Assessment of factors of regional economic stability using the XGBoost algorithm

Yulia Valentinovna Granitsa[1]

Lobachevsky State University, Center for Macro and Microeconomics, Nizhny Novgorod, Russia

Abstract. In modern conditions, the regions of Russia acquire a special role and are evaluated as independent economic entities. In a pandemic, the key factor for successful development at the meso-level is not so much the growth of well-being, the quality of life, but the preservation of the stability of economic systems, their ability to withstand external influences. Thus, in our study, we equate the economic stability and economic security of the regions. To assess economic stability, it is advisable, in our opinion, to use a group of indicators characterising resource provision, investment climate and the efficiency of functioning of regions. We assigned the rank of economic security to all regions, on the basis of which the regions were divided into two classes – economically safe and economically unsafe. The ensemble machine learning algorithm XGBoost was chosen as a method for factor analysis of economic security. The constructed classification model was interpreted by us using the Shap algorithm, which assumes the analysis of Shapley values for each economic determinant. The applied algorithms allowed us to identify significant factors that determine stable regions. These factors include investment risk, human development index and the ratio of the balanced financial result to the gross regional product.
Keywords: XGBoost, region, economic stability, inflation expectations, economic security, Shap algorithm

1 Introduction

During a global pandemic, there is an urgent need to ensure the independence of Russian regions as separate economic entities. The disproportionate development of regional structures, which are normally offset by injections of the federal budget funds, is only progressing in the context of a shortage of budgetary resources.

For that reason, the method of highlighting the competitive advantages of regional economies, which ensure the preservation and strengthening of the region's social and economic status and its resistance to external challenges, is becoming increasingly relevant. In our opinion, during the period of impact of external shocks on the economy of the sub-federal level (meso-level), the concept of economic stability is thus identified with economic security.

The regional stability and security criteria are analysed in a great number of publications by both domestic and overseas researchers, among which we note the works of Leonidova

[1] Corresponding author: ygranica@yandex.ru

G.E, Rumyantsev N.M. [1], Abramova M. A., Igonina L. L. [2], Nazarov A.I. [3], Kurbatova M.V. [4], Malkina M.Yu. [5], Maslennikov, D. A. [6], Pechenskaya M.A. [7], Granitsa, Y.V [8], Billon M. [9], Fornaro L., Wolf M. [10], Fritsch M., Wyrwich M. [11], Guerrieri V. [12].

The determinants of economic stability and security, that we have identified, are presented in Table 1. Each of them characterises the financial, economic, investment and innovative potential of a particular region.

Table 1. Regional Sustainability Indicators

Indicator group	Names of regional sustainability indicators
Investment climate indicators of a region	investment potential of a region (R), investment risk (Risk)
Resource endowment indicators	the proportion of the urban population in the total population (Urb), consumption per capita (Pt), proportion of small and medium enterprises in the gross regional product (MSP), regional budget expenditures per capita (Bd), dependency ratio (DN), proportion of products of high-tech and knowledge-intensive industries in the gross regional product (HT), proportion of public roads of regional or intermunicipal significance that meet regulatory requirements (NR), level of education (Ed), Gini coefficient (Gini), employment rate (UZ), the ratio of the net financial result to GRP (SFR), the proportion of shipped goods, works and services of the manufacturing industry in GRP (MI), the ratio of the volume of shipped goods of extractive industries to GRP (EI), the proportion of innovative goods in the total volume of shipped goods (In), industrial production index (IPP), consumer price index (IPC), debt ratio loans in roubles provided by credit institutions to the region's population (Zd), Ratio of investment in fixed assets to gross regional product (INV), human development index (HDI)
Performance indicators	The ratio of the balanced financial result to the gross regional product (SFR)

The websites of the Analytics Rating Agency, which collects statistics on the investment climate parameters [13], and the State Statistics for Russian Regions [14], the analysed periods are 2019 and 2020.

2 Methods

During the research, regions were clustered for each of the groups of economic stability indicators, three clusters were identified in each group, and a rank was assigned to each region depending on belonging to a cluster. As a result of this ranking, we divided all regions into two classes, i.e. economically safe and economically unsafe.

A factor analysis of the regional economic security was carried out with a machine learning algorithm based on a decision search tree, XGBoost. The advantages of using the algorithm in data mining are described in the works of Donchenko D. [15], Jabeur S.B. [16], Lundberg [17], Moeller K. [18], Scott M. [19], Xia Z. [20].

This algorithm uses the idea of gradient boosting based on adding the error of the current model to the subsequent model of the ensemble, so new classification trees to correct errors in the prediction of the already created model are created.

The Shap algorithm was used to interpret the classification model results. The importance of the determinants in the algorithm is determined based on the calculation of the Shapley value (vector) calculated as the difference between the prediction of the model without the studied factor and the prediction with the inclusion of information on the factor.

The research proposes a methodology for including inflation expectations and sentiments of economic entities in the regional economic security model.

The issues of assessing the impact of inflation expectations and sentiments as a channel of the transmission mechanism of monetary policy on economic systems are studied in the works of Karlova N.A. [21], Polyakova E.V., Vymyatnina Yu.V. [22], Balatsky E.V. and Yurevich M.A. [23].

In studying the dynamics of inflation expectations and subjective sentiment indicators from 2003 to 2021, we carried out a forecast of indicators with the Holt-Winters model, which resulted in an assessment of inflation expectations factors according to a binary scale.

3 Results and discussion

We have built a regional classification model in terms of economic security in the Jupyter Notebook environment.

Let us evaluate the influence of factors on model predictions with the Shap machine learning algorithm.

The diagram shown in Figure 1 illustrates how the determinants influence the final prediction of the model. The base value is the average value of the economic security class in the model. Highlighted in bold in the figure is the predictive value of the class in the model.

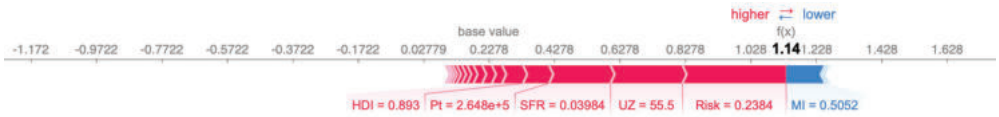

Fig. 1. Interpretation of the Regional Economic Security Model with the Shapley value.

The Shapley value highlighted in pink shifts the predictive estimate of the economic security class to one, and that in blue to zero.

Thus, the most significant for ensuring the region's stability (economic security) are such determinants as investment risk, employment level, net financial result, the proportion of shipped goods, works, manufacturing services in GRP, consumption and human development index.

Figure 2 shows a histogram demonstrating the ranking of variables by significance in the model.

Figure 3 shows a diagram of the Shapley density for each economic determinant. The features are sorted by the sum of the magnitudes of the Shapley values for the data being examined.

The diagram randomly selected the predictive value of the economic security class, which is highlighted with a bold pin along the ordinate axis, and the corresponding indicators of determinants such as risk, potential, employment level, net financial result, the proportion of manufacturing industries in the gross regional product and the human development index. At the same time, the investment potential and risk reduce the predictive indicator of the safety class, being in the blue zone, obviously due to the fact that the indicators do not even reach their median level. On the contrary, the values of the balanced financial result, employment level and human development index are high and correspond to the upper quartile of their distribution.

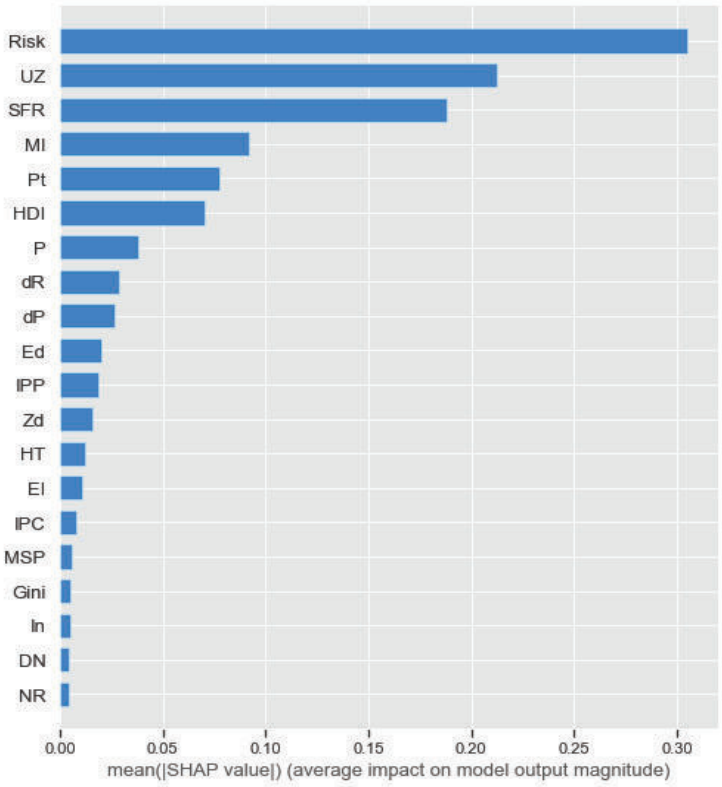

Fig. 2. Significance of variables in the economic stability interpretation model.

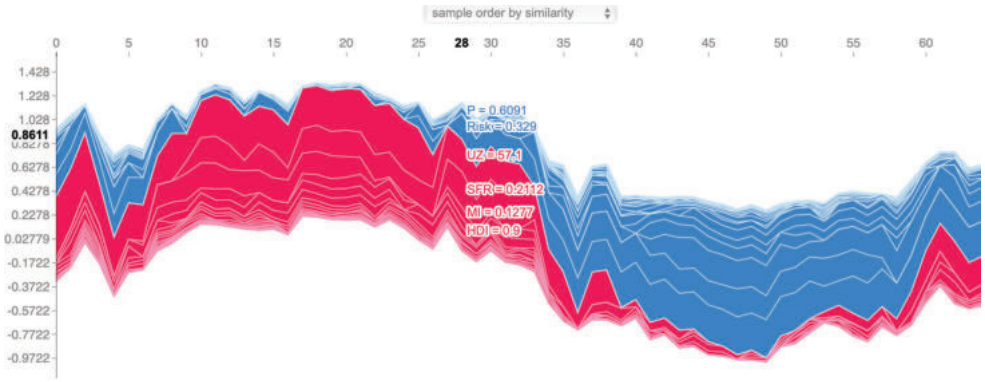

Fig. 3. Shapley density scatterplot.

We also made an attempt to include in the model the classification of regions according to the level of stability of factors that take into account the expectations and moods of agents.

Surveys commissioned by the Bank of Russia are used as a source of information on expectations and consumer sentiment [24].

In studying the dynamics of inflation expectations and sentiment indicators from 2003 to 2021, we compared the forecast values of expectation factors with the actual ones, and if

the actual indicator is more optimistic than the forecast one, the factor was assigned a positive class, otherwise a negative one class (Table 2)

Table 2. Characteristics of some inflation expectations and consumer sentiment indicators

Indicator	Actual value versus model value	Binary scale score
Core consumer price index	higher than predicted	0
Average disposable income per capita	higher than predicted	1
Bank lending tightness	higher than predicted	1
Observed inflation	lower than predicted	1
Consumer sentiment index	lower than predicted	0
Expectations index	higher than predicted	0
Major purchase Index	meets the predicted	0
Assessment of the prospects for production levels	lower than predicted	0

Having carried out a factor analysis of the economic stability taking into account inflation expectations and consumer sentiment, we obtained a ranking of determinants (Fig. 4)

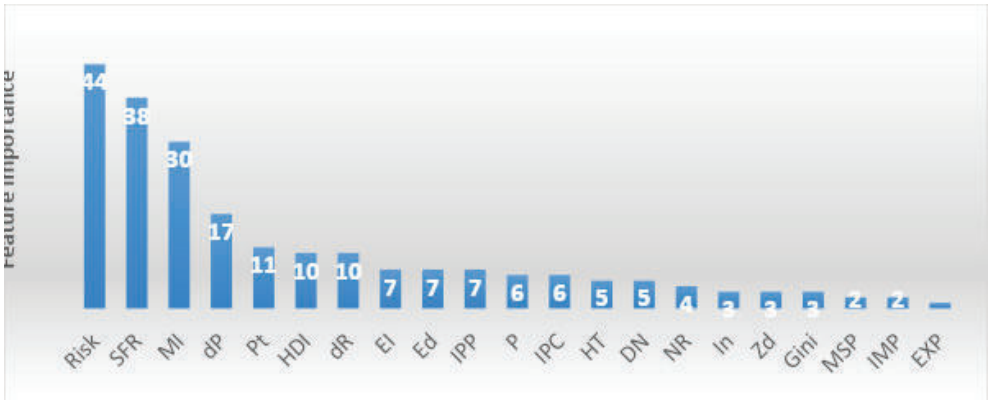

Fig. 4. Ranking of economic security determinants by degree of importance.

In Figure 4, the abscissa reflects F measure, which is a comprehensive classifier quality assessment calculated according to (1):

$$\text{F-measure} = \frac{2*Precision*Recall}{(Pecision+Recall)}, \quad (1)$$

where *Precision* – accuracy of the classifier;

Recall – completeness of the classifier.

It should be noted that all factors of inflation expectations and sentiment turned out to be insignificant due to the fact that at the moment only statistics at the federal level are available; that is, there is no data disaggregation by Russian regions.

4 Conclusion

During the research, we managed to identify the economic determinants that ensure regional stability over the period in question. We can see the prospect of further research, firstly, in studying the regional statistics related to the post-pandemic period, in order to draw a conclusion about the preservation or change in the regional development trends; secondly, in the analysis of inflation expectations as a channel for the monetary policy transmission mechanism in the regional aspect.

The study was carried out within the framework of the basic part of the state assignment of the Ministry of Education and Science of the Russian Federation, project 0729-2020-0056 "Modern methods and models for diagnosing, monitoring, preventing and overcoming crisis phenomena in the economy in the context of digitalization as a way to ensure the economic security of the Russian Federation"

References

1. G.E. Leonidova, N.M. Rumyantsev, Probl. Territ. Devel., **1(105)**, 52-63 (2020). https://doi.org/ 10.15838/ptd.2020.1.105.4
2. M.A. Abramova, L.L. Igonina, Fin.: Theory Pract., **22(1)**, 128-143 (2018). https://doi.org/10.26794/2587-5671-2018-22-1-128–143
3. A.I. Nazarov, Econ. Sec., **2(3)**, 255-263 (2019). https://doi.org/10.18334/ecsec.2.3.100656
4. M.V. Kurbatova, S.N. Levin, E.S. Kagan, D.V. Kislitsyn, Terra Econ., **17(3)**, 89-106 (2019). https://doi.org/10.23683/2073-6606-2019-17-3-89-106
5. M.Yu. Malkina, Spatial Econ., **3**, 88-114 (2018). https://doi.org/10.14530/se.2018.3.088-114
6. D.A. Maslennikov, S.N. Mityakov, L.Yu. Kataeva, T.A. Fedoseeva, Econ. Reg., **15(3)**, 707-719 (2019). https://doi.org/10.17059/2019-3-7
7. M.A. Pechenskaya, Econ. Soc. Changes: Facts, Trends, Forecast, **11(5)**, 61-73 (2018). https://doi.org/10.15838/esc.2018.5.59.4
8. Yu. Granitsa, *Analysis of the Interrelations of Economic Indicators as a tool for Predicting Regional Financial Instability*, in Proceedings of the International Scientific Conference "Far East Con" (ISCFEC 2020), 2772-2779 (2020). https://doi.org/10.2991/aebmr.k.200312.394
9. M. Billon, R. Marco, F. Lera-Lopez, Empir. Econ., **53(3)**, 1083-1108 (2017). https://doi.org/10.1007/s00181-016-1153-x
10. L. Fornaro, M. Wolf, COVID-19 Coronavirus and Macroeconomic Policy. Barcelona Graduate School of Economics. Working Paper No. 1168 (2020)
11. M. Fritsch, M. Wyrwich, Foresight STI Govern., **13(2)**, 62-71 (2019). https://doi.org/10.17323/2500-2597.2019.2.62.71
12. V. Guerrieri, G. Lorenzoni, L. Straub, I. Werning, Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages? Working Paper No. 26918 (2020). https://dx.doi.org/10.3386/w26918
13. Reiting investitsionnoi privlekatelnosti regionov RAEX za 2020 god [RAEX Regions Investament Attractiveness Rating for 2020]. Accessed on: March 01, 2022. [Online]. Available: https://raex-a.ru/ratings/regions/2020
14. Regiony Rossii. Sotsialno-ekonomicheskie pokazateli [Regions of Russia. Socio-economic indicators]. Accessed on: March 01, 2022. [Online]. Available: https://rosstat.gov.ru/folder/210/document/13204

15. D. Donchenko, N. Sadovnikova, D. Parygin, *Prediction of Road Accidents' Severity on Russian Roads Using Machine Learning Techniques*, in A. Radionov, O. Kravchenko, V. Guzeev, Y. Rozhdestvenskiy (eds.), Proceedings of the 5th International Conference on Industrial Engineering (ICIE 2019). Lecture Notes in Mechanical Engineering, 1493-1501 (Springer, Cham., 2020). https://doi.org/10.1007/978-3-030-22063-1_157
16. S.B. Jabeur, S. Mefteh-Wali, J.L. Ann Oper Res (2021). https://doi.org/10.1007/s10479-021-04187-w
17. S.M. Lundberg, G.E. Gabriel, S.-I. Lee, Consistent Individualized Feature Attribution for Tree Ensembles (2018). arXiv:1802.03888
18. K. Moeller, J. Reg. Sci., **58(4)**, 728-758 (2018). https://doi.org/10.1111/jors.12383
19. S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), 4768-4777 (Curran Associates Inc., Red Hook, NY, 2017)
20. Z. Xia, S. Xue, L. Wu et al., Distrib Parallel Databases **38**, 713-738 (2020). https://doi.org/10.1007/s10619-020-07294-y
21. N.A. Karlova, E.V. Puzanova, I.V. Bogacheva, Stud. Rus. Econ. Dev., **31**, 522-532 (2020). https://doi.org/10.1134/S1075700720050093
22. E.V. Polyakova, Yu.V. Vymyatnina, St Petersburg Univ. J. Econ. Stud., **37(3)**, 442-473 (2021). https://doi.org/10.21638/spbu05.2021.304
23. E.V. Balatsky, M.A. Yurevich, St Petersburg Univ. J. Econ. Stud., **34(4)**, 534-552 (2018). https://doi.org/10.21638/spbu05.2018.403
24. Inflyatsionnye ozhidaniya [Inflation expectations]. Accessed on: March 01, 2022. [Online]. Available: https://cbr.ru/statistics/ddkp/inflationary_expectations/