“The relationship between sovereign credit rating and trends of macroeconomic indicators”

AUTHORS
Jana Kotěšovcová https://orcid.org/0000-0002-3003-7352
Jiří Mihola
Petr Budinský

ARTICLE INFO

DOI
http://dx.doi.org/10.21511/imfi.16(3).2019.26

RELEASED ON
Friday, 04 October 2019

RECEIVED ON
Tuesday, 28 May 2019

ACCEPTED ON
Wednesday, 25 September 2019

LICENSE
This work is licensed under a Creative Commons Attribution 4.0 International License

JOURNAL
"Investment Management and Financial Innovations"

ISSN PRINT
1810-4967

ISSN ONLINE
1812-9358

PUBLISHER
LLC “Consulting Publishing Company “Business Perspectives”

FOUNDER
LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES
40

NUMBER OF FIGURES
3

NUMBER OF TABLES
5

© The author(s) 2019. This publication is an open access article.
Abstract
The sovereign credit rating provides information about the creditworthiness of a country and thereby serves as a tool for investors in order to make right decisions concerning financial assets worth investments. Thus, determination of a sovereign credit rating is a highly complex and challenging activity. Specialized agencies are involved in rating assessment. So, it's essential to analyze the efficiency of their work and seek out easily accessible tools for generating assessments of such ratings. The objective of this article is to find out whether sovereign credit rating can be reliably estimated using trends of selected macroeconomic indicators, despite the fact that sovereign credit rating is most likely influenced by non-economic factors. This can be used for strategic considerations at national and multinational levels. The relationships between sovereign credit rating and the trends of macroeconomic indicators were examined using statistical methods, linear multiple regression analysis, cumulative correlation coefficient, and multicollinearity test. The data source used is comprised of selected World Bank indicators meeting the conditions of completeness and representativeness. The data set has shown a cumulative correlation coefficient value greater than 95%, however at 100% multicollinearity. This is followed by the gradual elimination of indicators, but even this did not allow achieving acceptable values. So, the conclusion is that rating levels are not explainable solely by the trends of economic indicators, but other influences, e.g. political. However, the fact that the statistical model yielded acceptable results for five and fewer indicators allowed a regression equation to be found that gives good estimates of a country's rating. This allows, for example, predicting of ratings relatively easy by forecasting the development of selected macroeconomic indicators.

Keywords
rating level, credit rating, rating methodology, macroeconomic indicators, sovereign rating

JEL Classification
E27, G17, G24

INTRODUCTION
The rating of a country is a tool that indicates the possible future default of debtors in a timely fashion to creditor states, such that they might have the time for a sufficient response. It is a qualified evaluation of the reliability of debtors. The rating in most cases replaces or confirms investors’ own analyses of solvency and substantially influences their view of a given country, and thereby their inclination to invest or refrain from investing. The rating of a country is an important tool in international relations, which would ideally be determined by each country or major business partner themselves. The problem is that it is a highly complicated and challenging activity worth unifying in some manner. For this reason, specialized international agencies have gradually gained importance.

This article offers a non-traditional perspective by exploring the substance of the assignment of rating marks using original statistics tools.

© Jana Kotěšovcová, Jiří Mihola, Petr Budinský, 2019

Jana Kotěšovcová, Ing., Ph.D., Department of Finance, Faculty of Economic Studies, University of Finance and Administration, Czech Republic.

Jiří Mihola, Ing. Bc., CSc., Department of Economics, Faculty of Economic Studies, University of Finance and Administration, Czech Republic.

Petr Budinský, Doc. RNDr., CSc., Department of Finance, Faculty of Economic Studies, University of Finance and Administration, Czech Republic.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

http://dx.doi.org/10.21511/imfi.16(3).2019.26
and derives a relatively simple tool for reliable rating estimates using a small number of macroeconomic indicators, regression and correlation analysis, and multicollinearity. The calculation of multicollinearity is original.

The first part of the article analyzes the current issues and presents key information sources. The methodology used is then described, particularly the original approach in the multicollinearity measurement. This is followed by the compilation of a multiple linear regression and correlation model as an analysis tool, whose results are described in detail and summarized in the requisite conclusions.

1. LITERATURE REVIEW

This article, which focuses on the dependency of ratings on macroeconomic indicators, follows on and builds on the research work of numerous global authors. The definition of a rating and its classification can be found in Kotěšovcová (2013). The control and responsibility of credit rating agencies in the United States are addressed in the article “Control and Responsibility of Credit Rating Agencies in the United States” (Pinto, 2006). Credit rating agencies and their regions of operation are assessed by Yalta, T. and Yalta, Y. in “Are credit rating agencies regionally biased?” (2018).

The analysis of the mutual relationship between credit rating and selected macroeconomic indicators builds on the research work of global authors and seeks to expand upon that work. This pertains to the field of credit rating and its influence on financial markets, e.g. “The relationship between credit default swap spreads, bond yields, and credit rating announcements” conducted by Hull, Predescu, and Hull et al. (2004) and the article on the influence of market-based indicators on credit ratings “Do Market-Based Indicators Anticipate Rating Agencies?” (Di Cesare, 2006). This work also draws partially on work from the field of bank ratings (King, Ongena, & Tarashev, 2011). This article also rests on the field of credit risk assessment (Peškauskaitė & Daiva, 2017) and the influence of capital flows on the credit ratings of developing countries (Setty & Dodd, 2003).

This article also builds on the study “Relationship between Sovereign Ratings and CDS Prices” (Budinsky, 2015), which analyzes the accuracy of determining the credit rating of a country on the basis of an implicit rating based on credit default swaps (CDS). Hull, Predescu, and White (2004), Ismailescu and Kazemi (2010), and Mili (2019) further explore the influence of sovereign credit ratings on credit default swaps in the article “The impact of tradeoff between risk and return on mean reversion in sovereign CDS markets”, as do Rodriguez, Dandapani, and Lawrence in their article “Measuring Sovereign Risk: Are CDS Spreads Better than Sovereign Credit Ratings?” (2019). The transfer of contagion from the rating of a country and the rating of a bank was examined by Cantero Sáiz, Azofra, and Olmo in “The single supervision mechanism and contagion between the bank and sovereign risk” (2019). Minenna and Aversa captured the extent to which the European stability mechanism utilizes shared risk from public debts and unified commercial cycles in the Eurozone in order to improve the economic situation, and thereby the credit rating of the state in “A Revised European Stability Mechanism to Realize Risk Sharing on Public Debts at Market Conditions and Realign Economic Cycles in the Euro Area” (2009). Riaz, Shehzad, and Umar explore the influence of credit rating changes on stock returns in “Pro-cyclical effect of sovereign rating changes on stock returns: a fact or factoid” (2019). Yields from government bonds in Europe and their influence on the credit rating of the state was addressed by Afonso and Jalles in “Quantitative easing and sovereign yield spreads Euro-area time-varying evidence” (2019). Bettendorf (2019) analyzed the impacts of transferring the risk of credit default in the Eurozone and their influence on the credit rating of the state. An unconventional currency policy and its influence on stock markets and credit rating in the Eurozone were addressed by Chebbi (2019). The market for corporate bonds in Europe and their influence on credit ratings was analyzed by Horny, Manganelli, and Mojon in

This research study will be focused on exploring the dependency of a sovereign credit rating on macroeconomic indicators. The area of relationships between macroeconomic magnitudes and the rating of a country has been the focus in particular of Canuto, Pereira, and Porto (2012). An analysis of factors of sovereign credit rating is assessed by Alfonso (2003), Cantor and Packer (1996), as well as Mellios and Paget-Blanc (2006), and Kou and Varotto (2008). The influence of fiscal and monetary policy and macroeconomic stability on sovereign credit rating is addressed in “Fiscal and Monetary Policy Coordination, Macroeconomic Stability, and Sovereign Risk Premia” (Bonam & Lukkezen, 2019). Tobera examines how the costs of public debt influence ratings in the countries of Central and Eastern Europe in “Credit Rating and the Cost of Public Debt Service in Central and Eastern European Countries from 2005 to 2017” (2019). The influence of direct foreign investment on credit rating was examined by Cai, Kim and Wu in “Foreign direct investments from emerging markets: The push-pull effects of sovereign credit ratings” (2019).

2. OBJECTIVE, EXISTING DATA AND USED METHODOLOGIES AND ALGORITHMS

The objective of this article is to explore whether sovereign credit rating can be reliably estimated using trends of selected macroeconomic indicators, despite the fact that sovereign credit rating is most likely influenced by factors other than economic factors. To fulfill this objective, empirical data is required about trends of both credit ratings and macroeconomic indicators. The relationship between the two will be examined using the standard tools of multiple regression and correlation analysis, extensively supplemented with measurement of multicollinearity. Original tools are used for measuring multicollinearity, and will, therefore, be covered in greater detail in the methodology description. Despite its relatively small computational demands and yet very informative capability, this methodology has not been widely used thus far, as it still has not come into the broader awareness of technical community, despite having been published in the corresponding technical literature. This is also the reason why these methods have been given increased attention in this article.

2.1. Existing data

For the purposes of this study, it is determined based on a review of various information sources that the most appropriate source of a timeline of macroeconomic indicators is the World Development Indicators database of the World Bank, which became the primary source of this information. In addition to macroeconomic data, there is also the need to obtain data about credit ratings from rating agencies. This is based on data from the three most notable credit rating agencies: Standard & Poor’s, Moody’s, and Fitch, covering a sufficiently sizable time period. Data about the sovereign credit ratings of the three international credit rating agencies are drawn from the data resource countryeconomy.com.

The criterion for selecting the countries is taken from a geographical perspective, namely countries in the territory of Europe (in other words, mostly European Union member states), including certain non-member states of the EU such as Switzerland and Ukraine, as well as Russia and Turkey, who are only partially located in Europe. The list does not include the Baltic countries and Malta, despite being member states of the EU, due to a lack of availability of data. In total, therefore, 28 member states were selected that are located in the geographical territory of Europe, albeit in some cases only partially.

The macroeconomic indicators used for calculating the rating were drawn from the World Development Indicators database, which includes a total of 1,519 indicators from the World Bank. The indicators were selected such that they would cover all important macroeconomic areas: indebtedness, unemployment, inflation, GDP growth, exports of goods and services, gross domestic product, household and government final consumption expenditure, gross savings, interest payments, real interest rate, etc. Originally, 23 indica-
tors were selected. These indicators were obtained for the period from the beginning of the new millennium, i.e. from 2000 to 2016 for 28 European countries.

After thorough consideration, particularly regarding the completeness of data, the following 17 indicators were selected:

- indicator No. 1: central government debt (% GDP);
- indicator No. 2: unemployment rate (% of the total labor force);
- indicator No. 3: inflation, consumer prices (annual, %);
- indicator No. 4: GDP growth rate in %;
- indicator No. 5 GDP per capita (s.c. 2011, PPS, USD);
- indicator No. 6: government expenditure (% GDP);
- indicator No. 7: exports of goods and services (% GDP);
- indicator No. 8: net exports of goods and services (% GDP);
- indicator No. 9: household final consumption expenditure (% GDP);
- indicator No. 10: government final consumption expenditure (% GDP);
- indicator No. 11: gross domestic product (% GDP);
- indicator No. 12: gross savings (% GDP);
- indicator No. 13: import of goods and services (% GDP);
- indicator No. 14: interest payments on external debt (% costs);
- indicator No. 15: government income (+)/government expense (−) as % of GDP;
- indicator No. 16: real interest rate (%);
- indicator No. 17: net trade of goods and services (% GDP).

Indicators were not selected according to their individual relevance to the credit rating levels. Nonetheless, the selection of indicators is not based on their material content but could be compared to that of, e.g., Cantor and Packer (1996). The current article used a set of 17 indicators, compared to the eight indicators used in this publication. The present study does not work with soft indicators (economic development and default history) that were used in the aforementioned study (Cantor & Packer, 1996). From the above, it may be seen that the set of 17 indicators used can be considered representative from the perspective of their material content. The representativeness of the set thus obtained can be demonstrated, e.g., by comparing it with a set of key macroeconomic indicators used by the Czech National Bank (Česká národní bank, 2018) in its regular Inflation Reports/III.

Apart from the economic indicators, it is necessary to obtain data about the credit rating marks of the three credit rating agencies included in the study: Standard & Poor’s, Moody’s, and Fitch, covering a sufficiently extensive timeline. Data about rating marks assigned by all three credit rating agencies were obtained from the data resource countryeconomy.com

2.2. Methodologies and algorithms used

There were used 17 macroeconomic indicators from 28 European countries to apply the multiple linear regression and correlation analysis. The correlation analysis included calculation of the cumulative correlation coefficient and multicollinearity. The high value of the cumulative correlation coefficient can be devalued by the high multicollinearity, particularly when close to 1. For this reason, a correlation matrix is used to gradually select indicators as candidates for elimination such that the value of the cumulative correlation coefficient $R^2_\chi$ did not drop with the required decrease in multicollinearity $M$, at least below 50%, or is not lower or higher than the highest bivariate correlation coefficient $r_{xy}^2$. In this manner, suitable
macroeconomic indicators were gradually sought for elimination from the original or previous set.

An important separate task is to analyze the differences in the ratings of the three international rating agencies Standard & Poor’s, Moody’s, and Fitch, which was adapted from Kotěšovcová (2018). This conclusion is to be used in the next process for determining the mutual dependencies of the trends of rating marks on the trends of the selected macroeconomic indicators. Given that the deviations in the ratings between the individual agencies were perceived as non-significant, only average values or medial rating levels for the three agencies will be used in future analyses. The separate problem of aggregating and transforming rating marks to numerical values will be addressed in the upcoming subsections.

2.2.1. Transformation of credit rating marks to numerical values

Given that credit rating marks are traditionally issued by specialized institutions using a combination of non-numeric symbols (moreover varying among the various institutions), this subsection examines the methods of their conversion to purely numerical values suitable for regression and correlation analysis. First, it will be considered whether this transformation should be linear or logarithmic. Given the presumed statistical non-significance of differences between the rating marks of the three major credit rating agencies, this will be followed by an examination of methods for their aggregation using the median, simple arithmetic mean, or merely a selection of the scores of one of the credit rating agencies included in the study.

The simplest approach, i.e. the linear assignment of numerical values to credit rating marks that correspond to the same ranking in all three agencies, is given in the fourth column of Table 1. The lowest ranking is assigned a value of 1. Because 21 rating levels are being examined, the highest ranking is therefore assigned a value of 21. The numerical expression of the rating score will be designated using the symbol $R_t$. In some places objections are raised to this linear assignment that the differences between the lowest rating marks are more significant the lower the score (or rather, the worse the score), or that more significant differences are in the central passage, i.e. for scores in which the letter “B” appears. Non-linear assignment, which emphasizes the differences between the lower rating marks such that the lowest scores reflect the greatest differences, may be easily carried out using logarithmic transformation. The rating marks after logarithmic transformation are listed in the sixth column of Table 1 marked ln($R_t$). It may be inferred from this that certain studies, for example, Svitil “Comparison of banking rating systems” (2019), assign rating marks numeric values in a non-linear fashion. These more or less consider differences between high rankings of category “A” less significant than the low scores of category “B”. Between the rankings of category “C”, on the other hand, there is practically no difference at all. In the following calculations, the variant chosen will be the one that brings the higher value of the cumulative correlation coefficient at the lowest possible multicollinearity.

Table 1. Non-linear assignment of numeric values to credit rating marks

<table>
<thead>
<tr>
<th>No.</th>
<th>Standard and Poor’s</th>
<th>Fitch</th>
<th>Moody’s</th>
<th>Rt</th>
<th>ln(Rt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AAA</td>
<td>AAA</td>
<td>Aaa</td>
<td>21</td>
<td>3.045</td>
</tr>
<tr>
<td>2</td>
<td>AA+</td>
<td>AA+</td>
<td>Aa1</td>
<td>20</td>
<td>2.996</td>
</tr>
<tr>
<td>3</td>
<td>AA</td>
<td>AA</td>
<td>Aa2</td>
<td>19</td>
<td>2.944</td>
</tr>
<tr>
<td>4</td>
<td>AA−</td>
<td>AA−</td>
<td>Aa3</td>
<td>18</td>
<td>2.890</td>
</tr>
<tr>
<td>5</td>
<td>A+</td>
<td>A+</td>
<td>A1</td>
<td>17</td>
<td>2.833</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>A</td>
<td>A2</td>
<td>16</td>
<td>2.773</td>
</tr>
<tr>
<td>7</td>
<td>A−</td>
<td>A−</td>
<td>A3</td>
<td>15</td>
<td>2.708</td>
</tr>
<tr>
<td>8</td>
<td>BBB+</td>
<td>BBB+</td>
<td>Baa1</td>
<td>14</td>
<td>2.639</td>
</tr>
<tr>
<td>9</td>
<td>BBB</td>
<td>BBB</td>
<td>Baa2</td>
<td>13</td>
<td>2.565</td>
</tr>
<tr>
<td>10</td>
<td>BBB−</td>
<td>BBB−</td>
<td>Baa3</td>
<td>12</td>
<td>2.485</td>
</tr>
<tr>
<td>11</td>
<td>BB+</td>
<td>BB+</td>
<td>Ba1</td>
<td>11</td>
<td>2.398</td>
</tr>
<tr>
<td>12</td>
<td>BB</td>
<td>BB</td>
<td>Ba2</td>
<td>10</td>
<td>2.303</td>
</tr>
<tr>
<td>13</td>
<td>BB−</td>
<td>BB−</td>
<td>Ba3</td>
<td>9</td>
<td>2.197</td>
</tr>
<tr>
<td>14</td>
<td>B+</td>
<td>B+</td>
<td>B1</td>
<td>8</td>
<td>2.079</td>
</tr>
<tr>
<td>15</td>
<td>B</td>
<td>B</td>
<td>B2</td>
<td>7</td>
<td>1.946</td>
</tr>
<tr>
<td>16</td>
<td>B−</td>
<td>B−</td>
<td>B3</td>
<td>6</td>
<td>1.792</td>
</tr>
<tr>
<td>17</td>
<td>CCC+</td>
<td>CCC+</td>
<td>Caa1</td>
<td>5</td>
<td>1.609</td>
</tr>
<tr>
<td>18</td>
<td>CCC</td>
<td>CCC</td>
<td>Caa2</td>
<td>4</td>
<td>1.386</td>
</tr>
<tr>
<td>19</td>
<td>CCC−</td>
<td>CCC−</td>
<td>Caa3</td>
<td>3</td>
<td>1.099</td>
</tr>
<tr>
<td>20</td>
<td>CC</td>
<td>CC, C</td>
<td>Ca</td>
<td>2</td>
<td>0.693</td>
</tr>
<tr>
<td>21</td>
<td>D</td>
<td>D, DD, DDD</td>
<td>C</td>
<td>1</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Both of the aforementioned transformations of credit rating marks were used in an alternating fashion to produce the following calculations. It is determined that ceteris paribus, a higher aggre-
gate tightness $R_j^2$ is demonstrated using logarithmic transformation. Meanwhile, the difference compared to linear transformation is significant. Combined functions were also tested, however, these did not demonstrate any greater advantage. The results stated here, therefore, use logarithmic transformation.

2.2.2. The methods of aggregating rating marks

A calculation is performed in the article of the statistical significance of mutual deviations of rating marks from the three major rating agencies, Standard and Poor's, Fitch, and Moody's. It demonstrated that these deviations are not statistically significant. For this reason, either the scores of a single company of choice or an appropriate aggregation may be used. Of the options for aggregation, the median or simple arithmetic mean proved to be most suitable. As a central value, the median has the advantage that it represents a specific existing magnitude; whereas it represents a central rating score of three agencies examined in the study. For this reason, it is also very easily interpreted. While the simple arithmetic means is a fictional value, it would also achieve its purpose for the needs of further analysis.

Credit rating marks were found for 28 European countries from all three rating agencies between the years of 2011 and 2017. These scores were assigned numerical linear and logarithmic expressions in Table 1. In Table 2, the average and median were calculated for rating marks of all three rating agencies. At the same time, the difference is calculated between these calculated averages and medians. To compare the median and simple arithmetic mean, the average difference between the median and mean is also calculated, and the proportion of this average deviation to the average of averages of the median and arithmetic mean.

From Table 2 it is clear that for the linear transformation of the credit rating marks from 2011 the average difference between the median and the averages of all countries studied is merely 0.01, which is only 0.08% of the corresponding mean. Similarly low values were demonstrated for 2011 when using logarithmic transformation both in 2011 and particularly in 2016.

Given the negligible values of the analyzed deviations in Table 2, the simple arithmetic mean will be used in subsequent calculations for aggregating the credit rating marks of all three rating agencies in the study.

Measuring tightness of a dependent variable and multicollinearity

To measure the tightness of the explained variable on all independent variables, the coefficient of determination $R_j^2$ dependency of the $j$-th variable on all other variables was used (see for example Garaj and Šujan (1980)).

$$R_j^2 = 1 - \frac{|R_j|}{|R_j^j|}.$$  (1)

In this relationship, $R_j^j$ is a minor of the correlation matrix $R$. The correlation matrix $R$ is square (18x18) and includes all 17 explaining variables and the explained variable, which is credit rating. $R_j$ is a minor of a correlation matrix, which is square (17x17) and can be obtained by releasing the explained variable $x_j$. Given that the determinant $|R|$ just like the determinant $|R_j|$ the newly implemented characteristic of total tightness $\Phi^2$ is always less or at most equal to 1, $|R| \leq |R_j|$ for $j = 1$. The equivalence occurs at the moment when $|R_j^j| = 1$.

Measuring the multicollinearity of the study operates in essence from the concept articulated in Hebák and Hustopecký (1987), where it is defined as a proportion of the total tightness of the set of dependent variables and the tightness of the entire system. The cumulative tightness is expressed by the relationship $1 - |R|$. Such characteristics have a very favorable quality and can be depicted along with the most important correlation characteristics, as shown in Figure 1. The cumulative tightness of the system can be expressed using the determinant of the correlation matrix in a manner similar to that of Hebák (2004).
\[ \Phi^2 = 1 - |R|. \]  

From the research, it is clear that this expression represents an analogy of the coefficient of determination, which corresponds to a normalized orthogonal basis. If the \( j \)-th variable is a dependent variable, it means that the independent variables are not correlated and the indicator of tightness \( \Phi^2 \) (defined as \( \Phi^2 \) for the basic set, however, after the \( j \)-th variable is eliminated from it) acquires a zero value. In this way, the interpretation is also obtained of \( \Phi^2 \) as the threshold limit of the coordinate coefficients of determination for uncorrelated independent variables, or zero multicollinearity.

Multicollinearity ranks among the possible serious problems for constructing multiple linear regression models. The easy rotating of the hyperplanes influences the efficiency of estimators of the regression parameters as far as the impossibility of their economic interpretation, impedes the separation of individual factors, and increases the sensitivity of the reaction to the number of measurements. For working with a specific model, it often does not suffice to state that the model is encumbered by multicollinearity, and for this reason, it is necessary if one can determine its level or size. From the analysis of various methods of measuring multicollinearity, the requirement can be determined that when measuring multicollinearity it is necessary to achieve the relativity of this term. Using the above-defined term of tightness \( \Phi^2 \), the indicator of multicollinearity \( M_i \) can be defined as the proportion of tightness of the sys-
tem not including and including the dependent variable with index \( j \), i.e. the relationship:

\[
M_j = \frac{\Phi^2_{-j}}{\Phi^2_j}.
\]  

(3)

Thus \( M_j \) may be defined when \( \Phi^2 \neq 0 \). For \( \Phi^2 = 0 \), \( M_j = 0 \) is also applied. The multicollinearity is thus defined in a range from 0 to 1 or in % from 0 to 100%. Čermáková and Mihola (1989) and Mihola and Bílková (2014, pp. 543-1549) address in detail the definition of multicollinearity using the above method. The aforementioned studies indicate that the links between the stated variables allow the construction of Figure 1, into which all four of the aforementioned cumulative quantities can be introduced at once.

Figure 1 has cumulative (overall) tightness values on the x-axis of the dependent and independent variables \( \Phi^2 \), whereas the y-axis shows the cumulative tightness of the arrangement of independent variables \( \Phi^2_{-j} \). There are also two clusters of lines representing isoquants. The cluster passing through the origin of the coordinates \([0; 0]\) represents isoquants of multicollinearity \( M_j \). The highest value 1 acquires multicollinearity at the diagonal representing identity, i.e. for the 45º bisector. Null multicollinearity is on the x-axis. The scale of multicollinearity in the graph is not expressed as it corresponds to the scale of cumulative tightness of the arrangement of independent variables \( \Phi^2_{-j} \) for the value \( \Phi^2 = 1 \).

In Figure 1, the second cluster beginning in point \([1; 1]\) forms the isozon of the cumulative correlation coefficient \( R^2_j \). The lowest value, 0, is occupied by the cumulative correlation coefficient on the 45º diagonal representing identity, i.e. on the axis of the I quadrant. The unit cumulative correlation coefficient \( R^2_j = 1 \) is parallel to axis y, with a value of \( \Phi^2 = 1 \). The scale of the cumulative correlation coefficient \( R^2_j \) is not depicted in the graph because it matches the scale of cumulative tightness \( \Phi^2 \) for null multicollinearity \( M_j = 0 \). The importance of this graph is that it allows displaying four variables simultaneously.

Realistically, points in Figure 1 can appear only under the diagonal where all magnitudes are defined. Each point represents the value of all four tracked variables. It is desirable for the solution to

Figure 1. Complete x graph evaluating multiple regression and correlation analysis

Source: Mihola and Bílková (2014).

http://dx.doi.org/10.21511/imfi.16(3).2019.26
give a high cumulative correlation coefficient at low multicollinearity. These points are shown at bottom right in the figure. While there can be a high cumulative correlation coefficient on the top right, the multicollinearity is also too high here and therefore the solution has no value, as the estimates of the majority of regression parameters (independent variables) have too great standard deviation (are too volatile) and thus cannot be relied on. Although the points on the left have low multicollinearity, they also have a fundamentally very low value of cumulative correlation coefficient.

When “treating” the multicollinearity by releasing or replacing or aggregating the independent variables, one can track the trend of the basis with a gradual movement of points representing new solutions in Figure 1, which is also used in this article.

3. INVESTIGATION USING REGRESSION AND CORRELATION MODEL

3.1. Creation of regression and correlation model

The regression and correlation model is a tool that is created for testing dependencies between rating levels (independent variable) and selected macroeconomic indicators. The first step in creating this model is the selection of indicators, commensurate checks of completeness of data and significance, and in certain cases also transformation of certain data using logarithms. Another step is testing the parameters using the LINEST function (Excel), where regression parameters were determined for the regression function. The correlation basis is then developed, as part of which the correlation coefficient is created, the cumulative correlation coefficient calculated, and the cumulative multicollinearity determined.

3.2. Analysis of links of credit rating marks and macroeconomic indicators

From the original data described in the first section, 17 indicators were chosen for subsequent analysis, which represents the independent variable.

The results of regression analysis for the 17 selected macroeconomic indicators are concentrated in Table 3, which contains estimates of regression parameters $b_1$ through $b_{17}$, and an estimate of the absolute term $b_0$. Regression parameters are sorted in the table by size. In the first line, the value is stated of the cumulative correlation coefficient $R^2$, which generates the appropriate function concurrently with regression coefficients. The cumulative correlation coefficient expresses that the regression scores were explained from 95.1% by the given indicators. However, certain very low values of regression parameters are problematic, such as $b_{11} = 0; b_{14} = 0.003; b_2 = 0.004$. It is likely that these indicators may be eliminated from the calculation, however only after the correlation basis is evaluated. After verifying as part of correlation analysis, the indicator $b_{11}$, gross domestic product (% GDP) is a significant candidate for elimination in the 1st round, i.e. when reducing to 16 indicators. It is shown that indicators $x_8$ through $x_{11}$ are functionally dependent.

A key basis for correlation analysis is a complete correlation matrix containing all bivariate correlation coefficient of all pairs of variables, including the dependent variable, i.e. $r_{ij}$.

The correlation matrix shown in Table 4 enables a significantly correlated pair of indicators and tightness of dependent variable links to be found, i.e. credit rating marks (precisely its logarithm) along with individual explaining variables. The tightness of the regression links

Table 3. Regression coefficients for the 17 indicators

<table>
<thead>
<tr>
<th>Regression parameters</th>
<th>$b_0$</th>
<th>$b_7$</th>
<th>$b_{12}$</th>
<th>$b_5$</th>
<th>$b_4$</th>
<th>$b_{12}$</th>
<th>$b_{10}$</th>
<th>$b_2$</th>
<th>$b_{14}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>12.459</td>
<td>8.567</td>
<td>8.438</td>
<td>0.467</td>
<td>0.039</td>
<td>0.030</td>
<td>0.020</td>
<td>0.004</td>
<td>0.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression parameters</th>
<th>$b_{11}$</th>
<th>$b_6$</th>
<th>$b_{15}$</th>
<th>$b_{16}$</th>
<th>$b_8$</th>
<th>$b_3$</th>
<th>$b_1$</th>
<th>$b_9$</th>
<th>$b_{17}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.000</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.08</td>
<td>-0.676</td>
<td>-17.0 04</td>
</tr>
</tbody>
</table>
between the dependent variable \( y \) and the independent explaining variables does not exceed the absolute value of 0.74, as shown by the last column of the correlation matrix. There are also low values here, e.g. for indicator 12, gross savings (% GDP), which yields a value of -0.01. Low values of bivariate correlation coefficients with regard to the explained variable are also accompanied by high values of certain bivariate correlation coefficients among the explained variables. Among the bivariate correlation coefficients in the correlation matrix, a value six times higher than 0.7 appears.

The values of the key determinants of the correlation matrix \( |R| \) and \( |R| \) are null, thereby the cumulative tightness of \( \Phi^2 \) and \( \Phi^2 \) acquires the value of 1 (calculations were performed to 14 decimal places). Multicollinearity \( M_j = 1 \). Value of cumulative correlation coefficient \( R^2 = 0.951 \ 89.2 \)

The multicollinearity values are also one, which confirms the preliminary conclusion that the basis shows substantially more mutual associations between the explaining variables than to the selected independent variable, which is the average value of the rating level. The calculated cumulative tightness of 0.951289 is sufficiently high and equal to the value of cumulative tightness in Table 3. The results create the impression that it has been possible to gather a sufficient amount of information to explain the rating levels, as they are more than 95% described by the model. However, this otherwise favorable outcome is practically devalued by the extreme multicollinearity value equal to one. These cases are typically resolved with the exchange or reduction of certain indicators. The reason is to improve the quality of the estimate of regression parameters. As part of the regression and correlation basis, therefore, individual indicators were gradually eliminated and any decrease in the cumulative regression coefficient is monitored during an anticipated decrease in multicollinearity.

The first to be eliminated is the indicator \( x_{11} \) gross domestic product, which had a zero regression parameter value of \( b_{11} = 0.00 \) and a high bivariate correlation coefficient for indicator 11.

### Table 4. Correlation matrix for 17 indicators

|   | \( x_1 \) | \( x_2 \) | \( x_3 \) | \( x_4 \) | \( x_5 \) | \( x_6 \) | \( x_7 \) | \( x_8 \) | \( x_9 \) | \( x_{10} \) | \( x_{11} \) | \( x_{12} \) | \( x_{13} \) | \( x_{14} \) | \( x_{15} \) | \( x_{16} \) | \( x_{17} \) | \( y \) |
| \( x_1 \) | 1 | 0.13 | -0.05 | -0.19 | 0.06 | 0.26 | 0.04 | -0.45 | 0.47 | 0.30 | 0.45 | -0.44 | 0.17 | -0.01 | -0.30 | 0.24 | 0.10 | -0.19 |
| \( x_2 \) | 0.13 | 1 | -0.05 | -0.67 | 0.31 | 0.2 | -0.06 | -0.26 | 0.48 | 0.25 | 0.26 | 0.59 | 0.01 | 0.15 | -0.34 | 0.41 | -0.03 | -0.07 |
| \( x_3 \) | -0.05 | -0.05 | 1 | 0.40 | -0.67 | -0.26 | -0.18 | -0.25 | -0.09 | -0.53 | 0.25 | 0.12 | -0.16 | 0.36 | 0.22 | -0.44 | -0.17 | -0.68 |
| \( x_4 \) | -0.19 | -0.67 | 0.40 | 1 | -0.60 | -0.46 | 0.02 | 0.25 | -0.66 | -0.50 | -0.25 | 0.70 | -0.07 | 0.01 | 0.61 | -0.45 | -0.02 | -0.08 |
| \( x_5 \) | 0.06 | 0.31 | -0.67 | -0.60 | 1 | 0.49 | 0.17 | 0.26 | 0.36 | 0.78 | -0.26 | -0.57 | 0.15 | -0.43 | -0.33 | 0.23 | 0.16 | 0.74 |
| \( x_6 \) | 0.26 | 0.22 | -0.26 | -0.46 | 0.49 | 1 | 0.40 | -0.05 | 0.32 | 0.60 | 0.05 | -0.56 | 0.46 | -0.32 | -0.38 | 0.32 | 0.43 | 0.12 |
| \( x_7 \) | 0.04 | -0.06 | -0.18 | 0.02 | 0.17 | 0.40 | 1 | 0.41 | -0.16 | 0.37 | -0.41 | 0.07 | 0.97 | -0.56 | 0.23 | 0.24 | 0.99 | 0.06 |
| \( x_8 \) | -0.45 | -0.26 | -0.25 | 0.25 | 0.26 | -0.05 | 0.41 | 1 | -0.54 | 0.29 | -1.00 | 0.34 | 0.19 | -0.40 | 0.49 | -0.35 | 0.31 | 0.43 |
| \( x_9 \) | 0.47 | 0.48 | -0.09 | -0.66 | 0.36 | 0.02 | -0.16 | -0.54 | 1 | 0.23 | 0.54 | -0.77 | -0.01 | 0.14 | -0.52 | 0.29 | -0.09 | -0.16 |
| \( x_{10} \) | 0.30 | 0.25 | -0.53 | -0.50 | 0.78 | 0.60 | 0.37 | 0.29 | 0.23 | 1 | -0.29 | -0.49 | 0.35 | -0.39 | -0.15 | 0.26 | 0.37 | 0.53 |
| \( x_{11} \) | 0.45 | 0.26 | 0.25 | -0.25 | 0.26 | 0.05 | -0.41 | -1.00 | 0.54 | -0.29 | 1 | -0.34 | -0.19 | 0.40 | -0.49 | 0.35 | -0.31 | -0.43 |
| \( x_{12} \) | -0.44 | 0.59 | 0.12 | 0.07 | -0.57 | -0.56 | 0.07 | 0.34 | -0.77 | -0.49 | -0.34 | 1 | -0.03 | 0.12 | 0.55 | -0.34 | 0.02 | -0.01 |
| \( x_{13} \) | 0.17 | 0.01 | -0.16 | -0.07 | 0.15 | 0.46 | 0.97 | 0.19 | -0.01 | 0.35 | -0.19 | -0.03 | 1 | -0.47 | 0.06 | 0.38 | 0.99 | -0.02 |
| \( x_{14} \) | -0.01 | 0.15 | 0.36 | 0.01 | -0.43 | -0.32 | -0.56 | -0.40 | 0.14 | -0.39 | 0.40 | 0.12 | -0.47 | 1 | -0.21 | 0.07 | -0.52 | -0.35 |
| \( x_{15} \) | -0.30 | -0.38 | 0.22 | 0.61 | -0.33 | -0.38 | 0.23 | 0.49 | -0.52 | -0.15 | -0.49 | 0.55 | 0.06 | -0.21 | 1 | -0.41 | 0.15 | -0.05 |
| \( x_{16} \) | 0.24 | 0.41 | -0.44 | -0.45 | 0.23 | 0.32 | 0.24 | -0.35 | 0.29 | 0.26 | 0.35 | -0.34 | 0.38 | 0.07 | -0.41 | 1 | 0.31 | 0.03 |
| \( x_{17} \) | 0.10 | -0.03 | -0.17 | -0.02 | 0.16 | 0.43 | 0.99 | 0.31 | -0.09 | 0.37 | -0.31 | 0.02 | 0.99 | -0.52 | 0.15 | 0.31 | 1 | 0.02 |
| \( y \) | -0.19 | -0.07 | -0.68 | -0.08 | 0.74 | 0.12 | 0.06 | 0.43 | -0.16 | 0.53 | -0.43 | -0.01 | -0.02 | -0.35 | -0.05 | 0.03 | 0.02 | 1 |

Source: Own processing, Databank, World Bank, World Development Indicators.
cator 8, net exports of goods and services (% GDP), which is 1.00. In this manner, new values for the next round of calculations were obtained. Other indicators are then eliminated in a similar manner until one arrived at the results listed in Table 5.

Given the unsatisfactory values of multicollinearity, which is still too high and at still tolerable values of the cumulative correlation coefficient, the article will continue in the same manner with the gradual reduction of indicators to two indicators, provided that the cumulative correlation coefficient does not drop below the value of the highest bivariate correlation coefficient of the dependent variable, i.e. the rating level of the arbitrary explained variable. As shown in Table 4, this is the value of 0.68, which links the dependent variable \(y\) with the indicator \(x_3\) – inflation, consumer prices (annual in %). The study only fell below the stated value of 0.68 when eliminating all indicators except the last two, which are \(x_5\), GDP per capita, and \(x_8\), net exports of goods and services (% GDP).

### Table 5. Trends of cumulative characteristics when gradually eliminating indicators from 17 to 2

<table>
<thead>
<tr>
<th>(M_j)</th>
<th>(R^2_j)</th>
<th>Number of indicators</th>
<th>Number of excluded indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000000</td>
<td>0.951289</td>
<td>17</td>
<td>x</td>
</tr>
<tr>
<td>1.000000</td>
<td>0.951289</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>1.000000</td>
<td>0.942769</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>0.999993</td>
<td>0.939616</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>0.999981</td>
<td>0.935906</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>0.999783</td>
<td>0.935659</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>0.999373</td>
<td>0.933413</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>0.998655</td>
<td>0.921470</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>0.995818</td>
<td>0.921433</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>0.985566</td>
<td>0.891102</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>0.972801</td>
<td>0.875501</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>0.892436</td>
<td>0.869311</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>0.841370</td>
<td>0.854041</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>0.793799</td>
<td>0.820074</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0.587582</td>
<td>0.663155</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>0.103191</td>
<td>0.611670</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Source: Own processing.

![Figure 2. Trends of cumulative characteristics when gradually eliminating indicators](http://dx.doi.org/10.21511/imfi.16(3).2019.26)
From Table 5 it is clear that the dropping value of the cumulative correlation coefficient is accompanied by a still high though decreasing multicollinearity. Most solutions have multicollinearity of close to 1. The multicollinearity of 0.84137 drops below the value of the cumulative correlation coefficient of 0.85404 only in the case of five indicators, i.e. after eliminating 12 indicators.

Only the last two cases for 3 or 2 indicators approach the desired zone with their cumulative correlation coefficient at least above 0.5 and multicollinearity of at least under 0.5. This is an important signal of the finding that economic influences on rating level are not as significant as political.

The above conclusions can be displayed in the graph of the complex depiction of the results of multiple regression and correlation analysis, which is described in detail in Figure 1. This figure is supplemented with a quarter circle designating the space of allowable solutions, i.e. those that have sufficiently low multicollinearity and at the same time sufficiently high cumulative correlation coefficient. All solutions for 17 through 8 indicators are clustered in the right peak of the triangle where the multicollinearity approaches one. Another three points representing the solutions for 6, 5, and 4 indicators only “reluctantly” separate themselves from this peak. Only the point for three indicators touches the desired zone. The lowest placed point is now suitable with its low multicollinearity of 0.103 and at the same time still tolerably high cumulative correlation coefficient of 0.612. This solution is now used only by two indicators, i.e. indicators 5, GDP per capita (constant 2011, PPP, USD), and 8, net exports of goods and services (% GDP).

3.3. Deriving an equation for estimating the sovereign credit rating

An illustrative method of verifying the acquisition of results is the reverse estimate of rating levels using the obtained regression equation. In order to calculate estimates of the credit ratings of European countries, a regression equation is used with five indicators, i.e. the one where the gradual elimination of indicators first led to the multicollinearity $M_j = 0.843$ dipping below the value of the cumulative correlation coefficient $R^2 = 0.854$. It is shown that very high-quality estimates of the credit ratings can be obtained using regression parameters for five indicators, namely $x_1$, central debt, $x_3$, inflation, $x_4$, GDP growth rate, $x_5$, GDP per capita, and $x_8$, net exports, where the country forms the observation (see Figure 3).

The relevant regression equation, therefore, has the following form:

\[
\begin{align*}
\text{Sovereign rating} &= a_0 + a_1 x_1 + a_2 x_3 + a_3 x_4 + a_4 x_5 + a_5 x_8 \\
\text{Estimate rating} &= b_0 + b_1 x_1 + b_2 x_3 + b_3 x_4 + b_4 x_5 + b_5 x_8
\end{align*}
\]

Source: Own processing and countryeconomy.com

Figure 3. Comparison of estimated and real average credit rating marks
The obtained regression coefficients represent the weights of selected transformed indicators of the equation. The greatest value has a coefficient $x_5$ GDP per capita (2011, PPS, USD). There are three more indicators: $x_1$ – central government debt and $x_4$ – GDP grown rate. The indicators $x_1$, $x_3$, and $x_4$ are ten times smaller than $x_5$. The least importance is given by the regression parameter 0.009 by the $x_8$ net of goods and services (% GDP). The absolute regression coefficient of −1.516 is rather of calibration significance.

Figure 3 documents the very high quality of the estimates, which for most countries are very close to the actual values of the rating levels. The estimate is substantially higher for Poland and Hungary, which are countries with a significantly agrarian focus. A significantly lower estimate compared to the credit rating was found for the Netherlands, Spain, Great Britain, and Portugal, i.e. for technologically more developed countries.

**CONCLUSION**

More frequent and higher volumes of international capital flows have created a need for an ongoing, responsive, and sophisticated measurement of the creditworthiness of a country. The complexity of this activity has given rise to specialized institutions that focus on this problem in the form of sovereign credit ratings assessment. The key input information for a sovereign credit rating determining is the trends of macroeconomic indicators. This logically raises the question of whether it would be possible or not to create a regression equation to estimate the credit ratings of countries using the trends of these indicators. The legitimate derivation of such an equation, however, must avoid a deeper analysis of the relationship between a sovereign rating and the trends of these indicators.

The objective of this article is to explore whether sovereign credit rating can be reliably estimated using trends of selected macroeconomic indicators, despite the fact that sovereign credit rating is most likely influenced by factors other than economic factors. This article has confirmed that multiple regression and correlation analysis are suitable for this purpose, but must be supplemented by the original method of multicollinearity measuring.

A separate conclusion about the relationships between the trends of macroeconomic indicators being more significant than the relationships between the trends of these indicators and the sovereign credit rating clearly indicates a high degree of multicollinearity that could be understood as bad news for the prospect of finding a regression equation that would provide quality estimates of sovereign rating on the basis of the trends of certain selected macroeconomic indicators. The only possibility remaining is to formulate such an equation for a few indicators that already offer more satisfactory model parameters. For example, for five indicators the multicollinearity is already smaller than the cumulative correlation coefficient. Surprisingly, such an equation can be used quite well for the sought-after estimates. The quality of estimates when compared to reality is illustrated in the article. The aim of the article is thereby achieved. All conclusions, of course, rely on the assumption on linear links for indicators or their logarithmic transformations. The methodology did not incorporate any special non-linear regressions. Future research should remove this limitation. It would also be desirable to verify the results obtained with either a larger sample of countries or a larger number of macroeconomic indicators.

**ACKNOWLEDGMENT**

The output has been created by solving the student project “Synergic effect as a result of optimization of production function showing economic reality”, a grant for the Specific university research at the University of Finance and Administration.
REFERENCES


