0.00	0.39
PL	0.39	0.34	0.00	0.57
PT	0.76	0.60	0.38	0.00
RO	0.49	0.24	0.48	0.19
SI	0.54	1.00	0.00	0.00
SK	0.47	1.00	0.00	0.00
FI	0.22	0.63	0.33	0.30
SE	0.37	1.00	0.00	0.00
UK	0.68	0.97	0.00	0.01

Table A.2.: Threshold values by country for transformed stress episodes

B. Thresholds by country based on logit models

	Threshold	signalling power	missed episodes	false alarm
BE	0.06	0.78	0.00	22.39
BG	0.13	0.00	100.00	0.00
CZ	0.07	0.81	0.00	19.40
DK	0.06	0.86	0.00	13.64
DE	0.06	0.56	0.00	43.94
EE	0.09	0.92	0.00	7.94
IE	0.08	0.71	7.69	21.43
EL	0.32	0.97	0.00	3.12
ES	0.16	0.98	0.00	1.54
FR	0.09	1.00	0.00	0.00
IT	0.09	1.00	0.00	0.00
CY	0.15	1.00	0.00	0.00
LV	0.12	0.91	0.00	8.93
LT	0.22	1.00	0.00	0.00
MT	0.10	1.00	0.00	0.00
NL	0.06	0.62	0.00	38.46
AT	0.07	0.95	0.00	4.55
PL	0.06	0.04	7.14	89.09
PT	0.13	0.68	23.08	8.93
RO	0.12	0.24	25.00	51.35
SI	0.14	1.00	0.00	0.00
SK	0.14	1.00	0.00	0.00
FI	0.05	0.67	0.00	33.33
SE	0.07	1.00	0.00	0.00
UK	0.12	1.00	0.00	0.00

Table B.1.: Threshold values by country for pooled model and observed stress episodes

	Threshold	signalling power	missed episodes	false alarm
BE	0.05	0.72	0.00	28.36
BG	0.03	0.05	0.00	94.74
CZ	0.05	1.00	0.00	0.00
DK	0.06	0.97	0.00	3.03
DE	0.03	0.52	0.00	48.48
EE	0.17	1.00	0.00	0.00
IE	0.23	0.96	0.00	3.57
EL	0.15	0.95	0.00	4.69
ES	0.21	0.97	0.00	3.08
FR	0.03	1.00	0.00	0.00
IT	0.12	1.00	0.00	0.00
CY	0.11	0.99	0.00	1.47
LV	0.22	1.00	0.00	0.00
LT	0.20	1.00	0.00	0.00
MT	0.05	1.00	0.00	0.00
NL	0.03	0.23	0.00	76.92
AT	0.06	0.91	0.00	9.09
PL	0.15	0.12	35.71	52.73
PT	0.16	0.96	0.00	3.57
RO	0.32	0.31	9.38	59.46
SI	0.08	1.00	0.00	0.00
SK	0.07	1.00	0.00	0.00
FI	0.03	0.52	0.00	48.48
SE	0.03	1.00	0.00	0.00
UK	0.07	1.00	0.00	0.00

Table B.2.: Threshold values by country for fixed effects model and observed stress episodes

	Threshold	signalling power	missed episodes	false alarm
BE	0.03	0.93	0.00	7.46
BG	0.03	0.17	16.67	66.67
CZ	0.04	0.78	0.00	22.39
DK	0.03	0.82	0.00	18.18
DE	0.03	0.47	0.00	53.03
EE	0.04	0.59	33.33	7.94
IE	0.05	0.57	30.77	12.50
EL	0.06	0.68	20.00	12.50
ES	0.05	0.97	0.00	3.08
FR	0.04	1.00	0.00	0.00
IT	0.04	0.95	0.00	4.62
CY	0.06	0.94	0.00	5.88
LV	0.04	0.55	30.77	14.29
LT	0.06	1.00	0.00	0.00
MT	0.05	1.00	0.00	0.00
NL	0.03	0.54	0.00	46.15
AT	0.03	0.79	0.00	21.21
PL	0.04	0.07	78.57	14.55
PT	0.04	0.53	38.46	8.93
RO	0.03	0.07	6.25	86.49
SI	0.06	1.00	0.00	0.00
SK	0.08	1.00	0.00	0.00
FI	0.03	0.92	0.00	7.58
SE	0.04	1.00	0.00	0.00
UK	0.05	0.99	0.00	1.47

Table B.3.: Threshold values by country for fixed effects model and lead stress episodes

	Threshold	signalling power	missed episodes	false alarm
BE	0.03	0.93	0.00	7.46
BG	0.02	0.04	16.67	78.95
CZ	0.03	0.78	0.00	22.39
DK	0.02	0.85	0.00	15.15
DE	0.01	0.47	0.00	53.03
EE	0.02	0.68	0.00	31.75
IE	0.10	0.55	30.77	14.29
EL	0.09	0.69	20.00	10.94
ES	0.11	0.97	0.00	3.08
FR	0.07	1.00	0.00	0.00
IT	0.06	0.95	0.00	4.62
CY	0.18	0.94	0.00	5.88
LV	0.02	0.62	0.00	37.50
LT	0.19	1.00	0.00	0.00
MT	0.08	1.00	0.00	0.00
NL	0.02	0.57	0.00	43.08
AT	0.03	0.76	0.00	24.24
PL	0.14	0.05	92.86	1.82
PT	0.14	0.50	46.15	3.57
RO	0.03	0.14	18.75	67.57
SI	0.08	1.00	0.00	0.00
SK	0.36	1.00	0.00	0.00
FI	0.03	0.92	0.00	7.58
SE	0.07	1.00	0.00	0.00
UK	0.12	0.97	0.00	2.94

Table B.4.: Threshold values by country for fixed effects model and lead stress episodes

C. ROC curves

Fig. C.1.: Observed stress episodes

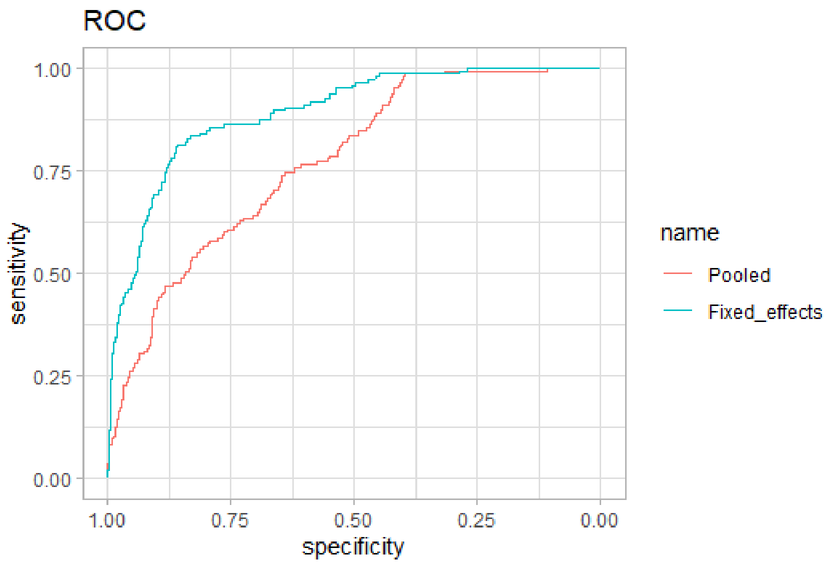
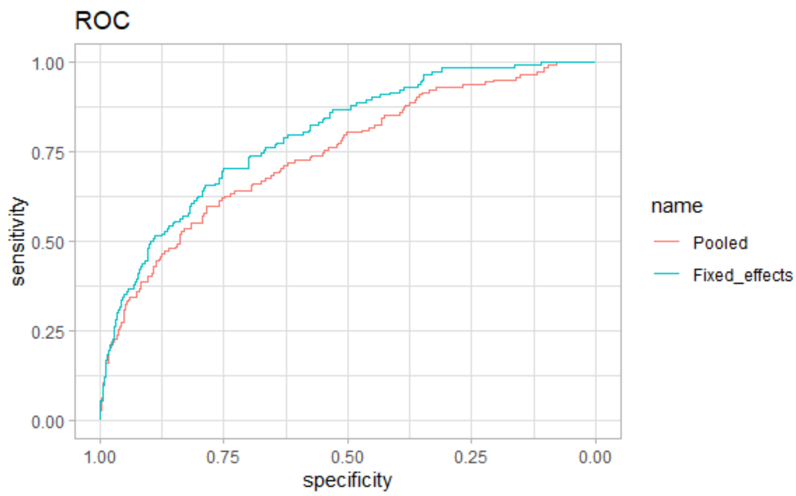


Fig. C.2.: Lead stress episodes



D. Definitions of Accuracy Measures[1]

	Yes	No
Predicted yes	True Positive (TP)	True Positive (FP), Type I error
Predicted no	False Negative (FN), Type II error	True Negative (TN)

Table D.1.: Confusion Matrix

Table D.1 shows a confusion matrix. The numbers along the major diagonal represent the correct decisions made, and the numbers of this diagonal represent.

Equations of several common metrics that can be calculated from it:

$$FP\ rate = \frac{FP}{No}$$
$$specificity = 1 - FP\ rate$$
$$TP\ rate = \frac{TP}{Yes}$$
$$sensitivity = TP \qquad rate$$

ROC graphs are two-dimensional graphs in which *tp rate* is plotted on the Y axis and *fp rate* is plotted on the X axis. An ROC graph depicts relative tradeoffs between benefits (true positives) and costs (false positives).

To measure performance of classification, a common method is to calculate the area under the ROC curve, abbreviated AUC. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1.0. However, because random guessing produces the diagonal line between (0, 0) and (1, 1), which has an area of 0.5, no realistic classifier should have an AUC less than 0.5.

Somers' Delta is is a measure of ordinal association between two possibly dependent random variables X and Y is usually used to quantify the quality binary choice models in econometrics. It is related above described area under ROC curve (AUC).

$$D_{XY} = 2(AUC - 0.5).$$

[1]Chapter based on Fawcett (2006) and Somers (1969)

E. List of variables and sources

Variable	Description	Source
Primary Balance	Overall balance, excluding interest payment	Eurostat, IMF
Short term GG debt	General government short term debt securities and short term loans	Eurostat
Gross GG debt	Gross debt of the general government	Eurostat
GDP	Gross Domestic product	Eurostat
Real GDP growth	GDP adjusted for inflation growth	Eurostat
Fertility rate	Average number of children per woman	Eurostat
Part of non-employed	Part of population excluded employed and unemployed	Eurostat
Nominal unit labour cost	Ratio of labour costs to labour productivity	Eurostat
Current account	Transactions of a country with the rest of the world	Eurostat
Balance	Net lending (+) or net borrowing (-)	Eurostat, IMF
Unemployment	Part of population which are not in paid employment or self-employment and are currently available for work during the reference period	Eurostat
Expenditure on pensions	Government expenditure on pensions	Eurostat
Expenditure on social protection on social protection	Government	Eurostat
GINI	Distribution of income across income percentiles in a population.	Eurostat

Variable	Description	Source
S80-S20	The ratio of total income received by the 20 % of the population with the highest income (top quintile) to that received by the 20 % of the population with the lowest income (lowest quintile).	Eurostat
Interest rate	Maastricht criterion bond yields.	Eurostat, Mac-robond
GG expenditure	Total expenditure of general government	Eurostat
GG final consumption	All government current expenditures for purchases of goods and services (including compensation of employees)	Eurostat
Interest paid	Interest payments	Eurostat
Debt, non-financial corporations	debt securities and loans of non-financial corporations	Eurostat
Short term debt, non-financial corporations	short term debt securities and short term loans of non-financial corporations	Eurostat
Inflation	Harmonised Index of Consumer Prices (HICP)	ECB