Machine Learning in Economics
(458657)

replication paper

Early Warning System for Fiscal Stress

comparing the traditional logistic regression approach with random forest

Department of Economics
University of Bern

Spring Semester 2022

submitted by Bela Tim Koch

Contents

1 Introduction and Literature Review

Since the latest Great Recession with its corresponding deterioration of public finances, the monitoring and prevention of fiscal crises has become increasingly prominent in the political debate, leading to an increasing demand for the development of reliable and early indicators that signal possible fiscal stress. In order to be able to assess countries' vulnerability to fiscal distress ex-ante, the literature is increasingly devoted to the development of early warning systems for fiscal stress, which build on early warning systems for banking and currency crises (Honda, Tapsoba, and Issifou 2022). The standard tool used in the literature for early warning systems are the signalling approach as well as discrete dependent variable models, such as logistic regression (Jarmulska 2020).

As an alternative to the traditional methods, early warning models based on machine learning techniques are proposed, claiming a possible improvement of prediction accuracy (Beutel, List, and Schweinitz 2019). As an example, when predicting the build-up of banking crises, the early warning system developed by Casabianca et al. (2019) which builds upon a supervised machine learning algorithm, i.e. Adaptive Boosting, outperforms the traditional approach of using a logit model. Another example which shows that using machine learning could drastically improve prediction accuracy is the early warning system developed by Samitas, Kampouris, and Kenourgios (2020), which reaches an accuracy of 98.8% when predicting the risk of contagion inside a financial network using a quadratic support vector machine.

However, a disadvantage many of these machine learning methods compared to more traditional approaches is the difficulty of understanding how a result was obtained and are therefore often referred to as a black box (Ghoddusi, Creamer, and Rafizadeh 2019). Consequently, the researcher is confronted with a trade-off between prediction or interpretation. Ghoddusi, Creamer, and Rafizadeh (2019) argue, that emphasis should be in interpretation for scientific research or policy decisions, since understanding the relationship and behavior among different variables is more important than prediction accuracy. In contrast, more emphasis is to be placed on prediction accuracy in specific industrial applications.

This paper aims to replicate some of the work done by Jarmulska (2020). Particular emphasis is placed on the comparison of the traditional method, i.e. a logit model with a least absolute shrinkage and selection operator, and a model based on machine learning, i.e. an implementation of the random forest algorithm. As the results are often criticized for being difficult to interpret, ways of interpreting the developed early warning model based on random forest are presented.

to do: noch ein paar beispielpapers hinzufügen also literature review ausbauen

2 Model Describtion

2.1 Performance Metrics

Jarmulska (2020) uses sensitivity, specificity, their average as well as the area under receiver operating curve (AUROC) as measures to assess the effectiveness of the early warning models. Since sensitivity corresponds to the proportion of stress episodes correctly classified whereby specificity corresponds to the proportion of tranquil episodes correctly classified, these metrics are dependent on the threshold, which determines whether a period is classified as stress or tranquil episode (Jarmulska 2020). In this paper, this threshold is specified by maximizing the weighted sum of sensitivity and specificity. In contrast, the AUROC is a robust measure, as all possible thresholds are taken into account in their calculation. This measure represents the area under the receiver operating curve, which displays the trade-off between the true positive rate (i.e. sensitivity) and the false positive rate (i.e. 1 - specificity). Theoretically, the AUROC can be between 0 and 1 (perfect classifier), whereby random guessing would result in to a value of 0.5 (Fawcett 2006).

2.2 Logit Model with LASSO penalisation

Jarmulska (2020) implemented two version of discrete dependent variable models (logit regression), first a standard logit model with ordinary least squares estimates and second a logit model with a least absolute shrinkage and selection operator (LASSO) penalization. These models are often used as the standard econometric approach, which is why they are used as the benchmark in this study.

Ordinary least squares estimates often have low bias but large variance, reducing prediction accuracy - the prediction accuracy can sometimes be improved by shrinking some coefficients towards zero to sacrifice bias in order to reduce variance of the predicted values and possibly improve overall prediction accuracy (Tibshirani 1996). To do so, LASSO penalization as proposed in Tibshirani (1996) can be applied. Because of this characteristics, only the logit LASSO model is considered in this replication.

Following Hastie et al. (2009), the LASSO problem in the Lagrangian form is given as follows:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{1}$$

whereby λ corresponds to the penalization parameter. As can be seen in Equation (1), the higher λ, the higher the number of coefficients shrunk to zero. Here, λ is chosen by 5-fold cross-validation, maximizing the AUROC.

2.3 Random Forest

As an alternative to the logit model, Jarmulska (2020) applied the random forest algorithm following Breiman (2001) as an ensemble of multiple classification and regression trees for binary classification. A single tree divides the predictor space into distinct and non-overlapping regions, whereby an observation is classified depending on the region it falls into, i.e. at the terminal node. Each non-terminal node corresponds to a question with a binary response, what determines the structure of the tree. Since each terminal node assigns the same class to all observations within this node, minimizing the classification error at the level of the terminal nodes results in a minimization of the tree's overall classification error (Jarmulska 2020). To measure the precision of the fit, the Gini Index, which is used as a loss function in classification and regression trees, can be utilised (Jarmulska 2020):

$$g(w) = \sum_{k \neq j} p_{wk} p_{wj} = \sum_{k} p_{wk}(1 - p_{wk}) \tag{2}$$

whereby p_{wj} corresponds to the probability distribution of class j in node w.

However, such decision trees might suffer from overfitting, what can be counteracted by bootstrapping the training set and averaging all the predictions (James et al. 2013). If these so called bagged trees are all influenced by some strong predictors, the single trees might be correlated resulting again in overfitting. To decorrelate the single trees, the non-terminal nodes can be forced to consider only a subset of the predictors, increasing the difference between the single trees, through which the overfitment problem can be reduced (James et al. 2013). This ensemble method corresponds to the random forest algorithm.

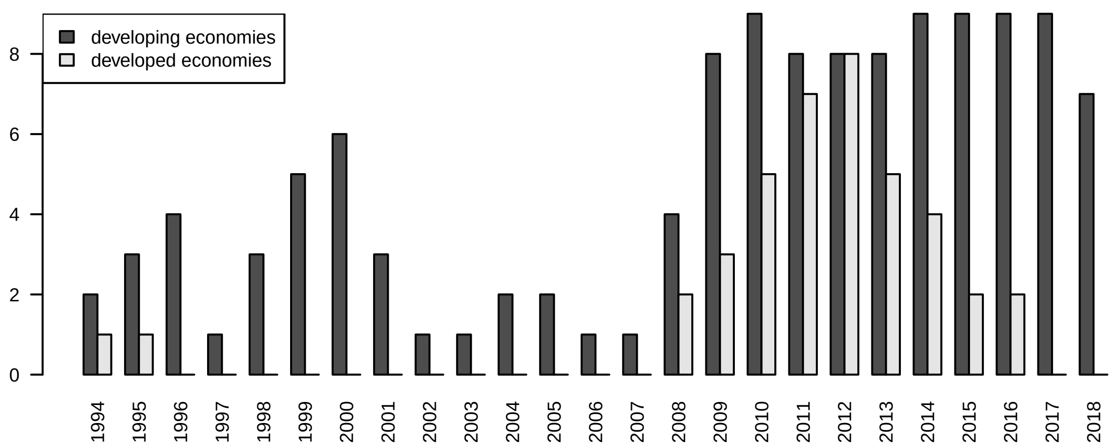
3 Data Describtion

3.1 Dependent Variable

The binary dependent variable to be predicted follows the definition in Dobrescu et al. (2011) takes the value of 1 in the case of a fiscal stress event in the next period and 0 otherwise. According to the definition provided by Dobrescu et al. (2011), an economy faces a fiscal stress event, if at least one of the following four conditions are fulfilled: (1) the economy fails to service debt as payments come due as well as if debt exchanges are distressed (2) the economy receives a large support program by the International Monetary Fund (3) hyperinflation is prevalent in the economy (inflation exceeding 35% for advanced economies, 500% for emerging economies) (4) the economy faces extreme financing constraints, i.e. the sovereign spread exceeds 1000 basis points or 2 standard deviations from the country average.

In the data used for this analysis, 16.5% of the recorded observation are classified as fiscal stress events, whereby these stress events are not equally distributed across country groups and over time. Figure 1 displays the distribution of the recorded stress events over time for developing and developed economies. The data shows that developing economies are more prone to fiscal stress events with 29% of all observations classified as fiscal stress events compared to 7.1% for developed economies.

Figure 1: Distribution of Stress Periods



3.2 Explanatory Variables

4 Empirical Results

Table 1: Means of Explanatory Variables

Variable	All periods	Tranquil periods	Stress periods	P-value	Significance
Competitiveness and domestic demand					
Current account balance	-0.52	0.28	-4.57	0.00	yes
CPI	4.26	3.68	7.18	0.00	yes
Credit to GDP change	1.42	1.58	0.58	0.40	no
Unemployment change	-0.04	-0.14	0.48	0.00	yes
Consumption dynamics	-4.21	-3.80	-6.31	0.00	yes
Export share dynamics	0.60	0.78	-0.36	0.08	no
Financial					
Fixed capital formation dynamics	7.82	8.00	6.96	0.34	no
FX rate dynamics	1.87	0.69	7.82	0.00	yes
Fiscal					
GDP dynamics	2.90	3.14	1.71	0.00	yes
China GDP dynamics	9.62	9.63	9.59	0.55	no
US GDP dynamics	2.46	2.58	1.82	0.00	yes
Labor market					
Labor productivity dynamics	1.76	1.90	1.05	0.00	yes
GDP per capita	26.13	28.08	16.28	0.00	yes
Macroeconomic and global economy					
Interest on debt	3.58	3.32	4.92	0.00	yes
US interest rates	4.27	4.40	3.64	0.00	yes
Net lending	-2.47	-2.06	-4.55	0.00	yes
Oil price dynamics	5.05	5.87	0.89	0.03	yes
Currency overvaluation	-33.76	-31.59	-44.73	0.00	yes
Public debt	58.51	56.69	67.71	0.00	yes
VIX	20.17	20.07	20.70	0.74	no

4.1 Performance

Table 2: Average prediction accuracy of early warning models for years 2009-2018

	Logit LASSO		Random Forest	
	advanced dummy	GDP per capita	advanced dummy	GDP per capita
% of correctly classified stress episodes	85.18	74.02	89.28	87.36
% of correctly classified tranquil episodes	54.04	70.95	63.18	69.65
Average	69.61	72.49	76.23	78.51
AUROC	0.84	0.85	0.88	0.89

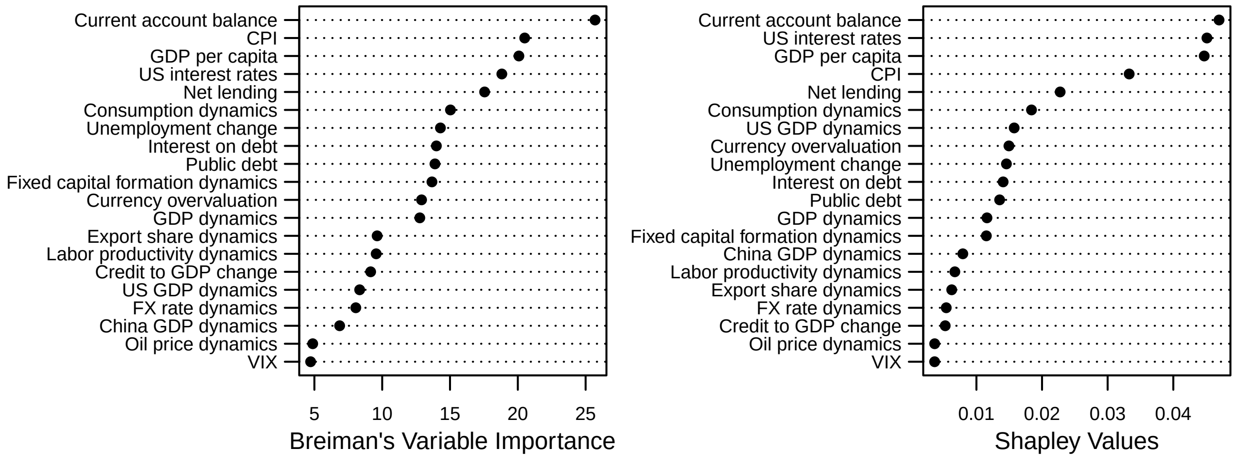
4.2 Interpretability of Random Forest Algorithm

4.2.1 Variable Importance and Shapley Values

4.2.2 Partial dependence plots and Accumulated local effects plots

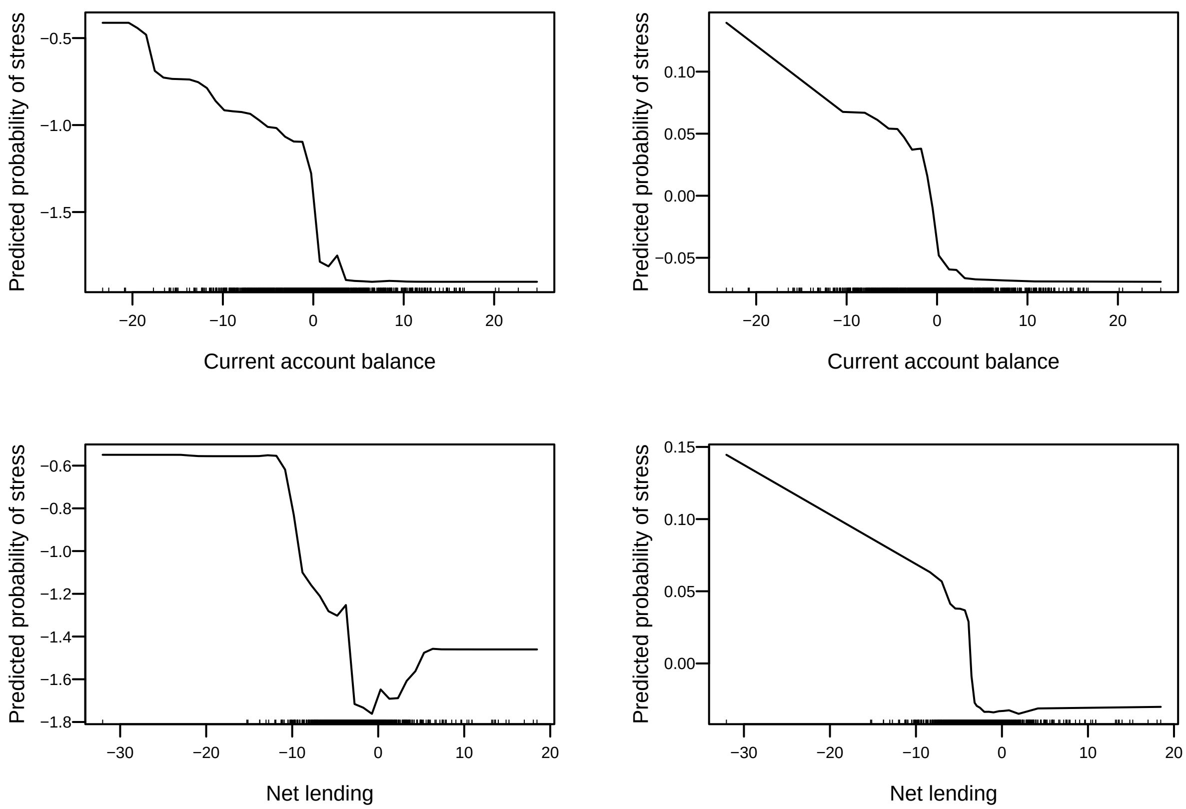
5 Conclusion

Figure 2: Variable Importance and Shapley Values of Predictors used



(#fig:rf.VI.SV)

Figure 3: Partial Dependence and Accumulated Local Effects Plots



(#fig:rf.pdp.alep)

6 References

The code and data used for this project can be found in the corresponding GitHub-repository: `https://github.com/bt-koch/ML-in-Economics`.

Beutel, Johannes, Sophia List, and Gregor von Schweinitz. 2019. "An evaluation of early warning models for systemic banking crises: Does machine learning improve predictions?" IWH Discussion Papers 2/2019. Halle Institute for Economic Research (IWH). https://ideas.repec.org/p/zbw/iwhdps/22019.html.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Casabianca, Elizabeth Jane, Michele Catalano, Lorenzo Forni, Elena Giarda, Simone Passeri, et al. 2019. "An Early Warning System for Banking Crises: From Regression-Based Analysis to Machine Learning Techniques." *EconPapers. Orebro: Orebro University*.

Dobrescu, Gabriela, Iva Petrova, Nazim Belhocine, and Emanuele Baldacci. 2011. "Assessing Fiscal Stress." *IMF Working Papers* 11: 100.

Fawcett, Tom. 2006. "An Introduction to ROC Analysis." *Pattern Recognition Letters* 27 (8): 861–74. https://doi.org/https://doi.org/10.1016/j.patrec.2005.10.010.

Ghoddusi, Hamed, Germán G. Creamer, and Nima Rafizadeh. 2019. "Machine Learning in Energy Economics and Finance: A Review." *Energy Economics* 81: 709–27. https://doi.org/https://doi.org/10.1016/j.eneco.2019.05.006.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Vol. 2. Springer.

Honda, Jiro, René Tapsoba, and Ismael Issifou. 2022. "When Do We Repair the Roof? Insights from Responses to Fiscal Crisis Early Warning Signals." *International Economics*. https://doi.org/https://doi.org/10.1016/j.inteco.2022.02.008.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.

Jarmulska, Barbara. 2020. "Random Forest Versus Logit Models: Which Offers Better Early Warning of Fiscal Stress?" *ECB Working Paper Series No 2408 / May 2020*.

Samitas, Aristeidis, Elias Kampouris, and Dimitris Kenourgios. 2020. "Machine Learning as an Early Warning System to Predict Financial Crisis." *International Review of Financial Analysis* 71: 101507. https://doi.org/https://doi.org/10.1016/j.irfa.2020.101507.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88.