ORIGINAL PAPER

DOI: 10.26794/2587-5671-2021-25-5-215-234
UDC 330.59(045)
JEL I10, I12, C53, C55

Prospects for the Integration of Google Trends Data and Official Statistics to Assess Social Comfort and Predict the Financial Situation of the Population

M.V. Shakleina[a] ✉, M.I. Volkova[b], K.I. Shaklein[c], S.R. Yakiro[d]
[a] Lomonosov Moscow State University, Moscow, Russia; [b] Plekhanov Russian University of Economics, Moscow, Russia; [c] OJSC "Russian Railways", Moscow, Russia; [d] JSC "SOGAZ", Moscow, Russia
[a] https://orcid.org/0000-0002-1947-8640; [b] https://orcid.org/0000-0001-8941-0548; [c] https://orcid.org/0000-0003-3508-7372; [d] https://orcid.org/0000-0003-2365-8043
✉ Corresponding author

ABSTRACT

This paper aims to develop a theory of statistical observation in terms of scientific and methodological approaches to processing big data and to determine the possibilities of integrating information resources of various types to measure complex latent categories (using the example of social comfort) and to apply this experience in practice through the use of the financial situation indicators in forecasting. The authors have built a social comfort model in which the choice of weights for its components is based on a modified **principal component analysis**. The assessment is based on Google Trends data and official statistics. Google Trends data analysis methods are based on the development of an integrated approach to the semantic search for information about the components of social comfort, which reduces the share of author's subjectivity; methodology of primary processing, considering the principles of comparability, homogeneity, consistency, relevance, description of functions and models necessary for the selection and adjustment of search queries. The proposed algorithm for working with big data allowed to determine the components of social comfort ("Education and Training", "Safety", "Leisure and free time"), for which it is necessary to directly integrate big data in the system of primary statistical accounting with further data processing and obtaining composite indicators. The authors **conclude** that a stable significant correlation has been found for the "Financial Situation" component, which makes it possible to use it for further calculations and extrapolation of financial indicators. The **scientific novelty** lies in the development of principles and directions for the integration of two alternative data sources when assessing complex latent categories. The findings and the results of the integral assessment of social comfort can be used by state statistics authorities to form a new type of continuous statistical observation based on the use of big data, as well as by executive authorities at the federal, regional and municipal levels in terms of determining the priorities of socio-economic policy development.

Keywords: social comfort; well-being; harmonization of information resources; official statistics; Google Trends; integral indicator

For citation: **Shakleina M.V., Volkova M.I., Shaklein K.I., Yakiro S.R. Prospects for the integration of Google Trends data and official statistics to assess social comfort and predict the financial situation of the population.** ***Finance: Theory and Practice.*** **2021;25(5):215-234. DOI: 10.26794/2587-5671-2021-25-5-215-234**

INTRODUCTION

Over the past two decades, the popularization of Internet use has increased significantly, which has contributed to an increase in the amount of stored information about user activities. Examples of big data are social media data, telephone records, websites, search engine data [1]. New trends have attracted academic interest in the use of big data in research.

Big data is gaining popularity for measuring human well-being as well as predicting financial performance. Based on the analysis of foreign and domestic periodical publication, the following main sources of big data can be identified:

1. Google Trends (GT). Y. Algan et al. [2] use search queries to construct an index of the well-being of Google in the United States, then the methodology for constructing an index of the social well-being of Russians is tested [3]. An article by Y. Algan et al. [2] identifies several aspects of well-being: material conditions (financial well-being), social aspects and health.

3. Articles in newspapers. E. Carlquist et al. [4] investigated occurrence for 1992–2014 of 39 words in the Norwegian press media describing well-being facets of the population. Four newspapers were selected to highlight socio-cultural and regional differences. The authors primarily sought to research words and phrases related to the financial situations in everyday Norwegian vocabulary.

2. Twitter. The data can also be used to study emotional well-being [5–9]. Twitter data, based on the analysis of published tweets, allows us to build either the sentiment index or the degree of satisfaction/dissatisfaction with life.

4. Facebook status. The authors [10–12] propose a method for predicting well-being using social networks. Based on the analysis of statuses, open messages, the semantic correlation of keywords in messages is determined, then an aggregated index of well-being is built.

Big data provides vast opportunities for understanding human interactions in societies with rich spatial and temporal dynamics, as well as for identifying complex interactions and nonlinearities between variables. One of the most popular big data resources is Google Trends.

According to a study [13], the following advantages of Google Trends data are noted:

1. High frequency/periodicity. User moods and preferences change can be observed every day.

2. Search queries are better at revealing the attitudes of the individual compared to traditional polls. Many respondents answer the questionnaire for altruistic reasons since there is no motive to answer the questions frankly and deeply [14]. Search queries can reveal more personal information. For example, the topic of losing a job can be very sensitive for respondents, and they may not want to discuss it.

On the other hand, the search volume for the words "find a job", "search for a job" shows the concern for this problem for a person. Hence, it is concluded that the data obtained from search queries are more objective.

Many researchers, heads of international organizations see in the future an objective replacement of the data of expensive official statistics, which reach users with a great delay, with big data in order to conduct various real-time monitoring of the living conditions of the population, achieve the UN Sustainable Development Goals, etc. One of the successful examples of using large data for monitoring purposes is a study by J. Ginsberg et al. [15], which shows that, based on Google Trends, it is possible to track and predict the spread of influenza before the Centers for Disease Control and Prevention.

However, there are a number of methodological difficulties in using big data. The author of the study [16] notes the incomparability of big data with official statistics due to the use of different methodologies and classifications.

However, bringing big data in line with the requirements of national and international recommendations will reduce its advantages in terms of efficiency of use, timeliness and relevance, which currently provides them with high economic efficiency.

A prerequisite for the use of big data is wide access of the population to the Internet. Despite the rapid development of the Internet in the last decade, the possibilities of big data in developed countries are higher than in developing ones [16]. So, according to official statistics, in 2019 in Russia the share of the population using the Internet on a daily basis was 73%, in Moscow — 82%. This fact can lead to biased estimates of the studied variables since the reliability of the results is guaranteed not only by a large number of observations but primarily by the representativeness of the sample population.

Serious methodological work and high risks of using big data form obstacles to their successful integration into official statistics. In Russia, there are some pilot projects on the use of Internet resources to improve consumer price statistics, data from mobile operators for tourism statistics, satellite communications monitoring for the development of environmental statistics [17]. However, we note the relatively narrow scope of their application and the lack of practical experience (no official publications are presented).

At the same time, there are examples of successful implementation of big data in official statistics in some developed countries. In 2015, Statistics Netherlands expanded transport statistics with the publication of indicators that were calculated on the basis of information received from sensors on the country's highways.[1] Real-time data made it possible to make timely decisions when ice formed in the northern part of the country. An example of such integration is the experience of Canada: forecasting the yield of agricultural crops is made not only on the basis of the results of surveys of farmers but also information on the state of land and climate obtained via satellite communications [18]. Also, active work is underway to attract big data as an alternative source of information on consumer prices in Denmark, the Netherlands, Italy, Norway, Australia, Switzerland, Belgium, New Zealand, Sweden.

Since the experience of introducing big data into the practice of state statistics bodies already exists, it is necessary to continue research in the field of the principles of integrating two information sources, and not be limited only to general projects to study the potential of big data.

In this regard, the aim of the study is to develop the theory of statistical observation in terms of scientific and methodological approaches to processing big data and to determine the possibilities of integrating information resources of various types in relation to measuring complex latent categories (using the example of social comfort).

The introduction of a new economic category "social comfort" is necessary to *"determine the dynamics of the real level of well-being of the population, assess the true quality of life of people"* [19]. Despite the novelty of the study of this process in Russian practice, foreign studies provide an analysis of comfortable conditions for an individual in the following areas: geography, sociology, medicine, psychology, economics, and finance. In [19], the axiomatics and composition of the introduced category are considered in detail. Incoming categories of social comfort: health and medical services, education and learning, social support and pensions, financial situation, employment, housing and living conditions, ethical norms and values, safety, political stability, rest and leisure, ecology and the environment, infrastructure.

[1] A13 busiest national motorway in the Netherlands. URL: https://www.cbs.nl/-/media/imported/documents/2015/31/a13-busiest-national-motorway-in-the-netherlands.pdf?la=en-gb (accessed on 02.09.2021).

In this work, based on the use of the resources of the Federal State Statistics Service (Rosstat) and big data, it is proposed to assess social comfort, to determine the degree of consistency of its components built on two data sources, and to identify possible directions of integration.

EMPIRICAL APPROACHES TO SEMANTIC SEARCH OF INFORMATION ON FINANCIAL AND ECONOMIC INDICATORS BASED ON WEB REQUESTS

Most of the studies conducted have demonstrated the promise of using big data. However, the problem of choosing keywords to determine user queries is still relevant. In many works, the formation of keywords is based on a small number of user queries based on the intuitive assumptions of the authors of the study about the importance of a particular query for a person. As noted earlier, the first work on introducing big data into statistical practice and calculations focused on financial aspects.

Thus, in [20], inflation forecasting is carried out on the basis of web requests. The author examines 75 search queries that are related to financial markets, as well as the interests of the population, economic and financial phenomena, and processes. These queries were selected among the most popular ones based on the analysis of correlation dependences with inflation.

The author of the study [21] analyzes inflation expectations using the keyword "inflation", on the basis of which all kinds of search queries are built. Further, the author makes a comparative assessment with the results of inflationary expectations of the population based on the data of sociological surveys.

In [2], for each index of subjective well-being, which is represented by an index of positive and negative emotions, a set of indirect variables from Google search queries (such as happiness, respect, stress, anxiety, etc.) is used reasonably. The choice is primarily determined by the direction of the evoked emotions: positive or negative.

The author of the study [22] uses the Google Trends service as a proxy to predict the volatility of energy prices. The authors begin their research with a set of 90 terms used in the energy sector. The most popular words on this topic from Google are added to terms gleaned from professional literature. Filtration of such a set of words occurs by building various models that best predict the volatility of prices for crude oil, fuel oil, gasoline, natural gas.

The author of the study [13] compiles a list of search queries that reveal links to economic conditions. To determine the search words as objectively as possible, the author starts working with a vocabulary of finance and text analytics [23] and selects words related to "economic" words that positively or negatively affect a person's mood. Studies [13, 23, 24] use Harvard IV-4 Dictionaries, which have several editions, as the rationale for the choice of words to create an aggregated index of investor sentiment. This dictionary was developed by Dexter Dunphy and colleagues [25–27].

The result is a list of 149 words such as inflation, recession, security, etc. Then each of the 149 words is entered into Google Trends, which selects the top ten (most popular) search queries for each word. For example, the word "deficit" taken from the dictionary leads to the following search queries: "budget deficit", "attention deficit", "trade deficit", etc. As a result, 149 words increased to 1490. The last stage excludes those words/phrases provided by Google that are not related to economic conditions or finance and have zero search volume.

After the procedure of statistical processing of the received time series of search queries, the financial and economic index is constructed as a new indicator for determining and predicting investor sentiment. The high predictive power of the index is noted and

possible prospects for practical application are indicated.

It is worth noting several works that, at the initial stage, take an arbitrary set of search terms, and then select the most informative ones using the Bayesian model averaging [3, 28].

In the study [29], the authors use three different types of information to identify the determinants of the saving behavior of the population in the EU countries: macroeconomic statistics (nominal effective exchange rate, nominal GDP, inflation indicators, etc.), Google search words (42 words), which reflect the mood of economic agents, behavioral, psychological factors influencing preferences, as well as the data of opinion polls reflecting expectations regarding the current and future financial and economic situation. The selection of all variables for analysis is based on the economic intuition of the authors. The Bayesian model averaging is used as a model tool. The authors attribute its use to the lack of Google's keyword selection strategy.

APPROACHES TO LARGE DATA PROCESSING

Big data is generated by the users themselves. Unlike official statistics, they are not collected according to a specially developed and approved methodology. In this regard, for their adequate use and integration into official statistics, it is necessary to develop a special methodology for collection and processing. We will analyze the existing experience in processing big data in foreign studies.

In a study [30], the authors analyze the frequency of search queries related to tourism in Germany based on Google Trends and suggest ways to cleanse the data to eliminate false predictions.

According to the source [31], Google Trends is formed as follows: the ratio of the number of Internet requests for a specific keyword at time *t* in region *r to the total number of requests* at time *t* in region *r* is determined, and the found ratio is multiplied by 100 to standardize.

The authors propose the following modification of the initial data: find the ratio of the number of web requests for a specific keyword at time *t* in region *r to the average value* of the volume of web requests for keywords at time *t* in region *r*. The resulting modified data are called averages divided by the analyzed categories. However, this transformation raises a number of problems:

- the time series has a stronger seasonality as a result of the peculiar seasonal variations of each time series separately;
- the time series has a larger number of outliers due to the increasing role of individual time series in the denominator.

To overcome the problems mentioned above, the authors propose to find the ratio of the number of web requests for a specific keyword at time *t* in region *r* to the average trend of web requests for keywords at time *t* in region *r*. The trend is determined according to the decomposition of the time series [32]. The modified time series is called division by the average trend.

Research [24] strives to confirm the potential for public sentiment-based searches to influence the Portuguese stock market. As a way of processing queries, the author proposes to logarithm the search query for a specific word for a week, and then find the first differences between the volumes of queries for a specific word over two periods of time. To ensure comparability and consistency of the data, the author proposes to amend for seasonality and heteroscedasticity. To remove outliers, the search sample is censored, that is, 5% of the sample with the smallest and largest search volume is cut off. To test the seasonality, the analysis of variance method is used, in which the hypothesis of equality of 12-month averages is tested. Queries with a pronounced seasonality were cleared by building regression models with 12 dummy variables (for each month), in which the residuals were found and used in subsequent iterations of the analysis. To

eliminate heteroscedasticity, a standardization procedure is used using the standard deviation.

In [33], it is proposed to combine the search queries for the standard deviation and the deseasonalization procedure for clearing seasonal fluctuations, performed through the seasonal package in the R language.

The author of the study [34] proposes to use the logarithm with the further taking of the first differences to bring the time series of search queries to a stationary form.

These procedures are followed to clear Google Trends search results [2]:

1. Standardization using Z-score, as a result of which the distribution of data for different search queries is reduced to one scale, which allows comparisons.
2. Elimination of sharp jumps in popularity with the help of a moving average.
3. Removal of search queries with a continuous zero search volume.
4. Clearing the time series from the trend by building regressions with a time trend and removing search queries with a coefficient of determination higher than 0.6.
5. Clearing the time series from seasonality by constructing regressions with monthly dummy variables and using further residuals of regression models.

This methodological approach to big data processing seems to be the most complex and consistent. The main stages of the described approach will be used in the framework of the study.

SELECTING KEYWORDS AND PROCESSING GOOGLE TRENDS SEARCH REQUIREMENTS

The main problem of most research on Google Trends is that the selection of keywords for the semantic disclosure of a particular socio-economic process or phenomenon is selected intuitively, based on the author's experience. Further, a number of econometric methods are used (Bayesian model averaging, factor analysis, correlation analysis, etc.) to select the most informative search queries that reflect the analyzed index or another indicator. A significant difference of the study is a well-grounded approach to semantic information retrieval (based on Google Trends search queries) about the components of social comfort, which consists in the use of the Harvard IV-4 Dictionaries. The peculiarity of this dictionary is that it helps to solve the problem of ambiguity in assigning words to certain categories. For example, the dictionary contains groupings of words according to the following categories: words of positive worldview, negative worldview, words of joy, pain, virtue, vice; words characterizing social categories (education, finance, labor, etc.); motivational words, etc.[2] In this regard, in this study, each of the twelve components of social comfort is filled with keywords from the Harvard dictionary, then an analysis of the degree of compliance with the realities of Russian reality is carried out (such words as canoe, cowboy, Thanksgiving Day, Independence Day, Constitutional Convention, Bill of Rights, Jury — have been removed).

In order to bring the set of keywords-queries of users closer to the real conditions of everyday Russian life, the following words were added: homeowners association, minimum wage, unified national exam, amendments to the Constitution, single voting day, housing and public utilities, compulsory medical insurance, voluntary medical insurance. Of course, the share of the author's subjectivity is not excluded, but it is less than 26.5% of the total number of words. In addition, it should be noted that the set of keywords did not include verb queries: get sick, seek, play, pray, etc., as well as those that have several lexical meanings: [vena] (Vienna) and [vena] (vein); [vera] (faith) and [Vera] (Vera — name); [zhelezo] (ferrum) and [zhelezo] (iron), etc.

An example of filling each block of social comfort with search queries can be seen in

[2] URL: http://www.wjh.harvard.edu/~inquirer/homecat.htm (accessed on 02.09.2021).

Table 1, which presents only some part of the generated set of keywords.

Big data processing starts with comparability.

At the first stage of big data analysis, as the experience of the conducted research shows, scaling is necessary to ensure the comparability and consistency of the initial data. One of the most common standardization methods is standard deviation correction [2, 33], while other researchers [24, 34] use logarithms. In this article, we will rely on the variant of normalization proposed by S.A. Ayvazyan [35] when constructing complex synthetic latent categories of the population's quality of life:

$$\tilde{x}_{j,t} = \frac{x_{j,t} - x_{j,min}}{x_{j,max} - x_{j,min}} N, \quad (1)$$

where $\tilde{x}_{j,t}$ — unified search query value ($j = 1,2,\ldots,p; t = 2010,2011,\ldots,2021$);

$x_{j,min}, x_{j,max}$ — maximum and minimum values of private indicators; $N = 10$.

$$\tilde{x}_{j,t} = \frac{x_{j,max} - x_{j,t}}{x_{j,max} - x_{j,min}} N. \quad (2)$$

The first option (1) of normalization is used in the case of a positive perception of the search query by an individual, the second option (2) of normalization is used in the case of a negative perception. Information on whether a word belongs to a positive or negative worldview is available in the Harvard IV-4 Dictionaries. Words that are not represented in the Harvard IV-4 Dictionaries were categorized by the authors themselves.

It is worth noting the following regularities between the growth of interest in a particular request and its positive/negative assessment (*Table 2*).

At the second stage, the risk of overfitting the model due to sharp jumps is reduced by using a moving average. For example: "Constitution" (sharp jumps caused by voting for amendments to the Constitution in 2020) or "football" (jumps caused by the 2018 World Cup), etc.

The moving average (MA) of order q is defined as follows:

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}. \quad (3)$$

In this case, the order q is calculated considering the number of previous values of random deviations $\varepsilon_{t-1},\ldots,\varepsilon_{t-q}$. In this study, the smoothing window will be equal to three months, respectively, the order q of the moving average is equal to three [MA (3)].

At the third stage, the time series is cleared from the trend. The need for this operation is due to the fact that a strong time trend can lead to an inadequate forecast of social comfort, and also when finding correlations in the process of grouping words in a block, the problem of compiling unreliable block indicators of social comfort may arise, since the words will have the same time trend in common and not a semantic load. In this regard, a regression equation is constructed for the dependence of the search query on the trend of the form:

$$y_t = \theta_1 t + \varepsilon_t, \quad (4)$$

where $\theta_1 t$ — trend; ε_t — a random variable characterizing the deviation of the level from the trend.

Based on the constructed equation, the corrected coefficient of determination is calculated. Search query models with a value of this coefficient above 0.6 will be removed [2].

The fourth stage is the deseasonalization of search queries. The presence of extreme seasonality in queries can lead to strong correlation between themselves, caused by the correspondence of the same seasonal pattern. For deseasonalization, the statsmodels Python module is used. This module includes many classes and functions for evaluating various statistical models, as well as for performing statistical tests and examining statistical data. In particular, this module includes

Table 1

Google Trends searches describing social comfort components

	Block	Examples of search queries	Number of words in a block
1	Health and medical services	Myopia, depression, healthy lifestyle, nerves, proper nutrition, ambulance, diet, etc.	38
2	Education and learning	Graduate, Unified National Exam, Bachelor's programme, mathematics, science, postgraduate study, time management, remote learning, tutors, etc.	42
3	Social support and pensions	Pension, social protection, seniority, allowance, social rehabilitation, nursing home, social support, etc.	23
4	Financial situation	Credit, wages, bankruptcy, deposit, inflation, property, crisis, obligation, poverty, currency, etc.	37
5	Employment and working conditions	HEADHUNTER, unemployed, high-paying job, career, professional development, labor market, labor relations, job fair, dismissal, etc.	53
6	Housing and living conditions	Receipt, homeowner's association, apartment rent, real estate market, apartment prices, neighbors, management company, etc.	13
7	Ethical norms and values	Atheism, the Bible, patriotism, morality, reputation, racism, freedom of speech, the meaning of life, Christianity, feminism, honor, etc.	52
8	Safety	Attack, security, discrimination, thieves, bullying, theft, fraud, punishment, accident, suicide, theft, etc.	60
9	Political stability	Bureaucracy, democracy, citizenship, hunger strike, constitution, parliament, officials, voice, legitimacy, bribes, extremism, etc.	53
10	Rest and leisure	Fishing, volleyball, vacation, seaside vacations, fiction, picnic, boarding house, zoo, museums, hotel, etc.	54
11	Ecology and the environment	Atmosphere, environment, waste recycling, ecology, water quality, nature, climate, greenhouse effect, industrial waste, treatment facilities, environmental education, forest resources, etc.	22
12	Infrastructure (transport, communications, Internet)	Mobile operator, online store, 5G, scooter, bicycle, track, vehicle, train, road, bridge, airport, car, etc.	28

Source: compiled by the authors based on Google Trends.

Table 2

Impact on social comfort of a decrease in interest in words of a positive and negative worldview

	Growth of interest (increase in search queries)	Decrease in interest (decrease in search queries)
Words of a positive worldview	Increase social comfort	Increases social comfort. Interest in words of a positive worldview can be reduced due to the achievement of a certain level of comfort that suits the individual in relation to this process, characterized by a search query. Example: the decline in searches for the word «internet» is related to the availability and good quality of the internet in recent years
Words of negative worldview	Decrease social comfort	Increase social comfort

Source: compiled by the authors based on Google Trends.

a Seasonal-Trend Decomposition Procedure Based on Loess Method (STL) — a method for decomposing a time series into seasonal and trend components, as well as residuals using a local regression method.

The STL method decomposes the time series into the components of the additive model:

$$Y_t = T_t + S_t + E_t, \quad (5)$$

where T_t — the trend component; S_t — seasonal component; E_t — the remainder.

Elimination of seasonality occurs by subtracting the seasonal component S_t from the time series.

As a result of sequential execution of the listed iterations of analysis and data processing, 475 words were selected (out of 574 words before processing) for the period January 2010–January 2021 (frequency — 1 month).

The proposed methodology for keyword search and Google Trends processing is implemented in Python code. It allows you to build statistical series in accordance with the basic principles that ensure the quality of statistics — comparability, consistency, accuracy and homogeneity of data.

BUILDING INTEGRATED SOCIAL COMFORT INDICATORS: OFFICIAL STATISTICS AND GOOGLE TRENDS

Large amounts of information require the use of special methods of aggregation and dimensionality reduction. The most popular are factor analysis and principal component analysis. In our study, we will use a modified principal component analysis, presented in more detail in the work of S.A. Ayvazyan [35]. According to this methodology, indicators within each of the 12 blocks[3] of social comfort are combined into block indicators, which are subsequently combined into a consolidated integral indicator of social comfort. Since the study used two types of data, respectively, at the output we got block indicators and

[3] The substantiation of the block indicators of social comfort is given in more detail in the study [19].

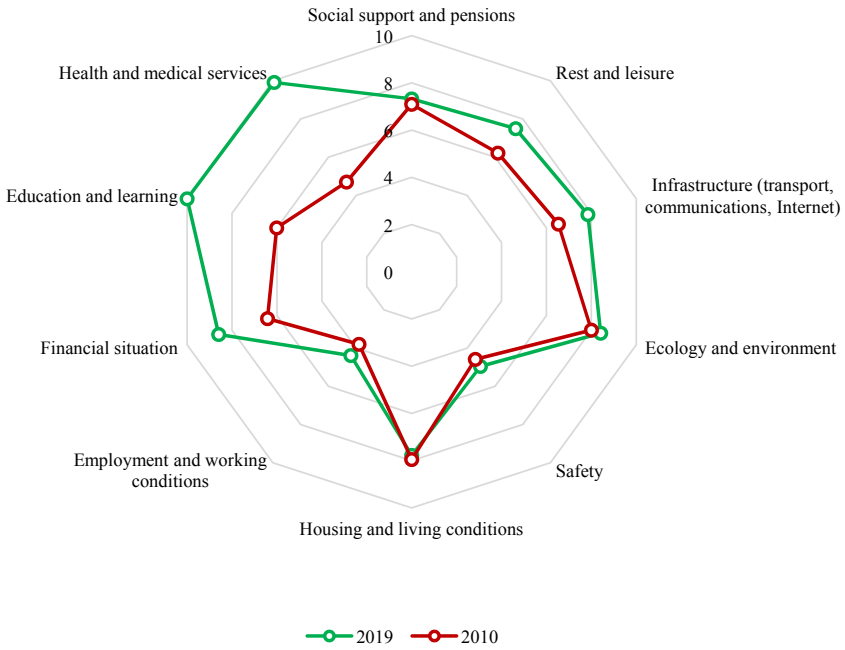

Fig. 1. **Block indicators of social comfort in Moscow for 2010–2019**

Source: compiled by the authors.

a composite integral indicator of social comfort, built according to Google Trends data and official statistics.

In order to harmonize and interconnect information resources of various types: GT and official statistics data presented by Rosstat, it is proposed to use normalization for all data, performed according to formulas (1) and (2), and for GT data, apply the smoothing processing procedures described above, trend exceptions, etc.

Further, we will discuss the results of modeling social comfort, obtained using various types of data, and assess the prospects for using GT in relation to complex latent categories (using the example of social comfort).

Simulation results based on Rosstat data

The information base for filling in the blocks of social comfort was the indicators of the socio-economic situation of the regions of Russia for 2010–2019, taken from the Rosstat website.[4] When choosing indicators, we were guided by the approach to the analysis of the contextual conditions of the Russian Federation and its regions (described in more detail in the study [19]), as well as the requirements for a set of particular criteria for the synthetic latent category [35]. Much attention was paid to the degree of conformity of the socio-economic content of the indicator to the directly measured hidden category ("Safety", "Housing and medical services", etc.), reliability, accessibility in the official source. In this regard, such blocks of social comfort as "Ethical norms and values", "Political stability" remained empty due to the high share of the subjectivity of the indicators included in them and the lack of information on the official resource of state statistics. The empirical base of the study includes a panel of 100 indicators for 2010–2019. The collected indicators are measured on a quantitative scale in accordance with a unified methodology and basic principles of statistical observation, which ensure the consistency and comparability of the objects of observation.

Since the normalization of the indicators included in the panel was carried out

[4] Federal State Statistics Service of the Russian Federation (Rosstat). URL: https://rosstat.gov.ru/ (accessed on 02.09.2021).

on a 10-point scale, then as a result of calculations using a modified analysis of principal component at the output, the values of the block indicators of social comfort will also belong to the interval from 0 to 10.

Let us consider, for example, the results of modeling the social comfort of Moscow for 2010–2019 (*Fig. 1*).

The consolidated integral indicator for Moscow in 2019 amounted to 6.815, which is the maximum value in 2019 among other regions of Russia and is 13% higher than the level of 2010. Such dynamics are explained by a slight change in the weight coefficients by the growth of block integral indicators.

For the analyzed period 2010–2019 Moscow has improved its position in eight out of ten presented components of social comfort. The most significant changes occurred in the components "Financial Situation" (+1.3 points), "Infrastructure" (+2.1 points), "Rest and Leisure" (+4 points), "Social Support and Pensions" (+ 5.2 points). There are no dynamics in the blocks "Health" and "Safety".

The reasons for the changes are as follows:

• "Financial Situation" block: the gross regional product per capita in Moscow is one of the highest in absolute terms and during the analyzed period has increased by about 2 times. In addition, the share of the population with monetary incomes below the subsistence level has significantly decreased (from 10 to 4%). At the same time, the structure of Moscow's GRP over 50% is formed at the expense of the service sector;

• "Infrastructure" block: Moscow is the first region of Russia where the use of mobile high-speed Internet LTE and 4G is widely used among the city's population, which in general contributed to the growth of digitalization of socio-economic processes. In particular, the share of the population using the Internet on a daily basis increased from 60% in 2014 to 82% in 2019;

• "Rest and Leisure" block: there are more than 18 thousand sports facilities in Moscow, which is 2 times higher than the level of 2010, and the all-Russian level increased by 11%;

• "Social Support and Pensions" block: 8.5% of the population of Russia lives in Moscow, this figure in the period 2010–2019 practically did not change, and the share of execution of the budgets of the Pension Fund and the Social Insurance Fund under the item "expenses" of Moscow in the structure of Russia increased by 1.5 and 4.2%, respectively.

According to the approach used for the convolution of multidimensional categories [35], the value of the aggregate integral indicator of social comfort will be determined by the formula:

$$Y_{t,ag.} = 10 - \sqrt{\sum_{j=1}^{12} v_j (y_{j,t} - 10)^2}, \qquad (6)$$

where v_j — these are normalized non-negative weights determined by the fraction of the explained variance of the first principal component of each of the 12 blocks;[5] $y_{j,t}$ — a block indicator of social comfort in year t.

The higher the weight of the block indicator, the more influence it has on the composite indicator of social comfort. Based on this method, it is possible to determine the priorities of socio-economic policy in order to improve social comfort.

We will analyze the performance of this method based on comparing the values of the block indicators and their weight in the composite indicator for two regions (the Republic of Buryatia and the Tula region).

Let us consider the growth rates of the values of block indicators of social comfort for 2010–2019 (*Fig. 2*), as well as the weight of each component in the composite indicator.[6] It can be noted that the priority areas for increasing social comfort are infrastructure

[5] According to official statistics, 10 blocks were built due to the lack of information on the blocks "Ethical norms and values", "Political stability", and according to Google Trends — 12 blocks.

[6] The weight of the block indicators is the average for 2010–2019 and constant for each object of observation.

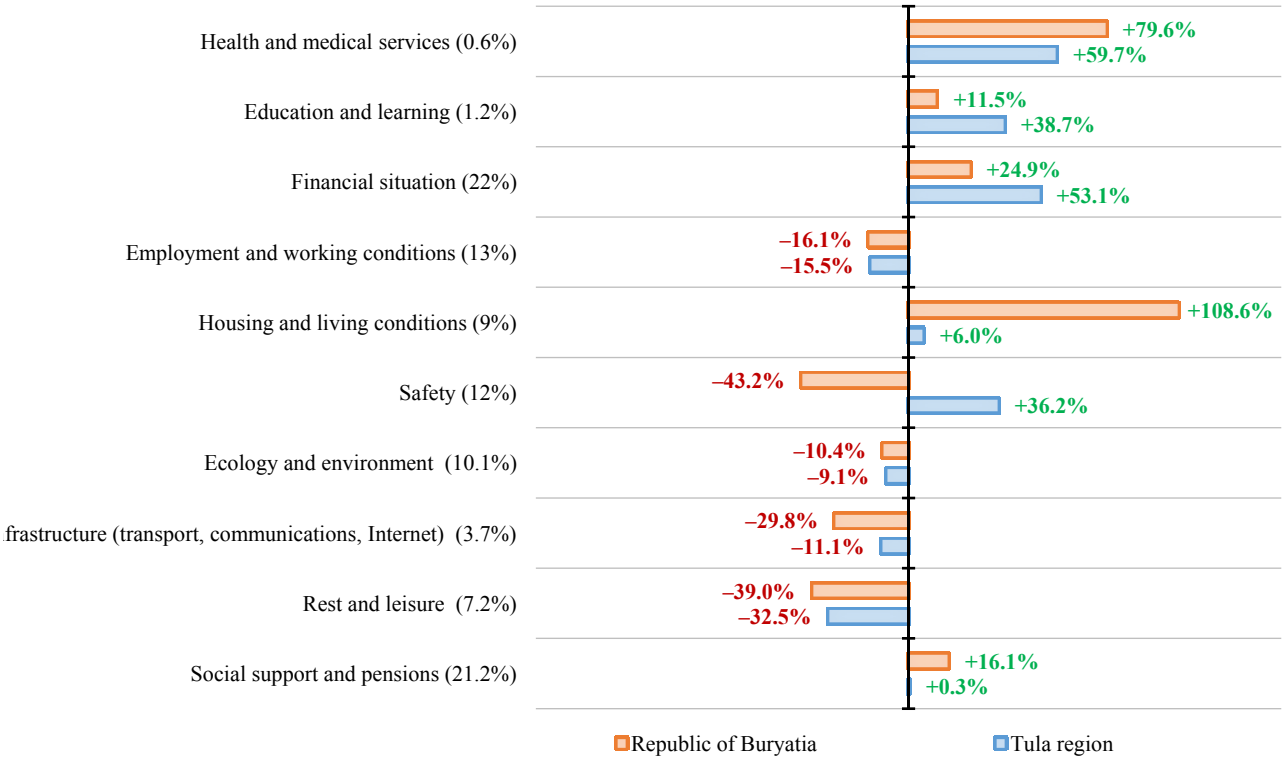

Fig. 2. **Composite indicator of social comfort in regions**

Source: compiled by the authors.

(weight in the aggregate indicator — 22%); health and medical services (21.2%); ecology and the environment (13%); housing and living conditions (12%). At the same time, the aggregate indicator of social comfort in the Tula region increased by 8.4%, and in the Republic of Buryatia — by only 0.5% due to the outstripping growth of block indicators of social comfort in the Tula region, which are a priority (infrastructure, housing). The data obtained can become the basis for monitoring social comfort and subsequent adjustments to the ongoing socio-economic policy of the region within the framework of priority factors.

Comparison of simulation results based on GT data and official statistics

According to the results of modeling based on Google Trends data, the composite integral indicator for Russia from 2010 to 2019 increased by 54% and amounted to 4.631 points. The subjective estimate of the population, aggregated based on search queries, is lower than the estimate based on Rosstat data (5.371).

Using the developed methodology for processing Google Trends (1)–(5), the values of block indicators of social comfort were calculated and a comparative analysis was carried out with similar indicators according to official statistics. *Table 3* shows correlations for block indicators of social comfort.

According to the analyzed table, a significant correlation of all block indicators of social comfort is obvious. At the same time, a stable positive linear relationship is observed for seven out of ten compared blocks of social comfort. There was a **strong** positive correlation of indicators of the block "Social support and pensions", "Ecology and environment"; **moderate** positive correlation — "Health and medical services", "Employment and working conditions",

"Infrastructure", "Housing and living conditions", "Financial situation".

"Financial situation", according to calculations, is formed mainly by such words as "credit", "deposit", "profit", "inflation", "accumulation", "cash", etc. Thus, the block "Financial situation" is determined mainly by words characterizing the processes of receipt of funds by the population.

At the same time, a similar component of social comfort in official statistics is assessed by the following indicators: "consumer spending", "fund ratio", CPI, the share of expenditures on food, etc. In general, official statistics primarily considers the population from the standpoint of forming the expenditures of the range of goods and services.

It is possible to increase the correlation by supplementing the list of indicators of official statistics that characterize the financial situation of the population in terms of the formation of its income, for example, the average profitability on bank deposits, income received from transactions in financial markets, the structure of the formation of disposable money income of the population, etc.

The debatable issue is the strong negative correlation of the blocks "Education and Learning", "Safety", "Rest and Leisure". Further, we will discuss the possible causes of existing dependencies.

The indicators of the **"Education"** block, according to official statistics, are formed by quantitative indicators of the coverage of secondary, secondary professional, and higher education. Most of the indicators of this block (for example, "The share of university students in the working-age population", "The number of graduate students", etc.) demonstrate negative dynamics. A similar trend, which began in 2010, is typical for most regions of Russia [36] and is associated with demographic problems in the country. In this regard, there is a negative dynamic of the block indicator "Education", which contradicts the dynamics of the indicator "Education" according to GT data. It should be noted that a significant drawback of the data on this block published by Rosstat is that they do not fully reflect the level of the intellectual development of the region/country, which ensures competitiveness, changes the standard of living and the level of social comfort. In this regard, the indicators of enrollment in education

Table 3

Pearson correlation coefficients of block indicators of social comfort based on official statistics and Google Trends data

	Pearson correlation coefficient	P-value
Health and medical services	0.59	0.0700
Education and learning	−0.90	0.0003
Social support and pensions	0.88	0.0007
Financial situation	0.46	0.0176
Employment and working conditions	0.58	0.0813
Housing and living conditions	0.35	0.0317
Safety	−0.77	0.0100
Rest and leisure	−0.76	0.0104
Ecology and environment	0.70	0.0250
Infrastructure (transport, communications, Internet)	0.41	0.0208

Source: compiled by the authors.

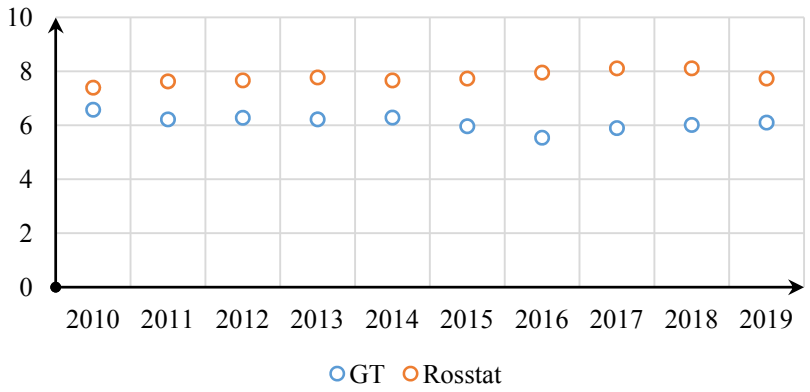

Fig. 3. **Comparative dynamics of the "Safety" block indicators based on official statistics and Google Trends data**
Source: compiled by the authors.

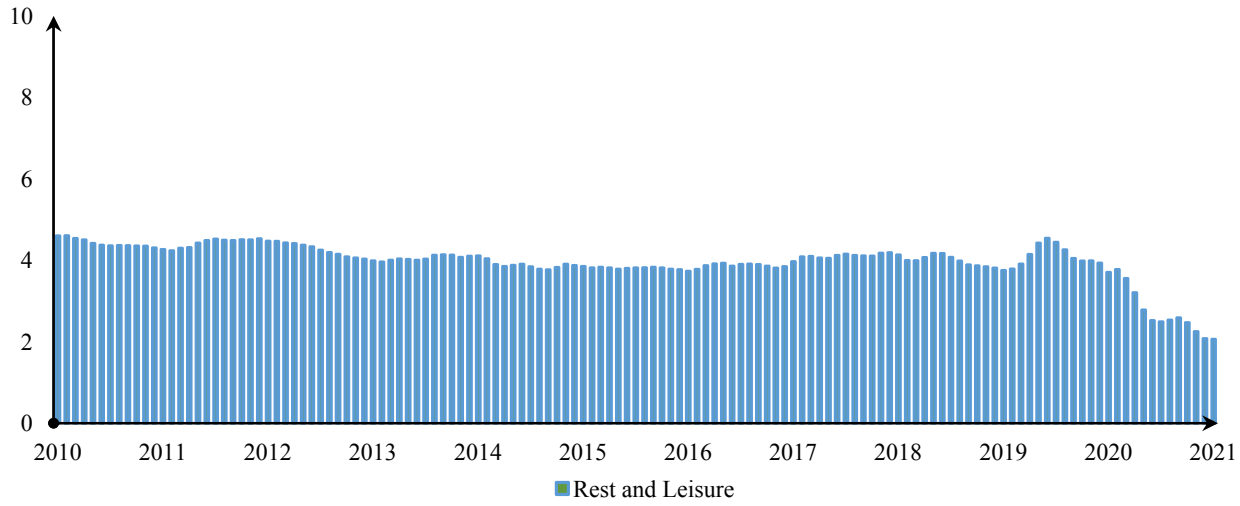

Fig. 4. **Dynamics of the component "Leisure and free time" based on Google Trends data**
Source: compiled by the authors.

should be supplemented with indicators characterizing the performance of schoolchildren, students, the quality of final/entrance exams, international tests (GMAT, IELTS, etc.); the number of academic competition winners; indicators of the attractiveness/openness/prestige of the university; availability of education (preschool, secondary, higher).

The work [37] details the advantages and objectivity of using a new approach to assessing the level of the country's intellectual capital based on the use of new indicators for assessing the level of education in the country. Using other sources of information: polls, GT data allow expanding the indicators of enrollment in education in Russia by qualitative characteristics.

An example of a strong inverse correlation indicates the need to expand education statistics with qualitative indicators, in particular, to consider the possibilities of using alternative sources — GT data, which will increase the objectivity and quality of the information provided. Thus, the analysis showed that the most popular and significant search queries in the "education" block are the words "remote learning", "English", "mathematics", "academic performance". The growing interest of the population on these topics indicates the growth of the intellectual capital of the population.

For the **"Safety"** block, the correlation is –0.77. The indicators of this block of social comfort should characterize the level of risk to the life and health of the population, and not the success of the law enforcement system. Hence, the safety indicators provided by Rosstat cannot be considered as sufficiently valid from the point of view of the characteristics of social comfort, since they are represented by the number of crimes of a different nature, robberies, and causing grievous bodily harm.

In [38], it is noted that the self-awareness of the safety of citizens is influenced by the personal attitudes of citizens and everyday practices (whether they have to carry weapons, gas canisters, etc.), and not the number and disclosure of crimes. In addition, in [39] it is shown that crimes that fall under the article "Murders" have a latency coefficient in Russia of 2.3. This means that the real number of crimes is 2.3 times higher than the indicators of official statistics. Данный факт подтверждается результатами проведенного исследования. This fact is confirmed by the results of the study. The numerical safety score according to GT data is much lower than the estimate according to official statistics: 6 points against 8 points (*Fig. 3*).

If we analyze the dynamics of the two indicators of the block in *Fig. 3*, it can be noted that during the unfavorable economic situation in 2014–2016, caused by the imposition of sanctions, the escalation of the geopolitical conflict, and, as a consequence, the depreciation of the national currency, there is a deviation of the indicator of the "Safety" block according to GT data. This is due to the concerns of citizens, a decrease in the sense of security and comfort. But the graph, built according to official statistics, in the period 2014–2016 demonstrates strong growth, which is contradictory. There is reason to believe that for an adequate assessment of the level of safety of the population, it is also advisable to integrate safety indicators based on GT data into crime statistics.

Block indicators "Rest and Leisure" also show an inverse linear relationship. The statistics of this block are represented by only three indicators: the number of sports institutions; the number of vouchers sold through travel agencies, and the number of Russian tourists served by travel agencies. At the same time, according to GT data, the analyzed block included about 54 indicators reflecting various aspects of an individual's rest and leisure (*Fig. 4*).

The dynamics of the analyzed component of GT are quite adequate to the realities of economic life: the consequences of the COVID-19 pandemic negatively affected the rest of the Russians, since they had to change their travel preferences. The consequence of this is a decrease in the values of this component by 31% within one year. Since the composite indicator of social comfort "rest" has a significant contribution — 10%, more attention should be paid to the development of new tourist destinations, active recreation of the population. The significant role of the indicator of the "Rest and Leisure" block in the formation of social comfort, as well as its weak representation in official statistics, justifies the need to use Google Trends in the development of a methodology for accounting for tourism and recreation statistics.

Thus, the components of social comfort, built on the basis of GT data and having a significant positive correlation with the components of official statistics, can be used to conduct operational monitoring of the living conditions of the population. For the components "Education and Learning", "Safety", "Rest and Leisure", a serious methodological study of options for integrating alternative sources of information into official statistics is required due to inconsistency of results and a poor reflection of the level of social comfort of the population.

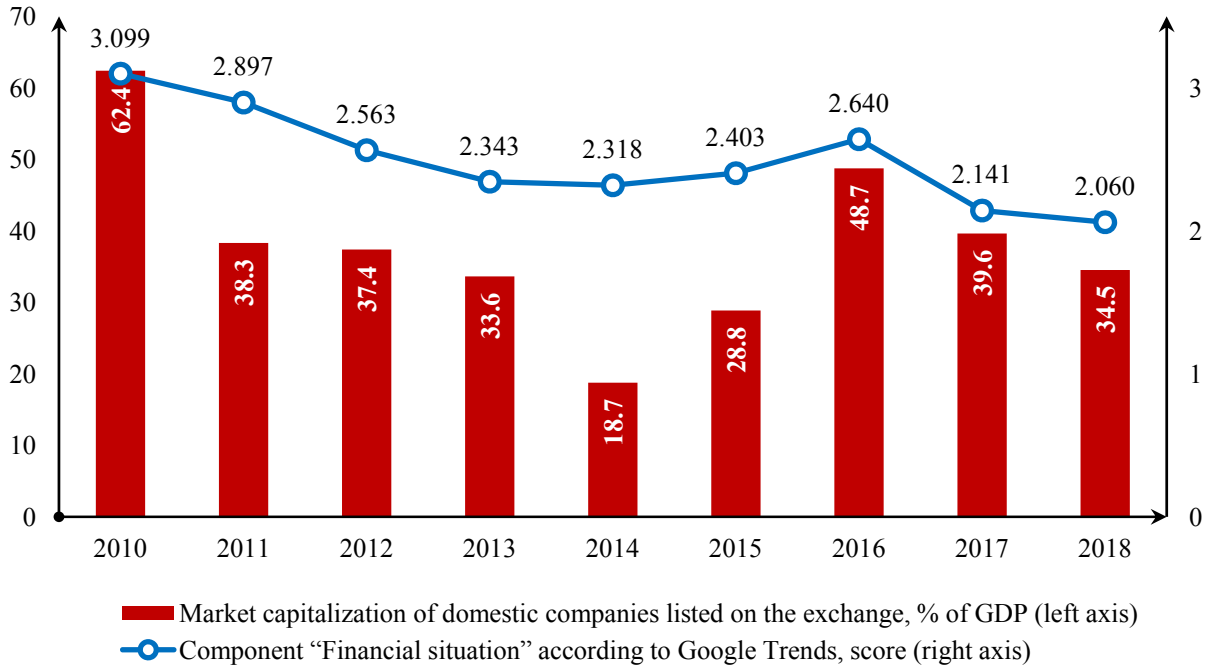

Fig. 5. **Dynamics of the component "Financial situation" and market capitalization of listed domestic companies**

Source: compiled by the authors based on Google Trends and World Bank Datebase.

PROSPECTS FOR USING BIG DATA TO FORECAST FINANCIAL PERFORMANCE

The calculations presented earlier demonstrate the good potential of using big data to predict 7 out of 10 components of social comfort, including the component "Financial situation". In connection with the deterioration of the financial situation of the population in Russia, the prospects for using search queries reflecting the financial mood of economic agents for forecasting economic indicators remain extremely relevant.

The financial sentiments of economic agents, aggregated in one indicator using Google Trends search queries, are becoming an important source of information about their preferences and behavior. To determine the prospects for using the calculated component as a proxy for indicators of financial condition, let us calculate the correlation with the indicator "Market capitalization of national companies whose shares are traded on the stock exchange". According to the published information of the World Bank, the banking system and the stock market are directly related to economic growth, which is the main factor affecting poverty reduction, which, in particular, is considered in the financial component of social comfort.

The high correlation with market cap indicates that search queries can be used to extrapolate financial health indices without using financial statistics resources, making the calculation easier and faster.

The reaction of Internet users' interest in relation to financial market indicators in response to changes in the economic indicator (GDP, MICEX capitalization index, inflation, deposit rates, etc.) social comfort depending on changes in the indicators of the financial situation of the population and the expected trends in economic growth.

CONCLUSIONS

Big data has more detailed statistical assessments of various phenomena and processes in society, which is a necessary argument in developing the provisions of the concept of the quality of life of the population as one of the most important categories of social and economic science. The introduction of the latent category of "social comfort" into

the scientific use deepens the theory of the quality of life of the population in terms of studying a person from the point of view of his inclusion in society, expanding the subjective aspect of measurement, which is explained by the need to use Google Trends.

In the proposed study, an integral assessment of social comfort is carried out using two sources of information: official statistics and Google Trends. The integral assessment of social comfort allows, in turn, to see a bigger picture of the development of the phenomenon in time, as well as to assess the ongoing socio-economic policy.

To minimize the author's subjectivity in assessing social comfort, according to Google Trends, a new approach to semantic search for information about the components of social comfort is used, based on the use of a specialized dictionary, which contains classifications of various processes and phenomena, as well as an analysis of the validity of each search query in terms of disclosure of social comfort and correlation with the realities of Russia.

In the process of modeling, the problem of harmonization and interconnection of different types of resources is solved: a set of econometric methods is used to diagnose data for the presence of a time trend, sharp jumps in the popularity of a query, the presence of a zero-search volume, extreme seasonality and bringing time series to a comparable form not only among themselves within the same information resource. The method used to standardize search queries allows for reliable estimates and modeling of composite categories based on different types of data.

On the basis of the applied methodology, in the analysis of social comfort, 475 Google Trends search queries and 100 indicators taken from official statistics were used, which were aggregated into block indicators of social comfort. Correlation analysis of block indicators showed a stable positive correlation between the components of social comfort built on the basis of GT and official statistics (7 out of 10 components), which indicates good prospects for using alternative sources of information (for example, GT) to assess social comfort for real-time monitoring without resorting to official statistics. There is a strong negative linear relationship for the three components "Education and Learning", "Safety", "Rest and Leisure", which is mainly explained by the weak reliability of statistical indicators for assessing social comfort and determines the primary need to integrate big data in these areas for sharing various sources of information in order to obtain more reliable estimates.

Thus, we can conclude that the use of big data in assessing latent categories gives good results, comparable to the data of official statistics, which opens up opportunities for their use in monitoring and forecasting the financial situation. However, the integration of the two data sources should be carried out sequentially when conducting possible verification with other sources, for example, with the data of opinion polls.

ACKNOWLEDGEMENTS

The reported study was funded by RFBR, project No. 20–310–70037 "Stability". Lomonosov Moscow State University, Moscow, Russia.

REFERENCES

1. Liu J., Li J., Li W., Wu J. Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016;115:134–142. DOI: 10.1016/j.isprsjprs.2015.11.006
2. Algan Y., Beasley E., Guyot F., Higa K., Murtin F., Senik C. Big data measures of well-being: Evidence from a Google well-being index in the United States. OECD Statistics Working Papers. 2016;(03). URL: https://www.oecd-ilibrary.org/docserver/5jlz9hpg0rd1-en.pdf?expires=1629818036&id=id&accname=guest&checksum=7ED855395D5B778D71E405ED1925ECE3

3. Fantazzini D., Shakleina M., Yuras N. Big data for computing social well-being indices of the Russian population. *Prikladnaya ekonometrika = Applied Econometrics*. 2018;50:43–66. (In Russ.).
4. Carlquist E., Nafstad H.E., Blakar R.M., Ulleberg P., Delle Fave A., Phelps J.M. Well-being vocabulary in media language: An analysis of changing word usage in Norwegian newspapers. *The Journal of Positive Psychology*. 2017;12(2):99–109. DOI: 10.1080/17439760.2016.1163411
5. Curini L., Iacus S., Canova L. Measuring idiosyncratic happiness through the analysis of Twitter: An application to the Italian case. *Social Indicators Research*. 2015;121(2):525–542. DOI: 10.1007/s11205–014–0646–2
6. Prata D.N., Soares K.P., Silva M.A., Trevisan D.Q., Letouze P. Social data analysis of Brazilian's mood from Twitter. *International Journal of Social Science and Humanity*. 2016;6(3):179–183. DOI: 10.7763/IJSSH.2016.V6.640
7. Nguyen Q.C., Kath S., Meng H.-W., Li D., Smith K.R., VanDerslice J.A., Wen M., Li F. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Applied Geography*. 2016;73(8):77–88. DOI: 10.1016/j.apgeog.2016.06.003
8. Yang C., Srinivasan P. Life satisfaction and the pursuit of happiness on Twitter. *PLoS ONE*. 2016;11(3):e0150881. DOI: 10.1371/journal.pone.0150881
9. Wang W., Hernancez I., Newman D.A., He J., Bian J. Twitter analysis: Studying US weekly trends in work stress and emotion. *Applied Psychology*. 2016;65(2):355–378. DOI: 10.1111/apps.12065
10. Liu P., Tov W., Kosinski M., Stillwell D.J., Qui L. Do Facebook status updates reflect subjective well-being? *Cyberpsychology, Behavior, and Social Networking*. 2015;18(7):373–379. DOI: 10.1089/cyber.2015.0022
11. LiKamWa R., Liu Y., Lane N.D., Zhong L. MoodScope: Building a mood sensor from smartphone usage patterns. In: Proc. 11th Annu. int. conf. on mobile systems, applications, and services (MobiSys). (Taipei, June 25–28). New York: ACM; 2013:389–402. DOI: 10.1145/2462456.2464449
12. Schwartz H.A., Sap M., Kern M.L., Eichstaedt J.C., Kapelner A., Agrawal M., Ungar L.H. et al. Predicting individual well-being through the language of social media. In: Proc. Pacific symp. on biocomputing (PSB). (Big Island of Hawaii, Jan. 4–8, 2016). Singapore: World Scientific Publishing Co.; 2016:516–527.
13. Da Z., Engelberg J., Gao P. The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*. 2015;28(1):1–32. DOI: 10.1093/rfs/hhu072
14. Singer E. The use of incentives to reduce nonresponse in household surveys. The University of Michigan. Survey Methodology Program. 2002;(051). URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.611.1597&rep=rep1&type=pdf
15. Ginsberg J., Mohebbi M.H., Patel R.S., Brammer L., Smolinski M.S., Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012–1014. DOI: 10.1038/nature07634
16. Upadhyaya S. Big data and official statistics. *Voprosy statistiki*. 2019;26(12):5–14. (In Russ.). DOI: 10.34023/2313–6383–2019–26–12–5–14
17. Oksenoyt G.K. Digital agenda, big data and official statistics. *Voprosy statistiki*. 2018;25(1):3–16. (In Russ.).
18. Plekhanov D.A. Bigdata and official statistics: A review of international experience with integration of new data sources. *Voprosy statistiki*. 2017;(12):49–60. (In Russ.).
19. Shakleina M.V., Volkova M.I., Shaklein K.I., Yakiro S.R. Theoretical and methodological problems of measuring social comfort: Results of empirical analysis based on Russian data. *Economic and Social Changes: Facts, Trends, Forecast*. 2020;13(5):135–152. DOI: 10.15838/esc.2020.5.71.8
20. Petrova D.A. Inflation forecasting based on Internet search queries. *Ekonomicheskoe razvitie Rossii = Russian Economic Developments*. 2019;26(11):55–62. (In Russ.).
21. Guzmán G. Internet search behavior as an economic forecasting tool: The case of inflation expectations. *Journal of Economic and Social Measurement*. 2011;36(3):119–167. DOI: 10.3233/JEM-2011–0342
22. Afkhami M., Cormack L., Ghoddusi H. Google search keywords that best predict energy price volatility. *Energy Economics*. 2017;67:17–27. DOI: 10.1016/j.eneco.2017.07.014

23. Tetlock P.C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*. 2007;62(3):1139–1168. DOI: 10.1111/j.1540–6261.2007.01232.x
24. Brochado A. Google search-based sentiment indexes. *IIMB Management Review*. 2020;32(3):325–335. DOI: 10.1016/j.iimb.2019.10.015
25. Dunphy D.C., Bullard C.G., Crossing E.E.M. Validation of the general inquirer Harvard IV dictionary. Cambridge, MA: Harvard University Library; 1974. 158 p.
26. Kelly E.F., Stone P.J. Computer recognition of English word senses. Amsterdam: North-Holland; 1975. 269 p.
27. Zuell C., Weber R.P., Mohler P.P. Computer-aided text classification for the social sciences: The General Inquirer III. Mannheim: ZUMA, Center for Surveys, Research and Methodology; 1989.
28. Scott S.L., Varian H.R. Bayesian variable selection for nowcasting economic time series. NBER Working Paper. 2013;(19567). URL: https://www.nber.org/system/files/working_papers/w19567/w19567.pdf
29. Kapounek S., Deltuvaitė V., Koráb P. Determinants of foreign currency savings: Evidence from Google search data. *Procedia — Social and Behavioral Sciences*. 2016;220:166–176. DOI: 10.1016/j.sbspro.2016.05.481
30. Bokelmann B., Lessmann S. Spurious patterns in Google Trends data — An analysis of the effects on tourism demand forecasting in Germany. *Tourism Management*. 2019;75:1–12. DOI: 10.1016/j.tourman.2019.04.015
31. Google Trends help — how Trends data is adjusted. Google 2018. URL: https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052 (accessed on 23.04.2018).
32. Cleveland R.B., Cleveland W.S., McRae J.E., Terpenning I. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*. 1990;6(1):3–73. URL: https://www.wessa.net/download/stl.pdf
33. Petrova D.A. Trunin P.V. Revealing the mood of economic agents based on search queries. *Prikladnaya ekonometrika = Applied Econometrics*. 2020;(3):71–87. (In Russ.). DOI: 10.22394/1993–7601–2020–59–71–87
34. Parker J., Cuthbertson C., Loveridge S., Skidmore M., Dyar W. Forecasting state-level premature deaths from alcohol, drugs, and suicides using Google Trends data. *Journal of Affective Disorders*. 2017;213:9–15. DOI: 10.1016/j.jad.2016.10.038
35. Ayvazyan S.A. Analysis of the quality and lifestyle of the population. Moscow: Nauka; 2012. 432 p. (In Russ.).
36. Mindeli L.E., Pashinceva N.I. Russian education system and how it is reflected in statistics. *Voprosy statistiki*. 2016;(11):67–84. (In Russ.).
37. Chan K.L. Intelligence Capital Index. 2017. URL: http://www.kailchan.ca/wp-content/uploads/2017/04/KC_Intelligence-Capital-Index-full-results-and-methodology_Apr-2017_v2.pdf
38. Satarov G.A., Blagoveshchenskii Yu.N. Statistical comparison of Russia and other countries. Civil Initiatives Committee. INDEM Foundation. URL: https://komitetgi.ru/upload/iblock/3cf/3cfcb375eced922f253c446a4b37645b.pdf (In Russ.).
39. Inshakov S.M. Theoretical foundations of research and analysis of latent crime. Moscow: UNITY-DANA; 2011. 839 p. (In Russ.).

ABOUT THE AUTHORS

Marina V. Shakleina — Cand. Sci. (Econ.), Assoc. Prof., Moscow School of Economics of the Lomonosov Moscow State University, Moscow, Russia
shakleina.mv@gmail.com

Mariya I. Volkova — Cand. Sci. (Econ.), Head of the Laboratory "Modeling of socio-economic systems", Plekhanov Russian University of Economics, Moscow, Russia
frauwulf@gmail.com

Konstantin I. Shaklein — Cand. Sci. (Econ.), Chief Specialist of Department of Economics OJSC "Russian Railways", Moscow, Russia
mrshaklein@gmail.com

Stanislav R. Yakiro — Chief Specialist of Department of Risk and Economic Performance Analysis JSC "SOGAZ", Moscow, Russia
yakirosr@yandex.ru

Authors' declared contribution:

Shakleina M.V. — introduction, relevance of the research topic and problem statement, analysis of literature sources, study of research problems, development of research methodology, interpretation of the results.

Volkova M.I. — justification of the choice of indicators for analysis; formation of conclusions and recommendations based on the results of the study.

Shaklein K.I. — statistical analysis of data, description of the calculation method used, analysis of the results, tabular and graphical representation of the results.

Yakiro S.R. — collection and processing of big data, assessment and forecasting.

The article was submitted on 13.05.2021; revised on 27.05.2021 and accepted for publication on 27.08.2021.
The authors read and approved the final version of the manuscript.