Zeszyty Naukowe
Wyższej Szkoły Bankowej w Poznaniu
2018, t. 79, nr 2
DOI: 10.26349/zn.wsb.w.poznaniu.0079.11

Sergiy Gerasymenko

National Academy of Statistics
Accounting and Audit, Kyiv
orcid.org/0000-0002-6522-3091
e-mail: serguey@wp.pl
phone: 797 101 677

The Role of Statistics in Banking Research

Abstract. *This article addresses the place and role of statistics in economic research. After giving a brief outline of the creation of economic statistics and problems they can help to solve, the author provides a detailed discussion of the use of statistical methods for estimating risk of banking activities.*

Keywords: *research tasks, applied statistics, statistical methods, risk associated with banking activities, banking statistics*

Introduction

The aim of the study is to define the place and the role of statistics in economic research. To reach this aim, it is necessary to identify the tasks of economic management, which requires the use of statistical methods. In particular, it is necessary to consider peculiarities of applying statistics for estimating investment attractiveness, which is one of the most important and complicated tasks in the financial market.

1. The current state of knowledge

The development of modern economic science requires high rates of scientific development. All areas of science must develop at equally high rates. For this

to happen, society needs to understand the importance of the development of each scientific discipline. Sciences are commonly divided into 2 categories – fundamental and applied. This leads to the following situation:

1) those pursuing fundamental sciences usually interpret applied sciences as the application of fundamental sciences to their own specific object of research,

2) applied scientists oversaturate their works with concepts taken from fundamental sciences fearing that otherwise they will not be considered scientific,

3) practitioners, who mainly rely on conclusions and recommendations provided by applied sciences to solve problems, cannot find proper concepts of applied sciences because of the complexity of fundamental terminology, and that is why they take the position of fundamental scientists.

And as a consequence:

1) applied scientists find it difficult to finance their research when obtained results do not have enough relevance for fundamental sciences, although, at is well-known, ideas which can be introduced in practice are made only based on results of applied research,

2) with the development of computers applied tasks can often be solved by making calculations involving various mathematical methods but without explaining the practical benefit of the received results,

3) inability to determine the practical aim of scientific research by most of fundamental scientists brings us back to the perspective of applied scientists [Leontief 1980: 126].

To confirm these theses it is quite enough to look through the works of the best known scientists in the field of the so-called "applied" science – economics. In the modern world Nobel laureates are considered to be such scientists. 65 prizes in economics have been awarded so far, but it is no use going into details of all their works. Some generalizations concerning their biographies and the form of their scientific publications justify the conclusion about the continued "priority of fundamental sciences" (i.e. mathematics) in determining the aims and tasks of economic research.

The first and the main consequence of this attitude is the desire on the part of the authors-laureates to substitute economic terms and concepts in their works with others. Very often these "new" terms are not understood even by their inventors. But the authority of famous scientists facilitates the spread of such inventions, whose meaning nobody tries to explain. One example of this can be a statement by Ragnar A.C. Frisch, who as chief editor of the journal "Econometrics", wrote in 1933 the following: "Econometrics is by no means the same as economic statistics. Nor is it identical with what we call general economic theory, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonymous with

the application of mathematics to economics.. Experience has shown that each of these three viewpoints, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient condition for a real understanding of the quantitative relations in modern economic life. It is the unification of all three that is powerful. And it is this unification that constitutes econometrics" [Frisch 1967: 134].

This long quote is included here in order to show that it is possible to use a lot of words but say nothing if there is nothing to say. Frisch, who is considered to be the author of the main concepts of econometrics, as we see, himself could not give a definition of this science. Neither could Trygve Haavelmo, who is called "the father of modern econometrics". All he could do was write that in his research he used the approach, which he called "the main law of econometrics: economic theory can be considered viable only after being checked by mathematical and statistical methods". He also writes about "statistical analysis of econometric models", and that he developed "a statistical theory for analyzing dynamic models" [Haavelmo 1954: 720].

As can be seen, the above mentioned authors, like many other Nobel laureates in economics, could not give an accurate definition or identify the aims, methods and results of their research. In many cases they used different notions to describe the same events and phenomena. For example, to analyze economic processes they suggested using such methods of analysis as mathematical, statistical, economic-mathematical, quantitative economic, econometric for the purpose of building models with corresponding names. However, they did not bother to explain the difference between them [Dovbenko 2005: 149].

The reason for this terminological mess produced by such famous scientists is quite simple: most Nobel laureates in economics were educated as mathematicians and mathematics was their first area of scientific interest! However, there have been no economists among laureates in mathematics, physics or chemistry!

Nobody doubts that the works of Nobel laureates in economics comply with the highest scientific standards. But their form is the main reason for the continued antagonism between fundamental and applied scientists: if there are few formulas, then a given study cannot be considered as scientific. It is important to say that the same laureates paid attention to the changing role of science and its results in modern society. For instance, George J. Stigler avoided using mathematics in his works giving priority to a simple literary style. This is why he gained general acceptance for accuracy and elegance of exposition and erudition. But he was an economist by education [Stigler 2000: 232].

What has been the result of excessive mathematisation of economic and statistical research? In short, the necessity of double work, as has been the case with Nobel laureates: first – the scientific (i.e. mathematical) formatting of the

results, and second – explaining the gist of the problem (its practical significance) and showing stages in which these results were achieved to practitioners of all levels.

All of what has been mentioned above concerns statistics, particularly, economic statistics. And the main reason why most practitioners interpret statistics not as a science but as a method of analysis is the creation of new "sciences" by borrowing some of their elements from statistics

For example, statistical formulas and methods used for deriving these formulas were called "mathematical statistics". Having found out that the interpretation of calculation results has a probabilistic character, the methods of probability theory were added and the new "science" was called "probability theory and mathematical statistics" [Fogel 1964: 34].

The statisticians were upset and in order not to lose the integral part of their science, in turn created a separate field of study called "the theory of statistics". As time went on, more subfields were created: "the theory of statistics" was divided into "descriptive" and "analytical". And based on the fact that economic phenomena and processes can be regarded both as functional and stochastic, "economic analysis" and "econometrics" were invented. But it did not stop there: different kinds of statistics appeared to cater for the needs of every branch and type of economic activity: economic, demographic, social, international and so on. Including the science of "simulation and forecasting", also separate from "Statistics".

As a result, "economics" does not consider statistics to be a science, but only a collection of mathematical methods for carrying out economic calculations. And in economic universities teach the subject called "mathematical economics", instead of "statistics" in the wide scientific sense of this word. And this despite the fact that all the concepts of this subject are based on the use of statistical indices and statistical methods.

Oversaturation of economic publications with mathematics in the 21st century is the legacy of the 19th and 20th centuries. In earlier times, the universalism of knowledge and skills was appreciated at all levels, not only in the scientific community. In the 21st century, this is no longer the case, since specialization gives better effects than an attempt to know everything and learn everything. And the manager (of any business, territory, any type of economic activity or even country), in order to make effective managerial decisions, must not only be able to make calculations but also to define tasks for those who can analyze, simulate and forecast. But if these three "sciences", as is the case in universities and in scientific publications, suggest that an economic process should be considered from their own point of view, then it is very difficult for them to be favorably received be practitioners [Stiglic 2003: 248].

And it is impossible to change the practice, established over 80 years, of dividing statistics into parts. But it can be done and must be done. Otherwise, in some time, the term "statistics" will go out of use among practitioners and will be transferred from the category of "applied sciencs" to the category of "fundamental sciences". But originally, "statistics" was created as exclusively "applied science".

2. The research procedure

In order to strengthen the status of statistics as a major science (and the role of statisticians) in ensuring the efficient management of economics, statisticians should define and use in their research two basic concepts – the aim and the tasks.

As a single science which "branched off", statistics never defined its object of study. The need to use statistics while researching those "mass phenomena and processes", investigated by modern economics at micro-, mezzo- and macro-level, requires no special proof. Simply speaking, "What is the benefit of using statistics in management?". "The benefit", in other words ,"the ultimate result", which the activity of an economic entity is designed to attain, is an growth in profit, which is achieved by increasing the efficiency of this activity. This means that the main task of the manager of an economic entity is to ensure that this aim is attained [Stiglic 2003: 58].

For that it is necessary:

– to estimate the results of the past work,

– to compare the extent to which specific factors contributed to reaching the result,

– to determine the future level of these factors taking into account investments which the entity can make,

– to calculate the forecasted level of profit using the forecasted level of the factors,

– to create two versions of the forecast: pessimistic and optimistic,

– all estimations and calculations should be done by persons responsible for the fulfillment of each solution.

It is impossible to make well-grounded decisions concerning each step without using statistics, which will be used to:

– collect necessary data,

– prepare them for analysis,

– carry out the analysis,

– draw conclusions about what happened in the past,

– create an information base for building and correcting the models of the entity's activity,

– calculate the forecasts.

Starting with the aims and tasks of users of statistics – managers of economic entities – one should proceed to defining the aim and tasks of statistics for economics:

1) the aim – to constantly update the information base, which is necessary for making effective managerial decisions,

2) the tasks – to provide answers to the following questions:

– What happened to the entity?

– Why did it happen?

– To what extent did certain employees contribute to what happened?

– To what extent will financial resources be required to change the level of the factors?

– What change in profit can be expected?

The aim and the above mentioned tasks can be achieved by performing well-known stages of statistical research – statistical observation, data processing and analysis, formulating the results, simulating and forecasting the process and phenomena. That is why, if statistical publications, in addition to statistical terminology, include applied economic terminology, most users – managers and analysts – will easily understand the advantages of the integrated use of one statistical science, instead of dozens of its derivations. These advantages will become apparent while formulating the tasks, which demand special statistic calculations to produce solutions which help to make effective managerial decisions [Tinbergen 1967: 293].

The banking system, in all its complexity – multiplicity and diversity of bank institutions, and the functions which they perform – generates a large flow of information. This information is used by bank institutions, their customers and partners in the country and abroad – in other words, by the world financial--economic system.

The demand for information is caused by the presence of risks in bank activities and the desire of those who use bank services to lessen the level of risk involved. A lot of scientists and practitioners have dedicated thousands of studies to the estimation of risks involved in banking activities. All of these studies make use of statistical methods. Because it is impossible to describe a bank institution, its operation and its customers using only 1 or 2 indices, methods of estimating risk proposed by the majority of these authors are cumbrous and unreliable.

The result of calculating the bank's ability to incur a risk-related loss consists in determining the limits both for the portfolio and for every item separately, while the probability of risk is calculated to determine a risk premium.

In banking practice, risks are calculated from two perspectives:

– the calculation of the bank's ability incur a risk-related loss without breaking its financial stability and reliability,

– the calculation of risk probability, which can be used to determine the expected profit of the bank, under conditions of responsible risk.

The wave of bank failures in 1990s prompted some agencies of bank supervision to strengthen their control over risks incurred by banks. The main tool of such control that the Basel Committee on Banking Supervision recommended in 1995 to central banks of the world was to use VaR-methodology for calculating the reserves necessary for covering possible risk-related losses.

The scope of application of the VaR-methodology is wide and includes various aspects. It is used as a tool:

– for inherent monitoring of the risks within the bank,

– for supervision by the central bank of capital adequacy necessary to cover risks (outward monitoring),

– for making decisions about the expediency of hedging risk operations (the comparison is done using VaR before the hedge and after it is carried out. If the difference between VaR before and after the hedge is not considerable, then the expediency of the hedging is doubtful),

– for determining the limits for dealers of the bank and for controlling the observance of these limits,

– for estimating different bank projects,

– for determining the efficiency of using the bank capital taking into account the risks,

– for estimating the efficiency of the bank's activity both as a whole and in each of its branches,

– for motivating bank dealers, as their reward is defined by taking into account the amount of received income from the bank's operations carried per one unit of VaR.

In comparison with other methods of risk calculation, the VaR methodology has some advantages:

– its wide scope, which means that it is possible to calculate risks for different markets including those characterised by high changeability, namely the markets of CIS,

– universality, which means that the risk is calculated not only for one position, but for the whole portfolio of the bank,

– simplicity of applying methods based on VaR-methodology,

– convenient form of user information: using one number expressing monetary value, it is possible to quantify the risk in the form of the maximum possible losses of the bank capital,

– takes into consideration the volatility of market securities, the value of risk position and the period of its supporting.

The VaR-methodology has also some disadvantages, in particular:

– it doesn't secure the accuracy of the result, which leads to the insufficiently correct estimation of the bank's risks, because calculations of VaR are based on the assumption of the normal distribution of random variables. In practice, it does not always hold true,

– it does not give indication as to the specific amount of risk-related losses,

– it is accompanied by rather big financial expenses, as it requires highly qualified staff and computerization of calculation procedures.

As is well-known, three main components need to be taken into account when calculating VaR:

– content and size of the bank's portfolio (risk position),

– time period for which it is calculated,

– the function of the distribution of the risk parameters (in the case of calculating the VaR-portfolio – the distribution of its current income).

Studies of calculation procedures in some banks have revealed a number of problems, in particular:

1. VaR for the bank's investment portfolio cannot be calculated because market prices of shares of most companies are not available.

2. There are some technical difficulties in calculations, when a bank's portfolio contains more then ten different financial tools, which requires the of use of a big correlation matrix. This poses a big problem in the conditions of partial automation [Sushko 2010: 10-17].

At present the banking system of Ukraine is in the stage of formation. The lack of knowledge and experience as well as corresponding normative demands on the part of the National Bank of Ukraine are the main reasons why the majority of banks do estimate risks.

The study of the practice of calculating risks by banks in Ukraine conducted by the author of this article indicates that in most cases the bank's ability to incur risk-related losses is calculated empirically, while risk probabilities are not calculated at all.

There are following reasons for this situation:

First – the instable macroeconomic situation in Ukraine, which makes it difficult to create an objective informational base that could serve as the basis for estimating and calculating risks, for predicting their occurrence and for determining their influence on banking activity.

Second – a very low level of methodological expertise and informational resources required for calculating bank risks. Taking into account the level of methodological expertise and practical experience in the estima-

tion and calculation of risks, the banks of Ukraine can be divided into three groups:

1. Subsidiaries of established foreign banks. It is necessary to stress that these banks have definite experience in the calculation of risks, but their experience was not elucidated through mass media.

2. Big banks, which are engaged in the process of creating a system of risk management using modern tools for calculating risks. Of the whole repertoire of modern methods of risk calculation (historical simulation, method of Monte-Carlo, test simulation, analytical method and so on) only the analytical method is sometimes used.

3. Mid-sized and small banks. These banks account for approximately 86% of all banks. The specific feature of this group is the lack of methodological expertise, informational resources and staff to carry out risk calculation. That is why they practically do not conduct quantitative estimation of risks using VaR-methodology.

Third – an insufficient level of IT technologies in banks, above all as far as software is concerned, which makes it impossible to completely automatize the process of risk calculation. This prevents the use of modern approaches to risk calculation in Ukraine. The main reason for this situation, in our estimation, is the high cost of software for risk calculation in relation to the level of profits earned by the banks of Ukraine, and also the insufficient economic advantage of introducing calculation procedures relative to the cost of their elaboration [Sytnikova, Hominich 2009: 17-25].

In some cases, multi-regression analysis can be used to determine the influence of various factors on the generalized estimation of risk. This gives users of bank services the possibility to compare the risk and profitability of specific bank operations and choose a given bank, to fix the cost of a specific bank operation, to prioritize a decision concerning a given bank branch and so on.

This explains the wide spread of comparative analysis for estimating risks of bank operations or risks faced by users of bank services and choosing ways of reducing them. In modern banking such comparisons are necessary:

– to help investors choose objects of investment,
– to help bank fix individual credit rates,
– to help clients choose a bank for cooperation,
– to estimate the financial conditions of the bank's branches, and so on.

Of course, comparisons are simultaneously made according to several indices. The main problem is to define the list of indices used for such comparisons. But this problem is not considered in this article.

Let's assume that a set of indices has been selected, then the problem is to determine the relative contribution of specific elements on the basis of these indices. The difficulty is that often the indices cannot be used to draw a simple

conclusion: values of some indices indicate that a given element should be included in the "best" group, while according to the values of others – it should be assigned to the "worst" group.

The generally known solution in this case is to calculate a multidimensional mean. The methods and peculiarities of its calculation are discussed in many scientific publications. But one of the most important questions about the grounds for the accuracy of conclusions reached in this way hasn't been discussed in applied statistical studies so far. It should be stressed that this observation refers only to applied studies, because the theory of statistics contains the description of the answer to this question. The separation of theory from practice in this case once more shows the distrust of economists-analysts to statistics and their inability (and very often – unwillingness) to operate with precise data [Hadjijev 2008: 23-31].

3. The methods of research

This article deals with the kernel of the problem of improving the precision of comparative analysis and its solution. A banker will say that the choice of the best (worst) branch must be done based on the "profitability of the assets", which is a general statement. But he must also take into consideration other indices, which contribute to the "profitability of assets" [Dovbenko 2005: 249].

For example:

– profits per one employee,
– profitability of credits,
– number of clients per one employee,
– proportion of interest-bearing income in the total sum.

To calculate multidimensional mean the following formula will be used:

$$\overline{P_i} = \frac{\sum p_{ik} \quad d_k}{k} \tag{1}$$

where:

p_{ik} – standardized values of original indices,
d_k – weights (the level of influence) of these indices,
k – number of indices.

The accuracy of the result depends, first of all, on substantiating the weights used for calculating $\overline{P_i}$. In practice, this is commonly done by relying on the opinion of experts.

Insufficient accuracy of determining the weights by relying on expert estimation is caused by the following:

1) the concepts such as "opinion" and "expert" are not statistical because they don't have generally accepted definitions,

2) an expert can determine the difference between the weights only in the form of conditional "unity", which has not got an economic interpretation,

3) if the conditions in which the event takes place change, another expert estimation is necessary.

At the same time the theory of statistics offers a very simple method of precisely determining the contribution of every factor in the variation of the result, which can be easily implemented using a computer.

As is known, the coefficient of multiple determination is given by the following formula:

$$R^2_{y12} = \frac{\sum (Y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \tag{2}$$

where:

Y_i denotes theoretical values of the indices calculated according to the equation of regression.

It is also known that in the case of correlation, for example, between three variables:

$$Y_i - \bar{y} = b_1(x_1 - \overline{x_1}) + b_2(x_2 - \overline{x_2}) \tag{3}$$

where:

b_k – the coefficients of regression.

Taking into account the formulas for calculating variances s_1^2 and s_2^2 , and also s_{12}, and substituting (2) with (1) we receive:

$$R^2_{y.12} = \frac{b_1^2 s_1^2 + 2b_1 b_2 s_{12} + b_2^2 s_2^2}{s_y^2} \tag{4}$$

If, instead of using original data, their standardized forms are used to derive the regression equation:

$$y' = \frac{y - \bar{y}}{s_y} \qquad\qquad x'_k = \frac{x_k - \overline{x_k}}{s_k} \tag{5, 6}$$

where:

s_y and s_k – standardized deviations

then the coefficient of regression b'_k will correlate with b_k in the following way:

$$b'_k = b_k \frac{s_k}{s_y} \tag{7}$$

Then:

$$R^2_{y12} = (b'_1) + (b'_2) + b'_1 b'_2 r_{12} \tag{8}$$

from this:

$$R^2_{y12} = R_{y1} b'_1 + R_{y2} b'_2 \tag{9}$$

and summarizing:

$$R^2_{y12...k} = R_{y1} b'_1 + R_{y2} b'_2 + ... + R_{yk} b'_k \tag{10}$$

where:

r_{yk} – bivariate correlation coefficient.

The sense of (7) lies in the following:

R^2_{y12} measures the influence of all x_k;

R_{yk} measures the influence of x_k on y including the indirect influence of other variables, if these other variables influence x_k: b'_k measures in terms of standardized deviations of y' the "clean" influence of x_k on y.

So, by multiplying r_{yk} by b'_k influence of x_k on y is corrected (the indirect influence of other factors is removed). It means that $r_{yk}\, b'_k$ become similar to the partial coefficient of determination, but, additionally, has its own advantages:

– it is much easier to calculate,

– it is always positive,

– it is additive.

Just this last property (7) makes it useful for determining weights in the calculation of $\overline{P_i}$:

$$d_k = \frac{r_{yk}\, b'_k}{R^2_{y12\ ...\ k}} \tag{11}$$

The above mentioned approach to banking research raises considerably the quality and relability of the results of the research.

Conclusions

In summary, the following conclusions can by made:

– correlation – regression analysis can be used to determine the individual contribution of each index to the estimation of the results of activities of branches,

– multidimensional mean can be used to account for the influence of each index in the form of one multidimensional index,

– the ranking of branches made on the basis of correlation – regression analysis provides an objective comparative estimation of the risks of activities conducted by the branches and helps to identify those most urgently in need of an intervention from the bank's top management in order to prevent a deterioration of the economic situation of the bank as a whole.

References

Dovbenko M.V., 2005, *Modern Economic Theory: Economic Nobeologie*, Kyiv: Academy.
Frisch R.A.C., 1967, Propagation Problems and Impulse Problem in Dynamic Economics, in: *Economic Essays in Honour of Gustav Cassel*, London: Taylor & Francis Group.
Fogel R.W., 1964, *Railroads and American Economic Growth*, Baltimore: John Hopkins Press.
Hadjijev F.V., 2008, Principles of Organization of Control of Current Risk of Banks, Kyiv, *Finances and Credit*, No. 8: 23-31.
Haavelmo T., 1954, *A Study in the Theory of Economic Evolution*, Amsterdam: North-Holland Publishing Company.
Leontief V.V., 1980, *Economic Essay*, Moskow: Politizdat.
Stigler G.J., 2000, *Monuments of Economic Idea*, Sankt Petersburg: Economic School.
Stiglic J., 2003, *Globalization and Its Burden*, Kyiv: Academy.
Sushko V.V., 2010, A Risk Management in Activity of Commercial Banks on a Base of VaR – Technologies, Kyiv, *Visnyk NBU*, No. 3: 10-17.
Sytnikova N.A., Hominich I.D., 2009, Revolution in a Risk Management, Kyiv, *Bank Services*, No. 32: 17-25.
Tinbergen J., 1967, *Mathematical Models in Economic Researches*, Moscow: Progress.

Rola statystyki w badaniach bankowości

Streszczenie. *W artykule zostało rozpatrzone miejsce i rola statystyki w badaniach ekonomicznych. Zaprezentowano krótki przegląd kształtowania się statystyki ekonomicznej i problemów, które są rozwiązywane za jej pomocą. Bardziej dokładnie omówiono stosowanie statystycznych metod w ocenie ryzyka działalności bankowej.*

Słowa kluczowe: *zadania badania, statystyka stosowana, metody statystyczne, ryzyko działalności bankowej, statystyka bankowa*