Economics Bulletin

Volume 42, Issue 4

Assessing the scoreboard of the EU macroeconomic imbalances procedure: (machine) learning from decisions

Tiago Alves
ISCTE-IUL

João Amador
Banco de Portugal and Nova School of Business and Economics

Francisco Gonçalves
Nova School of Business and Economics

Abstract

This paper uses machine learning methods to identify the macroeconomic variables that are most relevant for the classification of countries along the categories of the EU Macroeconomic Imbalances Procedure (MIP). The random forest algorithm considers the 14 headline indicators of the MIP scoreboard and the set of past decisions taken by the European Commission when classifying countries along the MIP categories. The algorithm identifies the unemployment rate, the current account balance, the private sector debt and the net international investment position as key variables in the classification process. We explain how high vs low values for these variables contribute to classifying countries inside or outside each MIP category.

The opinions expressed in this article are those of the authors and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. The authors thank António R. dos Santos, two anonymous referees and the Editor for useful comments and suggestions. Code and data are available at https://gitlab.com/alves.tiago/scoreboard. Any remaining errors and omissions are the sole responsibility of the authors.
Citation: Tiago Alves and João Amador and Francisco Gonçalves, (2022) "Assessing the scoreboard of the EU macroeconomic imbalances procedure: (machine) learning from decisions", *Economics Bulletin*, Volume 42, Issue 4, pages 2257-2266
Contact: Tiago Alves - alves.tiago@posteo.net, João Amador - jamador@bportugal.pt, Francisco Gonçalves - francisco.goncalves@posteo.co.uk.
Submitted: June 22, 2021. **Published:** December 30, 2022.

1. Introduction

The euro area sovereign debt crisis of 2010-2011 demonstrated the need for stronger economic governance and enhanced policy coordination between EU Member-states to avoid the accumulation of serious imbalances, with an impact on overall macroeconomic stability. Until the sovereign debt crisis, different economic policy coordination procedures were implemented without articulation. Later, Member-states were asked to synchronize their timetables and existing processes were streamlined in order to align national fiscal, growth and employment policies. This materialized into the designated "European Semester". Under this setup, the supervision and coordination of macroeconomic policies was expanded to include external imbalances, as well as labour market and credit developments. The designated "Macroeconomic Imbalances Procedure" (MIP) operationalized this inherently complex process (European Commission 2016). Overall, the main goal of the European Commission is to classify countries in terms of the seriousness of macroeconomic imbalances, using a set of common macroeconomic variables. Although initiatives to revisit the economic governance of the EU were launched recently, notably in the context of the response to the COVID 19 pandemics, the MIP remains a key surveillance tool (European Commission 2020, 2021).

Random forest algorithms are one of the most powerful machine learning methods and are suited for exercises of classification of observations into categories. Recent methods make it possible to identify the most important variables to the MIP classification, while also accounting for their positive or negative impact on the classification depending on the values they assume. Nevertheless, the Commission's MIP classification for each country in a specific year cannot be confirmed or dismissed by any objective economic criterion because the real situation of countries is not directly observable. Therefore, the methodology only considers the underlying decision process of the Commission, which is necessarily complex and includes non-economic information. This said, our research question is to unveil the key macroeconomic variables in the underlying decision process and how they contribute to the classification. We also describe how high or low values of each scoreboard variable contribute to selecting a country in or out of a MIP category, thus allowing for some interpretation of results.

Other papers have looked into the classification of countries in the MIP and the scoreboard. For example, Knedlik (2014) takes a political economy perspective and examines policy-maker's decisions to assess their preferences. They conclude that the Commission shows a higher relative preference for avoiding type I errors (false positives) and that the scoreboard is quite strict. From a different perspective, Sondermann and Zorell (2019) employs a multivariate discrete choice model to match past crises with patterns of macroeconomic imbalances spreading over a sample of 32 OECD and EU economies over almost 40 years. The exercise identifies the current account balance and export market share growth as good predictors of future crises.

2. The Macroeconomic Imbalances Procedure

As stated by the European Commission, the MIP is part of the annual European multilateral surveillance cycle and aims at identifying, preventing and eliminating excessive macroeconomic imbalances that are likely to affect economic stability in an individual Member-state, the euro area or the EU as a whole. The legal framework is based on regulations 1176 and 1174 of 2011, which are legislative pieces of the so-called "six-pack". Discussions regarding improvements of the MIP as part of a reform of the EU economic governance are ongoing. A recent thoughtful contribution to this debate is Koll and Watt (2022).

The MIP is based on a scoreboard with 14 macroeconomic indicators (plus 25 auxiliary indicators) where the situation in each country is compared with pre-established thresholds. The assessment leads the Commission to classify countries in a four-tier scale starting in the category "no imbalance" and ending with "excessive imbalance with corrective action".

It is relevant to note that labour market indicators (activity rate, long-term unemployment and youth unemployment) were added to the procedure at a later stage and thus do not play an equivalent role in the assessment of imbalances. Nevertheless, our methodology is precisely about assessing the relevance of the scoreboard variables in the decision process and is not affected by the inclusion of this latter group. As a robustness test, we have replicated the exercise excluding these three variables and the ranking of importance for the remaining variables does not change significantly.

3. Data

We use data regarding the scoreboard indicators and Commission's decisions since the beginning of the procedure in 2011. Two important features are worth noting. Firstly, only the 14 headline indicators of the MIP scoreboard are used. Secondly, we label the decisions regarding imbalances in four categories: 1- no imbalance (where we include the status of "no in-depth review"), 2- imbalance, 3- excessive imbalance, 4- excessive imbalance with corrective action (program country).

The final dataset comprises 220 observations ranging from 2011 to 2018, comprising 101, 73, 28, and 18 records in categories 1, 2, 3, and 4, respectively. The methodological procedure requires splitting the dataset into training and validation sets. The latter block represents 20 per cent of the observations available.

4. Methodology

The methodology consists of three steps. Firstly, we standardize the variables in the sample, so that their scale does not interfere with the results. As in Pedregosa et al. (2011), the *standard scaler* involves subtracting each variable by the mean and dividing by the variance. Secondly, we rebalance the number of observations per category. The library by Lemaître et al. (2017) enables the over-representation of minority classes, according to a percentage chosen to maximize the predictive power of the model. The samples picked to compensate for the under-representation are not merely copied and pasted. Instead we apply the SMOTE Chawla et al. (2002) method, which adds noise to each sample to make it slightly different from the original. Thirdly, we use a learning algorithm, that tries to unveil the implicit decision rules used by the Commission in the MIP classification.

In order to add as little discretionary decisions as possible, the type of scaling, the rebalancing strategy and even the classification algorithm, as well as the hyperparameter decisions in each of these three steps are selected after a probabilistic search across 20,000 possible combinations. The combination that promises the best results is automatically decided. Finally, in order to understand how decisions are made, we train a model with the LightGBM Ke et al. (2017) algorithm, based on *random forests* but designed to be faster. The choice of a model that clusters decision trees is good because the Commission's classification also results from aggregating the beliefs of those that take part in the decision process.

The methodology is validated in its predictions as the model is able to classify correctly in 74 per cent of the cases. In Table 1, we identify which categories are most difficult to differentiate. The groups with less severe macroeconomic conditions (i.e., no imbalance and imbalance) appear to be less distinguishable and therefore more susceptible to erroneous classifications.

A random forest is a complex object whose interpretation is simple but laborious. It can be used to solve regression and classification problems where the dependent variable is categorical, as in our case. Since our goal is to identify the variables that emerge as most important in the classification process, the model is used as a feature selection tool. The use of the SHAP library, which takes Shapley values to explain the result of any machine learning model, brings some interpretation and transparency to the results. The Shapley value is a cooperative game theory concept and corresponds to

Table 1: Confusion matrix

		Reference			
		1	2	3	4
Prediction	1	14	2	0	1
	2	5	7	1	0
	3	0	0	5	0
	4	0	1	0	3

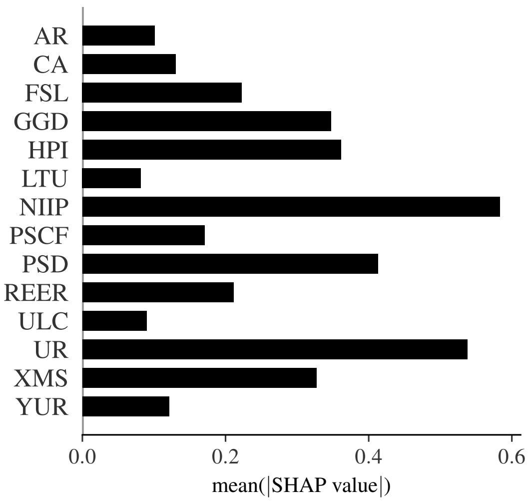
Note: 1- no imbalance (including the status of "no in-depth review"); 2- imbalance, 3- excessive imbalance, 4- excessive imbalance with corrective action. Actual classification of countries in columns and the corresponding classification by the algorithm in rows.

the expected average marginal contribution of a variable after all possible combinations have been considered. This method explains the local importance of the variable and how it changes with lower or higher sample values. Importantly, SHAP provides the importance of each indicator in the forecast, but does not evaluate the quality of the forecast itself.

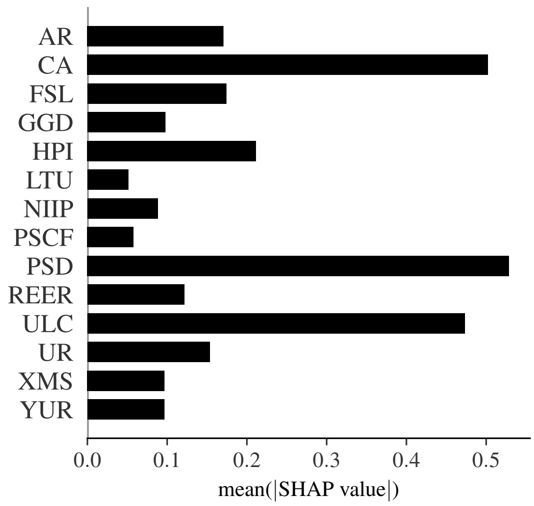
5. Results

Figure 1 shows how each variable affects prediction in the four MIP categories by taking the absolute SHAP value into account, i.e., not considering if the variable affects the prediction in a positive or negative way. For the "no imbalance" category (panel a) the most important variables are the net international investment position as a percentage of GDP with a SHAP value near 0.6, closely followed by the unemployment rate. As for the "imbalance" category, there are three main relevant variables, namely the consolidated private sector debt as a percentage of GDP, the current account balance and the 3-year percentage change in the unit labour cost. The unemployment rate is overwhelmingly important for the "excessive imbalance", with a SHAP value close to 1.5. Finally, in what regards the category "excessive imbalance with corrective action" the current account balance stands out as the most significant indicator, followed at large distance by the net international investment position.

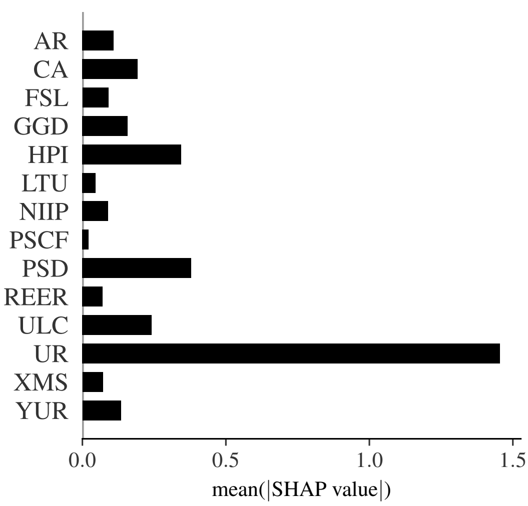
Figure 1: Relevance of scoreboard indicators (SHAP values)



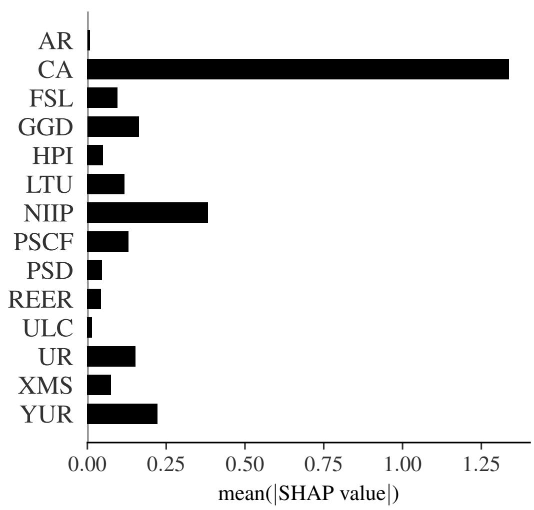
(a) No imbalance



(b) Imbalance



(c) Excessive imbalance



(d) Excessive imbalance with corrective action

Note: CA: 3-year backward moving average of the current account balance in % of GDP; NIIP: Net international investment position in % of GDP; LTU: 3-year change in p.p. of the long-term unemployment rate; UR: 3-year backward moving average of unemployment rate; PSD: Private sector debt (consolidated) in % of GDP; XMS: 5-year percentage change of export market shares; HPI: year-on-year changes in house price index; ULC: 3-year percentage change in nominal unit labour cost; REER: 3-year percentage change of the real effective exchange rates based on HICP/CPI deflators, relative to 41 other industrial countries; PSCF: Private sector credit flow in % of GDP; GGD: General government sector debt in % of GDP; FSL: Year-on-year changes in total financial sector liabilities; AR: 3-year change in p.p. of the activity rate; YUR: 3-year change in p.p. of the youth unemployment rate.

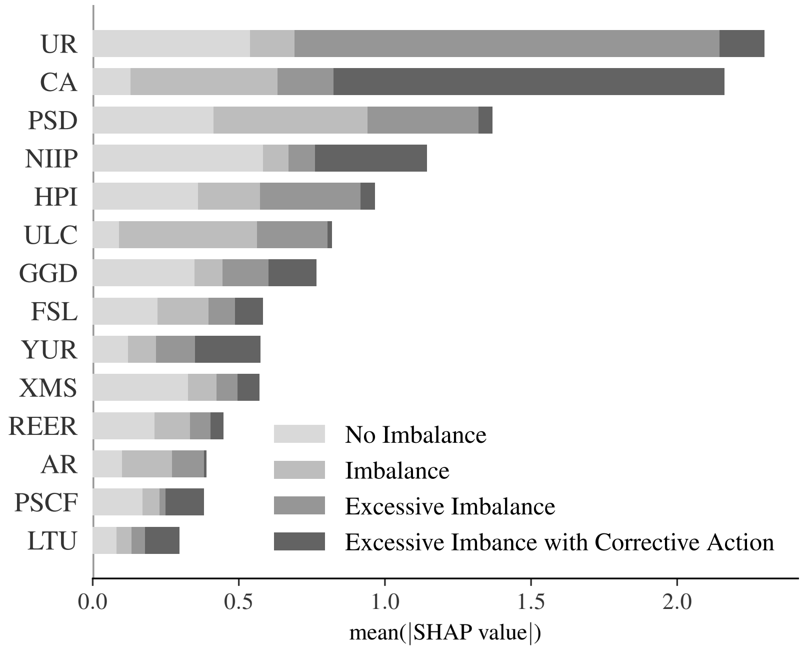
It is interesting to note that the unemployment rate plays a rather important role to identify “excessive imbalances” but the decrease in accuracy from its exclusion drops significantly when it comes to the identification of “excessive imbalances with corrective action”. A somewhat inverse pattern is found for the current account and the net international investment position. If a very strong macroeconomic crisis emerges - possibly with a sudden stop in external financing - the variables related to the balance of payments drive decisions. Conversely, the cyclical position of the economy, proxied by the unemployment rate, exacerbates existing problems and signals excessive imbalances.

Figure 2 sums the results presented in the four panels of Figure 1 thus providing an overall notion of the importance of each scoreboard variable. It highlights the relevance of the cyclical position of the economy, proxied by the unemployment rate, the current account and two indebtedness variables (private sector debt and the international investment position). On the lower end of importance stand the 3-year change in the activity rate, the private sector credit flow and the 3-year change of the long-term unemployment rate. Noticeably, the three more recent labour market scoreboard variables stand in the second half of this importance rank.

Up to now the analysis only took into account the absolute SHAP values. However, it is essential to know whether each variable affects the prediction in a positive or negative way. The panels of Figure 3 present the local variable contribution for each MIP category by scattering the corresponding bee-swarm plot. The y-axis indicates the variable and the x-axis the value of its mean SHAP value. Each dot stands for the positive (negative) contribution of each (country-year) observation to place a country inside (outside) the category and its colour defines whether it is influential with a high (dark) or low (light) value.

The first panel of Figure 3 presents a high dispersion of dots on the net international investment position, with the light coloured ones with a negative SHAP value and dark ones with a positive SHAP. This means that low international investment positions have a negative contribution, i.e., ruling out a country from the category of “no imbalance”, while high ones post a positive contribution. Indeed, it is expectable that countries with a comfortable international investment position signal an underlying solid macroeconomic situation and thus “no imbalance”. In addition, low unemployment rates, low growth of house prices and low general government debt (light coloured dots) signal a country as part of the “no imbalances” category. As for the “imbalance” category in panel b), high current account balances signal a country as part of this category. This seemingly strange result is justifiable by the fact that large persistent current account surpluses are deemed undesirable by the MIP and constitute an im-

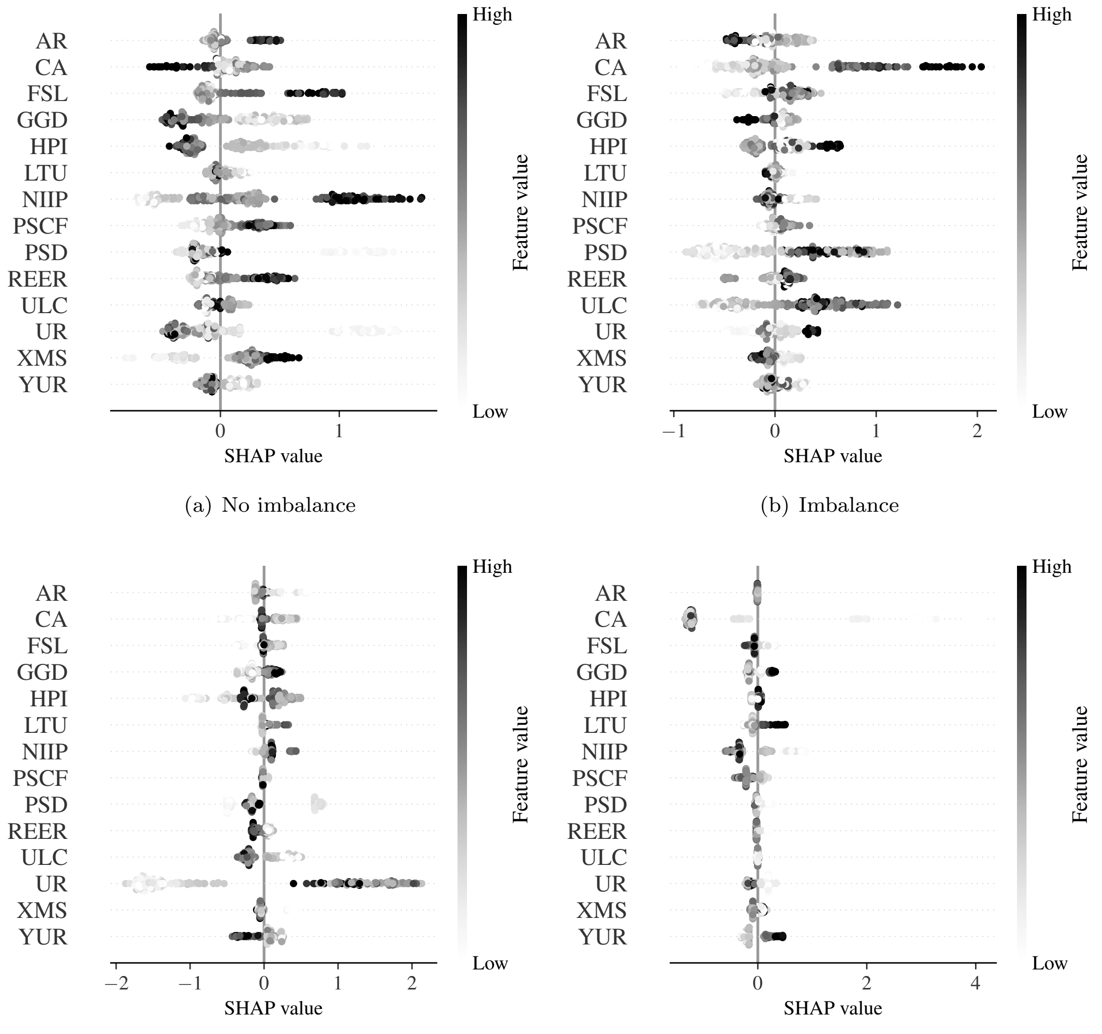
Figure 2: Relevance of scoreboard indicators (SHAP values)



Note: CA: 3-year backward moving average of the current account balance in % of GDP; NIIP: Net international investment position in % of GDP; LTU: 3-year change in p.p. of the long-term unemployment rate; UR: 3-year backward moving average of unemployment rate; PSD: Private sector debt (consolidated) in % of GDP; XMS: 5-year percentage change of export market shares; HPI: year-on-year changes in house price index; ULC: 3-year percentage change in nominal unit labour cost; REER: 3-year percentage change of the real effective exchange rates based on HICP/CPI deflators, relative to 41 other industrial countries; PSCF: Private sector credit flow in % of GDP; GGD: General government sector debt in % of GDP; FSL: Year-on-year changes in total financial sector liabilities; AR: 3-year change in p.p. of the activity rate; YUR: 3-year change in p.p. of the youth unemployment rate.

balance. High unit labour costs and private sector debt also strongly contribute to identify a country as part of this category. The important role of the unemployment rate in selecting countries as part of the "excessive imbalance" category is visible by the numerous highly positive and dark dots and many strongly negative and light dots. Finally, panel d) highlights the key role of very low (light coloured) current account balances and, to a lesser extent, low net international investment positions to identify situations of "excessive imbalance with corrective action". Reassuringly, these results are all in accordance with standard economic reasoning.

Figure 3: Relevance of scoreboard indicators considering individual contributions (SHAP values)



(a) No imbalance

(b) Imbalance

(c) Excessive imbalance

(d) Excessive imbalance with corrective action

Note: CA: 3-year backward moving average of the current account balance in % of GDP; NIIP: Net international investment position in % of GDP; LTU: 3-year change in p.p. of the long-term unemployment rate; UR: 3-year backward moving average of unemployment rate; PSD: Private sector debt (consolidated) in % of GDP; XMS: 5-year percentage change of export market shares; HPI: year-on-year changes in house price index; ULC: 3-year percentage change in nominal unit labour cost; REER: 3-year percentage change of the real effective exchange rates based on HICP/CPI deflators, relative to 41 other industrial countries; PSCF: Private sector credit flow in % of GDP; GGD: General government sector debt in % of GDP; FSL: Year-on-year changes in total financial sector liabilities; AR: 3-year change in p.p. of the activity rate; YUR: 3-year change in p.p. of the youth unemployment rate.

6. Concluding remarks

We conclude that the cyclical position of the economy, proxied by the unemployment rate, two indebtedness variables (private sector debt and the international investment position) and the current account are the key variables in classification of countries along the four MIP categories. Conversely, the 3-year change of the activity rate, the private sector credit flow and 3-year change of the long-term unemployment rate do not seem to be useful.

Although the EU economic surveillance procedures have been adjusted to accommodate the challenges posed during the pandemic crisis, in the future they will be at least as important as they were in the past. Our findings contribute to inform policy-makers regarding the improvement of the MIP scoreboard. Focusing on the most informative macroeconomic variables streamlines the economic surveillance procedures, while increasing their effectiveness. Although relying on a limited dataset, the article also illustrates the utilization of advanced machine learning methods to address economic issues.

References

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002), "SMOTE: synthetic minority over-sampling technique", *Journal of artificial intelligence research* **16**, 321–357.

European Commission (2016), The Macroeconomic Imbalance Procedure. Rationale, process, application: A compendium, Institutional Paper 039, European Commission.

European Commission (2020), Economic governance review, communication, Technical report, European Commission.

European Commission (2021), The EU economy after COVID-19: Implications for economic governance, communication, Technical Report COM(2021) 662 final, European Commission.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017), "Lightgbm: A highly efficient gradient boosting decision tree", *Advances in neural information processing systems* **30**.

Knedlik, T. (2014), "The impact of preferences on early warning systems: The case of the European Commission's Scoreboard", *European Journal of Political Economy* **34**, 157–166.

Koll, W. and Watt, A. (2022), "The macroeconomic imbalance procedure at the heart of EU economic governance reform", *Intereconomics* **57**, 56-62.

Lemaˆıtre, G., Nogueira, F. and Aridas, C. K. (2017), "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning", *Journal of Machine Learning Research* **18**(17), 1–5.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R. and Dubourg, V. (2011), "Scikit-learn: Machine learning in Python", *the Journal of machine Learning research* **12**, 2825–2830.

Sondermann, D. and Zorell, N. (2019), A macroeconomic vulnerability model for the euro area, Working Paper Series 2306, European Central Bank.