GUS

Indeks 371262
e-ISSN 2657-9545
ISSN 0033-2372

# Przegląd Statystyczny

# Statistical Review

Vol. 73 No. 1 2026

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

# INFORMATION FOR AUTHORS

*Przegląd Statystyczny. Statistical Review* publishes original research papers on theoretical and empirical topics in statistics, econometrics, mathematical economics, operational research, decision science and data analysis. The manuscripts considered for publication should significantly contribute to the theoretical aspects of the aforementioned fields or shed new light on the practical applications of these aspects. Manuscripts presenting important results of research projects are particularly welcome. Review papers, shorter papers reporting on major conferences in the field, and reviews of seminal monographs are eligible for publication based on the Editor-in-Chief's decision.

Since 1st May 2019, the journal has been publishing articles in English.

Any spelling style is acceptable as long as it is consistent within the manuscript.

All works should be submitted to the journal through the Editorial System (https://www.editorialsystem.com/pst).

For details of the submission process and editorial requirements, please visit https://ps.stat.gov.pl/ForAuthors.

# Przegląd Statystyczny
# Statistical Review

Vol. 73 No. 1 2026

**EDITORIAL OFFICE ADDRESS**

Statistics Poland (GUS), Al. Niepodległości 208, 00-925 Warsaw, Poland

Language editing: Scientific Publications Division, Analyses and Dissemination Department, Statistics Poland
Technical editing and typesetting: Statistical Publishing Establishment, Statistical Computing Center – team supervised by Mariusz Męcina

Centrum Informatyki Statystycznej

Printed and bound by: Statistical Computing Center
Al. Niepodległości 208, 00-925 Warsaw, Poland, cis.stat.gov.pl

**Website: ps.stat.gov.pl**



**ISSN 0033-2372**
**e-ISSN 2657-9545**
**Index 371262**

Information on the sales of the journal: Statistical Computing Center
Phone no.: +48 22 608 32 10, +48 22 608 38 10

Order no. 136/2026

# CONTENTS

# LETTER FROM THE EDITOR

Dear Readers,

It is with great pleasure that we present to you the first issue of *Przegląd Statystyczny. Statistical Review* in 2026. Reflection on quantitative research and its role in an increasingly complex and rapidly changing world remains an important part of our editorial perspective.

This past year has once again highlighted the growing importance of statistics as a fundamental tool for analysing social, economic and technological phenomena. In an era of expanding data availability and rapidly developing analytical methods, issues such as data quality, methodological transparency and research integrity have become more critical than ever. Our journal remains committed to serving as a platform for the exchange of scientific ideas, supporting both methodological advancements and their practical applications across disciplines.

We warmly invite researchers to submit their manuscripts to *Przegląd Statystyczny. Statistical Review*. We particularly welcome contributions that offer innovative methodological approaches as well as high-quality empirical studies that advance quantitative knowledge and its applications.

At the same time, we would like to express our sincere gratitude to all Authors and Reviewers for their invaluable contributions, dedication and commitment to maintaining the high academic standards of our journal. Your work is essential to its continued development and reputation.

I would like to invite you to explore the current issue and hope it offers valuable and inspiring insights.

On behalf of the Editorial Board,
Krzysztof Echaust
Editor-in-Chief
*Przegląd Statystyczny. Statistical Review*

Przegląd Statystyczny. Statistical Review, 2026, vol. 73, 1, 1–29 https://doi.org/10.59139/ps.2026.01.1
Received: 16.02.2026, revised: 31.03.2026, accepted: 31.03.2026

# Stochasticity, non-linearity and non-parametricity: statistical challenges in approximating socio-economic processes

Józef Pociecha[a]

**Abstract.** The aim of this article is to indicate the relevance of the challenges faced by statisticians in the analysis of socio-economic processes, which requires breaking many of the well-trodden paths of analysis. Until now a preference for linear, parametric methods was observed, supposedly allowing the discovery of cause-and-effect relationships. According to Kant's philosophy, 'a thing in itself' is not knowable and from this perspective, the main task of statistics is to approximate reality. In statistical studies, only stochastic approximation is made and determining the stochastic structure of processes is as important as studying their regression.

The presented review of the literature leads to the conclusion that stochasticity and non-linearity are the primary features of socio-economic processes and that nowadays their analysis is most effective when based on non-parametric methods. Thus, the paper presents a basic catalogue of methods used for studying non-linearity, the stochastic character and approximation of economic processes using non-parametric methods.

An additional aim of this paper is to emphasize the importance of fundamental statistical monographs to the development of statistical research methodology. The contribution of Polish publications to the advancement of contemporary knowledge is also discussed.

**Keywords:** approximation approach, stochastic approach, non-linear methods, non-parametric methods

**JEL:** C00, C10, C14

## 1. Introduction

The purpose of statistical research is to provide numerical characteristics of the reality surrounding us. When we conduct such research, it is also necessary to reflect on its philosophical and methodological foundations. An important question thus arises whether it is possible to arrive at the truth by statistical means. Statisticians believe that it is possible (Pociecha, 2023). Generally speaking, statistics is a method of learning from experience and making decisions under conditions of uncertainty. In this view, the truth can be understood as a judgment, with an acceptable probability (error), consistent with reality. Methods of statistical inference constitute one of the tools used for finding the truth in statistical terms.

[a] Krakow University of Economics, College of Economics, Finance and Law, Institute of Quantitative Methods in Social Sciences, Department of Statistics, ul. Rakowicka 27, 31–510 Kraków, Poland, e-mail: pociecha@uek.krakow.pl, ORCID: https://orcid.org/0000-0003-3140-481X.

Two outstanding mathematicians and statisticians of the 20th century, Richard von Mises and C. Radhakrishna Rao, emphasised that mathematical statistics is the universal way of arriving at the truth understood in probabilistic terms. In his book, *Probability, Statistics and Truth* (von Mises, 1957), von Mises discusses the key issues of probability calculus and mathematical statistics in the light of the foundations of mathematics and the philosophy of science. He discusses the definition and essence of randomness and considers the definitions of probability, starting from the classical one, through the geometric, subjective concept of probability, to the theory of frequentist interpretation of probability. He then asks the question: What is statistics? The author then proceeds to offer comprehensive possible answers to it. Von Mises also pays attention to exploring the nature of causality. He perceives the transition from probability to knowledge of the real world as follows: 'Owing the original relation between the basic concepts and the observed primary phenomena, this theoretical structure permits us to draw conclusions concerning the world of reality. In order to allow a rationally justified application of this probability theory to reality, a quantitative probability concept must be defined in terms of potentially unlimited sequences of observations or experiments' (von Mises, 1957, pp. 7–8).

Rao, likewise, sees statistics as a tool for discovering the truth. In his book, *Statistics and Truth* (Rao, 1997), the author uses examples from the natural sciences to show the scientific shift from determinism to probabilism. Quoting Max Born's thought, he writes: 'Today, the order has been reversed: chance has become the primary notion, mechanics an expression of its quantitative laws, and the overwhelming evidence of causality with all its attributes in the realm of ordinary experience is satisfactorily explained by the statistical laws of large numbers' (Rao, 1997, p. 20). Throughout the book, Rao emphasizes the view that statistics has become a tool for finding the truth in modern scientific research.

In addition to the views of eminent mathematicians and statisticians cited above, what inspired this article is a statement by Andrew Briggs, one of the pioneers of quantum technologies, professor at the University of Oxford. In an interview published in Poland, he states that 'life is nonlinear and stochastic' (Briggs et al., 2025, p. 46), which is developed in the book *Human Flourishing: Scientific insight and spiritual wisdom in uncertain times* (Briggs & Reiss, 2021). The authors emphasize that the statement 'life is nonlinear and stochastic' reflects the fact that life events and processes do not follow a straight, predictable path, but are influenced by chance and complex disproportionate relationships. Biological systems and individual human experiences exhibit non-linear dynamics, where small changes can lead to large, disproportionate effects. These systems and experiences are subject to stochasticity characterised by an inherent randomness and probability of events occurring over time.

Here, it should be noted that the full understanding of the reality around us is not possible. In philosophical terms, this was most fully expressed by Immanuel Kant, who formulated the concept of 'the thing in itself' as a reality that exists, but although inaccessible to human consciousness, its presence is manifested through phenomena that are the source of impressions (Tatarkiewicz, 1978). We can observe these phenomena with greater or lesser accuracy, i.e. we are only able to approximate the reality under study. As previously mentioned, socio-economic processes are stochastic in nature, therefore, the stochastic approximation method is the most general statistical method for searching for the truth about the reality.

The aim of this article is to recall the philosophical and methodological foundations of conducting research and statistical analyses and, in this context, to explore the challenges faced by those conducting such type of research. The most important contemporary challenges facing statistics are as follows:

- The shift from discovering the regularity of the course of socio-economic processes to analysing the stochastic structure of the models constructed to represent them. This will allow a more precise determination of the uncertainty about their future course and thereby, the risk of developmental anomalies;
- Emphasizing that statistical or econometric modelling is only a form of approximation of the studied socio-economic reality, where the acceptable accuracy of the approximation is up to us;
- Assuming that non-linearity is an essential feature of the course of socio-economic processes. Their linear course is only a rare case, and not a reference point for conducting empirical research. Therefore, the basic methods of statistical analysis should be non-linear methods;
- Difficulty in modelling non-linear processes using parametric methods, as it requires defining the analytical form of the non-linear functions and determining the number of parameters involved, resulting in their computational complexity. Thus, non-parametric statistical modelling is more useful;
- The use of statistical learning methods for non-parametric process modelling.

The secondary aim of the work focuses on Polish statistical and economic thought and presents works that made a significant contribution to the world achievements in the field of statistical methods and were published in Polish. The aforementioned challenges will be discussed along with an indication of the recent research achievements in this area and suggestions for the application of the statistical learning paradigm in conducting socio-economic analyses.

## 2. Cause and chance

Causality is a philosophical category that describes the universal and necessary relationship between phenomena that take place in the objective reality, where one factor, which is the cause, leads to the occurrence of another factor, which is its effect. Causality describes the relationship between variables where a change in the cause leads to a change in the effect. The principle of causality is a fundamental philosophical and scientific concept proclaiming that every phenomenon has its cause and its effect, which describes the universal interdependence of phenomena in the surrounding reality. According to this principle, phenomena are related to each other in a causal relationship, where the cause precedes the effect as a set of conditions necessary and sufficient for its occurrence (Jadacki, 2020). Aristotle had formulated criteria essential to establishing a cause-and-effect relationship (Aristotle's theory of four causes). These conditions are: logical temporal order, correlation, lack of alternative explanations (elimination of third factors) and a reliable causal mechanism (Nowakowski, 2017).

The philosophical concept which assumes that all events are inevitable and logical consequences of those preceding them is determinism. According to this idea, each state of the universe is shaped by its previous states and laws of nature, which excludes the influence of chance and free will. There are many types of determinism. The most important are: physical determinism, biological determinism and social determinism. According to physical determinism, all events and states of the universe are predetermined by previous causes and immutable laws of nature. Thus, any future is the only possible one when we know the initial state. According to biological determinism, human behaviours and traits are largely or entirely determined by genes and physiology. Social determinism assumes that an individual's behaviour and development are largely shaped by social factors such as the environment, culture, social class and interactions with other people.

In the history of human thought, determinism was the original philosophical position, according to which all events in nature are predetermined. In antiquity, the most famous supporter of deterministic causality was Democritus, known primarily as the creator of the ancient theory of atomistic matter. Atomists tried to explain the phenomena of the surrounding world through a causal framework. They claimed that nothing happens without a reason, but everything happens for some reason and out of necessity. They understood the nature of causes in a material and mechanical way.

Determinism can take two forms, depending on the degree of determination: extreme and moderate (Pociecha, 2023). In the history of philosophy, the most extreme deterministic view was that of Laplace's 'mathematical demon'. As a spirit endowed with unlimited mathematical deduction, it could predict all future events knowing all the quantities that characterise the present state. In moderate determinism, on the other hand, certain restrictions are imposed on the functioning of cause-and-effect relationships. An example of such a restriction is the view that determinism is basically an intellectual construction, while the development of nature, which is spontaneous and creative, is conditioned by internal force and life momentum, inhibited by inert matter (Lemańska, 1998).

In philosophy, necessity is opposed to contingency. The word ‘chance’ refers to an unpredictable event or situation. In statistics, instead of the term chance, the term ‘randomness’ is used. It describes a situation in which we are not able to give a purpose, cause, order or predictable course of a given phenomenon. A random process, on the other hand, is a process whose results cannot be accurately predicted, but can only be described by its distribution. The mathematical equivalent of the concept of chance is a random event. In the calculus of probability, random event A is defined as a measurable subset of a set of elementary events Ω of a given random experience. Each subset contains single or any number of elements, called elementary events. The measurability requirement implies that possible events must form a sigma-body on Ω.

A philosophical position that recognises the objective existence of chance is indeterminism. It assumes that the same causes do not necessarily lead to the same effects. Thus, it makes predicting later phenomena on the basis of earlier ones impossible and negates the strict conditioning of all phenomena. It is based on the common observation that random events are common in the world around us. Science nowadays recognises physical, biological and social indeterminism.

Newton’s classical physics, or classical mechanics, was an example of a deterministic understanding of the world. However, in as early as the second half of the 19th century, an Austrian physicist, Ludwig Boltzmann, provided the foundations of statistical mechanics, giving a statistical explanation of the second law of thermodynamics. Werner Heisenberg, a German physicist and philosopher of science, co-creator of the theory of quantum mechanics, winner of the Nobel Prize in physics, formulated the uncertainty principle. It states that there are pairs of quantities that cannot be measured with arbitrary accuracy at the same time. The act of measuring one quantity affects the system so that some of the information about the other quantity is lost. Heisenberg’s uncertainty principle does not arise from the imperfection of measurement methods or instruments, but from the very nature of reality. Heisenberg proclaimed that in a world of the smallest material particles there is no causal conditioning; he pointed to the ‘freedom of the will of the electron’. This position has shifted to the realm of philosophy, proving the existence of indeterminism in nature. Thus, the deterministic approach, which was the original philosophical understanding of the world, has been replaced by its indeterministic counterpart, which is now the dominant modern scientific view.

In biology, the theory of evolution is based on indeterministic behaviour, which explains the observed diversity of life through the accumulation of random mutations in the gene pool of populations that have not been in contact with each other. Influenced by various selection pressures, a different set of traits is selected from random mutations and, as a result, over time, populations become less and less similar to each other.

Social indeterminism is a philosophical and sociological view that denies the existence of strict, necessary laws governing social phenomena, which means that events in society are not strictly determined by the conditions that precede them. He rejects the idea that society is subject to the laws of determinism and instead emphasizes the role of chance, free will and other factors in shaping human behaviour and societies. Works exposing the indeterministic factor in the social sciences are puzzling. These include the work *Dynamics and Indeterminism in Developmental and Social Processes* by Fogel et al. (2014), in which the authors critically analyse the role of indeterminism in human psychological and behavioural development. They review the concepts of indeterminism and determinism in terms of dynamic systems thinking. They apply these general ideas to non-verbal communication systems, emphasizing the vagueness inherent in symbols and the creation of meanings in social systems. They also discuss indeterministic processes occurring within the individual, related to emotional, social and cognitive development.

Economic processes are also indeterministic in nature. Indeterminism in the economy assumes that economic events are not strictly determined by previous causes and their course is influenced by many unpredictable factors, including the free will to make decisions by the market participants. This means that even under identical initial conditions, different economic outcomes are possible. Intriguing are the works *The Concept of Indeterminism and Its Applications: Economics, Social Systems, Ethics, Artificial Intelligence, and Aesthetics* (Katsenelinboigen, 1997) and *Indeterministic Economics* (Katsenelinboigen, 1992) by Aron Katsenelinboigen, a professor at Wharton School, University of Pennsylvania. In these studies, the author uses the concept of indeterminism to analyse various systems, including economics, social, ethics, artificial intelligence, and aesthetics systems. Katsenelinboigen argues that indeterminism is a fundamental aspect of complex systems, separate from uncertainty and requiring specialist understanding. He promotes his own 'theory of predisposition', which assumes that indeterminism can be understood as a tendency of a system to evolve in a certain way. An interesting book, written in Polish, on the role of indeterminism in economics is by Leszek J. Jasiński, titled *Analysis and Interpretation of Economic Research* (Jasiński, 2017), where the author focuses on the role of induction in economic research, the indeterministic nature of economic processes, the ways of overcoming uncertainty and the functioning paradigms in economics.

All the works cited here indicate the need to emphasize the stochastic character of socio-economic processes. When conducting statistical modelling of such processes, we should not only strive to discover the regularities of their course, but characterise the stochastic range of their variability as well. Equally important is to measure the uncertainty of economic processes resulting in uncertainty about future events or the outcomes of actions, inherently present in the economy. Uncertainty is a key element influencing decision-making and a prerequisite for determining risk. However, risk is a broader concept, encompassing situations in which we know the potential effects of our actions and the probability of such situations occurring.

In light of the statements above, the challenge facing statistics involves treating uncertainty and the related risk measurement as precisely as possible, while uncovering the patterns of past processes and determining their expected course in the future. From a mathematical point of view, this is a problem of optimising the stochastic approximation of the studied socio-economic reality.

## 3. Approximation as a method of achieving the cognitive goal of science

The main function of science is to acquire knowledge about the surrounding reality, and the cognitive goals and values indicated in the philosophy and methodology of science should serve this function. Witold Strawiński defines the goals of science as follows: 'The main social function of science is the cognitive function. Thus, the main goals and values pursued in scientific research are primarily cognitive goals and values. Other functions of science – educational, innovative, 'emancipatory' – are derived from its main function' (Strawiński, 2011, p. 323).

The goal of scientific cognition is to obtain knowledge that is general, yet precise and simple. In terms of cognition, scientific knowledge is the best type of knowledge, which most adequately describes reality. Science owes its high cognitive status to the methods and language it uses. The use of scientific language and the application of generally accepted scientific rules determine whether or not a given statement is of a scientific nature. Consequently, the acquisition of knowledge about the reality that surrounds us is governed by the rules of scientific methodology, which is itself a metascience (Grobler, 2006). There are many typologies and classifications of scientific disciplines known in the literature (see, for example, Kamiński, 1981).

The essence and paths of scientific cognition are presented in more detail, e.g. in the work by Bogdanienko (2018).

The cognitive goals pursued in scientific research depend on what area of knowledge a given scientific discipline represents. The sciences are generally divided into two basic classes: theoretical sciences and empirical sciences. Empirical sciences, based on the principle of a posterior inference, include the social sciences and economics (Sagan, 2016). It should be noted, however, that economics also uses elements of *a priori* inference, an example of which is mathematical economics. An exhaustive overview of the issues related to the philosophy and methodology of economic sciences was compiled by Burnewicz (2021). Their cognitive function consists in identifying, describing and explaining economic phenomena and processes on a macro and micro scale, as well as in predicting their course (Kuciński, 2010). The position and specificity of the humanities and socio-economic sciences in the entire typology of sciences is emphasized in the book by Sosenko (2008).

The basic tool of quantitative economic research is the econometric model. Lawrence R. Klein, the winner of the 1980 Nobel Prize in Economics, defines the econometric model as follows: 'A model is a schematic simplification, omitting irrelevant aspects in order to explain the inner workings, form, or construction of a more complex mechanism. Social systems are extremely complex, sometimes so complex that we are not able to fully explain all aspects of them at once. In such cases, we break the problem down into parts, but sometimes even that is not enough to fully understand it. That is why modelling is an important stage. The social model consists of simplistic assumptions, approximate but understandable dependencies, and a certain explanation of reality. It is not reality, but only a simplified image of reality that man is able to understand' (Klein, 1982, p. 15).

However, Professor Zbigniew Czerwiński asks further questions and proposes some answers to them. 'Is a model a description of facts, a stated regularity (of what degree of generality?), the formulation of a hypothesis or something else? It depends on how you understand the model. Regardless of what definition we take, it will not be harmful to econometrics to say that it has not been able to detect universal regularities (independent of the coordinates of time and space), accurate regularities (which work within the limits of measurement error), and non-trivial regularities (beyond common knowledge). Unfortunately, this is the difference between econometrics (and economics in general) from natural sciences such as physics, chemistry, biology, although many economists wanted to imitate these sciences. In explaining this state of affairs, one can refer to various circumstances: to the extraordinary complexity of economic phenomena; the variability of the environment in which these phenomena occur; the impossibility of experimentation (the study of the course of phenomena in artificially created conditions) to the extent that it is possible in the natural sciences; stochastic nature of economic (and more broadly: social) phenomena. These circumstances may explain the difference in the degree of development of the natural and social sciences, without violating the principle of the unity of science. However, the reasons for this difference can be found in the fundamental difference between the objects studied by both sciences. There is nature, there is society, and thus ultimately man, who has certain features that do not occur in inanimate nature and are incomparably less developed in plants and animals: memory (including the memory of the effects of previous actions), assessing (positively or negatively) the situation in which he finds himself, the ability to formulate goals of action and select means to achieve goals. Due to such circumstances, there is a thesis about the fundamental difference in tasks, cognitive capabilities and methods of proceeding in the social sciences and in the natural sciences' (Czerwiński, 2002, pp. 443–444).

In light of the circumstances cited above, the question is to what extent can we achieve the cognitive goal of the socio-economic sciences. We must be content with the fact that our knowledge of socio-economic processes can only be a better or worse approximation of the surrounding socio-economic reality. Therefore, we must also formulate general criteria when to consider this approximation satisfactory. Such criteria are based on probabilistic grounds. Therefore, the most general method of cognition is the approximation method, because our cognition is always a simplification of the examined reality. The approximation of reality consists in simplifying it in such a way as to preserve all its essential features. It should be noted here that in its most general sense, it is a philosophical concept, because we cannot fully know the reality around us and only approximate it.

The concept of approximation is commonly known as a mathematical term. It means replacing 'true' function $f(x)$ by a simpler function $F(x)$, defined in the same area, whose values depend on a certain number of parameters. An approximated function, also called an approximate, is replaced by a function called an approximant. The approximant brings us closer to an unknown primary function. Approximation is used in situations where we do not know or there is no analytical form of the approximate which would allow us to determine the value for any of its arguments; at the same time, the values of this unknown function are known for a certain set of its arguments. The approximation of an approximate by a selected approximant is always associated with the risk of approximation errors. The estimated magnitude of these errors is the criterion for choosing the best approximation method. If the set on which we measure the approximation error is discrete, then we speak of point approximation; however, if a set is specified on a range of real numbers, then we speak of integral approximation.

Generalised polynomials are most often used as approximant functions, which constitute the base function $F(x)$, otherwise called the approximation base. If these are additive linear functions, then this type of approximation is called linear approximation. Numerical methods, as a branch of applied mathematics, offer many ways of approximating an unknown or undefined approximate using appropriate computational techniques, formulated as approximation algorithms. These include, for example, rational approximation, which is the quotient of two generalised polynomials sharing the same analytical form. These polynomials, however, differ in terms of their parameter values. The interpolating function takes the same values as the original function at the interpolation nodes.

This leads us to another type of approximation, i.e. interpolation, which consists in finding an interpolation function which takes the same values as the original function in the interpolation nodes and determines the approximate values of this function at points that are not nodes. Here, the simplest approach is to use polynomials as the base function. In practice, the Lagrange interpolation formulas, the Newtonian interpolation formulas or splined functions are most commonly used. A spline is an actual smooth function for which a family of subintervals exist in the domain of that function, such that the function is a polynomial on each of these intervals. In practice, splined functions of the third degree are often used (Fortuna et al., 1993).

An important approximation method is the mean-square approximation. It is used when approximated function $f(x)$ is known only on the discrete set of its $n$ arguments. We assume that $\varphi_j(x)$, $j = 1, \ldots, m$ is a system of basic functions. Then we look for a function $\varphi(x)$ that approximates the given real function and provides its smoothing, with a mean-squared error given by the following general formula:

$$R = \sum_{i=1}^{n} [\varphi(x_i) - f(x_i)]^2 \tag{1}$$

Thus, we are looking for a function $\varphi(x)$ that minimises the value of the error above. This leads to a known system of normal equations after substituting the values of the nodal points $x_i$ ($i = 1, \ldots, n$). If we take polynomials as the base functions, then we obtain a polynomial approximation. However, as the degree of the polynomial increases, the calculations become increasingly more laborious, while the results obtained more uncertain. These difficulties can be eliminated by using orthogonal polynomials in the approximation process (Fortuna et al., 1993).

In the process of approximation, a situation tends to occur that the real function $f(x)$ is a periodic function. Such a function must be approximated by trigonometric polynomials and this type of approximation is called trigonometric approximation. If $f(x)$ is a continuous function with a period of 2π, the trigonometric polynomial takes the following form:

$$Q_n = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx), \tag{2}$$

where $a_k$, $b_k$ are the trigonometric Fourier coefficients of function $f(x)$ with respect to the orthogonal system of base functions (Fortuna et al., 1993).

Deterministic approximation methods are an important segment of numerical methods, widely used in practice. Their mathematical formalisation, detailed description and computational algorithms can be found e.g. in Kordecki and Selwat (2020). Approximation methods are also used to design deterministic self-learning systems (Wawrzyński, 2021).

## 4. Stochastic approximation as a cognitive method for studying reality

In the classical approach to approximation, the considered variables are not random variables. The actual approximated function is not known, but only its values in a finite number of points, usually called nodal points. However, if the considered variable is treated as a random variable, then its approximation is called stochastic approximation. A random variable is a function that assigns numbers to elementary events, i.e. a function that allows for a mapping that transfers probability studies from an inconvenient probabilistic space to a well-known Euclidean space. In stochastic approximation, the approximation error is also random.

The concept of stochastic approximation was introduced in the early 1950s by American mathematicians, Herbert Robins and Sutton Monro (Robins & Monro, 1951). They presented an iterative method for finding the root of an unknown real function in cases where for each established value of an argument it is possible to obtain an estimate of the value of this function, but with a random error of a zero expected value. Inspired by this publication, American statisticians, Jack Kiefer and Jacob Wolfowitz (Kiefer & Wolfowitz, 1952), proposed an iterative method that would stochastically estimate the local maximum of a certain real function. Thus, two basic problems of stochastic approximation were formulated in the cited scientific articles along with a proposal of methods of solving them. Both the problems and the methods for their solving have been generalised to the multidimensional case, i.e. for finding the root of a function defined in $k$-dimensional Euclidean space and for searching for the local extrema of the real function of *k*-variables (Koronacki, 1989); these are known as RM (Robins-Monro) and KW (Kiefer-Wolfowitz) algorithms.

The stochastic approximation problems formulated above are the stochastic equivalents of deterministic problems involving the solution of systems of non-linear equations and optimisation in finite-dimensional spaces. However, what connects stochastic approximation with deterministic approximation is their iterative nature based on the successive approximations method. The RM and KW methods have gained great popularity, forming the theoretical basis for many stochastic approximation algorithms. Their formalisations and areas of application are comprehensively presented in a book by Koronacki (1989).

Soon after the idea of stochastic approximation in the field of regression research appeared in the world literature, it was transferred to Polish literature thanks to an important and yet somewhat forgotten book by Professor Zdzisław Hellwig, titled *Stochastic Approximation* (Hellwig, 1965). In the introduction to this work, the author writes that the subject of his interests is 'statistical and mathematical methods of research that can be applied in economics'. He continues stating that his paper focuses on 'multivariate random variables with particular emphasis on some aspects of the correlation and regression theory, especially important in economic applications. Some theoretical and practical shortcomings of the traditional way of studying and describing the relationship between statistical quantities prompted the author to attempt to frame the 'outline of the theory of stochastic approximation' presented in this paper. The main idea of the paper is the thesis that the regression function is nothing more than an approximation function, so in the theory of regression and correlation, the basic theorems and some calculation methods provided by the theory of approximation can be used' (Hellwig, 1965, p. 7).

In this work, the basic issue of approximation was formulated for the first time in Polish literature. Suppose we have two continuous variables X and Y, which are known to have some functional dependence, i.e.:

$$y = \psi(x), \tag{3}$$

where $\psi(x)$ is an element of space C of continuous functions. The $\psi(x)$ function is an approximate and it is unknown. In order to guess its form, an experiment is carried out through which $n$ observations of the following type are obtained:

$$(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n). \tag{4}$$

They are called experimental points or random points, because it is assumed that variables X and Y are random variables. Having experimental points, one should guess the form of function $\psi(x)$ or at least find such an $h(x)$ function that could slightly differ from $\psi(x)$ in a certain interval $[a, b]$. The $h(x)$ function is called the approximant. For the first time in the Polish literature, the quoted book provided a mathematical formalisation of the basic problem of approximation and the formalisation of the idea of stochastic approximation in the regression range. Let us assume that function $\psi(x)$ is continuous and differentiable over the entire interval $[a, b]$. Let us then consider a set of $H_k$ functions $h(x, a_1, a_2, \dots a_k)$, where $a_1, a_2, \dots a_k$, are parameters that can assume a value of zero in certain cases. If function $h(x) \in H_k$, then at each point of interval $[a, b]$, the following relation occurs:

$$|\psi(x) - h(x)| < \Delta, \tag{5}$$

where $\Delta$ is a positive real number.

This approach to the basic issue of approximation (commonly referred to as the mathematical approach to approximation) disregards the debate as to which of the two competing functions, $h_1(x)$ or $h_2(x)$ better approximates function $\psi(x)$. They are both considered approximants and indistinguishable from the point of view of the approximation criterion (5).

The mathematical formalisation of the regression problem in terms of stochastic approximation is as follows: the starting point is a certain sequence of numbers given by formula (4), which are observations of a two-dimensional random variable (X, Y). Let $a$, $b$ be two numbers, so chosen such that:

$$\int_{-\infty}^{\infty} \int_a^b f\,(x,y)\,dydx > 1 - \alpha, \tag{6}$$

where $\alpha$ is a predetermined, small number in the interval [0,1].

Let $R_k$ denote a set of functions $\psi\,(x, \alpha_1, , \alpha_2 \dots , \alpha_k)$ such that for a sufficiently large positive number $\Delta$, the following relation is satisfied:

$$P\,[|Y - \psi(x)| < \Delta] = \int_{-\infty}^{\infty} \int_a^b f\,(x,y)\,dydx > 1 - \alpha. \tag{7}$$

Set $R_k$ consists of approximated functions or regression. And the functions belonging to this set are approximates or regression functions in the population. Number $\Delta$ is called the tolerance or acceptable approximation error, while number $1 - \alpha$ is called the likelihood of the approximation.

On the other hand, if $H_k$ denotes a set of functions $h(x)$ with a number of parameters equal to $k$, such that $h(x) \in H_k$, then and only then the number of $m$ points $(x_i, y_i)$, chosen from sequence (4) and satisfying the relation:

$$|y_i - h(x_i)| < \Delta \tag{8}$$

makes up for inequality:

$$\frac{m}{n} \geq 1 - \alpha, \tag{9}$$

where $n$ is assumed to be large enough.

Set $H_k$ is called the set of approximants, and functions $h(x) \in H_k$ are called regression functions in the sample. Relation (9) implements that values $y_i$ can differ only randomly from their approximated values $h(x_i)$. Formula (7) can be expressed more conveniently in a slightly different form:

$$P[\psi(x) - \Delta < \mathrm{Y} < \psi(x) + \Delta] \geq 1 - \alpha. \tag{10}$$

This formula resembles the well-known expression for confidence intervals.
Inequality

$$\psi(x) - \Delta < \mathrm{Y} < \psi(x) + \Delta; x \in [a, b] \tag{11}$$

represents a certain area bounded by the lines $\psi(x) \mp \Delta$ and $x = a$; $x = b$. This area has the property that the probability of a random event consisting in the fact that a random point $(x,y)$ falling outside it is less than $\alpha$. This area is called the tolerance area, and its complement is the confidential area. The presented issue of stochastic approximation is not about determining the tolerance area itself, but about relating it to the concept of regression. In approximate terms, 'regression' does not mean a single approximation function, but a whole family of such functions.

The tolerance of an approximation can be a function of $x$. Then:

$$P\,[v_1(x) < \mathrm{Y} < v_2(x)] \geq 1 - \alpha\,; x \in [a, b], \tag{12}$$

where:

$$v_1(x) \equiv \psi(x) - \Delta(x);\ v_2(x) \equiv \psi(x) - \Delta(x). \tag{13}$$

This means that the tolerance range can be different for different $x$. Relation (12) can be assigned an even more general form if the following condition is imposed on functions $v_1(x)$ and $v_2(x)$:

$$v_1(x) < v_2(x). \tag{14}$$

Then the normal space Q bounded by curves $v_1(x)$ and $v_2(x)$ is called the tolerance area if:

$$P[(X, Y) \in Q] \geq 1 - \alpha. \tag{15}$$

Complementing the tolerance area is of critical importance.

To sum up, Hellwig (1965, p. 34) notes that an approach to the basic issue of approximation, where the search for the $h(x)$ function as the best approximant of $\psi(x)$ is abandoned, and the concept of the $H_k$ set with indistinguishable elements, each of which can play the role of an approximant, is generally and theoretically correct because:

1. Each given function $\psi(x)$ corresponds to an infinite $H_k$ set of functions $h(x)$, if the number of $k$ parameters occurring in these functions can be arbitrarily large.
2. In the $H_k$ set there is always an $h_0(x)$ function such that for any sequence (4),

$$\max_i \mid y_i - h_0(x_i) \mid = 0. \tag{16}$$

It is enough to assume that $h_0(x)$ is a polynomial of degree $n$ – 1, because for every sequence (4), it is possible to draw a polynomial of degree at most $n$ – 1, which takes the values of $y_1$, $y_2$, ... $y_n$ at points $x_1$, $x_2$, ... $x_n$. However, $n$ must be predetermined and cannot depend on the length of the experimental data series.

3. If $n$ could be determined in advance, then one can look for such an $h_0(x)$ function in set $H_k$, which would meet the following condition:

$$\max_i \mid y_i - h_0(x_i) \mid = inf \max_i \mid y_i - h_0(x_i) \mid; h(x) \in H_k. \tag{17}$$

Nevertheless, for computational reasons, the number of $k$ parameters of function $h(x) \in H_k$ should be significantly lower than the number of data $n$. However, there are no objective criteria for determining the number $k$, so we are forced to choose it arbitrarily.

All methods of solving the basic problem of approximation are accompanied by various forms of arbitrariness; therefore, all these methods should be considered equivalent (Hellwig, 1965, p. 35). In this situation, the statistical interpretation of the stochastic approximation is as follows: we are looking for a family $R_k$ of such $k$-parameter functions $\psi\ (x)$, which would satisfy the relation:

$$P[|\mathrm{Y} - \ \psi\ (x)| < \Delta] \geq 1 - \alpha \tag{18}$$

for the predetermined number $\alpha$ and Δ.

The general formalisation of the basic issue of approximation and the formalisation of stochastic approximation in a regression approach indicate that statistical regression is one of the forms of the approximation of real relations occurring in socio-economic processes. When approximating this reality, it is necessary to determine the tolerance that we accept, i.e. the acceptable discrepancy between the actual relationship between the explanatory variables we define and the variable we explain and its chosen approximant, in statistics called the permissible (average) error of estimation.

The above-mentioned general formulation of regression in the approximation approach, defining the confidence of approximation 1 – $\alpha$ suggests that in regression studies, the methods of interval estimation should be used rather than point estimation, because in statistics, the degree of credibility is defined as the confidence factor.

Consequently, the approximation approach to regression indicates that the approximation function chosen on the basis of the sample is only one of many possible regression functions belonging to the set of approximants. What is more, the individual functions belonging to this set are indistinguishable in terms of the accepted level of tolerance. Thus, in econometric modelling, it is unjustified to label a given model 'the winner' simply because other models fall within an acceptable tolerance range. The approximation theory, followed by statistics, provides many forms of analytic functions for use as approximation models. As previously indicated, generalised polynomials, trigonometric polynomials or splines are the most commonly assumed. However, it should be noted that these models require the number of parameters to be determined, at least initially. The approximation approach to regression is thus formalised mainly in terms of parametric statistics.

Currently, computer science provides tools based on artificial intelligence in the form of machine learning, which is the equivalent of the deterministic approach to approximation. It is utilised by numerical and statistical learning methods, equivalent to stochastic approximation. Learning methods thus extend the range of statistical methods to both parametric and non-parametric approaches.

## 5. Non-linearity of socio-economic processes

Economists and statisticians have long observed that socio-economic processes do not evolve in a continuous or linear way, but are shaped by sudden changes in the economic and political situation, entailing disruptions that lead to changes in developmental directions. This is caused not only by macroeconomic factors, but also by people's decisions and emotional behaviours which are not always rational. When analysing such processes, one cannot rely only on simple, linear models, but their complexity, dynamics and unpredictability need to be taken into account.

The non-linearity of socio-economic processes is caused by many factors. These include:

- The concurrence and interdependencies between different spheres of socio-economic life, politics and culture, which create dynamic and often unpredictable mechanisms of social behaviour;
- The presence of feedback effects between these spheres. The actions of one element of the system can affect others, which in turn affects the first element in a way that strengthens or weakens its influence;
- The risk of unpredictable, sudden changes occurring in the conditions of these processes, which is symbolised by the concept of a 'black swan'. Sudden shifts such as financial crises, technological revolutions or pandemics can radically change the course of socio-economic processes in unexpected ways;
- The possibility of the 'butterfly effect': minor initial deviations can lead to significant, divergent effects in the long term;
- The stochastic nature of socio-economic processes. These processes are usually not deterministic in nature. Human decisions, free will and cultural factors play a key role here, introducing an element of unpredictability.

The literature on the non-linearity of the course of socio-economic processes is very extensive. The beginnings of non-linear thinking in economics can be traced back to Adam Smith's theory, which assumes that growth has its limits caused by decreasing rates of return on investments. Formalised non-linear models of macroeconomic phenomena emerged with the works of several eminent 20th-century economists. These concerned the non-linear growth model presented by Nicholas Kaldor as the Kaldor-Hicks efficiency model, Richard Goodwin's macroeconomic model of economic growth and business cycles, and the classic non-linear relationship between basic macroeconomic processes, as the Phillips curve (Orzeszko, 2016). Significant examples of the use of non-linear functions in microeconomic research are the Cobb-Douglas production function and the Tornquist consumption functions introduced into the economic literature in as early as the 1920s.

The non-linear nature of socio-economic processes has long been recognised in statistical and econometric research and has led to the successive introduction of non-linear macroeconomic models. It particularly concerns the study of the non-linearity of economic growth and, more broadly, socio-economic development. A current review of this type of work is included in Orlando et al. (2021). Specialist journals are also devoted to the study of non-linearity, an example of which is the *Journal of Nonlinear Sciences and Applications*. It publishes scientific articles on non-linear mathematical methods and their applications, mainly in the field of natural sciences. A current overview of the methods and their applications in the study of non-linear dynamics of economic processes can be found in Gardini et al. (2021).

The non-linearity of processes is revealed in time series analysis based on stochastic models. One of the monumental figures in this area of statistical research, along with G.E.P. Box and G.M. Jenkins, is C.W.J. Granger, winner of the 2003 Alfred Nobel Prize in Economics. He is the author of the statement that the world is almost certainly non-linear (Granger, 1989) and the author of the definition of causality (known as the Granger causality). It states that X causes Y if and only if including explanatory variable X in the model predicting explanatory variable Y increases the accuracy of the prediction. He is also the author of the concept of cointegration introduced in dynamic econometrics, initially in a linear version (Granger, 1981), and ten years later in a non-linear version (Granger & Hallman, 1991). Several generalisations of the concept of cointegration to cases of non-linear relationships have been proposed in the econometric literature based on Granger's proposal. The book by Fan & Yao (2005) still remains a fundamental work in the field of non-linear time series. One of the most recent monographs in this field is a work by Maitra (2025). This book is a comprehensive and accessible guide allowing for the understanding of advanced econometric methods and their use to analyse real data in the form of economic and financial time series, skilfully combining theory with practice. The work presents data filtering techniques, highlighting the role of the Kalman filter in improving model accuracy. Volatility modelling is also discussed, addressing common challenges in measuring and interpreting the variance of financial data. Moreover, hybrid approaches combining GARCH models with neural networks and dynamic volatility models for option pricing are presented, describing both the theoretical foundations and providing practical tools for their application. Finally, mode-switching models, including MSAR and STAR are discussed to show the non-linear behaviours and structural shifts of data in time series. The monograph presents a coherent framework for modelling the dynamic behaviour of financial time series, with a particular emphasis on volatility and structural change. It may be particularly useful for finance professionals and data scientists.

Many works on non-linearity have also been published in Polish. A monograph on non-linear processes and long-term relationships in economics by Bruzda (2007) is one of the most comprehensive studies in this field. Its scope includes the identification of stationary non-linear processes and verification of long-term relationships based on the Granger and Hallman method, and a parametric estimation and verification of non-linear long-term relationships. Moreover, it explores non-linear error correction models, threshold cointegration and it provides many examples of applications of non-linear cointegration analysis. Another extensive work devoted to the identification of non-linearities in financial and economic time series is Orzeszko (2016).

This section outlines issues related to the non-linearity of economic processes, particularly focusing on their financial aspects. It shows how economics and financial econometrics developed from non-linear models describing the functioning of economic systems, formulated around the middle of the 20th century by such eminent economists as N. Kaldor, J. Hicks, W. Leontief, R. Goodwin, P. Samuelson, to the latest achievements of non-linear financial econometrics. The published results of both theoretical and empirical research indicate enormous progress in this area of economics. Nonetheless, linear approximation of these processes still prevails. It is thus time to acknowledge that socio-economic processes are inherently non-linear and their linearity is either a special case of non-linearity or their oversimplification. A challenge faced by statistics is to improve the tools for detecting and describing the non-linearity of the studied processes.

## 6. Non-parametric modelling

Non-parametric modelling is a method that does not assume a specific analytical form for the population distribution. Instead, it relies on the shape of the data distribution in a sample or a training set, which allows the structure of the model to be determined based on the data. It is a more flexible method than parametric modelling, which requires a defined analytical form of the approximated function and a fixed number of parameters. Non-parametric modelling is an important segment of statistical methods. It is especially useful when the data do not follow a normal distribution or when it is unknown, and when the studied features are measured on weak, ordinal or nominal scales. Common applications of the non-parametric approach include the Mann-Whitney U test, the Kruskal-Wallis test or the chi-square test, which serve as an alternative to parametric tests.

The above-mentioned problems indicate that their statistical description and modelling in a parametric form is computationally complex and requires the adoption of many restrictive assumptions. An alternative is the non-parametric approach. The non-parametric regression model is the most widely used type. With the help of non-parametric regression, the relationship between the considered variables may be analysed without a predetermined analytical form of the regression function. The non-parametric regression model derives directly from the data. Estimating a non-parametric regression function involves selecting a function of a specific class that is flexible enough to fit well to the data set. For this reason, non-parametric regression models require a much larger number of observations than their parametric counterparts. We approximate not only the values of their parameters, but also the nature of the relationships between the considered variables. Thus, as the dimensionality of the model increases, the number of observations necessary for a reliable estimation of the non-parametric regression function increases exponentially. This phenomenon, described as the curse of dimensionality, may be circumvented through semiparametric modelling (Pagan & Ullah, 2009) or the multimodelling approach (Gatnar, 2008).

Non-parametric regression methods are implemented within the statistical learning framework. In the general understanding of statistical learning, we consider variable Y (response variable), also understood as the explanatory (dependent) variable, and $k$ of the explanatory variables (predictors) $X_1, X_2,..., X_k$. We assume that between Y and $X = (X_1, X_2,..., X_k)$ there is a certain relationship, which can generally be described as:

$$Y = f(X) + \xi, \tag{19}$$

where $f$ is an unknown function that binds Y to X and $\xi$ is a random component. The essence of statistical learning is to guess the actual function $f$ by means of function $h$. It is one of the hypotheses concerning this unknown function $f$ , which belongs to the hypothesis space $H$ where $f$ is located (Hastie et al., 2009).

There are two main reasons why we try to guess function $f$. The first one is of a practical nature and involves the prediction of Y based on the knowledge of X. The second one is of a cognitive nature and involves the inference of Y based on X. The task of prediction arises when a set of predictors X is available, but the corresponding values of response variable Y are not known. We then make a prediction of Y using the equation:

$$\hat{Y} = \hat{f}(X), \tag{20}$$

where $\hat{f}$ is one of the functions within the hypothetical space $H$. In this context, $\hat{f}$ is treated as a black box, where uncovering a specific character $\hat{f}$ is not the main objective, but rather achieving an as accurate as possible prediction of Y.

The main idea of statistical learning is to know the actual function $f$. We estimate it on the basis of a dataset, called a training dataset, which contains information about input and output data ($x_i$, $y_i$). In other words, we are looking for a function $\hat{f}$, for which Y ≈ $\hat{f}$(X) for any pair of observations from the (X,Y) set. For this purpose, we can use a parametric or non-parametric approach. Non-parametric statistical learning methods do not make clear assumptions about the analytical form of the functions concerning *f*. Instead, they seek a form of function *f* that fits the data from the training set as best as possible. The non-parametric approach may have a great advantage over the parametric approach. because by avoiding the adoption of a specific analytical form of function *f*, it can fit the empirical data more accurately. However, the non-parametric approach has the disadvantage that it does not reduce the number of the estimated parameters to only the relevant ones; thus, it requires a much larger teaching set (James et al., 2013).

Selecting the best $\hat{f}$ function that belongs to hypothetical space *H* should be based on a specific criterion of the matching quality of function $\hat{f}$ with the actual function *f*. From among the many methods of measuring the effectiveness of statistical learning at a specific $\hat{f}$, the mean squared estimation error (MSE) is the most commonly used:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2, \tag{21}$$

where:
$\hat{f}(x_i)$ is the prediction of the actual *f* for the $i$-th observation.

If the estimation error is calculated for the data from the training set, then it is a measure of the goodness of fit for function $f$ to empirical data $y_i$.

One of the simplest and most well-known non-parametric methods is the $k$-nearest neighbours method, which can be used for both classification and regression purposes. The $k$-nearest neighbours (KNN) regression for the $k$-considered number nearest neighbours and for the $x_0$ from the test set, first identifies $k$ observations from the training set, the nearest $x_0$, denoted by $n_0$, and then it estimates $\hat{f}(x_0)$ using the average of all the values of $y_i$ from the training set, belonging to $n_0$, i.e.:

$$\hat{f}(x_0) = \frac{1}{k}\sum_{x_i \in n_0} y_i, \tag{22}$$

where:
$\hat{f}(x_0)$ is the $k$-nearest neighbour estimator of the non-parametric regression function.

More precise non-parametric regression methods allow modelling non-linear relationships between data by using kernel estimators to transform data in empirical distribution classes. A kernel density estimator is designed to determine the distribution density of a random variable based on the obtained sample or training set. It is given by the formula:

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(t_i), \tag{23}$$

where:

$n$ – the sample size,

$K_h$ – function of the kernel at smoothing parameter $h$,

$t_i$ – standardised values of random variable X.

The Gaussian function is most often chosen as the kernel function; other functions are: the Epanecznikov kernel, uniform kernel, two-weight kernel, triangular kernel, radial kernel or product kernel. It is also possible to use a linear combination of kernels (Kulczycki, 2005).

A specific class of models in the decision-making theory are decision trees, which can also be used for regression or classification purposes. A regression tree is used to show hierarchically the regression of a continuous explanatory variable relative to its explanatory variables. On the other hand, if the explanatory variable is measured on a nominal or ordinal scale, then a classification tree is employed. More precise methods have been developed on the basis of decision trees such as bagging, boosting methods or random forests. Each of them involves creating multiple trees which are then combined to obtain a more precise prediction of the explanatory variable.

Non-hierarchical methods can also be used to study non-parametric regression. Among them, the Support Vectors Machines (SVM) method is widely used, both for classification and regression purposes. In the case of regression, it is called a support vector regression (SVR) (Vapnik, 2000). The key concept of SVR is the use of support vectors, i.e. hyperplane observations in a multidimensional space. The selection of the best hyperplane is based on the method of an adaptive enrichment of the observation space and the search for a discriminatory hyperplane in such new spaces. Vapnik's idea is to solve an optimisation problem with a quadratic target function and linear constraints (square-linear optimisation) in Hilbert space (Koronacki & Ćwik, 2005).

The most popular non-parametric models are neural networks. These are computational models inspired by the functioning of the human brain, made up of connected artificial neurons arranged in layers that process information and learn to recognise patterns from data. The application of networks to solve regression problems is common. The user of the neural model expects that the created network will be able to describe the relationships between variables, while assuming that the explained variable is continuous. Proponents of neural networks as a tool for conducting economic research point to their advantage over classical regression models, discrimination models and classical trend models. However, whether a neural network function provides the correct solution to the posed problem depends on two basic factors:

- the values of the weight coefficients of the neurons that make up the network;
- the structure (topology) of the network, which is determined by the number of layers, the number of neurons in the individual layers, the way neurons are connected and the adopted model of the neuron (the aggregation of input data, type of the activation function used) (Lula, 1999).

Many interesting books on non-parametric modelling have been published in Polish. These include works by Bruzda (2007), Gatnar (2001) or a monograph combining problems related to non-linearity with a non-parametric approach by Orzeszko (2016). A representative example of many applications of non-parametric regression methods in socio-economic research can be found in the article by Trzęsiok (2013).

The world literature on non-parametric methods is extensive. A seminal study presenting the traditional approach to non-parametric methods is by Hollander et al. (2014). A comprehensive overview of non-parametric and semi-parametric models can be found in Härdle et al. (2004). Non-parametric modelling is also the subject of specialised scientific journals, among which the *Journal of Nonparametric Statistics* is well-known. It focuses on non-parametric statistics and related areas, including modelling, estimation, analysis and testing of statistical methods and algorithms. Regression and classification methods based on decision trees are central to the statistical learning paradigm. The basic monographs dedicated to these issues are by Hastie et al. (2009) and James et al. (2013).

The rapidly developing statistical methods include new branches of machine learning, such as deep learning and causal machine learning, which many recently published monographs are devoted to. One of the most interesting works in the field of deep learning is by Goodfellow et al. (2016), while in causal machine learning by Cunningham (2021).

## 7. Conclusions

The conducted extensive literature review along with the author's 50 years of research experience in the field of statistical analyses of socio-economic processes allow the formulation of the following remarks on the current challenges facing statistics:

- Statistical methods are tools for learning the truth about the reality that surrounds us. However, the truth discovered by statistical means must be understood in probabilistic terms. Thus, truth is a judgment, with an acceptable probability (error), consistent with reality;

- The surrounding reality is manifested through phenomena that are a source of impressions. We can observe these phenomena with greater or lesser accuracy and thus, only approximate the studied reality. Socio-economic processes are stochastic in nature, so the most general statistical method of searching for the truth about the reality is the stochastic approximation method;
- Throughout the history of human thought, the initial philosophical position was determinism, which assumed that all events in nature are predetermined. However, Werner Heisenberg demonstrated that in a world of the smallest material particles, causal conditioning does not exist, pointing to the 'free will of the electron'. The shift from physics to philosophy proved the existence of indeterminism in nature. Thus, the deterministic approach, which was the original philosophical position regarding the understanding of the world, has been replaced by its indeterministic counterpart, which is the dominant modern scientific view. In biology, the theory of evolution is based on indeterministic behaviour. Social indeterminism is a philosophical and sociological view that denies the existence of strict, necessary laws governing social phenomena. It is based on the belief that events in society are not strictly determined by the conditions that precede them, emphasising the role of chance, free will and other factors in shaping human behaviour. Indeterminism in economics is based on the assumption that economic events are not strictly determined by previous causes and their course is influenced by many unpredictable factors, including the free will to make decisions by market participants. This means that even under identical initial conditions, different business results may be achieved;
- The challenge for statisticians is to balance uncertainty and the related risk measurement with the discovery of the course of past processes and forecasting their course in the future. From a mathematical point of view, it involves optimising the stochastic approximation of the studied socio-economic reality;
- Learning about socio-economic processes can only be a better or worse way to get closer to the socio-economic reality that surrounds us. Therefore, it is necessary to formulate general criteria for this approximation to be satisfactory. Such criteria are based on probabilistic foundations. Therefore, approximation remains the most general method of the cognition of reality, as human cognition is inherently its simplification;
- Soon after the emergence of stochastic approximation in the field of regression research in the world literature, it was introduced into the Polish academia by Zdzisław Hellwig in his book on stochastic approximation (Hellwig, 1965). The author deals with multidimensional random variables, particularly some aspects

of correlation and the regression theory, which are especially important in economic applications. The main thesis is that the regression function is nothing more than an approximation function; therefore, the basic theorems and some computational methods provided by the theory of approximation can be used in the theory of regression and correlation;

- The problem of stochastic approximation is expressed by the following relation: $P[|\mathrm{Y} - \psi(x)| < \Delta] \geq 1 - \alpha$, where $\psi(x)$ is the approximant of an unknown function combining the components of a two-dimensional variable (X,Y), $\Delta$ is the tolerance or acceptable error of approximation, while $1-\alpha$ is the credibility of the approximation. Statistical regression is one of the forms of the approximation of real relations occurring in socio-economic processes. When approximating this reality, it is necessary to determine tolerance $\Delta$ for the discrepancy between the actual relationship among the variables we define and its approximant we choose;
- The approximation approach to regression indicates that the approximation function chosen on the basis of the sample is only one of many possible regression functions belonging to the set of approximation functions. What is more, the individual functions belonging to this set are indistinguishable in terms of the accepted level of tolerance. In econometric modelling, stating that a given model 'wins' simply because others are within an acceptable tolerance range is unjustified;
- Socio-economic processes do not develop in a continuous and linear manner, but tend to be disrupted and shaped by sudden changes in the economic and political situation. This is due not only to macroeconomic factors, but also to people's decisions and emotional behaviours, which are not always rational. The analysis of such processes cannot rely solely on simple, linear models, but their complexity, dynamics and unpredictability need to be considered as well. The linearity of the course of socio-economic processes is not a rule, but a special case of their course;
- Process non-linearity is revealed in time series analysis through stochastic modelling. According to Clive Granger, the world is almost certainly non-linear. Despite significant advancements in financial economics and econometrics associated with the modelling of non-linear socio-economic processes, improving the tools for detecting and describing the non-linearity of the studied processes remains a challenge for statisticians;
- Statistical description and parametric modelling of non-linear economic processes are computationally complex and require several restrictive assumptions. The non-parametric approach is a viable alternative; it does not require the assumption of a specific analytical form of the distribution in the population; instead, it relies on the shape of the data distribution in the sample or

training set. It allows the structure of the model to be determined based on the data. Non-parametric modelling is thus an important segment of statistical methods, especially useful when the distribution of the data is non-normal or unknown, and when the studied characteristics are measured on weak, ordinal or nominal scales;

- Non-parametric methods, particularly non-parametric regression, represent a significant step towards the development of multivariate statistical methods based on artificial intelligence. Despite their increased flexibility, these methods are not free from limitations that affect model adequacy and interpretability, including their sensitivity to the quality of the data in the training sets. Measurement errors, data gaps or outliers can significantly distort the results of the model. The lack of restrictive assumptions can only be apparent, as they can be transferred to a lower, 'technical' level. Moreover, as the complexity of statistical learning models grows, problems with their substantive interpretation intensify;

In conclusion, the author maintains that the rapid development of statistical learning methods should be based on a philosophical foundation, containing axioms concerning the stochastic, non-linear and non-parametric nature of socio-economic processes, effectively modelled by stochastic approximation methods.

## References

Bogdanienko, J. (2018). *Istota i problemy poznania naukowego*. CeDeWu.

Briggs, A., Łuczewski, M., & Puczkarski, P. (2025). Kwant bożego królestwa. *Tygodnik Powszechny*, (32), 45–47.

Briggs, A., & Reiss, M. J. (2021). *Human flourishing. Scientific insight and spiritual wisdom in uncertain times*. Oxford University Press.

Bruzda, J. (2007). *Procesy nieliniowe i zależności długookresowe w ekonomii. Analiza kointegracji nieliniowej*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Burnewicz, J. (2021). *Filozofia i metodologia nauk ekonomicznych*. Wydawnictwo Naukowe PWN.

Cunningham, S. (2021). *Causal Inference. The Mixtape*. Yale University Press.

Czerwiński, Z. (2002). *Moje zmagania z ekonomią*. Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Fan, J., & Yao, Q. (2005). *Nonlinear Time Series. Nonparametric and Parametric Methods*. Springer.

Fogel, A., Lyra, M. C. D. P., & Valsiner, J. (Eds.). (2014). *Dynamics and Indeterminism in Developmental and Social Processes*. Psychology Press.

Fortuna, Z., Macukow, B., & Wąsowski J. (1993). *Metody numeryczne*. Wydawnictwa Naukowo-Techniczne.

Gardini, L., Lamantia, F., Szidarovszky, F., & Tramontana, F. (2021). Nonlinear dynamics in economic modeling. *Decisions in Economics and Finance*, *44*, 485–487. https://doi.org/10.1007/s10203-021-00353-8.

Gatnar, E. (2001). *Nieparametryczna metoda dyskryminacji i regresji*. Wydawnictwo Naukowe PWN.

Gatnar, E. (2008). *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*. Wydawnictwo Naukowe PWN.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Granger, C. W. J. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics, 16*(1), 121–130. https://doi.org/10.1016/0304-4076(81)90079-8.

Granger, C. W. J. (1989). *Forecasting in Business and Economics*. Accademic Press. https://doi.org/10.1016/C2013-0-10755-6.

Granger, C. W. J., & Hallman, J. (1991). Long Memory Series with Attractors. *Oxford Bulletin of Economics and Statistics*, *53*(1), 11–26. https://doi.org/10.1111/j.1468-0084.1991.mp53001002.x.

Grobler, A. (2006). *Metodologia nauk*. Aureus, Znak.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer. https://doi.org/10.1007/978-3-642-17146-8.

Hellwig, Z. (1965). *Aproksymacja stochastyczna*. Państwowe Wydawnictwo Ekonomiczne.

Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparmetric Statistical Methods*. Wiley. https://doi.org/10.1002/9781119196037.

Jadacki, J. (2020). Zagadka przyczynowości. *Przegląd Filozoficzny – Nowa Seria*, *29*(2), 17–51. https://doi.org/10.24425/pfns.2020.133136.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Application in R*. Springer.

Jasiński, L. J. (2017). *Analiza i interpretacja badań ekonomicznych*. Oficyna Wydawnicza Politechniki Warszawskiej.

Kamiński, S. (1981). *Pojęcie nauki i klasyfikacja nauk*. Towarzystwo Naukowe Katolickiego Uniwersytetu Lubelskiego.

Katsenelinboigen, A. (1992). *Indeterministic Economics*. Bloomsbury Publishing.

Katsenelinboigen, A. (1997). *The Concept of Indeterminism and its Applications. Economics, Social Systems, Ethics, Artificial Intelligence, and Aesthetics*. Bloomsbury Publishing.

Kiefer, J., & Wolfowitz, J. (1952). Stochastic Estimation of the Maximum of a Regression Function. *The Annals of Mathematical Statistics*, *23*(3), 462–466. https://doi.org/10.1214/aoms/1177729392.

Klein, L. R. (1982). *Wykłady z ekonometrii*. Państwowe Wydawnictwo Ekonomiczne.

Kordecki, W., & Selwat, K. (2020). *Metody numeryczne dla informatyków*. Helion.

Koronacki, J. (1989). *Aproksymacja stochastyczna. Metody optymalizacji w warunkach losowych*. Wydawnictwa Naukowo-Techniczne.

Koronacki, J., & Ćwik, J. (2005). *Statystyczne systemy uczące się*. Akademicka Oficyna Wydawnicza EXIT.

Kuciński, K. (Ed.). (2010). *Metodologia nauk ekonomicznych. Dylematy i wyzwania*. Difin.

Kulczycki, P. (2005). *Estymatory jądrowe w analizie systemowej*. Wydawnictwa Naukowo-Techniczne.

Lemańska, A. (1998). *Filozofia przyrody a nauki przyrodnicze*. Akademia Teologii Katolickiej.

Lula, P. (1999). *Jednokierunkowe sieci neuronowe w modelowaniu zjawisk ekonomicznych.* Wydawnictwo Akademii Ekonomicznej w Krakowie.

Maitra, S. (2025). *Non-Linearity in Econometric Modelling: Vol. 1. A Practical Approach.* Springer, Cham. https://doi.org/10.1007/978-3-032-06462-2.

von Mises, R. (1957). *Probability, Statistics and Truth*. John Allen & Unwin.

Nowakowski, A. (2017). Przyczynowa wizja świata. *Przegląd Filozoficzny – Nowa Seria*, *26*(1), 53–68. https://journals.pan.pl/Content/100487/PDF/P.Filoz.%201-17%203-A.Nowakowski.pdf?handler=pdf.

Orlando, G., Pisarchik, A. N., & Stoop, R. (Eds.). (2021). *Nonlinearities in Economics. An Interdisciplinary Approach to Economic Dynamics, Growth and Cycles*. Springer. https://doi.org/10.1007/978-3-030-70982-2.

Orzeszko, W. (2016). *Nieparametryczna identyfikacja nieliniowości w finansowych i ekonomicznych szeregach czasowych*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Pagan, A., & Ullah, A. (2009). *Nonparametric Econometrics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612503.

Pociecha, J. (2023). *Filozoficzne i metodologiczne podstawy współczesnych analiz statystycznych*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Rao, C. R. (1997). *Statistics and Truth. Putting Chance to Work*. World Scientific Publishing.

Robins, H., & Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, *22*(3), 400–407. https://doi.org/10.1214/aoms/1177729586.

Sagan, A. (2016). *Metodologia badań ekonomicznych*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Sosenko, K. (2008). *Problemy filozofii i metodologii nauk dla ekonomistów*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Strawiński, W. (2011). Funkcja i cele nauki – zarys problematyki metodologicznej. *Zagadnienia Naukoznawstwa*, (3), 323–335. https://journals.pan.pl/dlibra/publication/108246/edition/93890/content.

Tatarkiewicz, W. (1978). *Historia filozofii* (Vol. 2). Wydawnictwo Naukowe PWN.

Trzęsiok, J. (2010). Wykorzystanie regresji nieparametrycznej do modelowania wielkości oszczędności gospodarstw domowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (102), 99–108. https://www.ue.katowice.pl/fileadmin/_migrated/content_uploads/7_J.Trzesiok_Wykorzystanie_regresji_nieparametrycznej....pdf.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory* (2nd ed.). Springer-Verlag. https://doi.org/10.1007/978-1-4757-3264-1.

Wawrzyński, P. (2021). *Uczące się systemy decyzyjne*. Oficyna Wydawnicza Politechniki Warszawskiej.

Przegląd Statystyczny. Statistical Review, 2026, vol. 73, 1, 30–55 https://doi.org/10.59139/ps.2026.01.2
Received: 9.01.2026, revised: 26.04.2026, accepted: 26.04.2026

# Disposable income and acquisition costs as drivers of CASCO insurance demand: an empirical analysis of the Polish market

Norbert Duczkowski,[a] Adam Śliwiński[b]

**Abstract.** This paper explores the determinants of demand for voluntary comprehensive car insurance (CASCO) in Poland. The study aims to identify the impact of acquisition costs and household disposable income on insurance uptake. Utilizing a multilayer perceptron neural network model, the research analyzes the relationship between motor insurance expenditures and two key variables: the disposable income of households and the acquisition costs ratio between voluntary CASCO and Mandatory Third-Party Liability insurance. The findings, validated by robust diagnostic tests, a linear ordinary least squares model as a benchmark and a leave-one-out cross-validation procedure, reveal that both factors are statistically significant predictors of demand. The results indicate that the expansion of the CASCO market in Poland is driven not only by rising consumer affluence but also by the commission ratio, which motivates intermediaries to prioritize voluntary products. This suggests a supply push mechanism where distribution costs shape market dynamics.

**Keywords:** motor insurance, CASCO, disposable income, acquisition costs, neural networks

**JEL:** G22, D12, C45, L14

## 1. Introduction

The motor insurance market represents a cornerstone of the non-life insurance sector in Poland, as gross written premiums (GWP) from motor insurance exceed 50% of the total gross written premiums within the non-life segment (Bankowy Fundusz Gwarancyjny [BFG], 2025). Its key components include Mandatory Third-Party Liability (MTPL) insurance and voluntary comprehensive car insurance (CASCO). While MTPL is legally required for all vehicle owners, the decision to purchase CASCO coverage remains discretionary, driven by a complex interplay of economic, psychological, and institutional factors (Dragos et al., 2023; Feyen et al., 2013). Understanding the drivers of CASCO uptake is crucial for insurers seeking to optimize their portfolios and for policymakers monitoring the financial resilience of households.

[a] SGH Warsaw School of Economics, Collegium of Management and Finance, Institute of Risk and Financial Markets, Department of Risk and Insurance, ul. Madalińskiego 6/8, 02–513 Warszawa, Poland, e-mail: nduczk@sgh.waw.pl, ORCID: https://orcid.org/0000-0003-4543-2678.

[b] SGH Warsaw School of Economics, Collegium of Management and Finance, Institute of Risk and Financial Markets, Department of Risk and Insurance, ul. Madalińskiego 6/8, 02–513 Warszawa, Poland, e-mail: asliwin@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-7817-0101.

Despite the growing maturity of the Polish insurance market, a significant disparity remains in the saturation levels (in terms of the number of policies) between mandatory and voluntary coverages. The classical economic theory suggests that the demand for insurance is primarily a function of risk aversion and income (Dąbrowski & Śliwiński, 2016; Śliwiński, 2019). However, within the specific context of the Polish market, two factors emerge as potentially decisive, namely the financial capacity of households and acquisition costs, which translate into incentive structures within distribution channels.

The first factor, disposable income, reflects the household's budget constraints. CASCO consumption is expected to be highly sensitive to fluctuations in disposable income, as this factor significantly shapes demand across both life and non-life insurance sectors (Dragos & Dragos, 2017; Duczkowski, 2022; Śliwiński, 2011; Śliwiński, 2016). The second, and perhaps more nuanced factor, relates to acquisition costs. In Poland, as in Western Europe, the distribution of motor insurance is heavily reliant on intermediaries (Lisowski & Zieniewicz, 2015). The remuneration of insurance agents, specifically the commission ratio of CASCO over MTPL, creates a relative incentive that may drive supply-side push effects. It could be assumed that when the financial reward for selling a CASCO policy significantly exceeds that of a mandatory one, intermediary behavior may become an important driver of market penetration.

While the existing literature often focuses on many variables from different groups (Dragos & Dragos, 2017; Dragos et al., 2023; Kumaga, 2016; Sherden, 1984), there is a lack of empirical research that accounts for both the consumer's ability to pay and the intermediary's incentive to sell. This paper aims to fill this gap by analyzing the role of disposable income and the acquisition cost ratio in shaping the demand for CASCO insurance. Using a neural network model, we demonstrate that both variables are statistically significant, providing new insights into the mechanisms of the Polish insurance market. The statistical analysis was also supported by the use of a linear ordinary least squares (OLS) model as a benchmark and the leave-one-out cross-validation (LOOCV) procedure.

## 2. The Polish motor insurance market

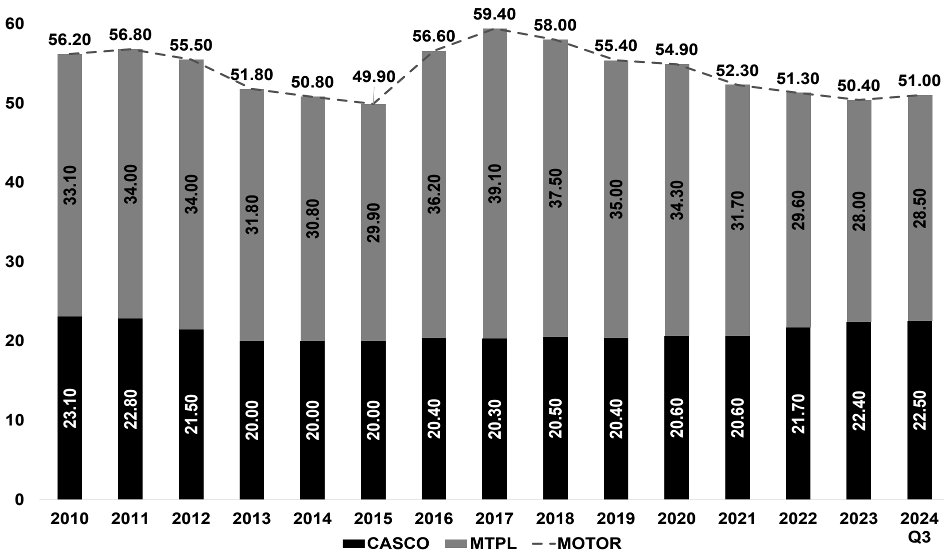
The motor insurance sector is the undisputed leader of the Polish non-life insurance market, as defined by Section II of the Act on Insurance and Reinsurance Activity (Pol. Ustawa z dnia 11 września 2015 r. o działalności ubezpieczeniowej i reasekuracyjnej). It consists of two primary segments: MTPL and CASCO. For this study, the CASCO segment is broadly defined and includes all motor insurance products other than MTPL, such as roadside assistance or tire insurance. The evolution of the motor insurance sector is deeply rooted in the rapid motorization of Polish society. By the end of 2023, the number of registered motor vehicles (passenger cars, goods vehicles and motorcycles) exceeded 32.9 million (BFG, 2025), providing a vast base for insurance penetration.

Motor insurance accounts for a dominant share of the total non-life insurance portfolio in Poland. As of Q3 2024, it represented 51.0% of the GWP from direct business (BFG, 2025) in a market characterized by the dominance of the Polish insurance company, PZU Group (Duczkowski, 2021a). However, the internal structure of this segment has undergone a significant transformation. While the share of MTPL in the total property insurance market decreased from 39.2% in 2017 to 28.46% in Q3 2024, the role of CASCO insurance steadily increased. CASCO accounted for 22.5% of the market in 2025 (BFG, 2025), reflecting a shift toward higher value protection products. This trend suggests that as the market matures, consumers are increasingly seeking protection beyond the minimum legal requirements. The market shares of motor insurance as a whole as well as the MTPL and CASCO segments within the non-life insurance sector are presented in Figure 1.

The Polish motor insurance market is characterized by ongoing development and increasing maturity. In the period between 2018 and Q3 2024, a consistent increase in the number of active insurance policies was observed. The MTPL portfolio grew by 14.5%, reaching 28.6 million policies (BFG, 2025), while the CASCO portfolio saw an even more dynamic increase of 15.5%, totaling 7.7 million policies (BFG, 2025). It is worth noting, however, that the number of CASCO policies is subject to certain fluctuations, having reached its peak in 2021 with the total count approaching 8.1 million units.

From a pricing perspective, the average MTPL premium has shown high volatility. After a period of aggressive price competition (insurance price fluctuations were characterized by a stable distribution (Duczkowski, 2021b) and declining rates (reaching a low point of PLN 500 in 2016), the market saw a reversal of the trend. By Q3 2024, the average MTPL premium rose to PLN 603 (BFG, 2025). Despite these nominal increases, motor insurance became more affordable in relative terms. The ratio of the average MTPL premium to the average national wage dropped significantly from 14.3% in 2017 to just 7.4% in 2024 (BFG, 2025). This increasing affordability, combined with growing disposable income, created a favorable environment for the cross-selling of CASCO policies.

**Figure 1.** Motor MTPL and CASCO share (in %) in the non-life insurance sector in Poland




Source: authors' work based on BFG (2025).

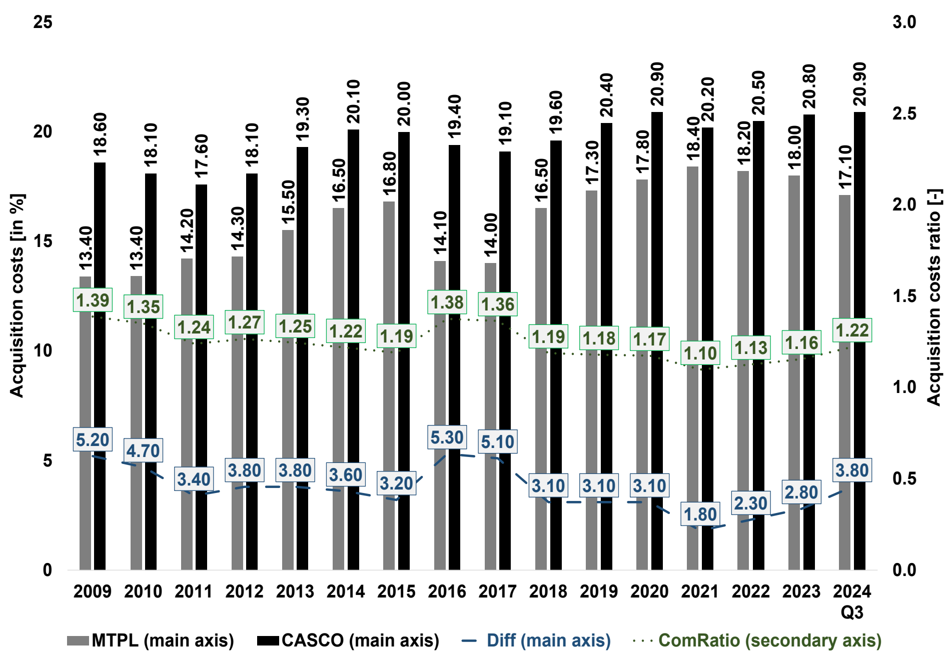
The analysis of the Polish motor insurance market must also consider the fact that the market is faced with mounting pressure from growing claim costs. The average payout for MTPL claims rose from PLN 6,865 in 2016 to PLN 10,533 in Q3 2024 (BFG, 2025). This increase was driven by rising spare parts prices, labor costs in repair shops and growing compensations for non-material damages. Interestingly, while the claim frequency for MTPL remained stable (between 3.1% and 4.5%), the CASCO segment recorded a much higher frequency, reaching 10.9% in 2024 (BFG, 2025).

The number of paid claims have a profound impact on the financial performance of insurance companies. The profitability of these two segments diverges sharply. The combined ratio (the ratio of claims paid and costs incurred to gross earned premium) for MTPL frequently exceeded the 100% threshold (reaching 108.8% in Q3 2024), indicating that insurers were generating technical losses on mandatory policies (BFG, 2025). In contrast, the CASCO segment remained the primary driver of technical profit for motor insurers, with a healthy combined ratio of 94.3% (BFG, 2025).

The divergence in technical results is closely linked to the distribution strategies employed by insurers. The analysis of operating expenses proves that acquisition costs, primarily consisting of intermediary commissions have a different structure for mandatory and voluntary products. According to the Polish Bank Guarantee

Fund (BFG, 2025) and the Polish Financial Supervision Authority (Pol. Komisja Nadzoru Finansowego – KNF) data from 2024, the Acquisition Cost Ratio (the ratio of acquisition costs to gross written premiums) for CASCO (20.9%) was significantly higher than for MTPL (17.1%). This commission spread or surplus is a deliberate market mechanism. Since MTPL is a mandatory product with high price sensitivity, insurers keep commissions lower to remain competitive. Conversely, CASCO is a higher-margin, voluntary product that requires a proactive selling strategy ('push') from insurance agents. Higher commissions for CASCO serve as a vital economic incentive for intermediaries to bundle voluntary coverage with the mandatory policy, leveraging their influence over the consumer's final decision. The acquisition cost ratios for MTPL, CASCO and the acquisition cost differential (*Diff*) between these insurance types, along with the ratio of CASCO acquisition costs to MTPL (*ComRatio*) are presented in Figure 2.

**Figure 2.** Acquisition costs for MTPL and CASCO, the acquisition cost differential (*Diff*) between them and the ratio of CASCO acquisition costs to MTPL (*ComRatio*)




Source: authors' work based on BFG (2025).

The observations presented above concerning the Polish motor insurance market, which serves as the primary motivation for this research, lead to the conclusion that the Polish motor insurance market is undergoing a significant structural transformation from a volume-driven expansion phase to a more sophisticated, value-driven strategic model. This shift is primarily catalyzed by the persistent technical losses recorded in the MTPL sector, which have reached a point where insurers are forced to aggressively seek sustainable profitability within the CASCO segment. Such a transition is underpinned by long-term macroeconomic stability and the consistently rising disposable income of Polish households, allowing consumers to pivot toward more comprehensive vehicle protection. However, as the empirical data suggest, the growth of the CASCO market is not merely a passive reflection of organic consumer demand. Instead, it is heavily facilitated and steered by the distribution network's complex incentive structure. Within this framework, a significant surplus of acquisition costs, specifically higher commission rates offered for voluntary products compared to mandatory ones, play a pivotal role. This commission gap effectively aligns the interests of insurance intermediaries with insurers' profitability goals, making the incentive-driven sales force a primary engine of market penetration (the number of CASCO policies increased by 15.5%, from 6,678,580 in 2018 to 7,711,817 in 2024) and one of the key determinants of the current portfolio mix in the Polish non-life insurance sector.

## 3. Factors determining the demand for motor insurance – literature review

The motor insurance market is the largest line of business within the non-life insurance sector, not only in Poland but also across Europe (BFG, 2025). Understanding the factors that drive the demand for these products is essential for insurers to optimize their distribution strategies and for policymakers to ensure market stability. It is therefore not surprising that demand for motor insurance, both compulsory MTPL and voluntary CASCO, has been the subject of numerous scientific studies, where researchers analyzed a large number of variables (Dragos & Dragos, 2017; Dragos et al., 2023; Kumaga, 2016; Sherden, 1984). Table 1 summarizes the primary determinants identified in the literature, categorizing them by their nature and the selected supporting research.

**Table 1.** Selected research papers on the factors dominating the demand for motor insurance (non-life insurance)[a]

| Category | Selected determinants | Supporting research / source |
|---|---|---|
| Vehicle specific | Vehicle age / year of manufacture | Awunyo-Vitor (2012); Hsu et al. (2014); Kumaga (2016) |
| | Size of car fleet / vehicle value | Awunyo-Vitor (2012); Dragos et al. (2023) |
| Economic and income | GDP *per capita* / household income | Awunyo-Vitor (2012); Beenstock et al. (1988); Park and Lemaire (2012) |
| | Ratio of income to car value | Dragos and Dragos (2017) |
| | Premium affordability and price sensitivity | Harrington and Niehaus (1998); Kumaga (2016); Sherden (1984) |
| Distribution and costs | Agent / broker commissions (acquisition costs) | Cummins and Doherty (2006); Inteliace Research (2014); Szymańska (2020); Wilder (2004) |
| | Remuneration models | Lisowski and Zieniewicz (2015); Reifner et al. (2013); Widura (2009); Ziemiak (2019) |
| Demographic | Education level of the insured (risk awareness) | Park and Lemaire (2012) |
| | Urbanization and population density | Dragos et al. (2023); Esho et al. (2004); Sherden (1984) |
| | Age and gender of the insured | Awunyo-Vitor (2012) |
| Psychological | Risk aversion / risk profile | Dragos and Dragos (2017); Kawiński and Szumlicz (2023) |
| Institutional | Rule of law and institutional quality | Dragos et al. (2023); Esho et al. (2004); Ziemiak (2019) |
| | Compulsory insurance regulations | Chen and Chen (2013); Marson and Ferris (2023) |

Source: authors' work.

In this research, the primary focus is on disposable income and the acquisition costs ratio as the most significant drivers of insurance demand. Therefore, a closer examination of these two groups of studies is presented below.

Income level, particularly disposable income, is a factor that significantly shapes demand across both life and non-life insurance (Dragos & Dragos, 2017; Duczkowski, 2022; Śliwiński, 2011; Śliwiński, 2016). It is therefore not surprising that this relationship holds for motor insurance, including both MTPL and CASCO. The level of disposable income remains one of the most fundamental determinants of demand for CASCO insurance. The literature suggests that as household wealth increases, the propensity to purchase optional coverage grows, as consumers seek to protect their assets against a wider range of risks than those covered by mandatory insurance (Sherden, 1984). Beyond this direct budgetary effect, income also exerts an indirect influence on insurance uptake. Higher income levels allow for the acquisition of newer and more valuable vehicles

(Dragos et al., 2023). Given that the financial loss associated with the theft or damage of such high-value assets is significantly higher, the economic rationale for purchasing comprehensive insurance becomes more compelling for affluent car owners (Kumaga, 2016). The positive correlation between income and insurance demand is extensively documented in the literature, which includes broader studies on non-life insurance (Beenstock et al., 1988; Browne et al., 2000; Esho et al., 2004; Park & Lemaire, 2012). On the other hand, a more detailed analysis should be devoted to acquisition costs and the specific dynamics of motor insurance distribution channels.

The acquisition costs ratio serves as a critical indicator of the competitive landscape, reflecting the intensity of distribution efforts within the motor insurance sector. The following section presents the key literature findings concerning this variable as they relate to the objectives of this research:

- Dominance of the agency channel – the agency channel remains the primary distribution force for non-life insurance, consistently accounting for over 50% of the gross written premium in markets such as Poland, where it significantly outperforms direct writing and bancassurance in this specific sector (Lisowski & Zieniewicz, 2015);
- Commission structures – remuneration for multi-agents is typically structured with higher commission rates for CASCO / package policies compared to MTPL insurance, reflecting the greater effort required for optional product placement (Inteliace Research, 2014);
- Commission significance – commission levels, including their fluctuations and regulatory limits, are identified as critical risk factors in the development of insurance distribution channels (Śliwiński et al., 2021);
- Conflicts of interest – commission-based remuneration models are frequently identified as a potential source of conflicts of interest, where intermediaries may be incentivized to prioritize short-term profit maximization over the long-term satisfaction or specific needs of the customer, both in the case of life and non-life insurance (Balewski & Janowski, 2008; Reifner et al., 2013);
- Informational role – intermediaries act as critical information conduits who alleviate adverse selection by providing insurers with higher-quality underwriting data about their clients than would otherwise be available through direct channels (Cummins & Doherty, 2006);
- Strategic consolidation – there is a visible market trend involving the decline of exclusive agency forces in favor of multi-agents, alongside a move by direct insurers to utilize intermediary networks as they realize that consumers remain strongly attached to face-to-face professional advice (Lisowski & Zieniewicz, 2015; New Direction, 2024);

- Time allocation incentives – the allocation of an agent's effort between acquiring new contracts and maintaining existing ones is highly sensitive to the ratio of new business commission rates to portfolio commission rates (Widura, 2009);
- Regulatory Safeguards – frameworks such as the Insurance Distribution Directive (IDD), prohibit distributors from using remuneration or sales targets that could incentivize products mis-selling or offering non-optimal policies to clients (Ziemiak, 2019);
- Risk Profile Assessment – integrating the driver's risk profile (measured through behavioral factors like saving versus borrowing habits) into the distribution process allows for a more comprehensive assessment of the policyholder's propensity to purchase voluntary motor insurance (Dragos & Dragos, 2017).

The aforementioned findings, particularly concerning commission structures, commission significance and time allocation incentives, provide the primary impetus for the study.

In conclusion, the interplay between consumer purchasing power (disposable income) and the economic incentives within distribution channels appears to be the primary mechanism driving the development of the voluntary motor insurance market. While disposable income determines the consumer's ability to afford premium protection, acquisition costs shape the intermediary's willingness to promote these products. The identification of these two factors as key statistical drivers in previous empirical studies has served as the core motivation for the research presented in this paper, aiming to further explore their impact within the specific context of the Polish insurance market.

## 4. Database and research method

The study utilizes secondary data sourced from the BFG, the KNF and Statistics Poland (GUS), representing the most authoritative and reliable data sources available in the country.

The following variables/time series were extracted from the GUS database:

- average monthly household expenditure on transportation-related insurance (*WUbtr*);
- average monthly disposable income *per capita* in households (*DisInc*).

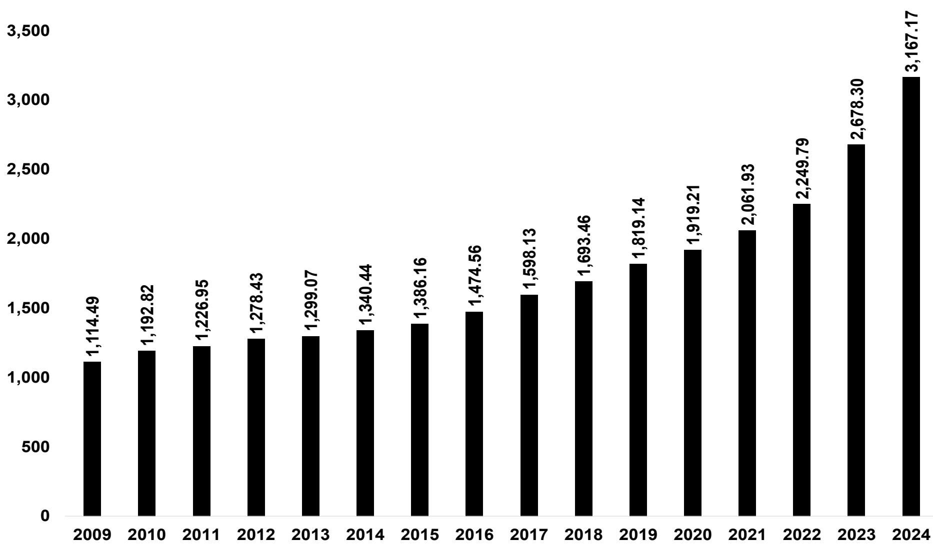
The average monthly disposable income *per capita* in households (*DisInc*), presented in Figure 3, was used as a direct input for further calculations.

The following variables / time series were extracted from the BFG and KNF materials and presentations:

- distribution costs ratio for MTPL insurance (*DisCostMTPL*);
- distribution costs ratio for CASCO insurance (*DisCostCASCO*);
- gross written premium for MTPL insurance (*GWPMTPL*);
- gross written premium for CASCO insurance (*GWPCASCO*).

Acquisition cost ratios (expressed as a percentage) for MTPL and CASCO were used to determine the variable applied in the model, namely the ratio of CASCO acquisition costs to MTPL (*ComRatio*) – the share acquisition costs of voluntary insurance relative to mandatory coverage. Variable *ComRatio* was presented in Figure 2.

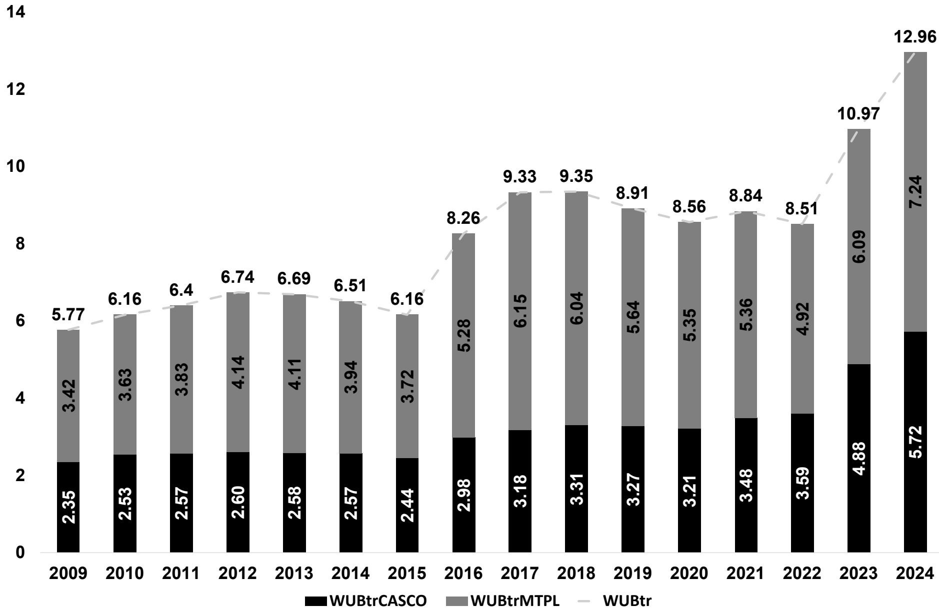
**Figure 3.** Average monthly disposable income (in PLN) *per capita* in households




Source: authors' work based on GUS (2010–2025).

GWP for CASCO and MTPL was utilized to calculate the proportion of total household expenditures on transport-related insurance, divided into voluntary (*WUbtrCASCO*) and mandatory (*WUbtrMTPL*) categories. This approach was necessitated by the fact that GUS does not publish data at such a granular level. Consequently, the study assumes that total household expenditures on transport insurance are proportional to the gross written premium within the voluntary and mandatory policy segments. It should be emphasized that allocating household transport insurance expenditures proportionally to GWP is an approximation, as total GWP for CASCO and MTPL also includes premiums for commercial vehicles and carriers. Nevertheless, it represents the best possible proxy for this analysis. Average monthly household expenditures on transport-related insurance (*WUbtr*), along with the estimated average monthly household expenditures on CASCO (*WUbtrCASCO*) and MTPL (*WUbtrMTPL*) insurance, are presented in Figure 4. Variable *WUbtrCASCO* was used as a direct input for further calculations.

**Figure 4.** Average monthly household expenditures on transport-related insurance (*WUbtr*), CASCO (*WUbtrCASCO*) and MTPL (*WUbtrMTPL*) in PLN




Source: authors' work based on GUS (2010–2025).

The first stage of the empirical analysis involved a statistical description of the variables to provide insight into their distribution and basic characteristics. Subsequently, to ensure the reliability of the econometric modelling, especially when comparing a neural network with a linear OLS benchmark, the stationarity of the time series was examined using the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. These two procedures were employed as complementary tools due to their different null hypotheses: the ADF test assumes the presence of a unit root (non-stationarity), while the KPSS test evaluates the null hypothesis of trend-stationarity. Using both tests simultaneously allows for a more robust verification of the properties of the series and helps avoid potential errors in identifying the order of integration.

In the next step of the study, an identical testing procedure was conducted on the first differences of the variables to achieve the stationarity required for further modelling.

The main part of the empirical analysis was conducted using an artificial neural network framework developed within the *RStudio* environment, specifically the *neuralnet* function from the *neuralnet* library. The network architecture was meticulously designed with a multi-layer perceptron structure (MLP) consisting of two hidden layers, defined by parameter *hidden = c(4, 2)*. The model was specified to examine the relationship between voluntary insurance penetration and its key drivers according to the formula:

$$\Delta WUbtrCASCO \sim \Delta ComRatio + ln(\Delta DisInc). \tag{1}$$

The model evaluates how the first differences of the CASCO to MTPL acquisition cost ratio ($\Delta ComRatio$) and the natural logarithm of changes (first differences) in disposable income ($\Delta DisInc$) influence the first differences of the CASCO insurance expenditures ($\Delta WUbtrCASCO$).

In addition, a Multiple Linear Regression model (namely OLS) was estimated in *RStudio* to serve as a benchmark for the results obtained from the neural network. To ensure full comparability and a rigorous baseline for the machine learning approach, the OLS estimation was conducted using the same set of variables and an identical equation (1). This procedure enabled a direct assessment of whether the non-linear MLP model provides significant added value relative to traditional econometric methods.

The multilayer perceptron architecture with a hidden layer structure of (4, 2) used as the main model for this study was carefully tailored to meet the specific constraints and requirements of the research. Given the relatively small size of the time series dataset, a more complex architecture with a higher number of neurons or layers would significantly increase the risk of overfitting. The model would then capture noise rather than the underlying economic patterns. The (4, 2) configuration provides a balanced degree of freedom, allowing the network to internalize non-linear relationships and interactions between the variables without losing its generalization capabilities. Additionally, neural networks handle non-stationarity or questionable stationarity of time series effectively. Furthermore, the preliminary results of the linear OLS benchmark presented in the Results section served as a critical justification for this choice. The limited explanatory power of the linear model compared to a neural network model suggests that the phenomenon of motor insurance demand is inherently very complex. Consequently, the MLP model was implemented to capture these complexities that traditional econometric tools might overlook.

To ensure the robustness and predictive accuracy of both the constructed network and the OLS benchmark, a comprehensive evaluation suite was employed. The goodness of fit was assessed using $R^2$ and Root Mean Square Error (RMSE) statistics. Furthermore, the model underwent rigorous econometric validation through the Durbin-Watson test for autocorrelation and the Shapiro-Wilk test for normality of residuals. Finally, a Neural Network Response Surface analysis was performed to visualize the non-linear interactions between the independent variables (*ln(ΔDisInc)* and Δ*ComRatio*) and the dependent variable (Δ*WUbtrCASCO*), providing deeper insights into the model's behavior.

Furthermore, the study utilized the leave-one-out cross-validation (LOOCV) procedure to evaluate the predictive performance of the models. This method was specifically chosen due to the limited size of the dataset, where a traditional split into training and testing sets would significantly reduce the amount of information available for the learning process. The LOOCV procedure is a specialized form of *k*-fold cross-validation (where $k=n$), consisting of $n$ iterations. In each step, a single observation is excluded from the dataset to serve as a test case, while the model is trained on the remaining $n-1$ observations. This process is repeated until every observation is used as a test set exactly once. This procedure enabled the calculation of out-of-sample performance metrics and the preparation of diagnostic charts, ensuring that the results are not biased by overfitting and that the model's stability is verified across the entire time horizon.

The decision to apply a natural logarithm to the change in disposable income, represented within the model as *ln(ΔDisInc)*, but not for variable Δ*ComRatio*, is rooted in the non-linear nature of agent motivation and market response. In the Polish motor insurance sector, small increases in the commission ratio between voluntary (CASCO) and mandatory (MTPL) products, resulting in a change in the *ComRatio* indicator, often yield disproportionately large shifts in sales efforts. By applying a natural logarithm transformation to the Δ*DisInc* variable, the model accounts for this convex relationship, where the marginal impact of the cost differential increases as the incentive becomes more attractive. This approach demonstrates the tipping point behavior of distribution networks, where a specific threshold in the acquisition cost surplus triggers a significant reallocation of resources and marketing focus toward CASCO policies, thereby enhancing market penetration more aggressively than linear variables would suggest.

Despite the statistical significance of the results, this study is subject to certain limitations. Firstly, a primary constraint arises from the lack of granular data published by GUS regarding the explicit breakdown of household expenditures on voluntary versus mandatory motor insurance. This necessitated the use of an estimation method based on GWP proportions to disaggregate the total transport insurance spending. While this approach is grounded in market-wide trends, it remains a proxy that may not reflect the full complexity of individual household-level decisions, as the total GWP for CASCO and MTPL includes premiums for commercial vehicles and carriers. Consequently, the potential differences in the pricing dynamics or commission structures between the retail and corporate sector may not be fully captured. Despite these constraints, the solution used represents the best possible proxy for this analysis. Secondly, the study is limited by the length of the available time series, covering the period from 2009 to 2024. Although this timeframe encompasses various economic cycles and regulatory shifts in the Polish market, a longer perspective is necessary to perform more advanced cointegration analyses or to capture more fully the long-term structural changes within the insurance sector.

## 5. Results

The results of the first stage of the empirical analysis involved a statistical description of the variables and stationarity tests for variable levels, presented in Table 2.

**Table 2.** Descriptive statistics and stationarity tests for variables in levels[a]

| Variable | Descriptive statistics (levels) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Min | Q1 | Median | Mean | Q3 | Max | Std. Dev. |
| ComRatio ...... | 1.098 | 1.178 | 1.220 | 1.236 | 1.287 | 1.388 | 0.091 |
| DisInc ............ | 1114.49 | 1293.91 | 1536.35 | 1718.75 | 1954.88 | 3167.17 | 580.28 |
| WUbtrCASCO | 2.35 | 2.57 | 3.08 | 3.2037 | 3.3525 | 5.72 | 0.9222 |
| Variable | Stationarity tests (levels) | | | | | | |
| | ADF stat | ADF p-value | Interpretation | KPSS stat | KPSS p-value | Interpretation | Conclusion |
| ComRatio ........ | −1.8315 | 0.6366 | Non-stationary | 0.4197 | 0.0686 | Stationary | Ambiguous |
| DisInc ............... | 1.5115 | 0.99 | Non-stationary | 0.6032 | 0.0223 | Non-stationary | Non-stationary |
| WUbtrCASCO | −0.5663 | 0.9695 | Non-stationary | 0.5665 | 0.0267 | Non-stationary | Non-stationary |

a The terms 'stationary' or 'non-stationary' serve as a concise notation indicating the rejection or failure to reject the null hypothesis of the respective statistical test.

Source: authors' work.

As evidenced by the data presented in Table 2, the analyzed variables are non-stationary. Although neural networks are inherently robust to non-stationary data, the transition to first differences was performed to maintain methodological consistency and ensure full comparability with the linear OLS benchmark. The results of the empirical analysis, including the statistical description of the variables and the stationarity tests for first differences, are presented in Table 3.

**Table 3.** Descriptive statistics and stationarity tests for first differences of variables[a]

| Variable | Descriptive statistics (first differences) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Min | Q1 | Median | Mean | Q3 | Max | Std. Dev. |
| ΔComRatio ....... | −0.1764 | −0.0325 | −0.0116 | −0.0111 | 0.0274 | 0.1854 | 0.0812 |
| ln(ΔDisInc) ........ | 3.0272 | 3.8819 | 4.5573 | 4.5433 | 4.8973 | 6.1921 | 0.8765 |
| ΔWUbtrCASCO | −0.1300 | −0.0150 | 0.1100 | 0.2247 | 0.2350 | 1.2900 | 0.3876 |
| Variable | Stationarity tests (first differences) | | | | | | |
| | ADF stat | ADF p-value | Interpretation | KPSS stat | KPSS p-value | Interpretation | Conclusion |
| ΔComRatio ....... | −1.6574 | 0.7029 | Non-stationary | 0.1526 | 0.1000 | Stationary | Ambiguous |
| ln(ΔDisInc) ........ | −4.5182 | 0.0100 | Stationary | 0.5205 | 0.0371 | Non-stationary | Ambiguous |
| ΔWUbtrCASCO | −1.3351 | 0.8257 | Non-stationary | 0.3385 | 0.1000 | Stationary | Ambiguous |

a The terms 'stationary' or 'non-stationary' serve as a concise notation indicating the rejection or failure to reject the null hypothesis of the respective statistical test.

Source: authors' work.

The descriptive statistics and stationarity tests for the first differences of the variables presented in Table 3 indicate that the transformation addressed the issues related to non-stationarity, thus providing a stable foundation for the comparative estimation of both the linear OLS model and the MLP neural network.

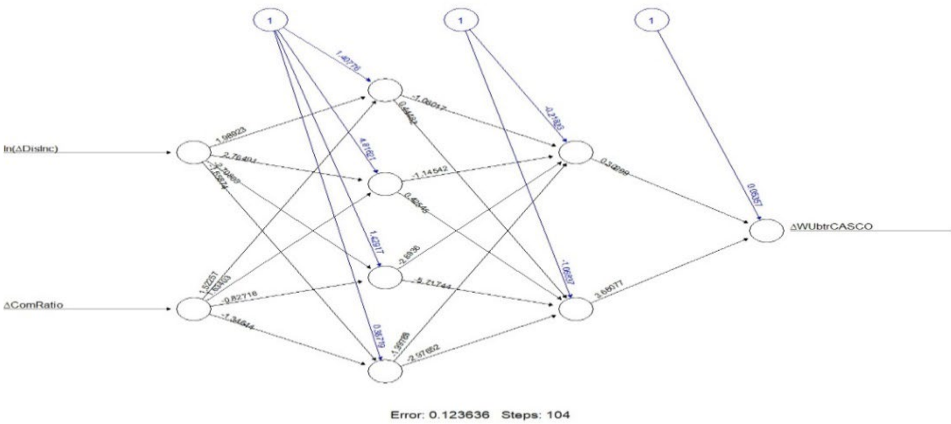
The results of the core empirical analysis, comprising the neural network model estimation (Table 4, Figures 5–6), the OLS linear model estimation (Table 5) and a comparison of their key performance metrics (Table 6), are presented below in accordance with the methodology described in section 4.

**Table 4.** MLP model estimation results

| Category | Parameter / Statistic | Value |
|---|---|---|
| Model architecture | Network type<br>Configuration (hidden layers)<br>Input nodes (variables)<br>Output node<br>Activation function<br>Output activation | multilayer perceptron (MLP)<br>c(4, 2) (two hidden layers)<br>2 (ln(ΔDisInc), ΔComRatio)<br>1 (ΔWUbtrCASCO)<br>Logistic (sigmoid)<br>Linear (identity) |
| Training process ...... | Learning algorithm<br>Convergence steps<br>Final error (SSE) | Resilient backpropagation (Rprop+)<br>104<br>0.1236 |
| Goodness-of-fit ........ | Coefficient of determination ($R^2$)<br>RMSE | 0.7631<br>0.1823 |
| Diagnostic tests ....... | Durbin-Watson (autocorrelation)<br>Shapiro-Wilk (normality) | DW=2.8855 (p-value=0.9681)<br>SW=0.9620 (p-value=0.7276) |

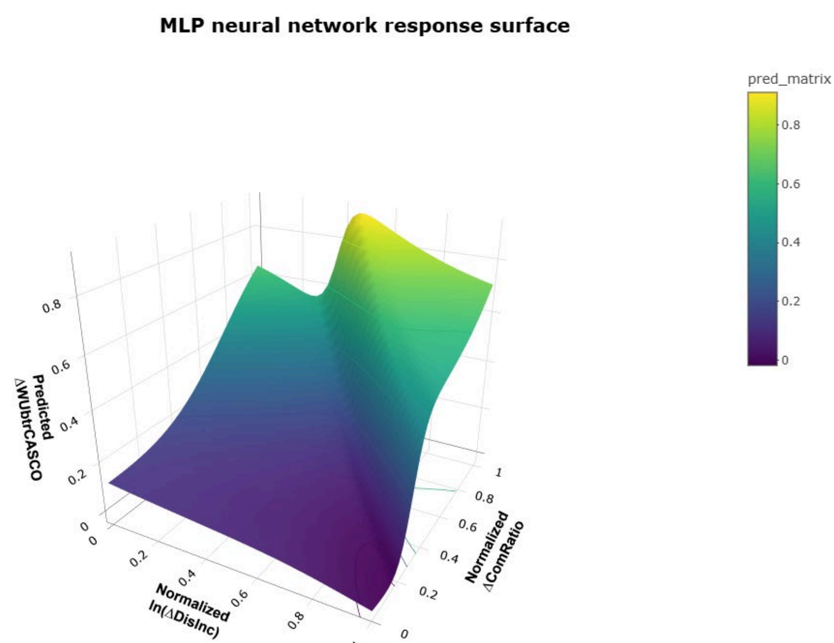
Source: authors' work.

**Figure 5.** Architecture of the MLP neural network model




Source: authors' work.

**Figure 6.** MLP neural network response surface for the analyzed neural network




Source: authors' work.

**Table 5.** The OLS model estimation results

| Category | Parameter / Test | Estimate | Std. Error | t- value | p-value |
|---|---|---|---|---|---|
| | ΔWUbtrCASCO | dependent variable | | | |
| Coefficients .... | ln(ΔDisInc) | 0.06158 | 0.01859 | 3.313 | 0.0056 |
| | ΔComRatio | 1.84524 | 1.08466 | 1.701 | 0.1127 |
| Residuals ......... | Min | –0.33434 | – | – | – |
| | 1Q | –0.26316 | – | – | – |
| | Median | –0.07812 | – | – | – |
| | 3Q | 0.06668 | – | – | – |
| | Max | 0.86296 | – | – | – |
| Model Fit ......... | RMSE | 0.3086 | – | – | – |
| | Coefficient of determination ($R^2$) | 0.5006 | – | – | – |
| | F-statistic | 6.514 | – | – | 0.01097 |
| Diagnostics ..... | Durbin-Watson (autocorrelation) | 1.6494 | – | – | 0.2458 |
| | Shapiro-Wilk (normality) | 0.8325 | – | – | 0.0099 |

Source: authors' work.

**Table 6.** Comparison of the MLP and OLS model estimation results

| Metric | OLS Model | MLP Model |
|---|---|---|
| Coefficient of determination ($R^2$) | 0.5006 | 0.7631 |
| RMSE | 0.3086 | 0.1823 |
| Durbin-Watson (statistic) | 1.6494 | 2.8855 |
| Durbin-Watson (p-value) | 0.2458 | 0.9681 |
| Shapiro-Wilk (statistic) | 0.8325 | 0.9620 |
| Shapiro-Wilk (p-value) | 0.0099 | 0.7276 |

Source: authors' work.

Based on the results of the estimation of the MLP neural network model presented in Table 4 and Figures 5–6, the OLS linear model in Table 5 and the comparative analysis of their key performance metrics in Table 6, it can be noted that:

- the MLP model achieved a high level of predictive accuracy, as evidenced by the coefficient of determination $R^2$ reaching 0.7631;
- the RMSE was maintained at a low level of 0.1823, indicating small deviations of the predicted values from the actual observations;
- the network reached a stable convergence in 104 steps, with a minimal final error (SSE) of 0.1236;
- the Durbin-Watson test statistic of 2.8855 (p-value = 0.9681) confirms the absence of autocorrelation among the residuals (standard significance level p-value = 0.05);
- the Shapiro-Wilk test results (SW = 0.9620, p-value = 0.7276) allow the acceptance of the null hypothesis regarding the normal distribution of the MLP model's residuals (standard significance level p-value = 0.05);

- the Neural Network Response Surface (Figure 7) reveals a clear, non-linear synergy between disposable income (*ln(ΔDisInc)*) and the incentive stimulus (*ΔComRatio*);
- the highest levels of CASCO insurance penetration are observed at the intersection of high household income and the maximum values of the acquisition cost differential ratio;
- the weights estimated for the hidden layers confirm that both the economic status of households and the distribution network's incentive structure are significant predictors in the model;
- the MLP model significantly outperformed the linear model (OLS) in terms of explanatory power, achieving $R^2 = 0.7631$ compared to $R^2 = 0.5006$ for the OLS estimation;
- the neural network reduced the prediction error (RMSE) by approximately 41% (from 0.3086 in the OLS model to 0.1823 in MLP), indicating a much better fit to the market data;
- the Shapiro-Wilk test results (SW = 0.9620, p-value = 0.7276) confirm that the MLP residuals follow a normal distribution, whereas the OLS model failed this diagnostic criterion (SW = 0.8325, p-value = 0.0099) at the standard significance level (p-value = 0.05);
- the Durbin-Watson test indicates that the MLP model (p-value = 0.9681) provides residuals that are more consistent with the assumption of no autocorrelation than the OLS model (p-value = 0.2458).

Based on the points mentioned above, it can be stated that the comparative analysis confirms that the neural network architecture is more effective at capturing the non-linear synergies between economic indicators and insurance expenses presented in these studies than traditional econometric methods.

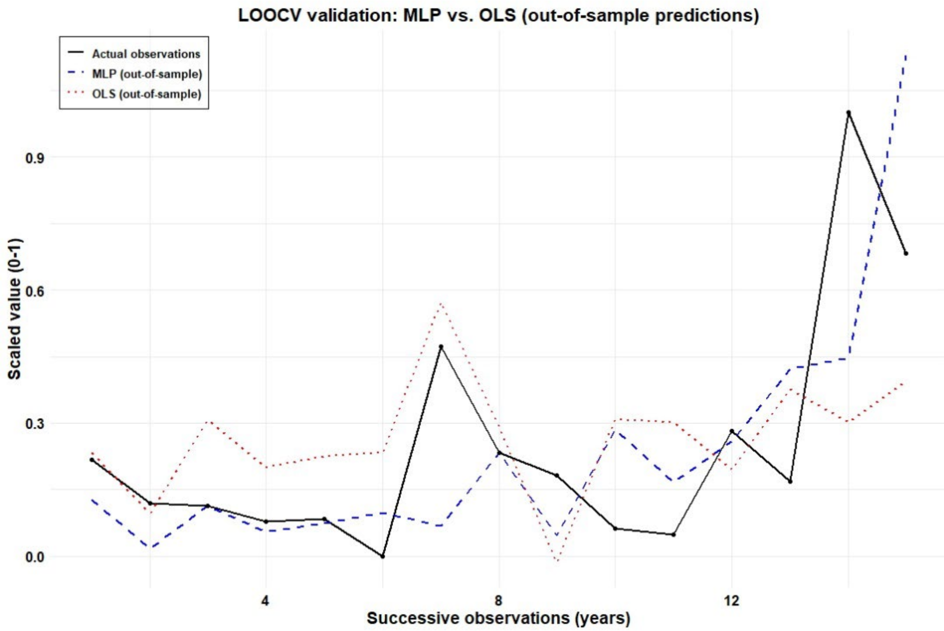
As indicated earlier, the study also utilized the LOOCV procedure to evaluate the predictive performance of the models. The results of this procedure, described in detail in section 4, are presented in Table 7 and Figures 7–9.

**Table 7.** Key summary statistics for the LOOCV procedure

| Metric | OLS Model | MLP Model | Improvement over the linear model |
|---|---|---|---|
| Coefficient of determination (R2) from the LOOCV procedure (out-of-sample) | 0.1135 | 0.1937 | 71% |
| Avg. RMSE from the LOOCV procedure | 0.2483 | 0.2368 | 5% |
| Avg. Mean Absolute Error (MAE) from the LOOCV procedure | 0.1907 | 0.1651 | 13% |
| Avg. Mean Absolute Percentage Error (MAPE) from the LOOCV procedure (in %) | –191.32 | –20.7 | 89% |

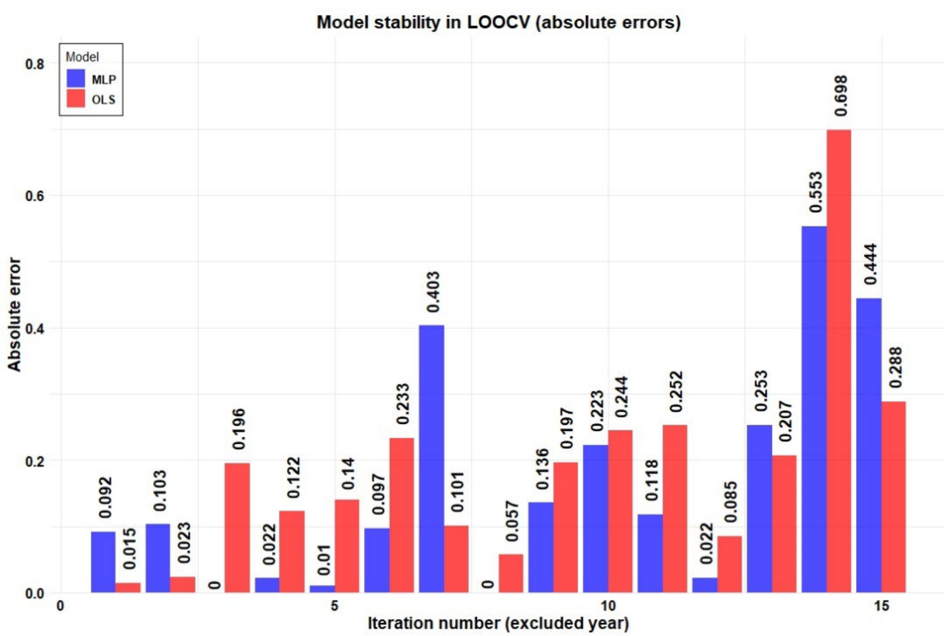
Source: authors' work.

**Figure 7.** Validation of the LOOCV procedure: comparison of the MLP and OLS model results



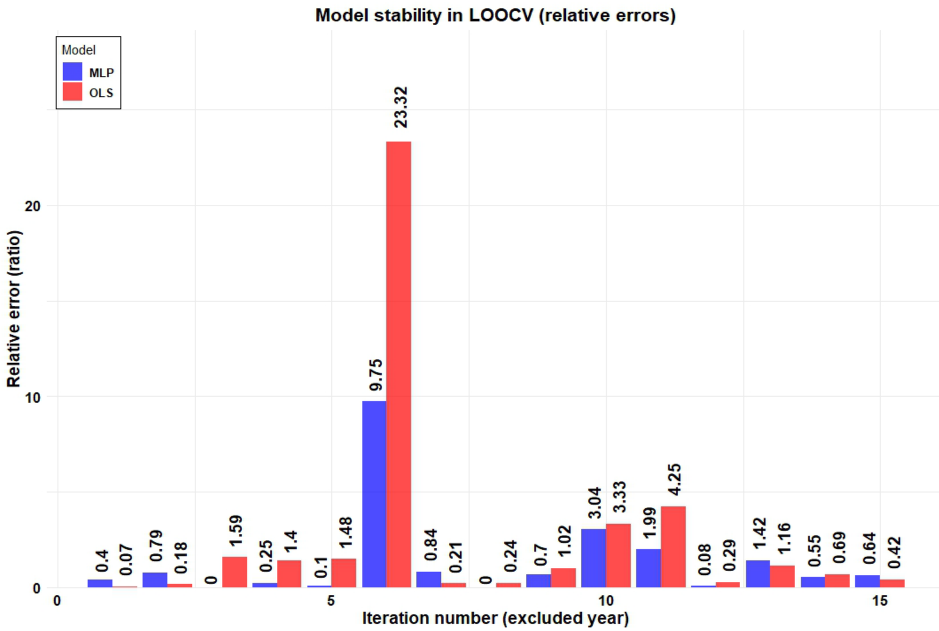

Source: authors' work.

**Figure 8.** Absolute errors observed in the LOOCV procedure




Source: authors' work.

**Figure 9.** Relative errors observed in the LOOCV procedure




Source: authors' work.

Based on the results of the LOOCV procedure, it can be noted that:

- the MLP model demonstrated a significant improvement in predictive stability compared to the linear approach, achieving an out-of-sample $R^2 = 0.1937$, which represents a 71% increase over the OLS model's result ($R^2 = 0.1135$);
- the average prediction error (RMSE) was reduced from 0.2483 in the linear model to 0.2368 in the MLP architecture, confirming the neural network's higher efficiency in generalizing patterns to unseen data;
- the application of the MLP model led to a 13% reduction in the MAE, lowering it to a level of 0.1651 compared to 0.1907 for the OLS estimation;
- the most substantial improvement was observed in the MAPE, where the MLP model achieved a significantly lower error rate of 20.7%, marking an 89% improvement over the unstable linear estimation;
- the LOOCV procedure results (Figures 7–9) confirm that the MLP model more closely follows the actual trend of CASCO insurance expenses, particularly in capturing the turning points that the OLS model failed to recognize.

Furthermore, it is important to underline that while the absolute values of the coefficients in the LOOCV procedure may appear relatively low (such as $R^2$), such results are typical and often expected when dealing with very small sample sizes and rigorous cross-validation techniques. In the LOOCV approach, the model is tested on individual, sometimes extreme observations, which naturally penalizes the goodness-of-fit metrics. Therefore, in this context, the primary indicator of the model's success is not the absolute value of $R^2$ itself, but the substantial relative improvement it offers over the traditional linear baseline, proving the neural network's superior robustness and stability.

To conclude, the empirical results of the study provide strong support for the research hypothesis, confirming that both the change in the acquisition cost ratio of CASCO over MTPL and the change of disposable income are significant determinants in modeling household expenditures on CASCO insurance. The statistical significance of these variables indicates that the demand for optional coverage in Poland is driven not only by the change in financial capacity of consumers (disposable income) but also by changes in the incentive structures within the distribution network, which prioritize the sale of higher-margin voluntary products.

## 6. Discussion and conclusions

The empirical results of this study clearly identify and confirm a dual-driver mechanism that shapes the Polish motor insurance market. The statistically significant relationship between disposable income and expenditures on CASCO policies confirms that voluntary insurance is a superior good. As household wealth (disposable income) increases, consumers not only have the financial capacity to purchase broader protection but also tend to acquire higher-value vehicles that necessitate such coverage (Dragos et al., 2023). As expected, the results obtained also corroborate findings from previous studies, indicating that disposable income is a factor that significantly shapes demand across both life and non-life insurance sectors (Dragos & Dragos, 2017; Duczkowski, 2022; Śliwiński, 2011; Śliwiński, 2016). This implies that along with increasing household wealth, a further expansion of the CASCO insurance market should be expected. However, the strength of these interpretive claims must be balanced against the study's methodological constraints. Since GWP data for CASCO and MTPL insurances were used to disaggregate total transport-related insurance expenditures published by GUS, the observed income elasticity represents a market-wide proxy rather than a granular reflection of individual household-level decisions. This is due to the fact that GWP includes data from both the private and commercial sectors. Despite these data-related constraints, the methodology employed represents the most robust proxy available for this analysis and does not undermine the overall validity of the identified trend.

Simultaneously, the model also highlights the role of the supply side, represented by the acquisition cost ratio variable (*ComRatio*). The persistent surplus of CASCO commissions over mandatory MTPL rates acts as a powerful economic stimulus for intermediaries. This suggests that the growth of the CASCO segment in Poland is not merely a passive result of rising consumer affluence but also an active outcome of a distribution strategy where agents are financially incentivized to push voluntary products to close the profitability gap caused by the low margins of the mandatory sector. It is important to note, however, that due to the method adopted to disaggregate total transport-related insurance expenditures published by GUS, potential differences in the influence of commission structures, specifically within the retail sector, may not be fully captured. This means that the magnitude of the 'commission push' effect requires further validation with micro-level data.

From a market perspective, these findings suggest that insurers must carefully balance their remuneration structures to ensure sustainable growth. For insurance companies, a high reliance on commission-driven sales for CASCO implies that any regulatory changes affecting intermediary compensation could lead to significant shifts in market penetration. This vulnerability is particularly relevant given the study's limited time horizon (2009–2024), which, while covering various cycles, may not fully capture the long-term structural evolution of the sector. On the other hand, it is worth emphasizing, once again, that the ratio of the average MTPL premium to the average national wage dropped significantly from 14.3% in 2017 to just 7.4% in 2024 (BFG, 2025). This increasing affordability of insurance relative to the average income creates a strategic possibility for insurers to bundle products more effectively (MTPL and CASCO or with banking products in the bancassurance channel).

Policy-wise, the results indicate that the Polish market is reaching a stage of maturity where competition is shifting from price-based (MTPL) to value-based (CASCO) coverage. Insurers should therefore focus more on long-term relationship management and cross-selling strategies that leverage the rising disposable income of households. This evolution necessitates moving beyond simple mandatory compliance, which has historically dominated the market, toward a model of comprehensive asset protection. Such a shift implies that insurance is no longer perceived by consumers merely as a legal burden or a fixed cost of vehicle ownership, but as a sophisticated tool for safeguarding household wealth. Consequently, the industry must transition from selling standardized, price-sensitive products to offering integrated financial security solutions. By emphasizing the protection of the vehicle's capital value rather than just fulfilling a statutory obligation, insurers can better align their value proposition with the growing financial aspirations and risk awareness of a more affluent society. This shift is already observable in the market, as evidenced by the proliferation of various assistance services bundled with insurance that provide significant perceived value to the customer beyond the core coverage. Nevertheless, the conclusion regarding market maturity should be interpreted with caution due to the fact that the current data in this study rely on proxy indicators. Thus, the speed and strength of the transition from mandatory compliance to assets protection remain to be fully confirmed by more granular research.

The strong correlation between acquisition costs and household expenditures on CASCO insurance raises serious concerns regarding the quality of distribution and the risk of market distortions. When the incentive gap between mandatory and voluntary products is too wide, there is an inherent danger of commission-driven sales, commonly referred to as mis-selling, where the intermediary's financial gain may take precedence over the actual needs of the policyholder. This is particularly risky in cases where CASCO coverage might be offered for low-value, older vehicles where the economic benefit to the client is marginal. Mis-selling may also occur in situations where the client is not informed about the details of the insurance coverage, its exclusions or other critical terms. While these findings are grounded in the observed statistical correlation, the limitations of proxy-based analysis presented in this research mean that mis-selling should be interpreted as a systemic risk. Consequently, these findings underscore the need for a more robust application of the Demands and Needs Test (Pol. Analiza Potrzeb Klienta – APK). Regulatory oversight should focus on ensuring that the sales process is driven by transparent advice rather than the commission push, protecting consumers from unnecessary costs and maintaining the long-term credibility of the insurance sector. It is worth emphasizing, however, that the KNF already drew attention to this, immediately after the mandatory nature of APK was introduced.

Future research should aim to incorporate micro-level data from individual insurance companies or household surveys to validate these findings. Additionally, future comparative analyses should be expanded to include other Central and Eastern European markets, especially in the context of analyzing the historical evolution of motor insurance. This would also allow deeper insights into the structural evolution of the sector. Such studies could determine whether the shift from mandatory coverage to value-added asset protection is a consistent regional trend driven by rising affluence.

In summary, this research demonstrates that the development of the CASCO motor insurance market in Poland is inextricably linked to both the rising households' disposable income and the economic motivations of the distribution network (acquisition costs). The identification of disposable income and commission differentials as key predictors provides a comprehensive framework for understanding market dynamics. The study concludes that while economic growth provides the foundation for market expansion, it is the structure of acquisition costs that often determines the actual direction and pace of growth. Ultimately, achieving a healthy and sustainable motor insurance market requires a delicate balance between aggressive sales incentives and the rigorous protection of consumer interests through transparent and needs-based distribution practices.

## References


Awunyo-Vitor, D. (2012). Comprehensive Motor Insurance Demand in Ghana: Evidence From Kumasi Metropolis. *Management*, *2*(4), 80–86. https://doi.org/10.5923/j.mm.20120204.01.

Balewski, B., & Janowski, A. (2008). Prowizyjny system wynagradzania agentów ubezpieczeń na życie – narzędzie motywacji czy patologia?. In Z. Jankowska (Ed.), *Dysfunkcje i patologie w sferze zarządzania zasobami ludzkimi* (vol. 2). Wydawnictwo Uniwersytetu Łódzkiego.

Bankowy Fundusz Gwarancyjny. (2025). *Polski rynek ubezpieczeń komunikacyjnych*. https://bfg.pl/wp-content/uploads/polski-rynek-ubezpieczen-komunikacyjnych-2024q3.pdf.

Beenstock, M., Dickinson, G., & Khajuria, S. (1988). The Relationship Between Property-Liability Insurance Premiums and Income: An International Analysis. *The Journal of Risk and Insurance*, *55*(2), 259–272. https://doi.org/10.2307/253327.

Browne, M. J., Chung, J., & Frees, E. W. (2000). International Property-Liability Insurance Consumption. *The Journal of Risk and Insurance*, *67*(1), 73–90. https://doi.org/10.2307/253677.

Chen, Y., & Chen, D. (2013). The review and analysis of compulsory insurance. *Insurance Markets and Companies: Analyses and Actuarial Computations*, *4*(1), 6–17. https://businessperspectives.org/index.php/journals/insurance-markets-and-companies/issue-201/the-review-and-analysis-of-compulsory-insurance.

Cummins, J. D., & Doherty, N. A. (2006). The Economics of Insurance Intermediaries. *The Journal of Risk and Insurance*, *73*(3), 359–396. https://doi.org/10.1111/j.1539-6975.2006.00180.x.

Dąbrowski, I., & Śliwiński, A. (2016). *Economics of Insurance*. Oficyna Wydawnicza SGH.

Dragos, S. L. (2014). Life and non-life insurance demand: the different effects of influence factors in emerging countries from Europe and Asia. *Economic Research – Ekonomska Istraživanja*, *27*(1), 169–180. https://doi.org/10.1080/1331677X.2014.952112.

Dragos, C. M., & Dragos, S. L. (2017). Estimating Consumers' Behaviour in Motor Insurance Using Discrete Choice Models. *E+M Ekonomie a Management*, *20*(4), 88–102. https://doi.org/10.15240/tul/001/2017-4-007.

Dragos, S. L., Mare, C., Mureșan, G. M., & Purcel, A. A. (2023). European motor insurance demand: a spatial approach of its effects and key determinants. *Economic Research – Ekonomska Istraživanja*, *36*(2), 1–20. https://doi.org/10.1080/1331677X.2022.2132348.

Duczkowski, N. (2021a). Wykorzystanie finansów behawioralnych do opisu wybranych zjawisk na polskim rynku ubezpieczeniowym. *Zeszyty Naukowe Polskiego Towarzystwa Ekonomicznego w Zielonej Górze*, (15), 84–102. https://doi.org/10.26366/PTE.ZG.2021.205.

Duczkowski, N. (2021b). Zastosowanie rozkładu α-stabilnego do modelowania zmian cen ubezpieczeń. *Wiadomości Ubezpieczeniowe*, (2), 109–127. https://doi.org/10.33995/wu2021.2.7.

Duczkowski, N. (2022). Application of reduced social welfare functions for estimating household insurance expenditures in Poland. *Wiadomości Ubezpieczeniowe*, (1), 75–96. https://doi.org/10.33995/wu2022.1.5.

Esho, N., Kirievsky, A., Ward, D., & Zurbruegg, R. (2004). Law and the Determinants of Property-Casualty Insurance. *The Journal of Risk and Insurance*, *71*(2), 265–283. https://doi.org/10.1111/j.0022-4367.2004.00089.x.

Feyen, E., Lester, R. R., & de Rezende Rocha, R. (2013). *What drives the development of the insurance sector? An empirical analysis based on a panel of developed and developing countries* (World Bank Policy Research Working Paper No. 5572). https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3076016_code2758723.pdf?abstractid=3076016&mirid=1&type=2.

Główny Urząd Statystyczny. (2010–2025). *Budżety gospodarstw domowych 2009–2024.*

Harrington, S. E., & Niehaus, G. (1998). Race, Redlining, and Automobile Insurance Prices. *The Journal of Business*, *71*(3), 439–469. https://doi.org/10.1086/209751.

Hsu, Y. C., Chou, P. L., Chen, Y. M. J. J., & Lin, J. J. (2014). Mixed logit model of voluntary selection of automobile insurance. *Journal of Information and Optimization Sciences*, *35*(5–6), 503–528. https://doi.org/10.1080/02522667.2014.961823.

Inteliace Research. (2014). *Raport z badania: Prowizje agentów w kanale małych i średnich multiagencji przy sprzedaży ubezpieczeń komunikacyjnych*. https://www.inteliace.com/files/2014/000123_PROWIZJE_MULTI-AGENTOW_UBEZPIECZENIA_KOMUNIKACYJNE.pdf.

Kawiński, M., & Szumlicz, T. (2023). Insurance Under- and Over-performance: The Case of Comprehensive Car Insurance in Poland. *Contemporary Economics*, *17*(4), 443–455. https://doi.org/10.5709/ce.1897-9254.521.

Kumaga, S. B. (2016). *Determinants of Comprehensive Motor Insurance Demand: Evidence from the Recent Premium Tariff Increase in Ghana*. University of Ghana. https://ugspace.ug.edu.gh/bitstreams/048c0192-e5d1-4d1f-91c1-61f39aa4d6d2/download.

Lisowski, J., & Zieniewicz, M. (2015). Zmiany w kanale agencyjnym na rynku ubezpieczeniowym w Polsce. *Annales Universitatis Mariae Curie-Skłodowska. Sectio H (Oeconomia)*, *49*(4), 335–345. https://doi.org/10.17951/h.2015.49.4.335.

Marson, J., & Ferris, K. (2023). The problem of VNUK and the EU response: a critique of the law on compulsory motor vehicle insurance. *Journal of Business Law*, *8*, 639–660. https://shura.shu.ac.uk/31252/.

New Direction. (2024). *Analyzing Changes in Insurance Distribution in Europe. A Comprehensive Study*. https://newdirection.online/publication/analyzing_changes_in_insurance_distribution_in_europe_a_comprehensive_study.

Park, S. C., & Lemaire, J. (2012). The impact of culture on the demand for non-life insurance. *ASTIN Bulletin: The Journal of the IAA*, *42*(2), 501–527. https://doi.org/10.2143/AST.42.2.2182806.

Reifner, U., Neuberger, D., Rissi, R., Riefa, C., Knobloch, M., Clerc-Renaud, S., & Finger, C. (2013). *Study on remuneration structures of financial services intermediaries and conflicts of interest (MARKT/2012/026/H) Final Report.* Institut für Finanzdienstleistungen e.V. https://finance.ec.europa.eu/system/files/2016-12/1311-remuneration-structures-study_en_0.pdf.

Sherden, W. A. (1984). An Analysis of the Determinants of the Demand for Automobile Insurance. *The Journal of Risk and Insurance*, *51*(1), 49–62. https://doi.org/10.2307/252800.

Szymańska, A. (2020). Udział kosztów działalności ubezpieczeniowej w składce zakładów działu II w Polsce. *Finanse i Prawo Finansowe. Journal of Finance and Financial Law*, (2), 115–128, https://doi.org/10.18778/2391-6478.2.26.08.

Śliwiński, A. (2011). Popyt na ubezpieczenia na życie w świetle badań światowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Finanse – nowe wyzwania teorii i praktyki. Ubezpieczenia*, (175), 150–160. https://www.dbc.wroc.pl/Content/73185/Sliwinski_Popyt_Na_Ubezpieczenia_Na_Zycie_w_Swietle.pdf.

Śliwiński, A. (2016). Popyt na ubezpieczenia na życie – przegląd badań światowych. In S. Nowak, A. Z. Nowak & A. Sopoćko (Eds.), *Polski rynek ubezpieczeń na tle kryzysów społeczno-gospodarczych* (pp. 110–125). Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego. https://doi.org/10.7172/978-83-65402-34-9.2016.wwz.11.

Śliwiński, A. (2019). *Rola ubezpieczeń w gospodarce*. Oficyna Wydawnicza SGH.

Śliwiński, A., Dropia, J., & Duczkowski, N. (2021). Risk Factors Affecting Bancassurance Development in Poland. *Risks*, *9*(7), 1–19. https://doi.org/10.3390/risks9070130.

Ustawa z dnia 11 września 2015 r. o działalności ubezpieczeniowej i reasekuracyjnej (Dz.U. 2015 poz. 1844).

Widura, R. (2009). *Commission-Based Pay – Three Essays on German Insurance Tied Agents* [Doctoral dissertation, Universität zu Köln]. https://kups.ub.uni-koeln.de/2933/1/Dissertation_Robert_Widura.pdf.

Wilder, J. M. (2004). *Competing for the Effort of a Common Agent: Contingency Fees in Commercial Insurance* (US Department of Justice Antitrust Division Economic Analysis Group Working Paper No. EAG03-4). https://dx.doi.org/10.2139/ssrn.418061.

Ziemiak, M. P. (2019). Wynagrodzenie pośrednika ubezpieczeniowego w kontekście ustawy o dystrybucji ubezpieczeń. Zagadnienia wybrane. *Wiadomości Ubezpieczeniowe*, (2), 35–49. https://doi.org/10.33995/wu2019.2.4.

Przegląd Statystyczny. Statistical Review, 2026, vol. 73, 1, 56–68 https://doi.org/10.59139/ps.2026.01.3

# 88 years of *Przegląd Statystyczny. Statistical Review*

Czesław Domański[a]

## 1. Introduction

The article is devoted to the history of the *Przegląd Statystyczny. Statistical Review* journal, which has been published for 88 years and has influenced the development of Polish official statistics. The history of the journal has been very complex, as has been Poland's in that period (see Domański, 2011). *Statistical Review* was established as the main organ of the Polish Statistical Association in 1937. In its long history, it was published by several organisations: the Polish Statistical Association, the Statistics Section of the Polish Economic Society, the Committee of Statistics and Econometrics of the Polish Academy of Sciences, and currently, since 2018, by Statistics Poland. The presented historical outline of *Przegląd Statystyczny. Statistical Review* demonstrates that the journal has always held a very high scientific rank. The list of authors and editors has encompassed many outstanding scientists representing mathematical and statistical disciplines.

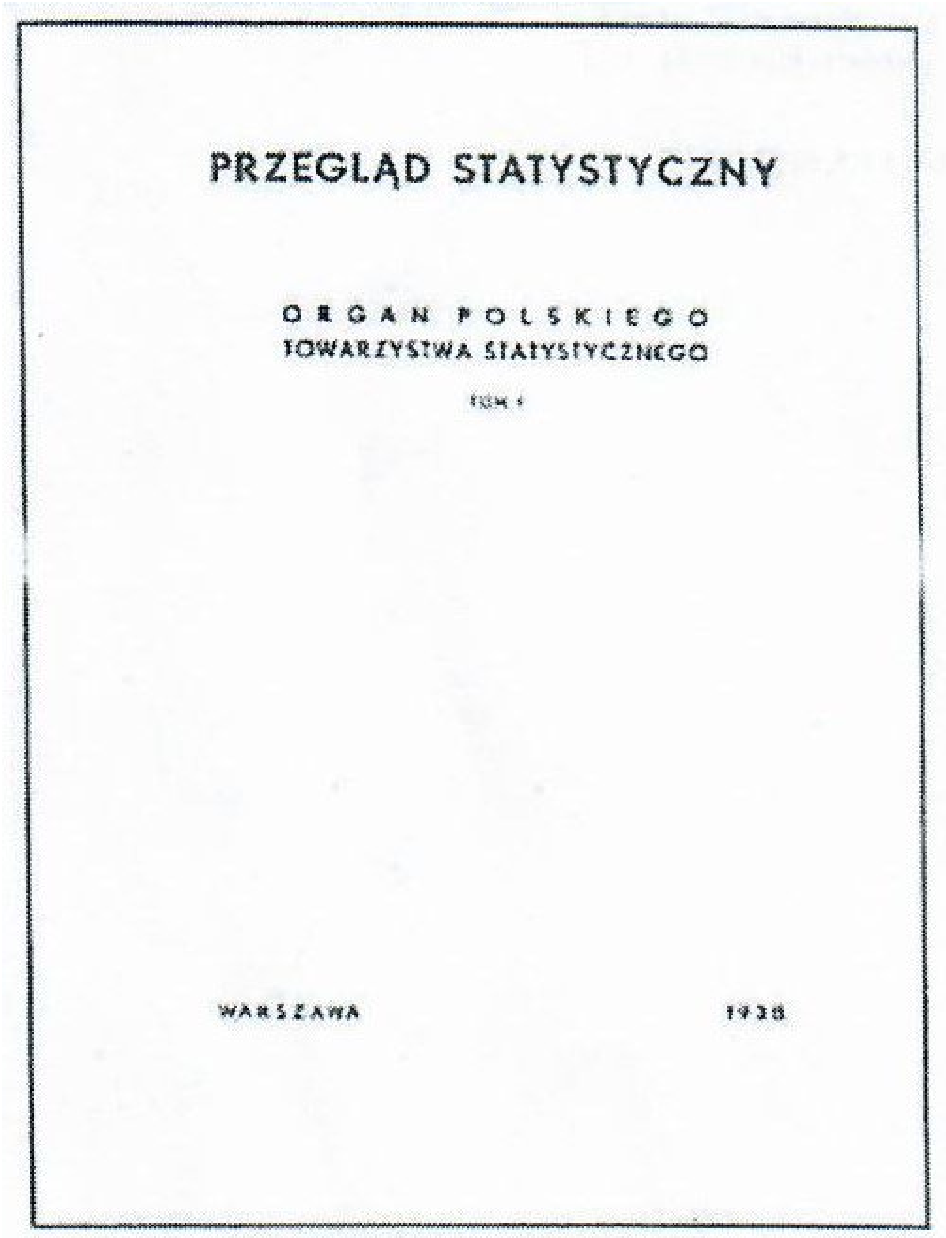
## 2. Early history

The journal *Ekonomista* (Eng. The Economist), established in 1864, might be regarded as a precursor of *Przegląd Statystyczny. Statistical Review*, which was established 73 years later, in 1937. On 1st November of that year, the Editorial Committee of *Przegląd Statystyczny. Statistical Review* was elected at a meeting of the Council of the Polish Statistical Association, consisting of: Zygmunt Limanowski (Chairman), Stefan Szulc (Deputy Chairman), Jan Wiśniewski (Secretary), Antoni Łomnicki and Jan Piekałkiewicz (Members).

The journal became the main organ of the Polish Statistical Association (cf. title page of Volume I). The authors of the articles in the first volume, published in 1938, included the following scholars: Edward Szturm de Sztrem, Jan Czechowski, Jan Wiśniewski, Egon Vielrose, Jerzy Spława-Neyman and Ludwik Krzywicki. We can say that for nearly 90 years, almost every statistician who published an article in *Statistical Review*, later become a well-recognised scholar popular in the community of Polish statisticians.



[a] University of Lodz, Faculty of Economics and Sociology, Institute of Statistics and Demography, ul. Rewolucji 1905 r. 41/43, 90–214 Łódź, Poland, e-mail: czeslaw.domanski@uni.lodz.pl, ORCID: https://orcid.org/0000-0001-6144-6231.


PRZEGLĄD STATYSTYCZNY

ORGAN POLSKIEGO
TOWARZYSTWA STATYSTYCZNEGO

TOM I

WARSZAWA 1938

It is worth quoting from the 'Introduction' to the first volume and the table of contents (*Słowo wstępne*, 1938, pp. 1–2):

The Polish Statistical Association has identified as one of its main tasks to publish a journal devoted to the theory and practical applications of statistics.

Administrative statistics, demography, economic and social life and life sciences all benefit from joint experiences and common statistical practice; they use the same methods, while perfecting and developing them. Statisticians' interest therefore reaches far beyond the scope of problems researched by them, which is often relatively narrow.

It is understandable, then, that we need a journal that would report on statistical work performed in different fields, thus facilitating the exchange of scientific thought and the coordination of efforts.

An especially strong need for such a journal in Poland became evident when the *Statistical Quarterly* (Pol. Kwartalnik Statystyczny) discontinued its operations. That journal, even though devoted mostly to administrative statistics, strove to satisfy more comprehensive needs.

*Przegląd Statystyczny*, as far as possible, will be issued four times a year.

It will feature: 1) original theoretical papers by Polish and foreign authors, 2) original papers in the field of applied statistics, important from the methodological point of view, 3) reports showing the current status and progress in selected branches of statistics, 4) reviews and critical analyses of statistical research, 5) a chronicle informing about the orgnisational and scientific life of the Polish Statistical Society, its sections and branches, and about similar foreign and international organisations, and 6) a detailed bibliography of Polish publications (...)

# SŁOWO WSTĘPNE

Polskie Towarzystwo Statystyczne postawiło sobie jako jedno z naczelnych zadań wydawanie czasopisma poświęconego teorii i praktyce statystycznej.

Statystyka administracyjna, demografia, życie gospodarcze i społeczne, nauki przyrodnicze korzystają ze wzajemnych doświadczeń i wspólnej praktyki statystycznej, korzystają z tych samych metod, doskonalą je i rozwijają. Zainteresowania statystyków sięgają więc daleko poza wąski nieraz zakres badanych przez nich zagadnień.

Zrozumiała jest w tych warunkach potrzeba czasopisma, które by informowało o pracach statystycznych, podejmowanych w tych różnych dziedzinach, ułatwiając wymianę myśli i koordynację wysiłków.

Potrzeba takiego czasopisma w Polsce zaznaczyła się szczególnie silnie z chwilą zawieszenia Kwartalnika Statystycznego, który choć poświęcony przede wszystkim statystyce administracyjnej, starał się zaspakajać i szersze potrzeby.

Przegląd Statystyczny w miarę możności ma się ukazywać 4 razy do roku.

Przegląd Statystyczny ogłaszać będzie: 1) oryginalne prace teoretyczne autorów polskich i obcych, 2) oryginalne prace z zastosowań statystyki, ważne ze względów metodologicznych, 3) referaty przedstawiające stan lub postępy pewnej gałęzi statystyki, 4) recenzje i krytyki dochodzeń statystycznych, 5) kronikę informującą o życiu organizacyjnym i naukowym P. T. S. oraz jego sekcyj i oddziałów, jak również pokrewnych instytucyj zagranicznych i międzynarodowych, 6) szczegółową bibliografię wydawnictw polskich.

Krótki czas, którym przy przygotowaniu niniejszego numeru rozporządzała Redakcja, nie pozwolił rozwinąć działu recenzji. Z tego samego powodu nie zdołano wykończyć opracowania szczegółowej bibliografii za 1937 r. Braki te zostaną wyrównane w następnych numerach.

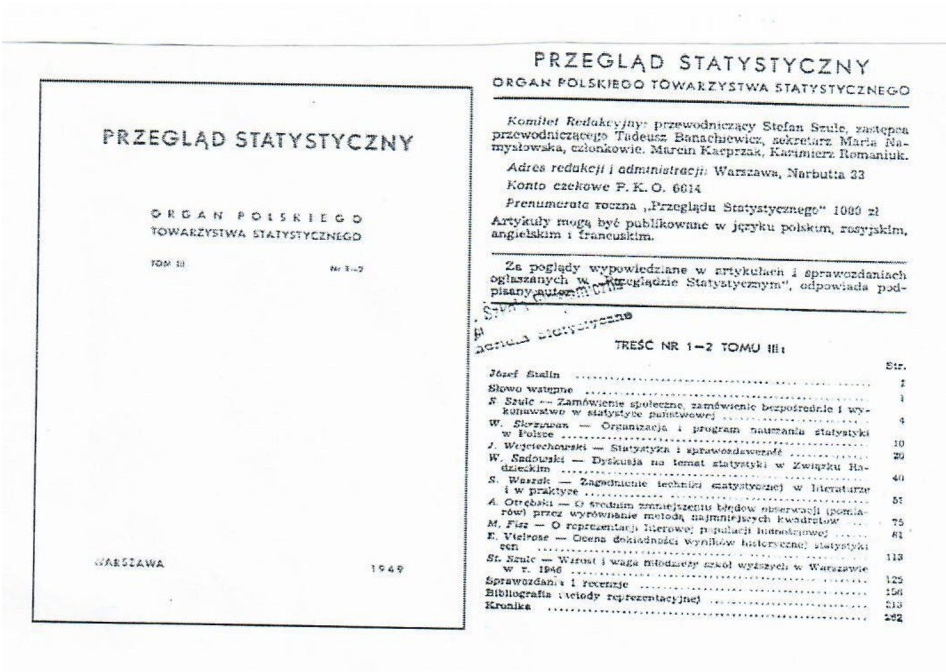
Oddając ten pierwszy numer Przeglądu Statystycznego do rąk czytelników, Redakcja sądzi, iż mimo swych usterek będzie on przez nich życzliwie przyjęty i znajdzie czynne poparcie wśród wszystkich osób pracujących na niwie statystycznej.

REDAKCJA

And this is what the Editorial Team wrote about Volume II of Statistical Review, published in 1939 (*Od Redakcji*, 1939, p. 1):

> As the date of the third census in Poland is approaching, the editors of *Statistical Review* decided to publish a series of papers discussing problems related to the census from various points of view. This issue is almost entirely devoted to these considerations; and more census-related articles will be published in subsequent issues. Until now, the following authors have confirmed they would write for us: Jerzy Heinrich, Zygmunt Limanowski, Wiktor Morawski, Edward Strzelecki, Edward Szturm de Sztrem and Stefan Szulc.
>
> Even issue 1 of *Statistical Review* published in 1938 featured an article related to censuses, namely *The Universal Agricultural Census* by Mieczysław Przypkowski (1938).
>
> This topic has already been incorporated into the agenda of the general scientific meeting of the Polish Statistical Association, scheduled for 2nd–3rd April 1939 in Warsaw, and the Congress of Municipal Statisticians (18th–19th March 1939). Detailed reports from both congresses will be published in issue 2 of *Statistical Review*.

The third volume of the journal was published in 1949 (the title page and contents are presented below). We can learn from them that articles sent to the journal at that time could be written in Polish, Russian, English and French. The topics of the articles in that issue were varied. They included: public procurement in official statistics, the organisation of statistical education in Poland, statistical reporting, statistical methodology, historical price statistics, official statistics in the Soviet Union, and the height and weight of students of higher schools in Warsaw in 1946.




PRZEGLĄD STATYSTYCZNY

ORGAN POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO

*Komitet Redakcyjny:* przewodniczący Stefan Szulc, zastępca przewodniczącego Tadeusz Banachiewicz, sekretarz Maria Namysłowska, członkowie: Marcin Kacprzak, Kazimierz Romaniuk.

*Adres redakcji i administracji:* Warszawa, Narbutta 33

*Konto czekowe* P. K. O. 6614

*Prenumerata* roczna „Przeglądu Statystycznego" 1000 zł

Artykuły mogą być publikowane w języku polskim, rosyjskim, angielskim i francuskim.

Za poglądy wypowiedziane w artykułach i sprawozdaniach ogłaszanych w „Przeglądzie Statystycznym", odpowiada podpisany autor.

TREŚĆ NR 1–2 TOMU III:



## 3. *Statistical Review* as the main organ of the Statistics Section of the Polish Economic Society

In the first issue of the 'new' *Statistical Review*, published in 1954, in the 'From the Editors' section, we have information about the procedure for establishing a new journal.

From 29th June to 21st July 1951, the First Congress of the Polish Science was held in Warsaw. The decisions made at the congress resulted in the transformation of the Polish science and started the process of monitoring it. The changes concerned both the organisational system and research methods, as well as the language and the way science was taught. In the wake of these transformations, the Polish Academy of Sciences (Pol. Polska Akademia Nauk – PAN) was established (Act of 30th October 1951, Journal of Laws of 1951 No. 57, item 391). It was designed as an institution supporting the political authorities and a tool of control over the scientific community. Also, in 1951, the Central Qualification Commission for Scientific Staff was set up, which had a significant impact on the personnel policy in the academic community. Many scientists had to leave their workplaces or found themselves on the margins of the scientific life. Jan Czekanowski, Adam Krzyżanowski, Marcin Nadomnik, Stefan Schmidt, Edward Strzelecki and Józef Wojtyniak were among the politically persecuted members of the management of the Polish Statistical Society (see Kruszka, 2012, p. 56).

In December 1952, the first National Conference of Faculties of Higher Schools of Economics and Universities was held in Warsaw. It was attended by faculties of sciences connected to statistics from the following higher schools and universities: the Warsaw School of Planning and Statistics, the Warsaw School of Foreign Service, the universities of economics in Częstochowa, Krakow, Łódź, Poznań, Sopot, Stalingrad, Szczecin and Wrocław, and the universities of Wrocław, Łódź, Warsaw and Poznań. Romaniuk (1954, p. 5) wrote:

> When considering the scientific and research program (...), we must first of all refer to the achievements of the First Congress of the Polish Science, and in particular to those of the Economic Sciences Section and the Subsection of Statistics, and then to the achievements of the National Conference of Faculties of Higher Schools of Economics (...).

At this conference, ignoring the formal existence of the Polish Statistical Association, its participants requested the establishment of a Statistics Section within the Polish Economic Society. It started operations in March 1953. It was the only such organisation accepted and endorsed by the then authorities, so the only place where Polish statisticians could relatively freely engage in social activity (see also Kruszka, 2012, p. 57). According to Romaniuk (1954, pp. 5–6):

> As scientific and research work in the field of statistics continues to develop, the forms of this work will also improve. The basic condition for the development of both the content and the form of all scientific and research work is its service to the needs of a man, the needs of practice, the needs of the implementation of socialism in our country.

The structure of the individual volumes of *Statistical Review* from the years 1938–1949 and the one established in 1954 was very similar, with the exception of the 'From the Editors' and the editorial articles.

Studying the publications in the journal (and other sources), we not only learn about the development of the Polish statistical thought, but also about the very interesting political, economic and social background. We can also see that the ethical standards of Statistics Poland's staff were very high. Let us quote from the article by Romaniuk entitled *The 1955 Statistical Yearbook was published* (1956 issue 1, pp. 239–241):

> The publication of the Statistical Yearbook after a six-year break is a cultural and political event of a great importance:
> Welcoming the publication of the 1955 Statistical Yearbook, we, the editors of *Statistical Review*, would at the same time like to emphasise the necessity of a thorough scientific

discussion about the Yearbook, concerning both its content and the methods of obtaining the published data. We would be happy to host such a discussion in our journal.

## 4. *Statistical Review* as the main organ of the Committee on Statistics and Econometrics of the Polish Academy of Sciences

On 11th April 1972, the Committee on Statistics and Econometrics was established by Resolution No. 14/72 of the Presidium of the Polish Academy of Sciences.

On 1st January 1974, *Statistical Review* became the main organ of the Committee on Statistics and Econometrics of the Polish Academy of Sciences, and therefore the authorities of the Polish Academy of Sciences appointed a new, expanded Editorial Committee composed of: Andrzej Baranowski, Czesław Bracha, Zbigniew Czerwiński (Deputy Editor-in-Chief), Wiesław Grabowski, Jerzy Greń, Jan Kordos, Adam Machnowski, Ireneusz Nykowski, Zbigniew Pawłowski (Deputy Editor-in-Chief), Kazimierz Romaniuk, Wiesław Sadowski (Editor-in-Chief), Oskar Starzeński, (Editorial Secretary), Krzysztof Zadora and Aleksander Zeliaś.

In May 1982, the Scientific Secretary of the Polish Academy of Sciences selected the following members of the Editorial Committee of *Statistical Review*: Zbigniew Czerwiński (Editor-in-Chief), Witold Jurek (Editorial Secretary), Wojciech Maciejewski, Ireneusz Nykowski (Deputy Editor-in-Chief), Antoni Smoluk, Andrzej Tomaszewicz, Krzysztof Zadora and Leszek Zienkowski. The journal's Editorial Board was also appointed. It consisted of: Zdzisław Hellwig, Tadeusz Kasprzak, Michał Kolupa, Mirosław Krzysztofiak, Kazimierz Romaniuk, Wiesław Sadowski (Chairman), Igor Timofiejuk, Władysław Welfe and Kazimierz Zając.

The new Committee began work in a situation where, for reasons independent of their predecessors, the publication of the 1980 and 1981 volumes was seriously delayed. It was already mid-1982 when the first issue from 1982 was published. It mostly consisted of articles submitted earlier (as indicated in the annotations to articles).

In the years 1993–2007, Michał Kolupa served as the Editor-in-Chief of *Statistical Review*. He held this position for 15 years, and he described this experience in the article entitled *From June 1993 to June 2007* (Kolupa, 2007, pp. 5–7).

The Editorial Board was then composed of: Maria Cieślak, Zbigniew Czerwiński, Zdzisław Hellwig, Jan Kordos, Mirosław Krzysztofiak, Wojciech Maciejewski, Wiesław Sadowski (Chairman), Władysław Welfe, Kazimierz Zając and Aleksander Zeliaś. The Editorial Committee was comprised of: Andrzej Barczak (Deputy Editor-in-Chief), Czesław Domański, Józef Hozer, Krzysztof Jajuga (Deputy Editor-in-Chief), Michał Kolupa (Editor-in-Chief), Teodor Kulawczuk, Emil Panek and Tadeusz Stanisz (Scientific Secretary).

As far as particular fields of knowledge were concerned, Editor-in-Chief Michał Kolupa was responsible for econometrics and its applications, Deputy Editor-in-Chief Andrzej Barczak dealt with statistics and its applications, Deputy Editor-in-Chief Krzysztof Jajuga specialised in econometrics, financial engineering and other financial applications of mathematics, while operational research, its applications and other applications of mathematics in economic research was the domain of Tadeusz Stanisz, the Secretary of the Editorial Board.

In volume 56 issued in 2009, the new composition of the Editorial Board and the Editorial Committee was presented. The Editorial Board was then composed of: Maria Cieślak, Zbigniew Czerwiński, Zdzisław Hellwig, Jan Kordos, Mirosław Krzysztofiak, Wojciech Maciejewski, Wiesław Sadowski (Chairman), Władysław Welfe, Kazimierz Zając and Aleksander Zeliaś. The members of the Editorial Committee were: Mirosław Szreder (Editor-in-Chief), Magdalena Osińska (Deputy Editor-in-Chief), Marek Walesiak (Deputy Editor-in-Chief) and Krzysztof Najman (Scientific Secretary).

That issue also featured a 'Thank you' letter to the former Editor-in-Chief Michał Kolupa. It read (*Podziękowanie*, 2009, p. 3):

> (...) We would like to express our sincere gratitude to Professor Michał Kolupa for his expert management of the Editorial Team of *Przegląd Statyczny. Statistical Review* over the past 15 years. Thanks to the efforts and commitment of Professor Kolupa, *Przegląd Statystyczny*, an important scientific journal of statisticians and econometricians, was published regularly, presenting research results and discussions conducted in both the above-mentioned, closely interwoven fields.
>
> We would like to thank all the members of the Editorial Committee for upholding high scientific standards of the journal, the topicality of the published studies, and for their efforts aimed at a greater integration of the communities of statisticians and econometricians in the research sphere. (...)
>
> We would like to express our hope for the future fruitful collaboration with the outgoing Editorial Committee and for their continuing active participation in the development of *Statistical Review*.

The letter was signed by Mirosław Szreder, PhD, DSc, ProfTit, Magdalena Osińska, PhD, DSc, ProfTit , Marek Walesiak, PhD, DSc, ProfTit, and Krzysztof Najman, PhD, DSc, ProfTit.

Let us also quote from the new Editorial Committee's 'From the Editors' letter emphasising the importance of the 55-year history of *Statistical Review* (*Od Redakcji*, 2009, p. 5).

> (...) Being aware of an over 55-year history of our journal as well as of its excellent scientific reputation, we (...) would like *Statistical Review* to continue being the first choice

> of Polish statisticians and econometricians for the exchange of scientific thought. In a developing market economy, this requires special attention to the adequate scientific standards of the journal, openness to the expectations of readers, and reaching out to new audiences, including foreign readers. We would like to use modern means of information to promote scientific work published in *Statistical Review* more boldly. We believe that in this way we will broaden and deepen the scope of scientific discussion on the most important aspects of the theory and applications of statistics and econometrics.
>
> We would like to declare our openness both to new scientific problems that should be discussed in *Statistical Review* and to any proposals of our colleagues and readers aimed at modernising the journal and raising its scientific rank.

The letter was signed by the same as above members of the Editorial Committee.

The Committee on Statistics and Econometrics of the Polish Academy of Sciences was still the publisher of *Statistical Review* when Volume 59 of the journal was issued in 2012. The composition of the Editorial Committee and message to the readers can be found in the 'From the Editors' article (*Od Redakcji*, 2012, p. 3):

> (...) The Committee on Statistics and Econometrics of the Polish Academy of Sciences, acting on behalf of the Chairman of the Committee, Professor Jacek Osiewalski, appointed a new Editorial Committee of *Statistical Review* and a new Editorial Board headed by Professor Władysław Milo.
>
> The new Editorial Committee consists of: Magdalena Osińska (Editor-in-Chief), Marek Walesiak (Deputy Editor-in-Chief), Piotr Fiszeder (Scientific Secretary of the Editorial Board), and Editors Anna Pajor and Michał Majsterek.
>
> We would like to express our gratitude to the outgoing Editor-in-Chief, Professor Mirosław Szreder, for his efforts and commitment to the editorial work, which resulted in maintaining high scientific standards of the journal. (...)
>
> We would also like to thank all the members of the Editorial Board (...). In this issue, we are publishing the memoir of the recently deceased Chairman of the Council, Professor Michał Kolupa.
>
> In its nearly 60-year history, our journal has always been the most important platform for the exchange of scientific thought between researchers dealing with the broadly understood quantitative methods, especially statistics and econometrics. Respecting the achievements of the journal to date, the new Editorial Committee will spare no efforts to raise its scientific level, to modernise the journal through electronic means of communication (including electronic contact with authors and reviewers, regular updating and expansion of the scope of information on the website, and seeking membership in more journal databases), and to upgrade the journal's parametric evaluation.

The increasing importance of *Statistical Review*'s internationalisation was emphsised in issue No. 63 from 2016 (*Od Redakcji*, 2016, pp. 5–6):

> (...) In 2015, the Editorial Board of *Przegląd Statystyczny. Statistical Review* published, as intended, four issues, most of the time with a quarterly frequency, including one issue in English. This was the second year when it was possible to publish an issue fully in English, which, coupled with the availability of the journal's electronic version, greatly expanded our audience.
>
> Publishing scientific journals in English is now an inevitable trend, and therefore we would like to encourage our authors to submit papers to *Statistical Review* in this language. In the current calendar year, which is already the 79th year of our journal's operations (...), our intention is to publish two issues entirely in English. The justification for popularising scientific results and conducting scientific discussion in English is not only that it is a fashion of the 21st century (...). We can also refer here to the first issue of *Statistical Review* published in 1938, in which the then Editorial Board clearly indicated in the introduction that the journal would publish 'original theoretical papers by Polish and foreign authors' (...), while providing the English name of the journal (*Statistical Review*), as well as the titles of articles in English and reviews of publications in foreign languages (mainly English and German). This is the tradition upon which we have been drawing until today. It is worth recalling that the creators of *Statistical Review* were outstanding Polish statisticians, who at the same time belonged to the group of the world's greatest statisticians, such as Professor Jerzy Neyman and Professor Jan Czekanowski.
>
> Over the years, however, significant changes have taken place. Statistics as a science has evolved in various directions, creating the foundations for econometrics, biometrics or psychometrics, while statistical research has increasingly often been used in various fields of knowledge, which frequently led to the emergence of new statistical methods and techniques. This creates ample possibilities for publishing research results, often different from the profile of *Statistical Review*. The number of journals publishing the results of the latest research in the field of statistics and econometrics has also significantly increased. In order to differentiate our journal from other similar periodicals and to achieve our goals and tasks, *Statistical Review* refers to the best traditions of Polish official statistics (...) by publishing the best submitted papers and performing an extremely reliable review procedure.
>
> I hope that the pursuit of excellence in publications will guide the entire community of Polish statisticians and econometricians, which will result in an increasing number of outstanding publications and in *Statistical Review*'s improved rating among the world's scientific journals.

The letter from the Editor-in-Chief was signed by Magdalena Osińska, PhD, DSc, ProfTit.

## 5. Statistics Poland becomes the publisher of *Statistical Review*

In 2018, Statistics Poland became the publisher of *Statistical Review*, and it has performed this role until today. This change resulted from the alterations in the publishing policy of the Polish Academy of Sciences, but, overall, did not affect the general profile of the journal. The Committee on Statistics and Econometrics of the Polish Academy of Sciences is still co-responsible for upkeeping high scientific standards of the journal. *Statistical Review* remains one of the leading Polish scientific periodicals in the field of econometrics, statistics, mathematical economics, operational research and other subdisciplines covering broadly understood quantitative methods in economics and their original applications.

At the beginning of this period, the journal was led by Paweł Miłobędzki of University of Gdańsk, PhD, DSc, Prof.Tit, who served as its Editor-in-Chief until 2025. Marek Walesiak of the Wroclaw University of Economics and Business, PhD, DSc, Prof.Tit was the journal's Deputy Editor-in-Chief, while Piotr Fiszeder (Nicolaus Copernicus University in Toruń), Maciej Nowak (University of Economics in Katowice), Emilia Tomczyk (SGH Warsaw School of Economics) and Łukasz Woźny (SGH Warsaw School of Economics, Poland) served as co-Editors. Dorota Ciołek of University of Gdańsk was the journal's Secretary and Managing Editor.

Krzysztof Echaust of Poznań University of Economics and Business, PhD, DSc, has served as the journal's Editor-in-Chief since March 2025. The co-Editors are recognised scholars both from Poland and abroad: Piotr Fiszeder (Nicolaus Copernicus University in Toruń), Michał Jakubczyk (SGH Warsaw School of Economics), Bogumił Kamiński (SGH Warsaw School of Economics), Gabor David Kiss (University of Szeged, Hungary), Aleksandra Łuczak (Poznań University of Life Sciences), Silvana Musti (University of Foggia, Italy), Maciej Nowak (University of Economics in Katowice), Monika Papież (Krakow University of Economics), Emilia Tomczyk (SGH Warsaw School of Economics) and Łukasz Woźny (SGH Warsaw School of Economics, Poland). Małgorzata Tymińska of Statistics Poland has been the Secretary and Managing Editor of the journal.

## 6. Conclusion

In 1937, *Przegląd Statystyczny. Statistical Review* quarterly was established as the main organ of the Polish Statistical Association. In the years 1950–1956, the total political control of the Communist apparatus over the social and economic life in Poland resulted in the discontinuation of the work of various scientific organisations, including the Polish Statistical Association, whose activities collided with the direction of the politics of Communist countries.

At that time, *Statistical Review* was 'adopted' by the Statistics Section of the Polish Economic Society. In the years 1954–1973, *Statistical Review* was published under the editorship of Prof. Kazimierz Romaniuk, as the main organ of the above-mentioned organisation.

In the years 1974–2018, *Statistical Review* started being published by the Committee on Statistics and Econometrics of the Polish Academy of Sciences. During this period, the journal was led by: Wiesław Sadowski, Zbigniew Czerwiński, Michał Kolupa, Mirosław Szreder, Magdalena Osińska and Paweł Miłobędzki. Since 2018, Statistics Poland has been the publisher of Statistical Review. Paweł Miłobędzki was succeeded by the current Editor-in-Chief Krzysztof Echaust in 2025.

Thanks to the endeavours of *Przegląd Statystyczny. Statistical Review*'s editors-in-chief and all the scientific community involved with the journal for over 88 years, despite several difficult moments in its history, *Statistical Review* has remained one of Poland's most recognised and esteemed statistical journals and one of the best platforms for exchanging scientific thought, especially in the field of statistics and econometrics.

## References


Domański, C. (2011). Losy *Przeglądu Statystycznego*. *Przegląd Statystyczny*, *58*(3–4), 348–357. https://ps.stat.gov.pl/Article/2011/3-4/348-357.

Kolupa, M. (2007). Od czerwca 1993 do czerwca 2007. *Przegląd Statystyczny*, *54*(1), 5–7.

Kruszka, K. (red.). (2012). Polskie Towarzystwo Statystyczne w latach 1947–1955. In *Polskie Towarzystwo Statystyczne 1912–2012* (pp. 49–57). https://pts.stat.gov.pl/media/pts1912-2012_ksiazka.pdf.

Oczapowski, J. B. (1874). Statystyka i drukowanie. Lekcja wprowadzająca do statystyki porównawczej na Uniwersytecie Jagiellońskim. *Ekonomista*, *8*, 1007.

Od redakcji. (1939). *Przegląd Statystyczny*, *2*(1), 1.

Od redakcji. (2009). *Przegląd Statystyczny*, *56*(1), 5. https://ps.stat.gov.pl/PS/2009/1/PS_2009_56_1_full.pdf.

Od redakcji. (2012). *Przegląd Statystyczny*, *59*(1), 3. https://ps.stat.gov.pl/PS/2012/1/PS_2012_59_1_full.pdf.

Od redakcji. (2016). *Przegląd Statystyczny*, *63*(1), 5–6. https://ps.stat.gov.pl/PS/2016/1/2016_63_1_from_the_editorial_team.pdf.

Podziękowanie. (2009). *Przegląd Statystyczny*, *56*(1), 3. https://ps.stat.gov.pl/PS/2009/1/PS_2009_56_1_full.pdf.

Przypkowski, M. (1938). Powszechny spis rolny. *Przegląd Statystyczny*, *1*(2), 3–41.

Romaniuk, K. (1954). W sprawie programu i organizacji pracy naukowo-badawczej w statystyce w Polsce. *Przegląd Statystyczny*, *17*(1–2), 5–20.

Romaniuk, K. (1956). Rocznik Statystyczny 1955 opublikowany. *Przegląd Statystyczny*, *19*(1), 239–241.

Romanowicz, T. (1877). Międzynarodowy Kongres Statystyczny w Peszcie. *Ateneum*, *1*, 328–356.

Słowo wstępne. (1938). *Przegląd Statystyczny*, *1*(1), 1–2.

Wodzicki, H. (1856). Statystyka i kongresy statystyczne. *Czas*, *1*, 624–639.

# Acknowledgements

## The Editorial Team of *Przegląd Statystyczny. Statistical Review* would like to thank the distinguished scholars whose names are listed below for devoting their time and expertise to reviewing the articles submitted for publication in 2025:

**Jan Acedański**, University of Economics in Katowice
**Sylwester Bejger**, Nicolaus Copernicus University in Torun
**Klara Cermakova**, Prague University of Economics and Business, Czechia
**Abdelhak Chouaf**, Djillali Liabes University, Algeria
**Marcin Czupryna**, Krakow University of Economics
**Krzysztof Dmytrów**, University of Szczecin
**Ewa Dziwok**, University of Economics in Katowice
**Viviana Fanelli**, University of Foggia, Italy
**Romana Głowicka**, Poznań University of Life Sciences
**Grzegorz Hałaj**, University of Gdańsk
**Małgorzata Just**, Poznań University of Life Sciences
**Agata Kliber**, Poznań University of Economics and Business
**Paweł Kliber**, Poznań University of Economics and Business
**Jerzy Korzeniewski**, University of Lodz
**Dominik Krężołek**, University of Economics in Katowice
**Paweł Kropiński**, Polish Academy of Sciences
**Joanna Landmesser-Rusek**, Warsaw University of Life Sciences
**Krzysztof Malaga**, Poznań University of Economics and Business
**Krzysztof Najman**, University of Gdańsk
**Anna Pajor**, University of Economics in Katowice
**Mateusz Pipień**, Krakow University of Economics
**Tomasz Potocki**, University of Rzeszów
**Grażyna Trzpiot**, University of Economics in Katowice
**Michał Rubaszek**, SGH Warsaw School of Economics
**Aleksandra Rutkowska**, Poznań University of Economics and Business
**Roma Ryś-Jurek**, Poznań University of Life Sciences
**Krzysztof Targiel**, University of Economics in Katowice
**Balázs Tóth**, University of Szeged, Hungary
**Justyna Wróblewska**, Krakow University of Economics