

Country Risk in International Investment Its' structure and methods of estimation

Alexey Ivkin¹

¹Student of Financial University under the Government of the Russian Federation, master program «International Finance», Moscow, Russia; vok-zal@mail.ru; +7 (916) 246–83–89.

Scientific Supervisor – A.I. Ilinsky, D. Sc. in engineering, professor of the Macroeconomic Regulation Faculty, Financial University under the Government of the Russian Federation, Moscow, Russia.

Abstract

In this paper, the issues of country risk assessment in economic security and sustainability context are investigated. The main object of research is country risk and its structural components. The scientific paper's main goal is to analyze valuation methods of country risk from different perspectives and suggest a model for country risk measurement that allows adequately evaluate country risk, economic security and economic sustainability level and dynamics, including structural components and their relationships. The paper approaches several main tasks. First, to highlight the importance of country risk evaluation and its assessment in growing global markets, analyzing causes and elements of country risk based on other scientific researches. Second, to explore and clarify advantages and disadvantages of the methods of country's risk assessment, as well as to investigate sources of country risk and ways how to manage the risk. Third, to apply quantitative and qualitative methods for analysis, formulate, create and present the model of country's risk assessment in economic security and sustainability context, which will identify factors, influencing country risk and determine their direct and indirect relationship between each other. The last task is to verify practical suitability of the model of country's risk assessment by performing empirical analysis all over the world, identifying directions for mitigating risk effects.

Keywords: country risk analysis; credit ratings; debt; structured qualitative methods; discriminant analysis; *k*-nearest neighbours (*k*-NN) classification; distance metrics; classification and regression trees (CART); support vector machine (SVM); discriminant analysis; missing data; Kolomogorov-Smirnov test; Anderson-Darling test; non-parametric methods.

JEL classification: C53, O16, O47

Each business operation causes some kind of risk. When business operations occur in international dimension, they bring additional risks, which are not typical for domestic operations. These additional risks are called country risks and usually include risks arising from a variety of national differences in policies, geography, economic structures, socio-political institutions, and currencies. Country risk analysis (CRA) tries to solve this problem by identifying the potential for these risks to decrease the expected return of cross-border investments.

Concept of “Country risk” began to be widely used in the 1970s. It was originally more professionally oriented in the sense that it aimed at addressing the concrete issue of a particular business

in a particular country and was generally used by the banking industry.

Every year it becomes more and more difficult to analyze and predict changes in the financial, economic and political sectors of business. The importance of country risk analysis is now more understandable and potential for it is growing by establishing more and more country risk rating agencies, which combine a wide range of qualitative and quantitative information regarding alternative measures of economic, financial, and political risk into associated composite risk ratings. However, the accuracy of any rating agency with regard to any or all of these measures is open to question.

Globalization, after undermining the old definition of economic security, is found at the centre of

a new definition that emphasizes the risks of unexpected shocks and economic volatility. The new definition must capture the causal consequences of globalization accurately and establish explicit benchmarks for assessing globalization's effects on economic security and country's economic sustainability.

Country risk assessment has been analyzed by different authors but in quite narrow way, in this paper the concept of country risk and influencing factors are presented in an extended view.

Following new scientific novelties in economics were discovered:

Expanded and consolidated overview of analyzes of country risk concept, its components and arising problems were analyzed in another angle which allowed to identify new possibilities and challenges for creating new model for assessment of country risk.

Broader analysis of country risk—includes not only political risk, but as well socio-economical aspects, presents clear and analyzed new concept which was not assumed in previous researches.

There are many studies related to country risk, its financial integration in a country, the impact on economics, and other aspects of country's welfare. To summarize the analysis of scientific literature about country risk, it is obvious that researchers are analyzing country risk approach only partially, not adapting the concept to growing globalization topic, which definitely makes changes in country risk approach. Country risk concept should be analyzed and understood in a broader way. This updated approach I will discuss further.

Methodologies of Country Risk and Methods for Its Valuation

Credit Ratings

Credit rating agencies (CRAs) play a key role in financial markets by helping to reduce the informative asymmetry between lenders and investors, on one side, and issuers on the other side, about the creditworthiness of companies or countries. CRAs' role has expanded with financial globalization and has received an additional boost from Basel II that incorporates the ratings of CRAs into the rules for setting weights for credit risk. Ratings tend to be sticky, lagging markets, and overreact when they do change. This overreaction may have aggravated financial

crises in the recent past, contributing to financial instability and cross-country contagion.

A credit rating is a current opinion and measure of the risk of an obligor with respect to a specific financial obligation based on all available information. For this purpose, S&P and Fitch define risk as the probability of default (PD), whereas Moody's define it as 'loss'.

The logic underlying the existence of CRAs is to solve the problem of the informative asymmetry between lenders and borrowers regarding the creditworthiness of the latter. Issuers with lower credit ratings pay higher interest rates embodying larger risk premiums than higher rated issuers. Moreover, ratings determine the eligibility of debt and other financial instruments for the portfolios of certain institutional investors due to national regulations that restrict investment in speculative-grade bonds.

Standard and Poor's ratings seek to capture only the forward-looking probability of the occurrence of default. They provide no assessment of the expected time of default or mode of default resolution and recovery values.

By contrast, Moody's ratings focus on the Expected Loss (EL) that is a function of both Probability of Default (PD) and the expected Recovery Rate (RE). Thus $EL = PD (1 - RE)$.

Fitch's ratings also focus on both PD and RE. They have a more explicitly hybrid character in that analysts are also reminded to be forward-looking and to be alert to possible discontinuities between past track records and future trends.

Models Used

A variety of statistical methods was employed to estimate models of debt rescheduling in the studies cited above. Since most authors chose their dependent variable to be a discrete binary variable, which took on the value one when a country 'rescheduled' within a given time-period and zero otherwise, the statistical methods used have been those designed for dichotomous dependent variables. These methods include discriminant analysis, linear-probability, probit, and logit models. In this section, we briefly describe each of these methods and then discuss criteria to use when choosing among them. Perforce, our discussion will be brief.

The methods used by the banks and other agencies for country risk analysis can broadly be classified as qualitative or quantitative. However,

many agencies amalgamate both qualitative and quantitative information into a single index or rating. The data was collected from various sources that include expert panel, survey, staff analysis, and published data sources. The country risk index could be either ordinal or scalar. A survey conducted by the US Export-Import Bank in 1976 categorized various methods of country risk appraisal used mainly by the banks into one of four types: (1) full qualitative method, (2) structured qualitative method, (3) checklist method, and (4) other quantitative method. Since our focus in this paper is on quantitative methods, we will only briefly discuss the other three categories.

Discriminant analysis. Discriminant analysis finds a set of prediction equations based on independent variables that are used to classify individuals into groups. There are two possible objectives in a discriminant analysis: finding a predictive equation for classifying new individuals or interpreting the predictive equation to better understand the relationships that may exist among the variables.

Objectives. The main objectives of discriminant analysis are:

Development of discriminant functions.

Examination of whether significant differences exist among the groups, in terms of the predictor variables.

Determination of which predictor variables contribute to most of the intergroup differences.

Evaluation of the accuracy of classification.

Discriminant analysis linear equation. A involves the determination of a linear equation like regression that will predict which group the case belongs to. The form of the equation or function is:

$$D = v_1X_1 + v_2X_2 + v_3X_3 + \dots + v_iX_i + a,$$

where

D = discriminate function

v = the discriminant coefficient or weight for that variable

X = respondent's score for that variable

a = a constant

i = the number of predictor variables.

Assumptions of discriminant analysis. The main assumptions of discriminant analysis are:

The observations are a random sample.

Each predictor variable is normally distributed.

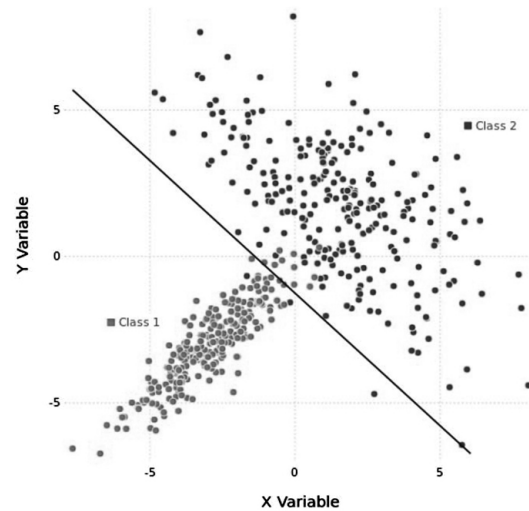


Figure 1. Example of linear discriminant analysis.

Each of the allocations for the dependent categories in the initial classification are correctly classified.

There must be at least two groups or categories, with each case belonging to only one group so that the groups are mutually exclusive and collectively exhaustive (all cases can be placed in a group).

Each group or category must be well defined, clearly differentiated from any other group(s) and natural.

The groups or categories should be defined before collecting the data.

The attribute(s) used to separate the groups should discriminate quite clearly between the groups so that group or category overlap is clearly non-existent or minimal.

Group sizes of the dependent should not be grossly different and should be at least five times the number of independent variables.

K-nearest neighbours (k-NN) algorithm. The k-NN (k-nearest neighbours) algorithm is a classification algorithm that can apply to question classification. However, its time complexity will increase linearly with the increase of training set size, which constrains the actual application effects of this algorithm. Simply stated, k-NN is an algorithm that classifies the new cases based on similarity measures or distance measures of pair of observations such as euclidean, cosine, etc. k-NN algorithm is a lazy learner i.e. it does not learn anything from the training tuples and simply uses the training data itself for classification. It is a non-parametric method used for classification and regression.

Different types of prediction using data mining techniques are:

(1) *Classification*: predicting into what category or class a case falls.

(2) *Regression*: predicting what number value a variable will have (if a variable varies with time, it is called 'time series' prediction).

Classification problems aim to identify the characteristics that indicate the group to which each case belongs. This pattern can be used both to understand the existing data and to predict how new instances will behave. Data mining creates classification models by examining already classified data (cases) and inductively finding a predictive pattern.

k-NN for classification. In pattern recognition, the *k-NN* algorithm is a method for classifying objects based on closest training examples in the feature space. *k-NN* is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification.

Figure 2 shows the *k-NN* decision rule for $K = 1$ and $K = 4$ for a set of samples divided into 2 classes. In Figure 2(a), an unknown sample is classified by using only one known sample; in Figure 2(b) more than one known sample is used. In the last case, the parameter K is set to 4, so that the closest four samples are considered for classifying the unknown one. Three of them belong to the same class, whereas only one belongs to the other class. In both cases, the unknown sample is classified as belonging to the class on the left.

Distance Metric. As mentioned before KNN makes predictions based on the outcome of the K neighbors closest to that point. Therefore, to make predictions with KNN, we need to define a metric for measuring the distance between the query point and cases from the examples sample. One of the most popular choices to measure this distance is known as Euclidean. Other measures include Euclidean squared, City-block, and Chebychev.

$$D(x, p) = \begin{cases} \sqrt{(x-p)^2} \\ (x-p)^2 \\ |x-p| \\ \text{Max}(|x-p|), \end{cases}$$

where x and p are the query point and a case from the examples sample, respectively.

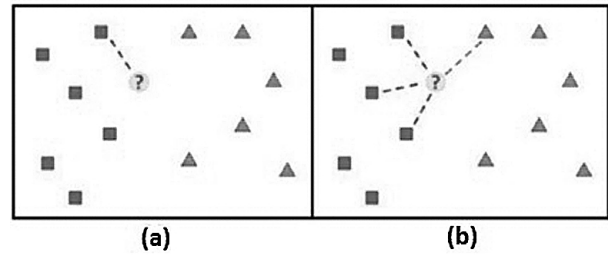


Figure 2. (a) The 1-NN decision rule: the point is assigned to the class on the left; (b) the KNN decision rule, with $K = 4$ the point is assigned to the class on the left as well.

The advantages of the *k-NN* method are as follows:

1. Analytically tractable.
2. Simple implementation.
3. Nearly optimal in the large sample limit ($NN \rightarrow \infty$).
4. Uses local information, which can yield highly adaptive behaviour.
5. Lends itself very easily to parallel implementations.

The drawbacks of the *k-NN* method are as follows:

1. *k-NN* algorithm is that it is a *lazy learner*, i.e. it simply uses the training data itself for classification.
2. Result of this is that the method does not learn anything from the training data, which can result in the algorithm not generalizing well. Further, changing K can change the resulting predicted class label.
3. Also algorithm may not be robust to noisy data.
4. To predict the label of a new instance the *kNN* algorithm will find the K closest neighbours to the new instance from the training data, the predicted class label will then be set as the most common label among the K closest neighbouring points.
5. The algorithm needs to compute the distance and sort all the cases at each prediction, which can be slow if there are a large number of training examples.
6. Large storage requirements.
7. Computationally intensive recall.
8. Highly susceptible to the curse of dimensionality.

Classification and regression trees (CART) method. CART is one method of machine learn-

ing which the exploration method is done by decision tree technique. The method developed by Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone is classification technique by using binary recursive partition algorithm.

Classification tree methods such as CART are convenient way to produce a prediction rule from a set of observations described in terms of a vector of features and a response value. The aim is to define a general prediction rule, which can be used to assign a response value to the cases solely on the bases of their predictor (explanatory) variables. Tree-structured classifications are not based on assumptions of normality and user-specified model statements, as are some conventional methods such as discriminant analysis and ordinary least square regression.

Tree based decision methods are statistical systems that mine data to predict or classify future observations based on a set of decision rules and are sometimes called rule induction methods because the reasoning process behind them is clearly evident when browsing the trees.

In CART, the observations are successively separated into two subsets based on associated variables significantly related to the response variable; this approach has an advantage of providing easily comprehensible decision strategies.

CART can be applied either as a classification tree or as a regressive tree depending on whether the response variable is categorical or continuous. Tree based methods are not based on any stringent assumptions. These methods can handle large number of variables, are resistant to outliers, non-parametric, more versatile, can handle categorical variables, though computationally more intensive.

CART methodology. For building decision trees, CART uses so-called learning set- a set of historical data with pre-assigned classes for all observations. An algorithm known as recursive partitioning is the key to the nonparametric statistical method of CART. It is a step-by-step process by which a decision tree is constructed by either splitting or not splitting each node on the tree into two daughter nodes.

An attractive feature of the CART methodology is that because the algorithm asks a sequence of hierarchical questions, it is relatively simple to

understand and interpret the results. The unique starting point of a classification tree is called a root node and consists of the entire learning set L at the top of the tree.

Steps in Cart. CART analysis consists of four basic steps. The first step consists of tree building, during which a tree is built using recursive splitting of nodes. Each resulting node is assigned a predicted class, based on the distribution of classes in the learning dataset which would occur in that node and the decision cost matrix. The assignment of a predicted class to each node occurs whether or not that node is subsequently split into child nodes.

The second step consists of stopping the tree building process. At this point a "maximal" tree has been produced, which probably greatly overfits the information contained within the learning dataset.

The third step consists of tree "pruning," which results in the creation of a sequence of simpler and simpler trees, through the cutting off of increasingly important nodes.

The fourth step consists of optimal tree selection, during which the tree which fits the information in the learning dataset, but does not overfit the information, is selected from among the sequence of pruned trees. Each of these steps will be discussed in more detail below.

The tree-building process starts by partitioning a sample or the root node into binary nodes based upon a very simple question of the form is $X \leq d$?, where X is a variable in the data set and d is a real number. Initially, all observations are placed in the root node. This node is impure or heterogenous because it contains observations of mixed classes. The goal is to devise a rule that will break up these observations and create groups or binary nodes that are internally more homogenous than the root node. Starting from the root node, and using, for example, the Gini diversity index as a splitting rule, the tree building process is as follows:

1. CART splits the first variable at all of its possible split points (at all of the values the variable assumes in the sample). At each possible split point of a variable, the sample splits into binary or two child nodes. Cases with a "yes" response to the question posed are sent to the left node and those with "no" responses are sent to the right node.

2. CART then applies its goodness-of-split criteria to each split point and evaluates the reduction in impurity that is achieved using the formula

$$\Delta i(s, t) = i(t) - p_L[i(t_L)] - p_R[i(t_R)],$$

3. CART selects the best split of the variable as that split for which the reduction in impurity is highest.

4. Steps 1–3 are repeated for each of the remaining variables at the root node.

5. CART then ranks all of the best splits on each variable according to the reduction in impurity achieved by each split.

6. It selects the variable and its split point that most reduced the impurity of the root or parent node.

7. CART then assigns classes to these nodes according to the rule that minimizes misclassification costs. CART has a built-in algorithm that takes into account user-defined variable misclassification costs during the splitting process. The default is unit or equal misclassification costs.

8. Because the CART procedure is recursive, steps 1–7 are repeatedly applied to each nonterminal child node at each successive stage.

9. CART continues the splitting process and builds a large tree. The largest tree is built if the splitting process continues until every observation constitutes a term in a node. Obviously, such a tree will have a large number of terminal nodes, which will be either pure or have very few cases (less than 10).

Summing up CART's strengths and weaknesses:

1. CART makes no distributional assumptions of any kind for dependent and independent variables. No variable in CART is assumed to follow any kind of statistical distribution.

2. The explanatory variables in CART can be a mixture of categorical and continuous.

3. CART has a built-in algorithm to deal with the missing values of a variable for a case, except when a linear combination of variables is used as a splitting rule.

4. CART is not at all affected by the outliers, collinearities, heteroskedasticity, or distributional error structures that affect parametric procedures. Outliers are isolated into a node and thus have no effect on splitting. Contrary to situations in parametric modeling, CART makes use of collinear variables in "surrogate" splits.

5. CART has the ability to detect and reveal variable interactions in the dataset.

6. CART does not vary under a monotone transformation of independent variables; that is, the transformation of explanatory variables to logarithms or squares or square roots has no effect on the tree produced.

7. In the absence of a theory that could guide a researcher, in a famine vulnerability study, for example, CART can be viewed as an exploratory, analytical tool. The results can reveal many important clues about the underlying structure of famine vulnerability.

8. CART's major advantage is that it deals effectively with large datasets and the issues of higher dimensionality; that is, it can produce useful results from a large number of variables submitted for analysis by using only a few important variables.

9. The inverted-tree-structure results generated from CART analysis are easy for anyone to understand in any discipline.

However, CART analysis does have some limitations. CART is a blunt instrument compared to many other statistical and analytical techniques. At each stage, the subdivision of data into two groups is based on only one value of only one of the potential explanatory variables. If a statistical model that appears to fit the data exists, and if its' basic assumptions appear to be satisfied, that model would be preferable, in general, to a CART tree.

A weakness of the CART method and, hence, of the conclusions it may yield is that it is not based on a probabilistic model. There is no probability level or confidence interval associated with predictions derived from a CART tree that could help classify a new set of data. The confidence that an analyst can have in the accuracy of the results produced by a given CART tree is based purely on that tree's historical accuracy — how well it has predicted the desired response in other, similar circumstances.

SVM model. Support vector machine (SVM) method which is widely used in the research area of artificial intelligence, is based on the theory of statistical learning which is concluded by combining VC dimension theory with principle of risk minimum. It is based on complexity and learning ability of the model to seek the best compromise, and finally get the best generalization ability.

Classification algorithm of SVM. SVM, which is under the condition of linear separable case, is

developed from the optimal separating hyperplane. In Figure 3, we show the basic principle of SVM. On the diagram, solid and hollow points on behalf of the two kinds of samples, H is the classification hyperplane, H_1 is sample that is most close to H and H_2 is hyperplanes which is parallel to H . They have equal distance to H , the distance between them is termed classification interval. The hyperplane is the one that separates the two types with biggest distance, for such an ill-posed problem of classifying two kinds of samples, the optimal separating hyperplane has maximum stability and high generalization ability.

Steps involved in the design of SVM are as follows:

1. Hyperplane acting as the decision surface is defined as

$$\sum_{i=1}^N \alpha_i d_i K(x, x_i) = 0$$

where $K(x, x_i) = \phi^T(x)\phi(x_i)$ represents the inner product of two vectors induced in the feature space by the input vector x and input pattern x_i pertaining to the example. This term is referred to as inner-product kernel.

Where

$$W = \sum_{i=1}^N \alpha_i d_i \phi(x_i)$$

$$\phi(x) = [\phi_0(x), \phi_1(x), \dots, \phi_{m_1}(x)]^T$$

$$\phi_0(x) = 1 \text{ for all } x$$

w_0 denotes the bias b .

2. The requirement of the kernel $K(x, x_i)$ is to satisfy Mercer's theorem. The kernel function is selected as a polynomial learning machine

$$K(x, x_i) = (1 + x^T x_i)^2.$$

3. The Lagrange multipliers $\{\alpha_i\}$ for $i = 1$ to N that maximize the objective function $Q(\alpha)$, denoted by α_0 , i is determined

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x, x_j).$$

Subject to the following constraints:

$$\sum_{i=1}^N \alpha_i d_i = 0$$

$$0 \leq \alpha_i \leq C \text{ for } i = 1, 2, \dots, N.$$

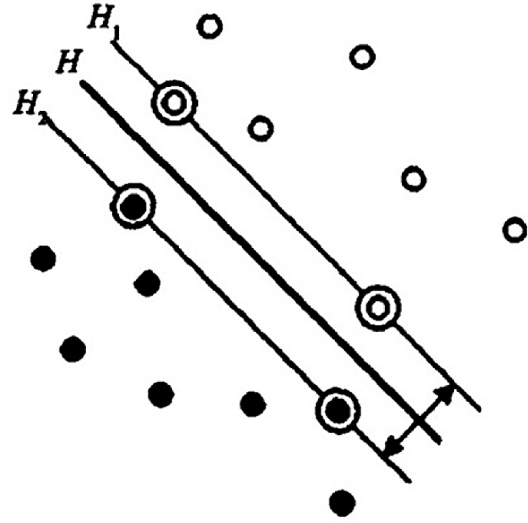


Figure 3. Optimal hyperplane for linearly separable patterns with category interval = $2/\|w\|$.

4. The linear weight vector w_0 corresponding to the optimum values of the Lagrange multipliers was determined using the following formula:

$$w_0 = \sum_{i=1}^N \alpha_{0,i} d_i \phi(x_i)$$

$\phi(x_i)$ is the image induced in the feature space due to x_i ; w_0 represents the optimum bias b_0 .

Application of the Building Models Process to the Countries' Conditions

Data source. The statistical data of the economic and financial variables considered in this paper come from the following sources:

International Monetary Fund (World Economic Outlook database) — <http://www.imf.org/external/index.htm>

The World Bank (World Development Indicators database) — <http://www.worldbank.org/>

Exogeneous variables. I used as an exogeneous variables 387 macroeconomic factors with the coverage period from 1980 up to 2016 for 217 countries, basing on which the model has been reproduced. Concerning the open source data, should be taken into account the various density of data filling while transitioning from one factor to another. Furthermore, fulfilling of the most factors was very low, which makes the issue of their usage problematic. There are completely filled factors for the entire time period, but some of the data is partially filled.

Table 1
Lists the different country risk levels or labels used by rating agencies

| Conversion of response into a numerical scale | | | | |
|---|------|---------|-------|---|
| TE | S&P | Moody's | Fitch | |
| 100 | AAA | Aaa | AAA | Prime |
| 95 | AA+ | Aa1 | AA+ | High grade |
| 90 | AA | Aa2 | AA | |
| 85 | AA- | Aa3 | AA- | |
| 80 | A+ | A1 | A+ | Upper medium grade |
| 75 | A | A2 | A | |
| 70 | A- | A3 | A- | |
| 65 | BBB+ | Baa1 | BBB+ | Lower medium grade |
| 60 | BBB | Baa2 | BBB | |
| 55 | BBB- | Baa3 | BBB- | |
| 50 | BB+ | Ba1 | BB+ | Non-investment grade |
| 45 | BB | Ba2 | BB | speculative |
| 40 | BB- | Ba3 | BB- | |
| 35 | B+ | B 1 | B+ | Highly speculative |
| 30 | B | B 2 | B | |
| 25 | B- | B 3 | B- | |
| 20 | CCC+ | Caa1 | CCC | Substantial risks |
| 15 | CCC | Caa2 | | Extremely speculative |
| 10 | CCC- | Caa3 | | In default with little prospect for recovery |
| | CC | Ca | | |
| 5 | C | C | | |
| 0 | D | / | DDD | In default |
| | | / | DD | |
| | | | D | |

Endogeneous variables. As a response, I used the credit ratings of countries from the leading rating agencies S&P, Moody's and Fitch. We have received a history of ratings since 1949. It is worth mentioning that in the initial period of time the process of assigning ratings, covered a limited number of countries. Only since 1990, the sample of ratings has become sufficiently representative to construct statistical classification models.

Conversion of response into a numerical scale. To quantify the data of the three rating agencies, to classify the classes, and to obtain the possibility of applying the models of classification with the models of generalized regression, the response was digitized.

We have converted the Standard & Poor rating scale (ranging from AAA to SD) into a numerical

scale (ranging from 0 to 100) and shall liberally refer to both of them as S&P ratings.

Table 1 lists the different country risk levels or labels used by Standard & Poor, Moody's and Fitch and provides also descriptions associated with these labels. Countries which are assigned a label inferior to BB+ are considered as non-investment grade (speculative) countries. Countries rated CCC+ or lower are regarded as presenting serious default risks. BB indicates the least degree of speculation and CC the highest. Ratings labeled from AA to CCC can be modified by the addition of a plus or minus sign to show relative standing within the major rating categories. We consider such subcategories as separate ratings in our analysis.

Data conversion. Using the Matlab software environment, we converted the data into the fol-

lowing form: Country, Year, Rating, and a set of all factors.

Missing data. The vital criterion is the availability of complete and reliable statistics. We eager to avoid difficulties related to missing ratings data that could reduce the statistical significance and the scope of our analysis. Further, we exclude all the rows from the training sample, in which the values of the ratings are not specified. Eventually, the volume of the training sample was reduced to 2395 points. A vast number of variables have a small coverage, hence we exclude those factors, in which the percentage of a missing data exceeds 40. Table 7 in the Appendix lists the factors' coverage.

Correlation analysis. We conducted a correlation analysis between the 164 factors and the response, in order to select the factors that will be used in further analysis in models' constuctioning. This correlation coefficient was calculated taking into account the presence of gaps in these factors. For this purpose, a special program MATLAB was used. Next, we sorted the factors, in descending order, by reducing the absolute value of the correlation of factors.

Factors' selection for the model construction. The criterion is the significance of variables for estimating a country's creditworthiness. We have performed an extensive literature review which played an important role in defining the set of candidate variables for inclusion in our model. Based on the correlation matrix, economic interpretation of factors and our representations, we selected a certain number of factors for models' constructioning. It worth mentioning that within the process of factors' selection, we have omitted those factors that had a close economic sense and, correspondingly, were too correlated with each other. Thus, we tried to avoid the problem associated with multicollinearity. The clarification of the factors, participated in the model constructioning are listed below. Furthermore, the correlation between factors and response has been clarified. Table 2 represents the correlation matrix between the factors and response which have been selected.

The interpretation of the correlation matrix coefficients. For the purpose of analyzing the correlation matrix, we introduce the notion of a coupling force. It is generally accepted that the strength of the correlation coefficient, as one indicator of a boundary measure, is differentiated into three levels for both positive and negative

correlations. First and foremost, we will commence with analyzing the correlation coefficients between predictors and response. As we can see from the correlation matrix, generally there are positive correlation coefficients as well as negative ones. According to the notion of the coupling force, we will determine the following factors as ones with the positive strong coupling force: Adjusted net national income per capita (constant 2010 US\$), Household final consumption expenditure per capita (constant 2010 US\$), GDP per capita (constant 2010 US\$). To make the analysis clearer and concise, we will clarify the correlation of each factor in the way that is more detailed.

If to consider the further factor, Adjusted net national income per capita (constant 2010 US\$), with the most strong correlation of 0.826, we may conclude, that with the slight increase of Adjusted net national income per capita, the S&P rating increases considerably. To make it easy for understanding, we will clarify that with the increase of net national income per capita, the consumption is increasing, the savings are also increasing, and therefore the economy of the country is boosted. Hence, the country becomes more reliable to invest, as the probability of debt repayment i.e. meeting the obligations increases. Therefore, the rating agencies assign more enhanced rating to this country.

Moving on to the Household final consumption expenditure per capita (constant 2010 US\$), with the correlation of 0.802, we may see the similar tendency. With the growth of this indicator, the assigned rating enhances in the same way. To make it clear, we will build the following boundaries. The Household final consumption expenditure increases, which stimulate the production, therefore the demand for goods and services is growing and the whole economy is flourishing. Hence, the country becomes more reliable to invest, as the probability of debt repayment i.e. meeting the obligations increases, which ultimately leads for country rating boost.

If to cover the last indicator in that group, namely, GDP per capita (constant 2010 US\$), it is obvious, that its correlation with the response is relatively less in comparison with the above mention pair of factors, but still strong — 0.781. To make it more clear, with an increase of GDP per capita, both the consumption, the demand, the savings, the living standards are expanding. Hence,

Table 2
Correlation matrix

| Factors | S&P rating | Adjusted net national income per capita (constant 2010 US\$) | Household final consumption expenditure per capita (constant 2010 US\$) | GDP per capita (constant 2010 US\$) | Household final consumption expenditure, etc. (% of GDP) | Adjusted savings: consumption of fixed capital (% of GNI) | Urban population (% of total) | Lending interest rate (%) | Exports of goods and services (constant 2010 US\$) | Broad money to total reserves ratio | General government final consumption expenditure (% of GDP) |
|---|------------|--|---|-------------------------------------|--|---|-------------------------------|---------------------------|--|-------------------------------------|---|
| Adjusted net national income per capita (constant 2010 US\$) | 0.826 | 1.000 | 0.971 | 0.983 | -0.520 | 0.514 | 0.558 | -0.420 | 0.466 | 0.321 | 0.436 |
| Household final consumption expenditure per capita (constant 2010 US\$) | 0.802 | | 1.000 | 0.938 | -0.386 | 0.528 | 0.546 | -0.387 | 0.490 | 0.375 | 0.426 |
| GDP per capita (constant 2010 US\$) | 0.781 | | | 1.000 | -0.541 | 0.470 | 0.559 | -0.387 | 0.396 | 0.287 | 0.359 |
| Household final consumption expenditure, etc. (% of GDP) | -0.563 | | | | 1.000 | -0.332 | -0.381 | 0.257 | -0.172 | -0.034 | -0.239 |
| Adjusted savings: consumption of fixed capital (% of GNI) | 0.515 | | | | | 1.000 | 0.368 | -0.196 | 0.266 | 0.187 | 0.417 |
| Urban population (% of total) | 0.470 | | | | | | 1.000 | -0.153 | 0.251 | 0.176 | 0.399 |
| Lending interest rate (%) | -0.450 | | | | | | | 1.000 | -0.280 | -0.202 | -0.094 |
| Exports of goods and services (constant 2010 US\$) | 0.446 | | | | | | | | 1.000 | 0.447 | 0.144 |
| Broad money to total reserves ratio | 0.380 | | | | | | | | | 1.000 | 0.090 |
| General government final consumption expenditure (% of GDP) | 0.379 | | | | | | | | | | 1.000 |

the country becomes more reliable to invest, as the probability of debt repayment i.e. meeting the obligations increases. Therefore, the rating agencies assign more enhanced rating to this country.

Moreover, among selected factors, we may refer the ones to the group with the positive moderate coupling force:

Adjusted savings: consumption of fixed capital (% of GNI), Urban population (% of total), Exports of goods and services (constant 2010 US\$), Broad money to total reserves ratio, General government final consumption expenditure (% of GDP).

The first one, exactly, Adjusted savings: consumption of fixed capital (% of GNI), has the strongest correlation between the response within this group of 0.515. Consequently, we may resume, that with the increase of this indicator, the S&P rating as well increases, but in more stepless way. In order to clarify, we will build up the whole chain. When the level of savings increases, hence the consumption (demand) may decrease, as well as the production and GDP. Ultimately, the country rating is less likely to increase.

The next one, in particular, Urban population (% of total), has slightly less correlation of 0.470 in comparison with the previous one. Subsequently, we may affirm that with the increase of this indicator, the rating will slightly increase as well. To make it more clear, it should be emphasized that the more urban population, the less rural one. If accept this phenomena, then we may see the logic that the overall country economy will be transferred from the agricultural to the industrial type. In other words, industrialization may occur, which in turn, is going to increase the production capacity, level of consumption, living standards, volume of savings and ultimately GDP. Hence, the country becomes more reliable to invest, as the probability of debt repayment i.e. meeting the obligations increases. Therefore, the rating agencies assign more enhanced rating to this country.

The following factor, Exports of goods and services (constant 2010 US\$), has the correlation of 0.446, which relatively the same as the previous one. If to clarify, with the expansion of Exports of goods and services in the country, the rating will experience minor change. To make it more clearly, we will clarify the overall chain. With the increase of Exports of goods and services, the production capacity increases, hence, the unemployment is decreasing, the living standards are enhancing and

lastly the GDP is increasing. Hence, the country becomes more reliable to invest, as the probability of debt repayment i.e. meeting the obligations increases. Therefore, the rating agencies assign more enhanced rating to this country.

The last two factors from this group have approximately the same correlation coefficients. The first one, Broad money to total reserves ratio, has relatively small correlation of 0.380. It means that with the small increase of this coefficient, the rating is less likely to be altered. To make it more clear and concise, we may clarify, that with the increase of this ratio the volume of the broad money is expanding, bringing about the inflation growth. Therefore, in response latter, the prices are going to be increased. Hence, the demand will go inverse, followed by the shrinkage of supply. Ultimately, the overall economy development is going to be imprompted, causing the ratings to be less likely enhanced.

The one and the last from this group, namely, General government final consumption expenditure (% of GDP), has the correlation of 0.379. Consequently, we may resume, that with the increase of this indicator, the S&P rating as well increases, but in more steeples way. In order to clarify, we will build up the whole chain of clarification. Due to an increase of this indicator, we may state that the general demand will undergo the growing phase. Thus, it will stipulate the increase of supply, which in turn will lead to the expansion of the sector's of economy, ending up the economy development. Hence, the country becomes more reliable to invest, as the probability of debt repayment i.e. meeting the obligations increases. Therefore, the rating agencies assign more enhanced rating to this country.

The ultimate group of factors, which ought to be mentioned has the negative moderate coupling force. Referring to the first one, Household final consumption expenditure, etc. (% of GDP), the correlation of which is standed for -0.563 . To be more consize, it means that with the increase of this indicator, the country rating will be diminished.

To make it clearly for understanding, if the consumption of the households is going to be increased, hence the less funds will be left for savings (putting money under deposits). Therefore, due to the shortage of the deposits, the banks will not have sufficient funds to make credit operations. Therefore, they will fall back on increasing

interest rates (in the case of the rejection of credit from the Central Bank), which will make the access for the funds more sophisticated. Ultimately, the volume of investments will be decreased harshly, slowing down the growth of the GDP. Therefore, the rating is less likely to be altered.

And, the ultimate factor, Lending interest rate (%), which has the correlation of -0.450 , means that with the increase of the Lending interest rate, the country rating is to be deteriorated. In other words, the increase of the key interest rate is accompanied with the tight monetary policy which is proclaimed by the Central Bank of country. Therefore, the level of cash distribution throughout the economic agents and the economy in the whole is significantly falling down. In turn, this phenomenon gives a rise to shrinkage of the volume of investment as within the country as well as from outside it.

It is worth mentioning, that according to the yield curve, increased yield is giving rise to the risk exposure, which is similar to our context with the interest rates. Ultimately, the economy development is deteriorating, country risks are going to rocket and in that case the probability of increased country rating is going down.

As we have already described the correlation between predictor and response, its time to analyze the correlation between factors as well.

We will give a rise to such notion as multicollinearity. It means that there is any mutual strong interconnection between factors within the sample. It is worth mentioning that if it exists, therefore it deteriorates the sample with the additional boundaries between factors. Ideally, the correlation between factors should be tended to vanish (seek a null position). In other words, they ought to be statistically independent.

If you analyze our sample, it should be clearly seen, that there is a high multicollinearity between the following factors:

1. Adjusted net national income per capita (constant 2010 US\$) and Household final consumption expenditure per capita (constant 2010 US\$).
2. Adjusted net national income per capita (constant 2010 US\$) and GDP per capita (constant 2010 US\$).
3. Household final consumption expenditure per capita (constant 2010 US\$) and GDP per capita (constant 2010 US\$).

The multicollinearity between these 3 pairs of factors is almost 1, which considerable dete-

riorates our sample. Conducting the correlation analysis of the data sample, we may conclude that the majority of factors have a high correlation with the response value. The highest value is presented by *Adjusted net national income per capita (constant 2010 US\$)* and the least by *General government final consumption expenditure (% of GDP)* respectively.

Moreover, it can be clearly seen, that there is a high Multicollinearity (almost 1) between the following factors:

1. Adjusted net national income per capita (constant 2010 US\$) and Household final consumption expenditure per capita (constant 2010 US\$).
2. Adjusted net national income per capita (constant 2010 US\$) and GDP per capita (constant 2010 US\$).
3. Household final consumption expenditure per capita (constant 2010 US\$) and GDP per capita (constant 2010 US\$).

Summing up the carried out descriptive statistics, namely, Group Means and Confidence Intervals, Descriptive Statistics for the group of factors, and Correlation matrix, we can conclude that:

On the one hand, there is the variety of advantages over the data sample, which are the following:

1. First and foremost, the majority of factors have relatively high correlation coefficients with the response value.
2. Moreover, across almost all factors the intense multicollinearity is not observed.

The analysis of the relationship between factors and rating based on mean group values and confidence intervals. Eventually, with the help of conducting the descriptive statistics concerning Group Means and Confidence Intervals, we may conclude that, almost all factors have a positive correlation with the response except Household final consumption expenditure, etc. (% of GDP) and Lending interest rate (%).

In addition, due to the corresponding of one value of the factor to various confidence intervals, countries can be assigned a different rating. Thus, we may conclude that for the factor levels corresponding to different confidence intervals, the value of the rating is adjusted by other factors.

It is obvious that low rating values are prescribed by a narrow confidence interval, due to a small spread of factor values between the minimum and maximum values. Meanwhile, as the

Table 3
 Discriptive statistics for the group of factors

| Factors | Mean | Min | Max | STD | Median | Mode | Skew-ness | Kur-tosis |
|---|--------------|-----------|---------------|--------------|-------------|-----------|-----------|-----------|
| Adjusted net national income per capita (constant 2010 US\$) | 14377.77 | 277.50 | 66462.98 | 15203.35 | 7048.62 | 277.50 | 1.24 | 3.60 |
| Household final consumption expenditure per capita (constant 2010 US\$) | 9496.61 | 288.74 | 41428.17 | 9469.80 | 5461.40 | 288.74 | 1.29 | 3.94 |
| GDP per capita (constant 2010 US\$) | 18633.51 | 335.86 | 91593.67 | 20066.35 | 9002.54 | 335.86 | 1.36 | 4.11 |
| Household final consumption expenditure, etc. (% of GDP) | 59.35 | 13.41 | 89.27 | 12.62 | 58.86 | 13.41 | -0.44 | 4.12 |
| Adjusted savings: consumption of fixed capital (% of GNI) | 13.31 | 2.07 | 24.13 | 4.63 | 13.63 | 2.07 | -0.11 | 2.38 |
| Urban population (% of total) | 66.34 | 8.55 | 100.00 | 21.03 | 72.63 | 100.00 | -0.78 | 2.95 |
| Lending interest rate (%) | 12.55 | 0.50 | 86.36 | 9.35 | 10.48 | 5.38 | 3.15 | 17.82 |
| Exports of goods and services (constant 2010 US\$) | 123303072753 | 494277379 | 2208569650600 | 192036214575 | 59391883047 | 494277379 | 4.72 | 40.77 |
| Broad money to total reserves ratio | 7.61 | 0.24 | 90.94 | 12.53 | 3.49 | 0.24 | 3.66 | 17.32 |
| General government final consumption expenditure (% of GDP) | 15.99 | 5.69 | 27.63 | 4.75 | 16.00 | 5.69 | 0.17 | 2.19 |

rating increases, the spread of factor values rises as well, which leads to wide confidence intervals.

By no means unimportant is that due to the peculiarity of the data sample, for a high rating values, namely from 80 to 90, there is a low number of observations for this factor, which leads to

fluctuations for both the mean value and confidence interval.

The summary of carrying out the discriptive statistics for the group of factors. Conducting the descriptive analysis for the group of factors, we may conclude that almost all factors has multi-

modal distribution (has two modes). Consequently, we may conclude that it is far from the normal distribution.

Besides, the sample has right-sided asymmetry, as evidenced by a positive Skewness. In other words, the mean value and the median one is greater than the mode value. This suggests that our data is shifted to the left, relative to the normal distribution. Consequently, there is an abnormality in the distribution of the values of factors.

It should not go unspoken about the fact, there is an excess in the data sample. For the major part of data sample, the peak values are higher than the normal distribution. This is confirmed by a value of Kurtosis more than 3 whereas the normal value is 3.

However, the following factors are different from the data sample: Household final consumption expenditure, etc. (% of GDP), Adjusted savings: consumption of fixed capital (% of GNI), Urban population (% of total).

The distribution is unimodal for this group of factors (has one mode). Consequently, we may conclude that it is close to the normal distribution.

However, the sample has left-sided asymmetry, as evidenced by a negative value of Skewness. In other words, in comparison with previous factors, the mean value and the median one is not much but still greater than the mode value. This suggests that our data is slightly shifted to the right, relative to the normal distribution. Consequently, there is a normality in the distribution of the values of factors.

Furthermore, there is an excess in the data sample. The peak values are less than normal distribution. This is confirmed by a value of kurtosis less than 3, whereas the normal value is 3.

Conducting the tests on normality. Based on our assumption of non-parametricity of data, we are going to test our data on normality, using the number of tests. The normality assumption is at the core of a majority of standard statistical procedures, and it is important to be able to test this assumption. In addition, showing that a sample does not come from a normally distributed population is sometimes of importance *per se*. Among the procedures used to test this assumption, one of the most well known is a Kolmogorov–Smirnov test.

First, we verify the null hypothesis concerning the normality of distribution using The Kolmogorov–

Smirnov test. In Matlab application, we used the *ks*-test function. It returns a test decision for the null hypothesis that the data in vector x comes from a standard normal distribution, against the alternative that it does not come from such a distribution, using the one-sample Kolmogorov–Smirnov test. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, or 0 otherwise.

The findings of the Kolmogorov–Smirnov test showed that throughout the sample, all factors reject null hypothesis. Therefore, we may conclude that there is an absence of normality and our sample has non-normal distribution.

Moreover, we conducted the additional test to verify the null hypothesis concerning the normality of distribution using the Anderson–Darling test. In Matlab application, we used the *ad*-test function. It returns a test decision for the null hypothesis that the data in vector x is from a population with a normal distribution, using the Anderson–Darling test. The alternative hypothesis is that x is not from a population with a normal distribution. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, or 0 otherwise.

The findings of the Anderson–Darling test showed the same results as Kolmogorov–Smirnov. To be more clear and concise, it confirmed that throughout the sample, all factors reject null hypothesis. Therefore, we may conclude that there is an absence of normality and our sample has non-normal distribution.

Construction and verification of models of classification. In our scientific paper, we have been using both the special application, namely Matlab Toolbox, and writing the code by ourself. Let us consider the Matlab Toolbox in more detail.

Construction of models, including missing data. After selecting and preparing a database of factors and responses, the first model was created using the “Classification Learner”. The latter allows solving classification problems and building models in an interactive mode.

Using this approach, the training sample size is 2395 points.

Method 1. Classification and regression trees. The methods of classification and regression trees showed the following results:

The accuracy of various types of classification and regression trees is shown in the table 4.

Based on the accuracy analysis of the methods, we can conclude that Bagged Trees showed the highest accuracy of 51.5%. Therefore, we take this type of classification and regression trees as a basis in subsequent analyzes and model constructions.

After matrix analyzing, we can conclude that this model is prone to overvaluation of the rating, if to be concrete, it overestimates the set of points and assigns them a ration equal to 100 or AAA. In addition, we see that for the critical cases, namely the countries of bankruptcy or close to them, the model does not determine it (0, 10, 15, 20).

Method 2. Discriminant Analysis. The methods of discriminant analysis showed the following results:

Proceeding from the table 5, we will draw the following conclusions:

1. Linear discriminant analysis showed a classification accuracy of 23.6%, which indicates about significant classification problems.

2. Unfortunately, quadratic discriminant analysis failed due to degeneracy of the covariance matrix which is caused by the strong collinearity between factors.

3. Subspace discriminant showed the best accuracy among other types of discriminant analysis at 28.5%, but this accuracy is too small for classification.

After matrix analyzing, we can conclude that this model is prone to overvaluation of the rating, if to be concrete, it overestimates the set of points and assigns them a ration equal to 100 or AAA. In addition, we see that for the critical cases, namely, the bankrupt countries or close to them, the model does not determine it (0, 10, 15, 20). A very strong deviation from the true value is also noticeable, which indicates a poor quality of classification. The general conclusion: discriminant analysis gives an unsatisfactory result.

Method 3. Support Vector Machine (SVM). The methods of SVM showed the following results:

Proceeding from the table 6, we will draw the following conclusions:

1. Based on the accuracy analysis of the methods, we can conclude that Cubic SVM showed the highest accuracy of 28.5%. Therefore, we take this type of SVM as a basis in subsequent analyzes and model constructions.

2. After matrix analyzing, we can conclude that this model is prone to undervaluation of the rating, if to be concrete, it underestimates the set of points and assigns them a ration equal to 0 (D).

Table 4
The accuracy of various types of classification and regression trees

| Type of Classification and Regression Tree | Accuracy (%) |
|--|--------------|
| Simple Tree | 24.6 |
| Medium Tree | 32.1 |
| Complex Tree | 38.9 |
| Boosted Trees | 33.4 |
| Bagged Trees | 51.5 |
| RUSBoosted Trees | 26.0 |

Table 5
The accuracy of various types of discriminant analysis

| Type of Discriminant Analysis | Accuracy (%) |
|-------------------------------|--------------|
| Linear Discriminant | 24.8 |
| Quadratic Discriminant | Failed |
| Subspace Discriminant | 28.5 |

Table 6
The accuracy of various types of SVM

| Type of SVM | Accuracy (%) |
|---------------------|--------------|
| Linear SVM | 20.7 |
| Quadratic SVM | 26.6 |
| Cubic SVM | 28.5 |
| Fine Gaussian SVM | 28.1 |
| Medium Gaussian SVM | 24.4 |
| Coarse Gaussian SVM | 16.8 |

Table 7
The accuracy of various types of k-NN

| Type of k-NN | Accuracy (%) |
|---------------|--------------|
| Fine k-NN | 41.7 |
| Medium k-NN | 33.9 |
| Coarse k-NN | 25.6 |
| Cosine k-NN | 33.4 |
| Cubic k-NN | 33.6 |
| Weighted k-NN | 40.9 |
| Subspace k-NN | 24.8 |

3. A very strong deviation from the mean value is also noticeable, which indicates a poor quality of classification.

The general conclusion: SVM gives an unsatisfactory result.

Method 4. k-nearest neighbors (k-NN) The methods of *k-NN* showed the following results.

Proceeding from the table 7, we will draw the following conclusions:

1. Based on the accuracy analysis of the methods, we can conclude that Fine *k-NN* showed the highest accuracy of 41.7%. Therefore, we take this type of *k-NN* as a basis in subsequent analyzes and model constructions.

2. After matrix analyzing, we can conclude that this model is prone to overvaluation of the rating, if to be concrete, it overestimates the set of points and assigns them a ration equal to 100 or AAA. In addition, we see that for the critical cases, namely the countries of bankruptcy or close to them, the model does not determine it (0, 10, 15, 20).

3. A very strong deviation from the true value is also noticeable, which indicates a poor quality of classification.

The general conclusion: the model sufficiently deviates from the true values, but the accuracy estimates are at least better than in the Discriminant Analysis and the Support Vector Mashine Method. As a result, it can be concluded that the majority of builded models with presence of missing data (NAN) in the original sample show insufficient accuracy; hence, this methodology cannot be applied.

However, it can be clearly seen that the method of classification and regression trees, as well as the method of the nearest neighbor, showed quite good results. Consequently, using the following approach, we will choose only the following methods, specifically, ones with the highest accuracy:

1. Classification and Regression Tree: Bagged Trees.

2. Discriminant Analysis: Subspace Discriminant.

3. Support Vector Machine: Cubic SVM.

4. *k*-nearest neighbors: Fine *k-NN*.

Building Models with Exclusion of the Missing Data

When using this methodology, we exclude all points in which missing data exists at least in one factor. As a result, our training sample shranked from 2395 to 1042 points. On

the one hand, we have got rid of missing data, which positively influences the construction of the models, but on the other, we lost a certain amount of information.

Method 1. Classification and regression trees: Bagged Trees. The results of building the model were shown as a Confusion Matrix Graph. If to compare with the previous approach, we can see a considerable improvement, particularly, the accuracy of the classification has increased from 51.5% to 70.5%. Also when analyzing the matrix, we see a significant enhancement in the accuracy of the classification. The points deviate much less from the true value. However, the model has a significant drawback. It was impossible to classify properly critical countries with low ratings (bankrupt).

Method 2. Discriminant Analysis: Subspace Discriminant. The results of building the model were shown as a Confusion Matrix Graph. If to compare with the previous approach, we see that the accuracy of the model has improved, but not considerably. The accuracy of classification has increased from 28.5% to 31.5%. When analyzing the matrix, we did not notice significant improvements in comparison with previously built models. Also as with classification and regression trees, the model has a significant drawback. It was impossible to classify properly critical countries with low ratings (bankrupt).

Method 3. Support Vector Mashine: Cubic SVM. The results of building the model were shown as a Confusion Matrix Graph. If we compare it with the previous approach, we can see a considerable improvement, particularly, the accuracy of the classification has increased from 28.5% to 68.2%. Also, when analyzing the matrix, we see a significant enhancement in the accuracy of the classification. The points deviate much less from the true value. However, the model has a significant drawback. It was impossible to classify properly critical countries with low ratings (bankrupt).

Method 4. k-nearest neighbors: Fine k-NN. The results of building the model were shown as a Confusion Matrix Graph. If to compare with the previous approach, we can see a considerable improvement, particularly, the accuracy of the classification has increased from 41.7% до 70.2%. Also, when analyzing the matrix, we see a significant enhancement in the accuracy of the classification.

Table 8
The accuracy of the models with and without missing data

| Type of method | Accuracy of the models including missing data (%) | Accuracy of the models with exclusion of the missing data (%) |
|-----------------------|---|---|
| Bagged Trees | 51.5 | 70.5 |
| Subspace Discriminant | 28.5 | 31.5 |
| Cubic SVM | 28.5 | 68.2 |
| Fine KNN | 41.7 | 70.2 |

The points deviate much less from the true value. However, the model has a significant drawback. It was impossible to classify properly critical countries with low ratings (bankrupt).

After conducting this approach, it can be concluded that all models have enhanced their results. The greatest increase in accuracy was obtained by the Bagged Trees, Cubic SVM and Fine KNN. Meanwhile Subspace Discriminant has not shown significant improvements.

Analysis of uniformity of distribution of sampling points by classes. It can be concluded that almost all models have enhanced their accuracy, but unfortunately, all of them unable to classify properly critical countries, ones with low ratings (bankrupt), which are vital for us.

We are convinced that this was due to the uneven presence of countries with various ratings within the training sample, specifically: the moments when countries are assigned low ratings are relatively rare, meanwhile for countries with the highest rating, almost all points are generally known for all factors. Consequently, points with high ratings dominate in our training sample.

Below is the histogram of the distribution of our sample by classes.

The histogram shows that in classes that correspond to countries with a low rating (from 0 to 25), there is a very small number of points, whereas in the class that corresponds to the highest rating (100), there is a prevailing number of points. The highlighted fact strongly distorts the construction of the model.

The method of grouping and converting data in manual mode. Applying this method, we eliminated 30 points from the group with the highest rating. Moreover, we also combined groups with a low rating in the range (from 0 to 25) to one group, and set its value to 10.

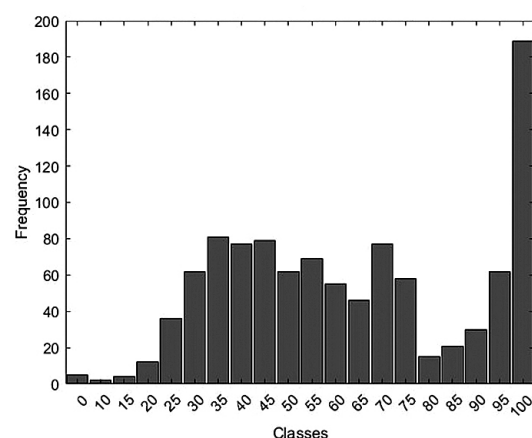


Figure 4. The histogram of the distribution of our sample by classes.

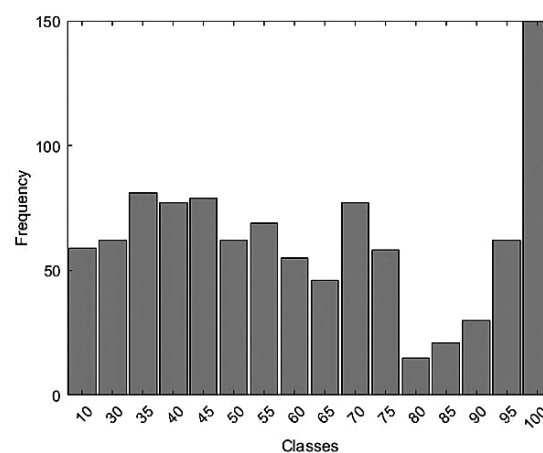


Figure 5. Classes distribution after aggregation.

The histogram clearly shows that the frequency of low-rated countries has increased significantly. Also, the quantity of points in the class with the highest rating was reduced up to 150.

Method 1. Classification and Regression Trees: Bagged Trees. The use of grouping and data conversion approach did not effect significantly on the accuracy improvement in the model. The accuracy of the model increased from 70.5% to 72%. However, the graph shows that the amalga-

Table 9

The accuracy of the models without missing data and with grouping and data conversion approach

| Type of method | Accuracy of the models with exclusion of the missing data (%) | Accuracy of the models with grouping and data conversion approach (%) |
|--|---|---|
| Bagged Trees | 70.5 | 72.0 |
| Subspace Discriminant/Quadratic Discriminant | 31.5 | 51.2 |
| Cubic SVM | 68.2 | 66.1 |
| Fine KNN | 70.2 | 72.4 |

mation of groups with a low rating (from 0 to 25), and a reduction in sampling points for countries with high ratings, had a positive effect on the model, specifically decreasing deviations from the true value. The model commenced assigning values to the class of countries with poor rating. In addition, the model has become less inclined to overestimate the values of the groups of countries with the highest rating.

Method 2. Discriminant Analysis: Quadratic Discriminant. The results of building the model are shown below as a Confusion Matrix Graph. Using this approach, the method of discriminant analysis has been changed from Subspace Discriminant on Quadratic Discriminant. This has led to significant improvement in the accuracy of the model, namely from 31.5% to 51.2%. However, the graph shows that the amalgamation of groups with a low rating (from 0 to 25), and a reduction in sampling points for countries with high ratings, had a positive effect on the model, specifically decreasing deviations from the true value. The model commenced assigning values to the class of countries with poor rating. In addition, the model has become less inclined to overestimate the values of the groups of countries with the highest rating.

Method 3. Support Vector Mashine: Cubic SVM. The results of building the model were shown below as a Confusion Matrix Graph. The use of grouping and data conversion approach did not effect significantly on the accuracy improvement in the model, but on the contrary to its deterioration. The accuracy of the model decreased from 68.2% to 66.1%. However, the graph shows that the amalgamation of groups with a low rating (from 0 to 25), and a reduction in sampling points for countries with high ratings, had a positive effect on the model, specifically decreasing deviations

from the true value. The model commenced assigning values to the class of countries with poor rating. In addition, the model has become less inclined to overestimate the values of the groups of countries with the highest rating.

Method 4. k-nearest neighbors: Fine k-NN. The results of building the model are shown below as a Confusion Matrix Graph. The use of grouping and data conversion approach did not effect significantly on the accuracy improvement in the model. The accuracy of the model has grown from 70.2% to 72.4%. However, the graph shows that the amalgamation of groups with a low rating (from 0 to 25), and a reduction in sampling points for countries with high ratings, had a positive effect on the model, specifically decreasing deviations from the true value. The model commenced assigning values to the class of countries with poor rating. In addition, the model has become less inclined to overestimate the values of the groups of countries with the highest rating.

Ultimetal, it should be concluded, that the use of grouping and data conversion approach did not effect significantly on the accuracy improvement throughout all models (Table 9). Approximety, in average, the accuracy has been increased by 3%. However, the amalgamation of groups with a low rating (from 0 to 25), and a reduction in sampling points for countries with high ratings, had a positive effect on the models, specifically decreasing deviations from the true value. Models commenced assigning values to the class of countries with poor rating. In addition, the models has become less inclined to overestimate the values of the groups of countries with the highest rating.

Logarithmic scaling of factors. Having carried out a logarithmic scaling of factors, it can be concluded that this method did not enhance the accuracy of the model.

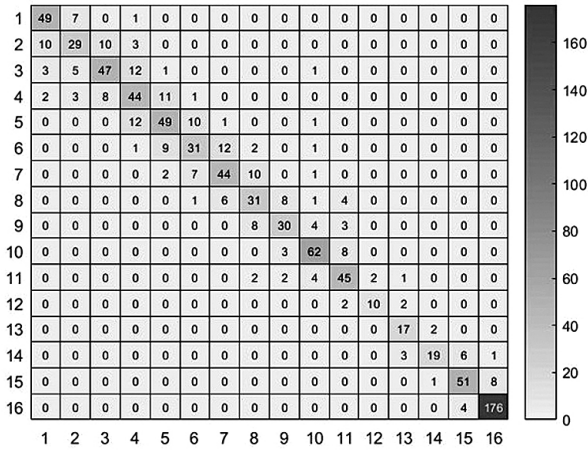


Figure 6. Confusion matrix – Ensemble tree model.

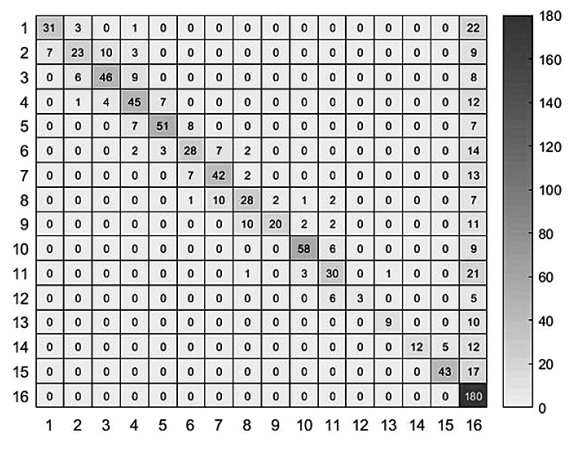


Figure 7. Confusion matrix for SVM classify model.

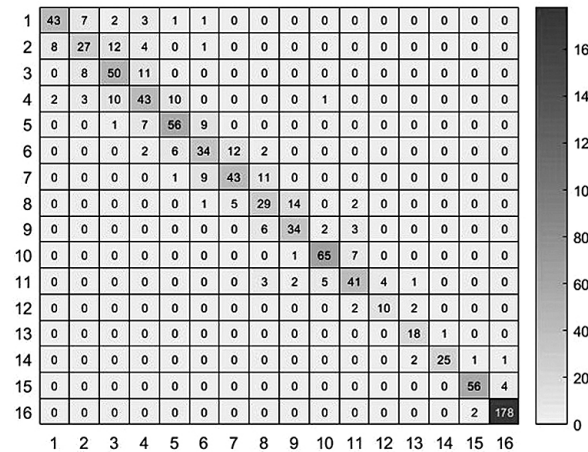


Figure 8. Confusion matrix for Fine k-NN model.

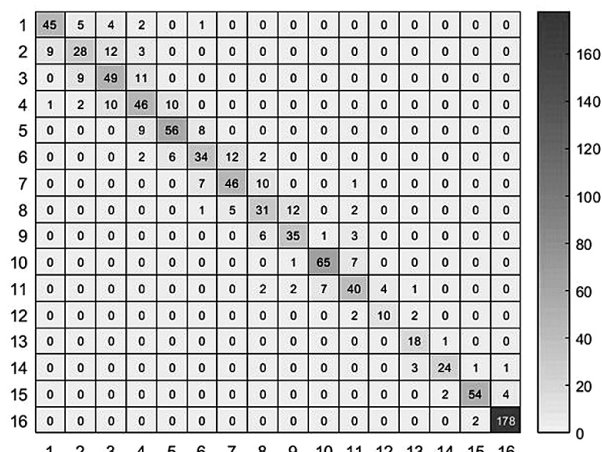


Figure 9. Confusion Matrix for enhanced k-NN model.

Dynamics of factors with the account of the previous year. So far, we have used cross-sectional models with data. In other words, the factors' data were fixed for the year of rating. Thus, we built static models. The following step in our studies was to analyze the impact of changes in factors over time on the country's rating.

For this purpose, we amalgamate the data of the current year with the data of the previous one, for the selected factors. As a result, each factor will represent a model with two vectors, the values of the current and previous year.

Moreover, we refused to use Matlab Toolbox due to some limitations and lack of functions. In our vision, Matlab Toolbox is limited in terms of determining the quality criteria of the model. In other words, Matlab Toolbox allows you to determine only the criterion of classification accuracy.

However, in the future we are going to conduct a thin comparison of the selected models.

For this purpose, we will need those criteria that will evaluate classification errors, for example Mean Absolute Error (MAE) and Mean Squared Error (MSE).

As a result, a special program was created, the purpose of which was to carry out a thin comparison of models using various criteria for both the accuracy of the classification and additional criteria MAE (Mean Absolute Error) and MSE (Mean Squared Error).

Method of Classification with the Account of the Previous Year

Method 1. Classification and Regression Trees: Bagged Trees. The results of building the model were shown in figure 6 as a Confusion Matrix Graph.

This approach, namely, taking into account the previous year, significantly improved the accuracy of the model. The accuracy of classification has increased from 72% to 75.9%. Also

Table 10
Comparative data for results of the models used

| Type of method | Accuracy of the models with grouping and data conversion approach (%) | Accuracy of the models taking into account the previous year (%) | Mean Absolute Error |
|------------------------|---|--|---------------------|
| Bagged Trees | 72.0 | 75.9 | 1.8666 |
| Quadratic Discriminant | 51.2 | NaN | NaN |
| Cubic SVM | 66.1 | 67.1 | 8.6970 |
| Fine KNN | 72.4 | 77.7 | 1.7839 |

when analyzing the matrix, we see a significant enhancement in the deviations. The points deviate much less from the true value. The absolute error was 1.8666, which does not exceed the conventional boundary between classes, equal to 5.

Method 3. Support Vector Machine: Cubic SVM. The results of building the model are shown in figure 7 as a Confusion Matrix Graph.

This approach failed to improve significantly the accuracy of the classification, which slightly increased from 66.1% to 67.1%. Also, when analyzing the matrix, we did not notice any obvious enhancement with respect to deviations. The absolute error was 8.6970, which indicates a significant deviation from the true value for an erroneous classification.

Method 4. k-nearest neighbors: Fine k-NN. The results of building the model were shown in figure 8 as a Confusion Matrix Graph.

This approach, namely, taking into account the previous year, significantly improved the accuracy of the model. The accuracy of classification has increased from 72.4% to 77.77%. Also when analyzing the matrix, we see a significant enhancement in the deviations. The points deviate much less from the true value. The absolute error was 1.7839, which does not exceed the conventional boundary between classes, equal to 5. Using the methodology with the account of the previous year, we may conclude, that generally, the models has been enhanced significantly.

Particularly, the following models showed perfect response:

Bagged Trees

The accuracy of the classification has increased significantly, namely by 6%. And moreover, using our own algorithm, we are aware of the value of mean absolute error, which equals 1,86. Therefore, we may con-

Table 11
Results for enhanced k-NN model with two distance metrics

| Metrics type | Accuracy | Mean Absolute Error |
|--------------|----------|---------------------|
| Euclidean | 0.7777 | 1.7839 |
| Manhattan | 0.7849 | 1.6339 |

clude that the deviation is too low, because it does not exceed the conventional boundary between classes, equal to 5.

Fine k-NN

The next and one model, which also showed perfect results is Fine k-NN. The accuracy of the classification has increased significantly, approximately by 6%. Moreover, using our own algorithm, we become aware of the value of mean absolute error, which equals 1.78. Therefore, we may conclude that the deviation is too low, because it does not exceed the conventional boundary between classes, equal to 5. Unfortunately, Quadratic Discriminant is failed this test.

Cubic SVM

Cubic SVM has also enhanced its results, but insignificantly. The accuracy of the classification has increased only by 1%, which can claim that this method does not fit to this approach. Furthermore, the value of mean absolute error is too high, particularly, 8.69, which exceeds the conventional boundary between classes, equal to 5.

Enhancing of k-NN methodology

In order to improve the accuracy of the classification, besides the Euclidean metrics, we verified, and other ones. As a result, the metric

CityBlock or Manhattan showed the best result, enhancing the accuracy of the model by 1%, which is a significant improvement. The results of building the model were shown in figure 9 as a Confusion Matrix Graph.

Conclusions

Carrying out the relatively vast number of tests and constructioning a significant quantity of models, we conclude that the most appropriate are the following two: Classification and Regression Trees (Bagged Trees) and k -nearest neighbors: Fine k -NN.

The quantitative characteristics of the building models are given below:

The best accuracy which has been shown by the k -nearest neighbors, (Fine k -NN), using Manhattan metrics is 0.7849. Therefore, the points almost do not deviate from the true value.

The best mean Absolute Error has been shown by the k -nearest neighbors. Its value is 1.6339, which does not exceed the conventional boundary between classes of 5.

The one and the last model, with relatively the same characteristics, is Classification and Regression Trees (Bagged Trees).

The accuracy shown by the Classification and Regression Trees (Bagged Trees) is 75.9%. Therefore, the points almost do not deviate from the true value.

The best mean Absolute Error has been shown by the Classification and Regression Trees (Bagged Trees) is 1.8666, which does not exceed the conventional boundary between classes of 5.

In addition, the models are not inclined to overestimate the values of the groups of countries with the highest rating.

Moreover, our models have a significant advantage. They are able to classify criti-

cal countries properly, ones with low ratings (bankrupt).

Furthermore, it is vital to enlist the characteristics of the data sample used to construct models.

There is a variety of advantages over the data sample, which are the following:

First, and foremost, the majority of factors have relatively high correlation coefficients with the response value. Second, across almost all factors the intense multicollinearity was not observed.

On the other hand, there are a number of significant drawbacks, which deteriorates the whole sample:

1. Almost all factors has multimodal distribution (has two modes). Consequently, we may conclude that our sample is far from the normal distribution.

2. The vital thing is that, the sample has right-sided asymmetry, as evidenced by a positive Skewness. In other words, the mean value and the median one is greater than the mode value. This suggests that our data is shifted to the left, relative to the normal distribution. Consequently, there is an abnormality in the distribution of the values of factors.

3. Equally important is the fact that there is an excess in the data sample. For the major part of data sample the peak values are higher than the normal distribution. This is confirmed by a value of kurtosis more than 3.

4. Based on the interpreted results of the descriptive statistics, we should sum up that our data is far from the normal distribution. In other words, the Gauss-Markov conditions are not metted.

To conclude, we may say, that such classification techniques and algorithms are recommended for analysis, in the absence of the insider information and small variety of factors (sample).

Sources

http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1073&context=econ_fac

https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Discriminant_Analysis.pdf

<http://download.portalgaruda.org/article.php?article=81950&val=160>

<http://www.ics.uci.edu/~welling/teaching/273ASpring10/AdaBoost4Stocks.pdf>

http://pdf.usaid.gov/pdf_docs/PNACH725.pdf

<http://www.tsijournals.com/articles/research-of-risk-assessment-and-prediction-based-on-support-vector-machine.pdf>

Страновой риск в международных инвестициях: структура и методы расчета

Алексей Ивкин¹

В данной статье исследуются вопросы оценки странового риска в разрезе экономической безопасности и устойчивости экономики. Основными объектами исследования являются страновой риск и его структурные компоненты. Основная цель научной статьи – проанализировать методы оценки странового риска с разных точек зрения и предложить модель для измерения странового риска, которая позволила бы адекватно оценить страновой риск, экономическую безопасность, уровень и динамику экономической устойчивости, включая структурные компоненты и их отношения. В статье выделено несколько основных задач:

- понять важность оценки странового риска в контексте развивающихся мировых рынков, учитывая причины и элементы странового риска на основе других научных исследований;
- изучить и выяснить преимущества и недостатки методов оценки страновых рисков, а также определить способы управления рисками;
- применить количественные и качественные методы анализа, сформулировать, создать и представить модель оценки странового риска в контексте экономической безопасности и устойчивости, которая будет определять факторы, влияющие на страновой риск, и определять их прямые и косвенные отношения между собой;
- проверить практическую пригодность модели оценки странового риска путем проведения эмпирического анализа во всем мире, определения направлений смягчения последствий странового риска.

Ключевые слова: анализ странового риска; кредитные рейтинги; долг; структурные качественные методы; дискриминантный анализ; регрессия; классификация; метод ближайших соседей; метрики расстояния; классификационные и регрессионные деревья; опорные вектора; дискриминантный анализ; недостающие данные; тест Колмогорова–Смирнова; тест Андерсона–Дарлинга; непараметрические методы

JEL classification: C53, O16, O47

¹ Студент Финансового университета, магистерская программа «Международные финансы», Москва, Россия; vok-zal@mail.ru

Научный руководитель – А.И. Ильинский, доктор технических наук, профессор кафедры «Макроэкономическое регулирование», Финансовый университет, Москва, Россия.